International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Comparing Segmentation Quality of Real-Time Image Segmentation Techniques Using Metrics

Sandeep Kumar Dubey[1] Bineet Kumar Gupta[2], Shobhit Sinha[3], Pratibha[4] , Sandip Vijay[5]

Submitted: 28/01/2024 **Revised**: 06/03/2024 **Accepted**: 14/03/2024

Abstract: The roles of real-time image segmentation and matching are of utmost importance, requiring the utilisation of efficient algorithms to promptly and precisely analyze images. The main goal of this research is to explore the procedure of obtaining real-time images from video footage using a camera, followed by the selection of a random image frame for the purpose of segmentation. Nevertheless, a notable obstacle arises when calculating metrics like Precision, Recall, F1 score, Accuracy, and SSIM by utilizing the segmented image and the Ground Truth image. To address these problems, a range of segmentation approaches are assessed in terms of their efficacy in computing the metrics described above. The segmentation technique employed in this study is the proposed method, which yields segmented images that exhibit successful outcomes. The paper presents a suggested methodology to augment the metrics of Precision, Recall, F1 score, Accuracy, and SSIM.

Keywords: *Image segmentation, foreground segmentation, Precision, Recall, F1-score, Accuracy*

1. Introduction

Foreground quest stands as a fundamental and extensively employed methodology for the segmentation of foreground objects in real-time applications. Among the three prevalent techniques utilized for detecting foreground objects in a video scene, namely optical flow, background subtraction and temporal differencing, latter is the recommended method due to its ability to accurately recognize the foreground and its computing efficiency.

In general, background subtraction approaches [1,4] include the establishment of a model that represents the stationary components of a video image as the backdrop. This model is then utilized to analyze the existing image in contrast to the background model to identify and differentiate foreground entities. The algorithms mentioned in reference [5] encompass a wide spectrum, ranging from basic frame subtraction techniques to advanced probabilistic models. Nevertheless, basic background subtraction techniques may encounter difficulties in reliably identifying foreground items in dynamic video sequences.

Dynamic video sceneries frequently encompass intricate background components, as seen by the presence of waving trees, fluctuating lighting conditions, and mobile items such as seats and escalators [6]. In such situations, conventional background subtraction techniques may prove inadequate in achieving precise foreground segmentation.

The accurate detection of foreground objects is of utmost importance in applications related to public safety [7] and traffic monitoring systems. Hence, developing robust foreground detection algorithms remains a critical task to ensure the effectiveness of these applications.

Despite numerous proposed algorithms [13,15], achieving illumination-invariant foreground detection remains an on-going challenge. An effective foreground detection algorithm should be capable of adapting to both subtle, gradual shifts and abrupt alterations in illumination. Moderate changes in illumination may occur due to factors like the movement of the sun throughout the day, while sudden changes may result from clouds passing over outdoor scenes or indoor lights being turned on or off. While existing algorithms can handle gradual illumination changes to some extent, sudden illuminations still pose difficulties.

Although many works have attempted to assess the efficacy of foreground detection algorithms [2,3], as far as our awareness extends, this article represents the inaugural effort to explore the adaptability and effectiveness of various algorithms in the context of diverse illumination variations. This paper offers a comprehensive examination of frequently employed foreground detection algorithms alongside a novel approach designed to identify moving objects in conditions of changing illumination. The findings derived from this investigation could prove valuable for engineers tasked with choosing the most suitable algorithms for precise foreground object detection in diverse video scenes.

[1]*Department of Computer Science and Engineering, Shri Ramswaroop Memorial University, Barabanki, Uttar Pradesh, India*
[2,3]*Department of Computer Science & Information Systems, Shri Ramswaroop Memorial University, Barabanki, Uttar Pradesh, India*
[4]*Electronics & Communication Engineering, Shri Ramswaroop Memorial College of Engineering and Management, Lucknow, Uttar Pradesh, India*
[5]*Electronics & Communication Engineering, Tula's Institute, Dehradun ,Uttarakhand, India*
*[1] *Corresponding Author Email: sandycs924@gmail.com*

1.1 Survey Outline

The subsequent segments of this paper are structured in the following manner: Section 2 furnishes an outline of pertinent literature, Section 3 explains the proposed methodology, Section 4discusses Comparative Evaluation Metrics for foreground segmentation, Section 5 the basic concepts of segmentation to provide a perspective, Section 6 unveils both the simulation outcomes and experimental findings, while Section 7 provides a concluding summary and delves into potential avenues for future research.

