

Adaptive Cluster based Model for Fast Video Background Subtraction

Muralikrishna SN¹, Balachandra Muniyal², U Dinesh Acharya³
Department of Computer Science & Engineering^{1,3}
Department of Information & Communication Technology²
Manipal Institute of Technology
Manipal Academy of Higher Education
Manipal, India, 576104

Abstract—Background subtraction (BGS) is one of the important steps in many automatic video analysis applications. Several researchers have attempted to address the challenges due to illumination variation, shadow, camouflage, dynamic changes in the background and bootstrapping requirement. In this paper, a method to perform BGS using dynamic clustering is proposed. A background model is generated using the K'-means algorithm. The normalized γ corrected distance values and an automatic threshold value is used to perform the background subtraction. The background models are updated online to handle slow illumination changes. The experiment was conducted on CDNet2014 dataset. The experimental results show that the proposed method is fast and performs well for baseline, camera-jitter and dynamic background categories of video.

Keywords—Background subtraction; Gaussian mixture model; K'-means; clustering; object detection; transform

I. INTRODUCTION

Background subtraction refers to the extraction of moving objects, which are of special interest, from a video frame by removing the stationary contents. Background subtraction is one of the key tasks in surveillance video analysis. There are several methods available in the literature to address the issues in background subtraction. Some of the well-known methods for background subtraction can be classified as supervised background subtraction and unsupervised background subtraction.

In a supervised background subtraction, the process of bootstrapping generates a mathematical representation of the background called the background model. Further, the BG subtraction is done using the BG model, by assigning a membership value to the pixels as belonging to the foreground or to the background. In an unsupervised setting, one uses the derivative approach. The difference values from previous frames to current frame and few future frames are considered for background subtraction. Sometimes, the additional spatial information is used to predict the foreground. The frame differencing is one such method where the difference in intensity between two successive frames is used. The double differencing method is an enhancement where the second-order difference is considered for foreground detection. However, these methods fail miserably in situations of dynamic changes in the background [10], and also pose a challenge of ghosts in the foreground detected [5]. In a supervised method, a few frames called the training frames are used to generate a probabilistic model such as in GMM. This method

works well in most conditions. However, fixing the number of mixture components k is a challenge and normally it is determined experimentally [32]. Similarly, one can consider $k = 1$ for representing each pixel in the background with a running average value. The soft computing technique like Self-Organizing Background Subtraction (SOBS) maintains a lookup map for each pixel [18]. This self-organizing map is updated subsequently to determine the foreground. SOBS works well in indoor conditions and also does not require bootstrapping.

There are several other background subtraction methods such as Eigen-background, Kernel Density Estimation (KDE), running average etc. All these methods have their own inherent limitations [5]. The major challenges involved in background subtraction are, dynamic changes in the background which may include additional stationary objects in the background, which was not a part of the learning, illumination changes, shadows, camouflage [14]etc. Illumination variation can be a sudden change or a slow change in the lighting condition. In an outdoor setting, this is a common problem. The method such as GMM is simple and can handle a lot of these issues, however, it ignores the lower order distribution in the background model of a pixel. This is due to the fact that in a dynamic background the number of components can vary drastically. In the simplest form, GMM can be implemented using the K-Means algorithm as mentioned in [24]. The major limitation of the K-Means algorithm is that the number of clusters is predetermined and fixed. So, to overcome this difficulty the K'-Means algorithm is implemented for background subtraction in the proposed method. The details of the methodology are given in section III. In section IV, the experimental setup and the results are presented.

II. REVIEW OF EXISTING TECHNIQUES

One of the techniques to detect a moving object is using background subtraction. In this, the tasks involved are background modeling and foreground segmentation. Some of the simplest methods for background subtraction are temporal differencing, double differencing, running average and optical flow. The most basic technique is the inter-frame difference with a global threshold. A few others are based on probabilistic methods. These methods often aim at making the detection process more robust to noise, camera jitter and background motion [1]. The motion detection is also achieved using object detection techniques [11]. In this case, the system is trained

to detect any objects of interest. A window sliding method is applied to detect the object of different scales. Normally, feature-based or template-based technique is used to represent the object. The classifiers such as Support Vector Machine (SVM), Naive Bayes, and Artificial Neural Network are used to detect the object. Once the object is detected they can be tracked as well. One of the widely used methods for object detection is Histogram of Oriented Gradients (HoG) descriptor-based method. The descriptor based methods are normally invariant to rotation, scaling or illumination changes. The main disadvantage of this method is that it requires training and also window-sliding which is normally time-consuming. These methods are comparatively slower than the background subtraction methods discussed earlier. Several variants of descriptors are available.