2 Related Works

Ran Jin et al. (2021) proposed The ESPNet (Efficient Spatial Pyramid of Dilated Convolutions for Semantic Segmentation) is enhanced using a multilabel classification approach through the following steps. Initially, the conventional convolutional process is substituted with the application of Receptive Field within the Deep Convolutional Neural Network's convolutional layer. This substitution ensures that each pixel within the covered region contributes to the ultimate feature reaction. Subsequently, the ASPP (Atrous Spatial Pyramid Pooling) module is refined by leveraging atrous convolutions, while DB-ASPP (Dilate Batch Normalization-ASPP) technique is introduced to mitigate gridding artifacts stemming from the multilayer atrous convolution. This approach achieves the amalgamation of multiscale information and the integration of feature details pertinent to the image dataset [18].

Tianfei Zhou et al. (2022) suggested Segmentation, which involves dividing video frames into various segments or objects, holds immense significance across a diverse array of real-world applications. Its utility spans from enriching visual effects within films, comprehending scenes within autonomous driving contexts, to fabricating virtual backgrounds for video conferencing purposes. [19].

H R Swathi, Shah Sohini et al. (2020) Our approach to video object segmentation combines a streamlined yet highly effective target appearance model with a segmentation network. Surprisingly, even with its straightforward design, a linear discriminative model proves adept at producing strong target predictions. These predictions are then translated into top-notch object segmentations by the segmentation network. Notably, our method efficiently trains the target model during inference, enabling high-speed operation. Demonstrating its prowess, our technique attains state-of-the-art results on the YouTube-VOS data files and delivers competitive outcomes on DAVIS 2017, even when trained on a restricted dataset. [20].

Highlighting its excellence, our approach achieves top-tier results on the YouTube-VOS dataset and produces competitive results on DAVIS 2017, even when it has been trained on a limited dataset.

Tonghao Chen; Derek Eager; Dwight Makaroff et al. (2023) proposes a We initially implement a pre-processing step to mitigate disparities in the urban dataset we've collected and a publicly available dataset. Following this, we employ the pre-processed dataset to up skill multiple segmentation models. These models' segmentation outcomes are then amalgamated into a single segmentation map based on a weighted rule. To assess the effectiveness of our method, we assembled a collection of unlabeled Google Street View images. We've developed a model by utilizing the cityscapes dataset that shares similarities with the foregrounds in the gathered images. Our evaluation encompasses both quantitative analysis, comparing the model's performance against ground truth annotations in the cityscapes dataset, and qualitative examination, involving user studies to assess the segmentation results of GSV data. Remarkably, our approach surpasses existing techniques, highlighting the capacity of computer vision technology to enable precise and efficient analysis of urban environments. [21].

Karthikeyan Panjappagounder Rajamanickamet al. (2019) suggested approach adjusts for alterations in lighting conditions by dynamically refining the background model, utilizing disparities in entropy values between the present and prior frames. This method of background modeling based on entropy proves adept at effectively managing abrupt as well as gradual shifts in illumination. To validate its efficacy, the newly devised algorithm is assessed across 6 video strings and contrasted against 4 other alternative algorithms, showcasing its proficiency through metrics such as F-score, similarity, and frame rate performance. [22].

Patel et al. (2015) presented an interaction framework designed to connect an illumination-invariant model with a color-based model. This innovative framework introduces an exceptional feedback mechanism that clusters attributes in a novel way, which strongly shapes model updates and indirectly contributes to object validation. The architecture prioritizes real-time performance, and its core principles can be adapted into a more streamlined real-time version by simplifying each module. [23].

P. R. Karthikeyan et al. (2018) suggested a methodology that assesses the entropy of a video scene in order to measure the degree of illumination changes. Subsequently, they select the most suitable updating model by comparing the disparities in entropy measurements. Various standard datasets, showcasing diverse and demanding illumination scenarios, have been employed to assess the effectiveness of the foreground perception techniques. Through exploratory investigations, the proposed algorithm's performance is demonstrated alongside multiple other algorithms, across a range of illumination conditions, highlighting its efficiency and minimal time complexity. [24].

Yuan-Ting Hu et al. (2018) an innovative algorithm for video object segmentation is introduced, utilizing a matching-oriented methodology. Unlike conventional classification techniques that rely on memorization, this novel approach involves training to correlate extracted features with a given template, avoiding the need to memorize object appearances. The efficacy and resilience of this method are confirmed through rigorous testing on demanding datasets including DAVIS-16, Youtube-Objects, DAVIS-17 and JumpCut. [25].

Yuan-Ting Hu et al. (2016) introduces a highly efficient multi-view learning method designed for foreground detection in traffic surveillance scenarios. The method comprises three key stages. Initially, a source backdrop image is formed using temporal median filtering. Subsequently, diverse range of characteristics such as illuminance, texture, chromaticity variations is extracted from the video sequence, with each feature representing a distinct perspective. Finally, a Multi-perspective learning technique has been devised for the real-time, continuous estimation of conditional probability distributions for foreground and background elements. [26].

3. Proposed Methodology

Conducting real-time image segmentation in research poses a formidable challenge that demands the deployment of efficient algorithms and methodologies too swiftly and accurately process images.

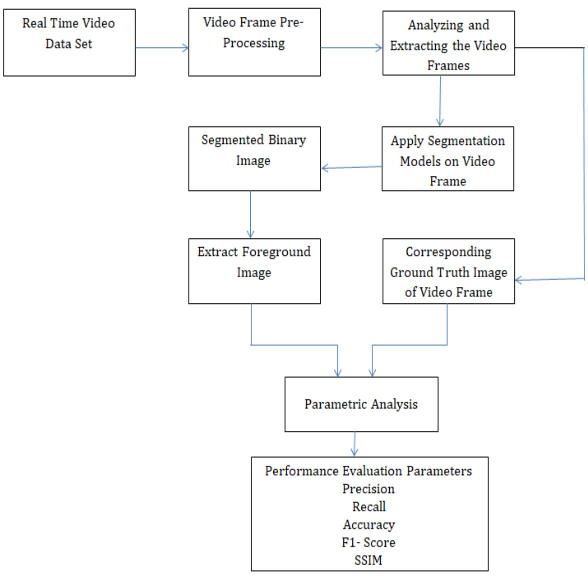

Fig 1- Block Diagram of Segmentation Performance Evaluation

The suggested algorithm aims to gauge the extent of illumination alterations within a video scene. It accomplishes this by quantifying the entropy of each frame, effectively assessing the degree of visual transformation in the arena. In this method, the initial frame of the video functions as the baseline background depiction. Subsequent frames' entropies are computed and juxtaposed against the entropy value from the initial frame. This comparison assists in deciphering the scale of change that transpired in the video scene.

When the ascertained alteration surpasses a predetermined threshold, it indicates a significant illumination shift. Consequently, the existing background model reverts to its original state. Conversely, if the detected alteration is within acceptable limits, the on-going background model is updated iteratively, akin to a single Gaussian model. The overarching approach here is twofold: swift alterations in illumination prompt the transition to the baseline background model, while gradual changes permit the on-going model to evolve incrementally. The empirically determined threshold value set to 0.07 facilitates this distinction.

The elegance of this algorithm lies in its straightforwardness. It achieves real-time operation thanks to its straightforward design and delivers commendable outcomes. This combination of efficiency and efficacy makes it a valuable tool for assessing illumination changes within video scene.

The provided algorithm includes a pseudo-code representation of the proposed model:

1. Input: Real Time Video File

- Import the necessary libraries.