The adaptive BGS uses dynamic updates in the background model. In the past decade, a lot of methods have been proposed for background subtraction using parametric and non-parametric background density estimates and spatial correlation. These methods are proven to be effective in background subtraction. Some of the methods from these categories are running Moving average, temporal median filter, KDE, GMM, Sequential KD approximation, Co-occurrence of image variations, Eigen-backgrounds and so on. Evaluations of background subtraction methods with respect to the challenges of video surveillance suffer from various shortcomings. To address this issue, the challenges of background subtraction in the field of video surveillance is studied in [3]. In [4], Conte et al. have conducted a thorough experimental comparison of different foreground detection algorithms on a large dataset of videos. It is concluded that both the techniques, GMM [24] and enhanced background subtraction (EBS), algorithms are adaptable and can be used effectively. Finally, it is given that the statistical background algorithm performs poorly when compared to the others. In some of the works presented in the literature, there have been attempts to capture the spatial dependency of pixels. Instead of using the pixel intensity other features like texture is also considered. In [25] several statistical features such as brightness, inverse contrast ratio, mixed contrast strength, integrated modal variability etc are used. In the training phase, these features are extracted and a bag-of-feature model is used. The background subtraction is done using a distance threshold from the bag-of-feature. This method provides a mechanism to update the background model. Also, this method makes use of majority voting scheme for label fusion. In most of the research presented in the literature, a number of general assumptions are considered as listed in [20].

In [23], a method for BGS is presented based on spatio-temporal binary feature and colour intensity of a pixel. This feature is good to represent the texture and allows to detect camouflaged objects and are not sensitive to illumination changes. The pixel-level feedback loop is maintained to update the model. The adjustments are based on the continuous monitoring, local segmentation noise levels. This approach outperforms most of the previously tested state-of-the-art methods on the Change Detection.net (CDNet) dataset.

The work presented in [22] uses multiple color spaces such as RGB, YCbCr, to create the background model. Unlike the existing techniques in this method multiple background

models, called the Background Model Bank (BMB), are used instead of a single background model. Each training image is treated as a background model. Then this set of initial models are clustered into a number of average background model in an iterative way. The absolute difference of the frame and the average background model is used as a clue to perform the final background subtraction. To handle the spatial dependency of pixels super-pixel segmentation and DBSCAN clustering are used. This makes use of color, texture and size information. Here the purpose of DBSCAN is to avoid over-segmentation. The performance of the algorithm is very good in terms of accuracy, however, the algorithm can process approximately 10 frames per second of an image of size 320×240 on a Intel core i5 PC with 8GB of RAM.

Recent developments in the field of convolution neural networks (CNN) and deep neural network (DNN) have contributed largely to background subtraction. Military applications require the detection of moving objects in camouflaged patterns. A detailed review of the existing methods for BGS using DNN is presented in [2]. The DNN based method to detect camouflaged people is presented in [30]. However, most of the existing techniques fail to extract the foreground in the camouflaged videos. BGS still remains a challenge as there are a few unsolved issues [12].

In brief, locating moving objects in a video sequence is the first step of many computer vision applications. Among the various motion detection techniques, background subtraction methods are commonly implemented, especially for applications with a fixed camera. In the second phase, after detecting the object, object classification, tracking is done [11][29][8]. The information from all these stages is crucial in any vision application. So, to implement a robust framework in any surveillance system requires a reliable background subtraction. This would help to speed-up the processing time of the entire system. In this paper, the most widely known algorithms such as SOBS, GMM, Simplified self organizing background subtraction, code-book based method and a few other algorithms are used for comparison. The GMM based method proposed by Zivkovic et al. in [31] addresses the same issue that is discussed in this paper. In the following section, the description of the proposed method is given.

III. PROPOSED WORK

Background modelling is an unsupervised learning task. Clustering techniques have been widely used in unsupervised learning. In the proposed method the key idea is to learn the different variations of pixel intensity values over time for every pixel in the video. This corresponds to the learning of multi-modal distribution of the pixel values. However, the number of components in the distribution is not known and difficult to estimate [22]. In this paper, an extension of the K-Means clustering [9] called the K^l -Means clustering [27] is implemented. This method helps to find the optimal number of cluster components $k' < k$ unlike the method given in [24].

In the proposed method, the background subtraction problem is formulated as an unsupervised learning task where the clusters are formed for each pixel location (p, q) varying over time. Given a video with n frames, the set of training frames is denoted by $\mathbf{H} = \{\mathbf{I}^1, \mathbf{I}^2, \dots, \mathbf{I}^t\}$ where $1 \leq t < n$. The intensity

or the colour information of the pixel at location (p, q) at time i is denoted by $\mathbf{I}_{p,q}^i \in \mathbb{R}^d$. For each location (p, q) , the clusters with $k'_{p,q}$ components are formed, considering the data $X_{p,q}^t = \{\mathbf{I}_{p,q}^1, \mathbf{I}_{p,q}^2, \dots, \mathbf{I}_{p,q}^t\}$. The colour information (R, G, B) or the (H, S, I) or any region descriptor can be used to form the background model. Now the data, $X_{p,q}^t$, is clustered to generate the background model. The process of model generation, and background subtraction is explained in the following sections.