2. Open Video File:

- Specify the path to the input video file.
- Create a video capture object and pass the video path.

3. Process Video Frames:

- Start a loop that iterates while the capture object is open.
- Read the next frame from the video.
- If the returned value (ret-variable assumed)is False, it means there are no more frames, so break the loop.

4. Convert Frame to Gray scale:

- Convert the current colour frame to

grayscale using conversion code.

5. Calculate Histogram:

- Calculate the histogram of pixel intensities in

the grayscale frame.

6. Calculate Probabilities:

- Calculate the probabilities of each intensity level by dividing the histogram values by the sum of all histogram values.

7. Calculate Entropy:

- Calculate the entropy of the frame using the formula:

"entropy = -np.sum (probabilities * np.log2(probabilities + 1e-10))"

$$H(s) = \sum_{i=1}^{n} p_i log_2 \frac{1}{p_i}$$

$$H(s) = -\sum_{i=1}^{n} p_i log_2 p_i$$

8. Print and Update Entropy:

- Print the calculated entropy value for the present frame.

Calculate the entropy of the present frame as Ecurrent then compare it with entropy of the previous frame, Eprevious.

If

(Ecurrent - Eprevious) is greater than the threshold (thresh):

Assign the starting background model to the present background model.

Else:

Upgrade the background model using a recursive filter.

Determine the foreground by contrasting the present frame with the background model.

Bring up to date the entropy value.

9. Release Resources:

- After the loop ends, release the video capture object to free up system resources.

An apt segmentation technique plays a pivotal role in computing essential values like Precision, Recall, F1 score, Accuracy, and SSIM. Through rigorous experimentation encompassing a range of 5 segmentation methods, an intriguing revelation emerged: proposed emerged as the most promising candidate for yielding superior outcomes concerning Precision, Recall, F1 score, Accuracy, and SSIM in the realm of real-time processing.

The deployment of the proposed segmentation technique unveils segmented images marked by F1 scores ranging from 0.2 to 0.5, coupled with accuracy levels spanning from 0.5 to 0.7. Building upon this foundation, a novel approach is introduced – the proposed method approach aims to refine the Precision, Recall, F1 score, Accuracy, and SSIM values, presenting an avenue for even more precise results.

Upon applying the proposed approach, the subsequent phase involves matching image frames extracted from a video stream. The culmination of this endeavour is the demonstration of an accuracy percentage pertaining to the successfully matched frames within the video. This intricate research initiative into real-time image segmentation and matching underscores the intricate interplay between cutting-edge algorithms and the imperative for swift, accurate image processing.

4. Comparative Evaluation Metrics for segmentation Methods

This paper primarily emphasizes the Segmentation part to make the foreground image extractable, as well as to satisfy the Quality for image data comparison. For comparing and choosing one of the best segmentation techniques for our research work, it is very crucial to know about the necessary evaluation metrics for segmentation methods.

4.1 Precision Score

The precision score serves stands as a valuable metric for assessing the effectiveness of predictions, particularly in situations where there is a significant class imbalance. In mathematical terms, it is denoted as the proportion of true positives in relation to the total of true positives and false positives.

Precision Score = TP / (FP + TP)

4.2 Recall Score

It quantifies the effectiveness of our machine learning model in correctly detecting all true positives among the total positives present within dataset. Recall is alternatively referred to as sensitivity or the true positive rate. Recall can also be called sensitivity or the true positive rate.

Recall Score = TP / (FN + TP)

4.3 Accuracy Score

Accuracy serves as a vital metric in the realm of machine learning, specifically in classification tasks. It is quantified as the proportion of true positives and true negatives divided by the sum of all positive and negative observations. Essentially, accuracy provides insight into the frequency with which our machine learning model accurately predicts outcomes among the total instances it makes predictions.

$$\text{Accuracy Score} = \frac{(\text{TP+TN})}{(\text{TP+FN+ TN+FP})}$$

4.4 F1-Score

The F1 score serves as a performance metric for assessing machine learning models by considering both precision and recall. Unlike accuracy, which relies on knowing the total number of observations, the F1 score offers a well-balanced assessment of a model's accuracy without requiring that information. It is often employed as a single, informative value that assesses the overall quality of a model's output.