A. Background Model Generation

The proposed method uses a two-phase algorithm for background model generation. The first phase is the K-Means algorithm governed by Equation 1.

$$\text{Minimize } J_{MSE} = \sum_{i=1}^k \sum_{\mathbf{x}_t \in \mathbf{c}_i} \|\mathbf{x}_t - \mathbf{c}_i\|^2 \quad (1)$$

Equation 1 is used to assign a point \mathbf{x}_t to a cluster centre \mathbf{c}_i with k clusters, where \mathbf{x}_t and $\mathbf{c}_i \in \mathbb{R}^d$.

The second phase reduces the number of clusters by minimizing information uncertainty in the clusters indicated by J_1 and the mean-square-error J_{MSE} . The modified cost function is given in Equation 2.

$$\text{Minimize } J = J_{MSE} + J_1 \quad (2)$$

The distance metric given in Equation 3 is used to assign a point \mathbf{x}_t to a cluster centre \mathbf{c}_i . The term $\log_2(p(\mathbf{c}_i))$ is the information conveyed by the cluster \mathbf{c}_i . The parameter E is a constant defined by the user. In the first phase, the initial k clusters are formed. Further, Equation 3 helps to detect the optimal number of clusters k' starting from a k where $k > k'$.

$$dm(\mathbf{x}_t, \mathbf{c}_i) = \|\mathbf{x}_t - \mathbf{c}_i\|^2 - E * \log_2(p(\mathbf{c}_i)), \quad E \in [a, 3a] \quad (3)$$

where $a = \text{average}(r) + \text{average}(d/2)$, r =radius of cluster, d =smallest distance between cluster centres. Any input data \mathbf{x}_t is assigned to a cluster i using Equation 4.

$$M(\mathbf{x}_t, i) = \begin{cases} 1 & \text{if } i = \text{argmin}(dm(\mathbf{x}_t, j)) \quad j = 1 \dots k \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

The clusters are formed from the data $X_{p,q}^t$, for each pixel (p, q) . The set of cluster centers $C_{p,q} = \{\mathbf{c}_{p,q}^i, i = 1 \dots k'\}$, for each location p, q in the video frame, and their variances $\sigma_{p,q}^i$ are used as the background model for a given video. It is important to adapt the background model for non-stationary data. The centroids are updated during background subtraction using online K-Means clustering of non-stationary data given in [24] [15]. The algorithm for background model generation is listed in Algorithm 1.

B. Background Subtraction

The background subtraction is done using 2. A distance matrix D_t is constructed for each frame at time instance t . A small distance value in the distance matrix indicates that the pixel belongs to the background and a high value for

Algorithm 1: Background Model Generation

Input: A sequence of spatially smoothed video frames
Parameters: Number of training frames t , the initial value of k , width w , and height h of video frame
Output: A set of cluster centroids $C_{p,q} = \{\mathbf{c}_{p,q}^i\}$, $i = 1, 2, \dots, k'$, for each location (p, q) in the video frame.
for $p \leftarrow 1$ **to** w **do**
 for $q \leftarrow 1$ **to** h **do**
 Initialize the cluster centres for the pixel (p, q)
 ;
 $C_{p,q} = \text{K'-Means}(X_{p,q}^t)$
 end
end

the foreground pixel. We address two cases in the proposed background subtraction algorithm. The first case is when there are no moving objects in the frame. The second case is when there are one or more moving objects. The first case is handled with a threshold value T which is set to p times the maximum cluster variance $\max_i\{\sigma_{p,q}^i\}$. In the second case, we use the Otsu binarization method [21] to automatically detect a threshold for the range normalized distance matrix. In the first case, the centroid of the corresponding model is updated online [24] [15].

Algorithm 2: The proposed method for Background Subtraction

Input: A sequence of spatially smoothed video frames
parameter: A threshold value T , Video Frame at time instance t , width w , and height h of video frame
Output: Background subtracted Image
for $p \leftarrow 1$ **to** w **do**
 for $q \leftarrow 1$ **to** h **do**
 $\text{dist} = \text{argmin}_i \text{getDistance}(\mathbf{c}_{p,q}^i, I_{p,q}^t)$
 if $\text{dist} < T$ **then**
 $\mathbf{D}_t(p, q) = \text{dist}$
 UpdateCentroid(p, q, i)
 else
 $\mathbf{D}_t(p, q) = \infty$
 end
 end
end
 $\mathbf{D}_t = \text{normalize}(\mathbf{D}_t)$;
 $\mathbf{D}_t = \text{power}(\mathbf{D}_t, \gamma)$;
Foreground $_t = \text{BinarizeDistanceMatrix}(\mathbf{D}_t)$;

Another important factor in the proposed method is the introduction of Gamma transformation. The transformation function is given by

$$\mathbf{D}_t = \text{Power}(\mathbf{D}_t, \gamma), \quad \gamma \geq 0 \quad (5)$$

The transformation function is shown in Figure 1. This helps to work with a fixed threshold value. The γ parameter is very helpful in the hardware-based implementation of the algorithm. The distance matrix can be transformed so as

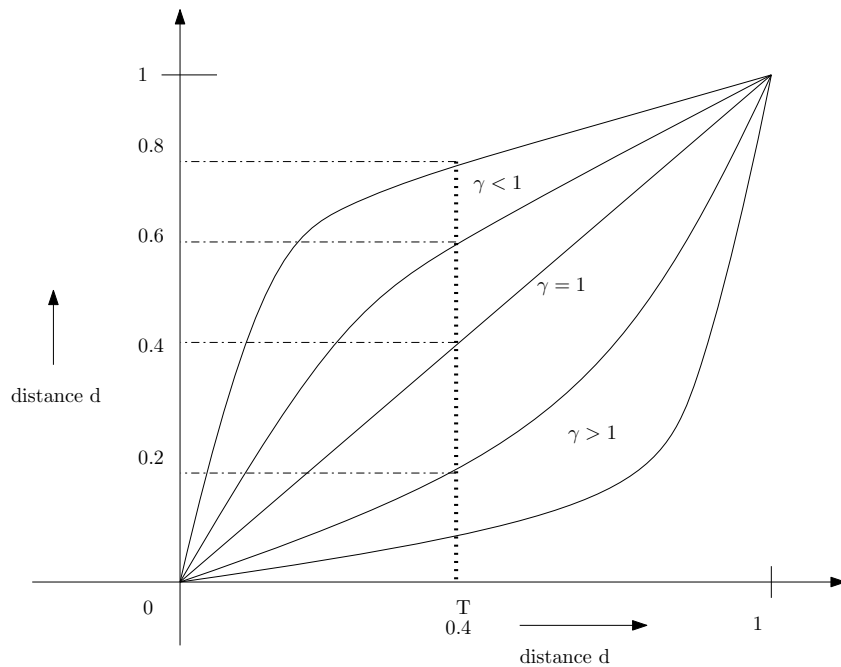


Fig. 1. The transformation using Gamma function on a distance matrix.

to include or exclude pixels which have an almost equal likelihood of belonging to the foreground or background. The value γ helps in adjusting the recall rate. Setting the γ to a lower value increases the recall rate. This can be observed in the example Figure 2. Experimentally it is observed that the γ is significant to fix an optimal threshold.

IV. EXPERIMENTS AND RESULTS

A good background subtraction algorithm must handle challenges such as gradual illumination changes, sudden illumination changes, dynamic background, camouflage, shadows, bootstrapping, and video noise. So the evaluation must be done using the dataset which has all these cases. Some of the most commonly used datasets are Wallflower [26], Performance Evaluation of Tracking and Surveillance (PETS), and Change Detection dataset [7] [6] etc. There are so many other datasets such as CAVIAR, Pedestrian detection dataset, IBM dataset etc.

The proposed method was evaluated quantitatively using the Change Detection (CDNet-2014) dataset [28]. The CDNet-2014 is one of the standard datasets with 11 categories of video. CDNet-2014 dataset contains more challenging cases like camera jitter, low frame rate etc. Each video category has four to six videos in it. The different categories of video are, Baseline, Dynamic Background, Camera-Jitter, Shadow, Intermittent Object Motion, Thermal, Challenging Weather, Low Frame-Rate, Night, PTZ, Air and Turbulence. Totally 53 videos having a resolution of each video frame varying from 160×120 to 720×576 . Duration of the videos is from 900 to 7,000 frames. To measure the performance of the algorithm there are several parameters such as precision, recall, specificity, false positive rate (FPR), false negative rate (FNR), the percentage of wrong classification (PWC), and F-measure.

The implementation of the proposed method was done

using C++ with OpenCV support. The development and testing were done on Intel(R) Core(TM) i5-6200U CPU @ 2.30GHz processor with 4GB RAM. The size of the video frames was down-sampled to half the original size in all the experiments to speed-up the processing. In the experiments, to perform background subtraction, the first $N = 100$ frames were used as the training data. We have used the parallel implementation of K-Means algorithm for the first phase of the algorithm. The initial k value was set to 10, using the rule of thumb on the best k in K-Means. The value of parameter E is set as 10. To evaluate the performance the threshold T is set as the 3.5 times the maximum standard deviation among the cluster components. The parameter γ is set to 0.85. These parameter values were retained for all the categories of the videos similar to the evaluation method used in the literature. The evaluation was done using the tool given by CDNet-2014 dataset providers.

The baseline category contains the most primitive challenges. The result of the baseline category is presented in Table I. The analysis shows that the proposed method works well on the baseline video category. It is better than some of the state-of-the-art methods like GMM, simplified Self-organized background subtraction, Multi-scale temporal model, DCB, and GraphCutDiff in terms of $F1$ -metric for baseline category video. This can be observed in Figure 3.