The F1 score is particularly valuable in situations where optimizing either precision or recall can negatively impact the model's performance.

$$\text{F1 Score} = \frac{2\times \text{Precision Score}\times\text{Recall Score}}{(\text{Precision Score} + \text{Recall Score})}$$

4.5 Structural Similarity

The Structural Similarity (SSIM) Index serves as a metric for assessing the quality of images. It involves the computation of the SSIM index for a given image in comparison to a reference image, typically one of pristine quality. This quantitative measure encompasses three key aspects: luminance, contrast, and structural information, all of which are used to calculate the SSIM value.

SSIM can serve as a reference standard for evaluating the efficacy of various image processing algorithms.

5 Segmentation Techniques

Foreground detection is a computer vision technique used to separate objects or regions of interest from the background in an image or video sequence. It is commonly used in applications like object tracking, surveillance, and video analytics. There are various methods and algorithms for foreground detection, and they can vary in complexity. One common approach is based on background subtraction.

Background subtraction is a simple and effective technique for foreground detection. It involves subtracting a background model through the current frame to obtain the foreground mask that highlights the in motion or dynamic objects in the scene. Here's a basic equation to represent background subtraction for foreground detection:

$$F_t(x,y)=|I_t(x,y)-B_t(x,y)| \qquad (1)$$

$F_t(x,y)$is foreground mask at t time with pixel coordinates (x,y).

$I_t(x,y)$ stands for intensity (or color) value of the pixel at position (x,y) within current frame at time t.

$B_t(x,y)$ represents the intensity (or color) value of the corresponding pixel within background model at t time.

5.1 MOG2:

The MOG2 algorithm models the background of an image by employing a combination of several Gaussian distributions, with each one capturing distinct color or intensity levels. When a new frame is captured, the algorithm updates these Gaussian distributions and calculates the likelihood of each pixel being associated with the background based on these distributions. If a pixel's likelihood falls beneath a specific threshold, it is classified as part of the foreground; otherwise, it is considered part of the background[12].

The specific formula for the probability calculation involves computing probability density function (PDF) of each Gaussian distribution and then combining PDFs to get the final probability for a pixel. The PDF for a Gaussian distribution is given by:

$$P(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}\, exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \qquad (2)$$

P(x; μ, σ) represents the probability density function describing a Gaussian distribution characterized by its mean (μ) and standard deviation (σ).

Here, 'x' denotes the pixel value.

μ signifies the mean value of the Gaussian distribution associated with this pixel.

σ denotes the standard deviation of the Gaussian distribution associated with this pixel.

5.2 KNN

K-Nearest Neighbors (KNN) stands as a supervised machine learning technique employed in both classification and regression undertakings. It presents a straightforward and easily graspable approach that hinges on data point resemblances. In classification tasks, KNN assigns a class label to a given data point by scrutinizing the majority class among its k-nearest neighbors. Meanwhile, in regression scenarios, it makes predictions of continuous values by computing the average or weighted average of the target values associated with its k-nearest neighbors [13].

The Euclidean distance emerges as the customary metric to gauge the separation between data points in the KNN methodology. Given two data points A and B with n features (dimensions), the Euclidean distance formula is:

$$Euclidean\ Distance\ (d) = \sqrt{\sum_{i=1}^{n}(A_i - B_i)^2} \qquad (3)$$

A_i and B_i are the values of the i-th feature for data points A and B, respectively.

∑ denotes summation across all features.

5.3 Chan VESE

The Chan-Vese model formulates image segmentation as a task of minimizing energy. Its primary goal is to partition the image into two regions: the object of interest and the background. The expression representing the energy functional of the CV model is as follows:

*E(u, C_1, C_2) = Length(Total Variation) + λ * (|I - C_1|^2 - |I - C_2|^2)* (4)

E(u, C_1, C_2) is energy functional to be minimized.

u(x, y) is a binary-valued function representing the segmentation mask, where u(x, y) = 1 inside the object of interest and u(x, y) = 0 in the background.