The dynamic background category is more challenging when compared to the baseline category. In the dynamic background category, the proposed method clearly performs better than all the listed methods in Table II except for SC-SOB [18] method. The PWC is observed to be ≈ 1.15 . The result is shown in Figure 4. The result of the experiment on camera-jitter category is presented in Table III. In this category, the performance of the proposed method is better than SC-SOBS and other techniques. The comparison is shown in Figure 5.

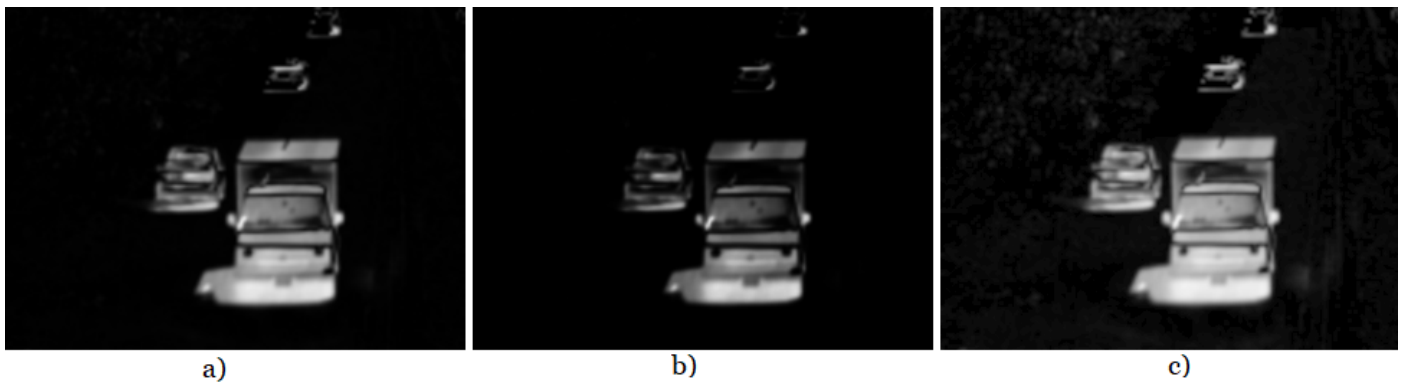


Fig. 2. Distance matrix of the video frame no. 268 from baseline category highway subcategory of changedetection (CDNet2014) dataset. a) Distance matrix with $\gamma = 1$ b) Transformed distance matrix with $\gamma = 0.6$ c) Transformed distance matrix with $\gamma = 1.6$

TABLE I. COMPARISON OF BGS METHODS ON BASELINE VIDEO CATEGORY.

Method	Average Sp	Average FPR	Average FNR	Average Recall	Average Precision	Average F-Measure
SC-SOBS [18]	0.9980	0.002	0.0673	0.9327	0.9341	0.9333
GMM — Zivkovic[31]	0.9972	0.0028	0.1915	0.8085	0.8993	0.8382
Multiscale Spatio-Temporal BG Model [17]	0.997	0.003	0.1863	0.8137	0.887	0.845
GraphCutDiff [19]	0.9960	0.004	0.2972	0.7028	0.8093	0.7147
Simplified Self-Organized Background Subtraction [13]	0.9965	0.0035	0.5106	0.4894	0.8419	0.6085
DCB [16]	0.9982	0.0018	0.2877	0.7123	0.907	0.7695
Proposed Method	0.9966	0.0034	0.1406	0.8994	0.9165	0.9027

TABLE II. COMPARISON OF BGS METHODS ON DYNAMIC BACKGROUND VIDEO CATEGORY.

Method	Average Sp	Average FPR	Average FNR	Average Recall	Average Precision	Average F-Measure
SC-SOBS [18]	0.9836	0.0164	0.1082	0.8918	0.6283	0.6686
GMM — Zivkovic [31]	0.9903	0.0097	0.1981	0.8019	0.6213	0.6328
GraphCutDiff [19]	0.9063	0.0937	0.2307	0.7693	0.5357	0.5391
DCB [16]	0.9991	0.0009	0.4197	0.5803	0.7632	0.6149
Simplified Self-Organized Background Subtraction [13]	0.9252	0.0748	0.4766	0.5234	0.1013	0.1613
Proposed Method	0.9721	0.0179	0.1774	0.8126	0.6273	0.6524

TABLE III. COMPARISON OF BGS METHODS ON CAMERA-JITTER VIDEO CATEGORY.

Method	Average Sp	Average FPR	Average FNR	Average Recall	Average Precision	Average F-Measure
DCB [16]	0.9969	0.0031	0.7204	0.2796	0.9107	0.3669
SC-SOBS [18]	0.9768	0.0232	0.1887	0.8113	0.6286	0.7051
GMM — Zivkovic [31]	0.9665	0.0335	0.3100	0.6900	0.4872	0.5670
Multiscale Spatio-Temporal BG Model [17]	0.9477	0.0523	0.2829	0.7171	0.3979	0.5073
GraphCutDiff [19]	0.9222	0.0778	0.3062	0.6938	0.5918	0.5489
Simplified Self-Organized Background Subtraction [13]	0.9373	0.0627	0.4192	0.5808	0.3411	0.4147
Proposed method	0.9889	0.0111	0.1650	0.8150	0.7313	0.7326

There is no single method which performs well in all categories of video. So, the proposed method performs well in primitive challenges. However, the overall performance is degraded because of the certain environmental conditions in the categories such as PTZ, badweather, and turbulence. The proposed method shows a high recall rate in most categories of the video indicating that the method is able to detect moving objects. The result of all the categories of videos is presented

in Table IV.

We have compared the performance of the proposed algorithm for its speed. The number of frames the algorithm is able to process per second is considered. The information available in the changedetection.net against each of the existing algorithm is used for comparison. The information is not complete for all resolutions. The proposed algorithm is much faster when compared to the other methods. This is important

TABLE IV. PERFORMANCE OF THE PROPOSED ALGORITHM ON ALL CATEGORIES OF VIDEO.

Method	Recall	Specificity	FPR	FNR	PBC	Precision	FMeasure
Baseline	0.8994	0.9966	0.0034	0.1406	0.4670	0.9165	0.9028
Dynamic background	0.6724	0.9721	0.0179	0.1774	1.1502	0.8126	0.6525
CameraJitter	0.8150	0.9889	0.0111	0.1650	1.7524	0.7313	0.7326
IntermittentObjectMotion	0.8423	0.8606	0.1394	0.1577	13.0122	0.4918	0.5149
LowFramerate	0.7143	0.8896	0.1104	0.2857	11.6353	0.4711	0.4040
NightVideo	0.3877	0.9758	0.0242	0.6123	3.4235	0.2676	0.2622
Shadow	0.6809	0.9479	0.0521	0.3191	6.2910	0.5227	0.4828
Thermal	0.8126	0.9116	0.0884	0.1874	9.3849	0.5287	0.5520
Turbulence	0.7589	0.9193	0.0807	0.2411	8.3041	0.0488	0.0756
PTZ	0.3109	0.7817	0.2183	0.6891	22.1236	0.0178	0.0335
BadWeather	0.7359	0.8474	0.1526	0.2641	15.7556	0.2347	0.2399
Overall:	0.6937	0.9174	0.0817	0.2945	8.4818	0.4585	0.4412

TABLE V. COMPARISON OF EXECUTION SPEED OF THE BACKGROUND SUBTRACTION ALGORITHMS

Method	Resolutions	Hardware	# Frames processed \approx
SOBS	320×240	C code on Core i3-330M for 2.13GHz	23
GraphCutDiff	320×240	video with C++ on a 2.4 Intel Core 2 Duo, single threading	7
Proposed Method	320×240	C++ Intel(R) Core(TM) i5-6200U CPU @ 2.30GHz	110
GMM — Zivkovic	720×480	C++ on Core i7 3.4GHz	49
GMM — Stauffer & Grimson	720×480	C++ on Core i7 3.4GHz	21
Proposed Method	720×480	C++ Intel(R) Core(TM) i5-6200U CPU @ 2.30GHz	49
SOBS	720×576	C code on Core i3-330M for 2.13GHz	4
Simplified Self-Organized Background Subtraction	720×576	core i7 laptop	0.06
Proposed Method	720×576	C++ Intel(R) Core(TM) i5-6200U CPU @ 2.30GHz	47

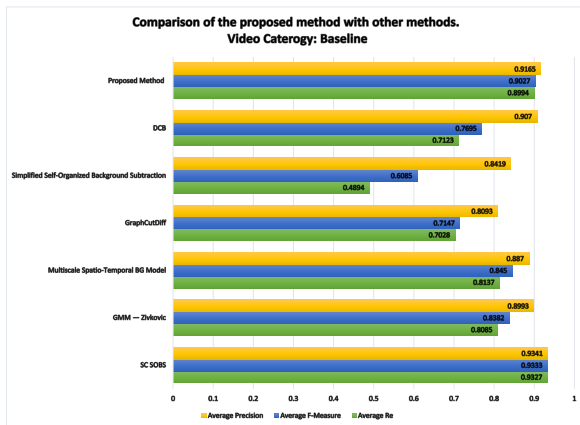


Fig. 3. Average precision, Average recall and Average F1-metric for the baseline video category