C_1 and C_2 denote the mean intensity values within the object and background regions, respectively.

I(x, y) represents the grayscale image being segmented.

λ serves as a standardization parameter which balances the length of total variation term and the data fitting term[8].

5.4 Single Gaussian

A Single Gaussian, often referred to as a univariate Gaussian or normal distribution represents a probability distribution that models a continuous random variable. This distribution is defined by two essential parameters: mean (μ) and standard deviation (σ). The probability density function (PDF) for a Single Gaussian is expressed through the following mathematical formula:

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (5)$$

x is the random variable for which you want to calculate the probability density.

μ often referred to as mean or average of distribution, which represents the center of distribution and gives its location.

σ (sigma), known as the standard deviation, quantifies the extent to which the data in the distribution is spread out or dispersed.

The function $f(x|\mu,\sigma)$ denotes the probability associated with the random variable x assuming a particular value within the distribution. The area under the curve within a range of values represents the probability of x falling within that range.

5.5 GMM

A Gaussian Mixture Model (GMM) is a probabilistic modeling technique employed for both clustering and density estimation. It characterizes a blend of multiple Gaussian (or normal) distributions, each distinguished by its unique set of parameters. GMMs find widespread application in the realms of machine learning and statistics, serving a variety of purposes including clustering, estimating data densities, and generating data [21].

The PDF of a univariate Gaussian distribution with mean (μ) and variance (σ^2) is expressed as follows:

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (6)$$

For a multivariate Gaussian distribution with a vector of means (μ) and a covariance matrix (Σ), the PDF is:

$$f(x|\mu,\Sigma) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)} \quad (7)$$

x is a D-dimensional input vector.

The magnitude of Σ represents the determinant of the covariance matrix.

Mixture Model:

In GMM, we represent the data as a combination of K Gaussian distributions. Probability density function for GMM is a weighted sum of these Gaussians:

$$f(x) = \sum_{i=1}^{K} \pi_i . f(x|\mu_i, \Sigma_i) \quad (8)$$

K represents the count of Gaussian components within the mixture.

πi denotes the mixing coefficient associated with the i-th Gaussian, signifying the likelihood of selecting that specific Gaussian

μi and Σi stand for the mean and covariance matrix for the i-th Gaussian component.

6 SIMULATION RESULTS AND ANALYSIS

The following analysis is presented in this paper:: (1) Capturing the real-time video footage and convert it into multiple frames (Like 30 frames per sec) then analyzing the dimensions of different frames as illustrated in Fig 2 and Fig 3.

(2) Performance segmentation of various segmentation techniques based on some of the metrics discussed above and that illustrates the Accuracy and F1-Score.

(3) The proposed segmentation has been shown below.

The segmentation's performance is emulated through the utilization of open-source software. Real-time acquisition of image frames is tested using software with webcam. The input image sequences considered for implementation are “real time self-images” with dimension 640×480 is converted from real time video. Many of the segmentation techniques can be applied on the frames which are selected for the further processing.

Data Set 1: Close to camera

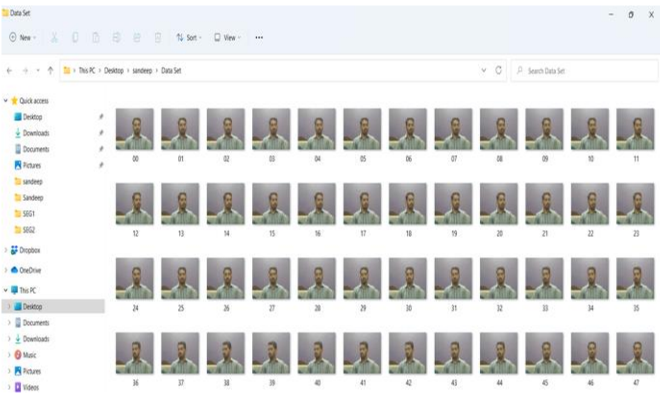

Fig 2- Capturing the close to camera real time video and convert it into multiple frames (approx. 30 frames per sec)

The best results are obtained for the proposed technique.