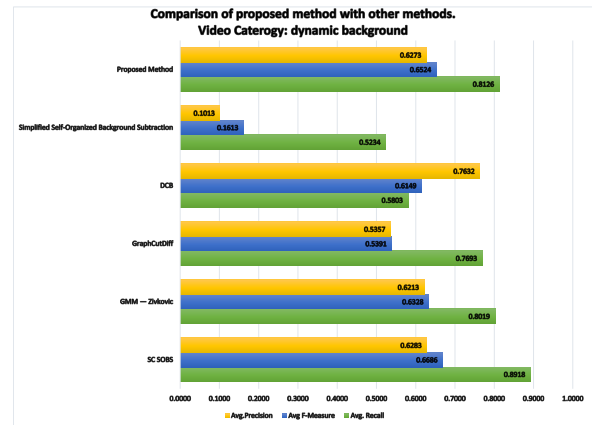


Fig. 4. Average precision, Average recall and Average F1-metric for the dynamic background video category

to develop real-time applications. The result is shown in Table V. The time required for training is not considered in the experiments. The qualitative result of the experiment is shown in Figure 6 and Figure 7.

V. CONCLUSIONS AND FUTURE DIRECTIONS

In this work, a fast background subtraction algorithm based on K'-Means algorithm is presented. The proposed method performs well in videos of baseline category and the camera-jitter, however, requires bootstrapping. Experiments were conducted extensively on the change detection dataset (CDNet2014) to demonstrate that the proposed method works

well in challenging conditions. The online centroid-update scheme helps to handle slow and gradual illumination changes. In the proposed work, building the background model is a time-consuming operation, however, the background subtraction can be done in almost real-time on a video of resolution 640×480. The results of the proposed method are compared with various other techniques. It is observed that the proposed method works better than some of the existing techniques in the baseline, dynamic background, and the camera-jitter category. Additional spatial features and texture feature can be used to enhance the results. The parallel implementation of the K'-Means algorithm can be used to speed up the training. The clustering method proposed in this paper is very simple to

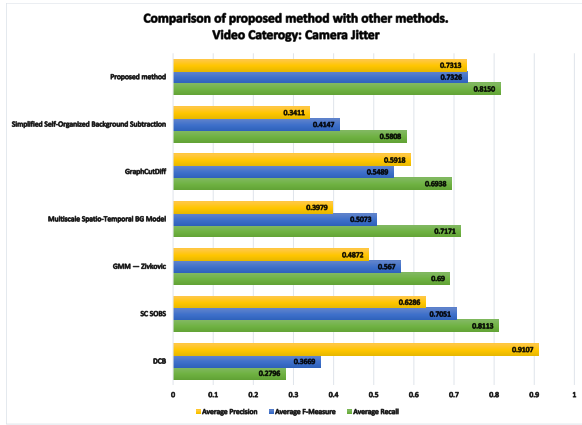


Fig. 5. Average precision, Average recall and Average F1-metric for the camera jitter

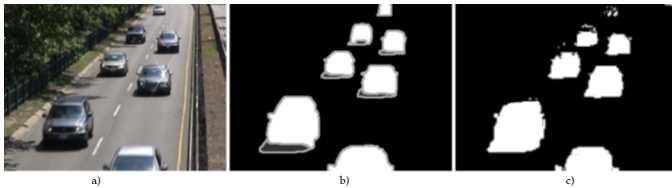


Fig. 6. BGS result of the proposed method Video category: baseline, subcategory: highway, Frame: # 838 a) Input Frame b) Ground truth c) Result

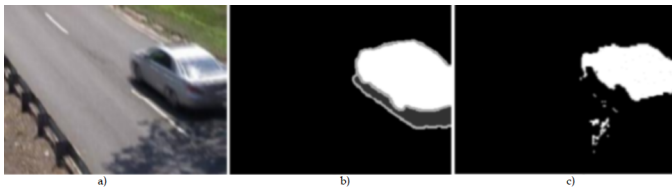


Fig. 7. BGS result of the proposed method Video category: camera-jitter, subcategory: traffic, Frame: # 1413, a) Input Frame b) Ground truth c) Result

implement and it is suitable for practical applications.

ACKNOWLEDGMENT

Authors would like to thank Manipal Academy of Higher Education for supporting this research work.

REFERENCES

- [1] Y. Benezeth, P.-M. Jodoin, B. Emile, H. Laurent, and C. Rosenberger. Review and evaluation of commonly-implemented background subtraction algorithms. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4, 2008.
- [2] Thierry Bouwmans, Sajid Javed, Maryam Sultana, and Soon Ki Jung. Deep neural network concepts for background subtraction: A systematic review and comparative evaluation, 2019.
- [3] S. Brutzer, B. Hoferlin, and G. Heidemann. Evaluation of background subtraction techniques for video surveillance. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 1937–1944, 2011.