The performance of the segmentation is analyzed by assessing parameters like Precision, Recall, Accuracy and F1 Score.

Data Set 2: Stay away from camera

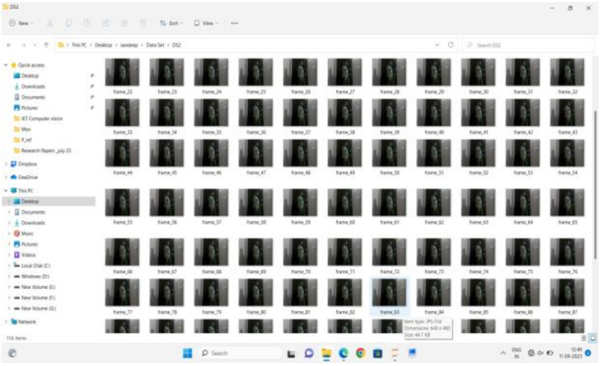

Fig 3- Capturing the Stay away from camera real time video and convert it into multiple frames (approx. 30 frames per sec)

The experiments are conducted for all Segmentation techniques on real time set of image frames.

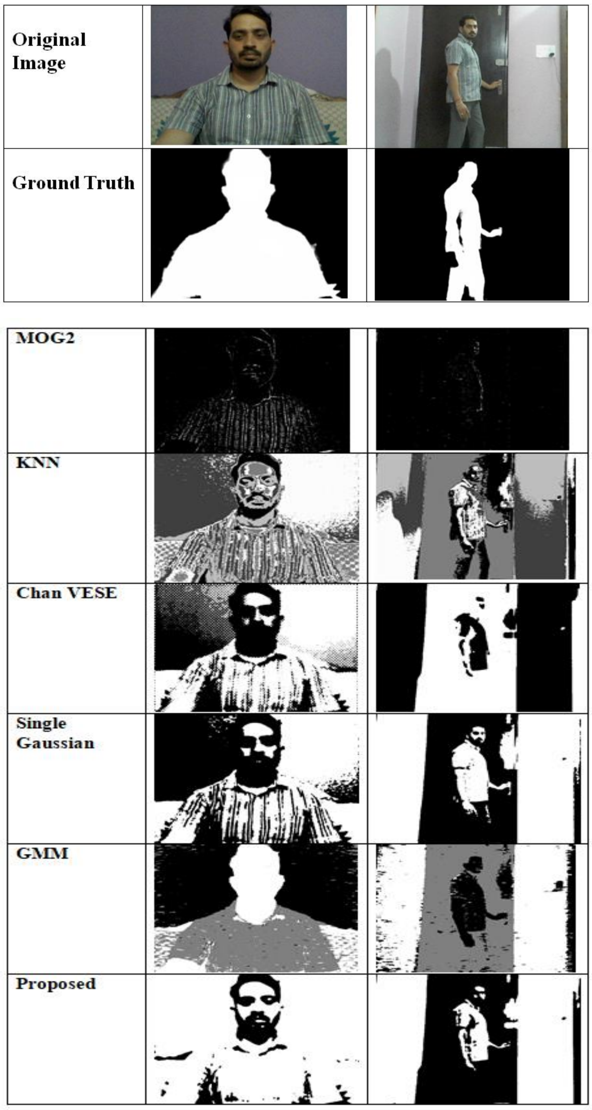

Fig 4 : Results of foreground detection for close to camera and stay away from camera video sequences

Table 1- Comparison of different segmentation techniques based on various metrics