- [4] Donatello Conte, Pasquale Foggia, Gennaro Percannella, Francesco Tufano, and Mario Vento. An experimental evaluation of foreground detection algorithms in real scenes. *EURASIP J. Adv. Signal Process.*, 2010:7:1–7:10, February 2010.
- [5] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati. Detecting moving objects, ghosts, and shadows in video streams. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10):1337–1342, Oct 2003.
- [6] N. Goyette, P. M. Jodoin, F. Porikli, J. Konrad, and P. Ishwar. Changetection.net: A new change detection benchmark dataset. In *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–8, June 2012.
- [7] N. Goyette, P. M. Jodoin, F. Porikli, J. Konrad, and P. Ishwar. A novel video dataset for change detection benchmarking. *IEEE Transactions on Image Processing*, 23(11):4663–4679, Nov 2014.
- [8] I. Haritaoglu, D. Harwood, and L.S. Davis. W4: real-time surveillance of people and their activities. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):809–830, 2000.
- [9] J. A. Hartigan and M. A. Wong. A k-means clustering algorithm. *JSTOR: Applied Statistics*, 28(1):100–108, 1979.
- [10] K.K. Hati, P.K. Sa, and B. Majhi. Intensity range based background subtraction for effective object detection. *Signal Processing Letters, IEEE*, 20(8):759–762, 2013.
- [11] Weiming Hu, Tieniu Tan, Liang Wang, and S. Maybank. A survey on visual surveillance of object motion and behaviors. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 34(3):334–352, 2004.
- [12] Pierre Marc Jodoin, Lucia Maddalena, Alfredo Petrosino, and Yi Wang. Extensive Benchmark and Survey of Modeling Methods for Scene Background Initialization. *IEEE Transactions on Image Processing*, 26(11):5244–5256, 2017.
- [13] Jean Meunier Kamal Sehairi, Fatima Chouireb. Comparative study of motion detection methods for video surveillance systems. *Journal of Electronic Imaging*, 26:26 – 26 – 29, 2017.
- [14] W. Kim and C. Jung. Illumination-invariant background subtraction: Comparative review, models, and prospects. *IEEE Access*, 5:8369–8384, 2017.
- [15] Angie King. Online k-means clustering of nonstationary data. 2012.
- [16] R. Krungkaew and W. Kusakunniran. Foreground segmentation in a video by using a novel dynamic codebook. In *2016 13th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, pages 1–6, June 2016.
- [17] X. Lu. A multiscale spatio-temporal background model for motion detection. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 3268–3271, Oct 2014.
- [18] L. Maddalena and A. Petrosino. The sobs algorithm: What are the limits? In *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 21–26, June 2012.
- [19] A. Miron and A. Badii. Change detection based on graph cuts. In *2015 International Conference on Systems, Signals and Image Processing (IWSSIP)*, pages 273–276, Sept 2015.
- [20] Thomas B. Moeslund and Erik Granum. A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding*, 81:231–268, 2001.
- [21] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, Jan 1979.
- [22] Hasan Sajid and Sen Ching Samson Cheung. Universal multimode background subtraction. *IEEE Transactions on Image Processing*, 26(7):3249–3260, 2017.
- [23] P. L. St-Charles, G. A. Bilodeau, and R. Bergevin. Subsense: A universal change detection method with local adaptive sensitivity. *IEEE Transactions on Image Processing*, 24(1):359–373, Jan 2015.
- [24] Chris Stauffer and W.E.L. Grimson. Learning patterns of activity using real-time tracking. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):747–757, 2000.
- [25] Badri Narayan Subudhi, Susmita Ghosh, Simon C.K. Shiu, and Ashish

- Ghosh. Statistical feature bag based background subtraction for local change detection. *Information Sciences*, 366:31–47, 2016.
- [26] Kentaro Toyama, John Krumm, Barry Brumitt, and Brian Meyers. Wallflower: Principles and practice of background maintenance. In *ICCV*, 1999.
- [27] Krista Rizman Žalik. An efficient k'-means clustering algorithm. *Pattern Recognition Letters*, 29(9):1385 – 1391, 2008.
- [28] Y. Wang, P. M. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, and P. Ishwar. Cdnet 2014: An expanded change detection benchmark dataset. In *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 393–400, June 2014.
- [29] Xin Xu, J. Tang, Xiaoming Liu, and Xiaolong Zhang. Human behavior understanding for video surveillance: Recent advance. In *Systems Man and Cybernetics (SMC), 2010 IEEE International Conference on*, pages 3867–3873, 2010.
- [30] Yunfei Zheng, Xiongwei Zhang, Feng Wang, Tiejong Cao, Meng Sun, and Xiaobing Wang. Detection of people with camouflage pattern via dense deconvolution network. *IEEE Signal Processing Letters*, 26(1):29–33, jan 2019.
- [31] Z. Zivkovic. Improved adaptive gaussian mixture model for background subtraction. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, volume 2, pages 28–31 Vol.2, Aug 2004.
- [32] Zoran Zivkovic and Ferdinand van der Heijden. Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern Recogn. Lett.*, 27(7):773–780, May 2006.