Distance Invariant Data	Metrics	Segmentation Techniques					
		MOG2	KNN	Chan VESE	Single Gaussian	GMM	Proposed Method
Data SET 1	Precision:	0.17	0.19	0.2	0.27	0.43	0.49
	Recall:	0.34	0.33	0.42	0.48	0.64	0.71
	Accuracy:	0.66	0.61	0.77	0.72	0.89	0.89
	F1 Score:	0.23	0.24	0.27	0.35	0.51	0.58
	SSIM:	0.43	0.16	0.49	0.29	0.57	0.61
Data SET 2	Precision:	0.16	0.15	0.19	0.56	0.25	0.43
	Recall:	0.19	0.23	0.38	0.24	0.76	0.73
	Accuracy:	0.59	0.47	0.74	0.65	0.75	0.76
	F1 Score:	0.17	0.18	0.25	0.34	0.38	0.54
	SSIM:	0.74	0.17	0.47	0.25	0.17	0.54

The table-1 presents theoretical metrics for different segmentation techniques (MOG2, KNN, Chan VESE, Single Gaussian, and GMM) applied to real-time images. Graphical representation of comparison of different metrics over various segmentation techniques is shown in figure-5 to 9 and based on the provided metrics; the following observations can be made:

To quantitatively evaluate the algorithms, we employed three key metrics: recall, precision, and F-measure. Table 1 presents the numerical evaluations for various methods applied to each tested video sequence. It's important to note that all these metrics fall within a range of 0 to 1, where a lower value indicates poorer performance, while a higher value signifies better performance.

For a visual representation of the F-measure results across three datasets, please refer to Figure 8. The figure clearly demonstrates that our proposed algorithm consistently performs well across all three datasets, whereas other approaches exhibit varying levels of performance across these datasets.

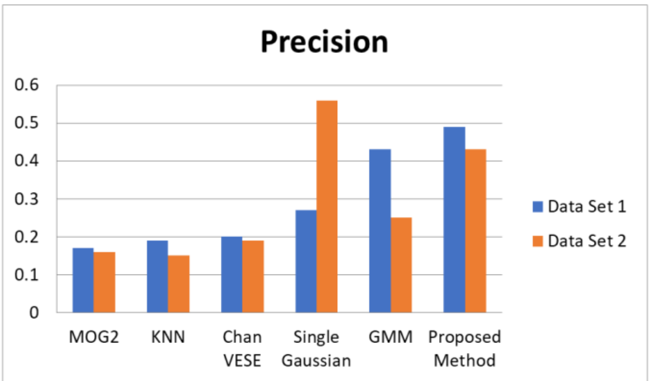

Fig 5- Comparison of precision value over different segmentation techniques

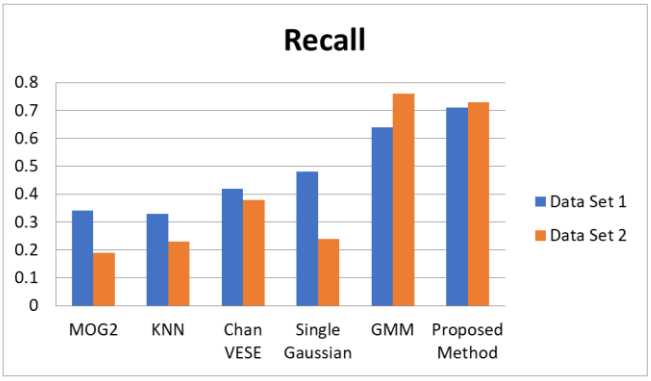

Fig 6- Comparison of recall value over different segmentation techniques

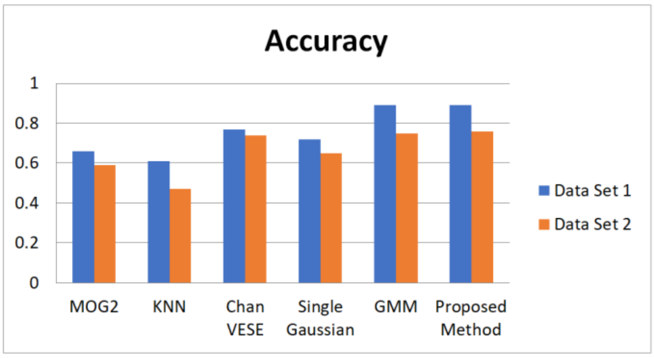

Fig 7- Comparison of accuracy value over different segmentation techniques

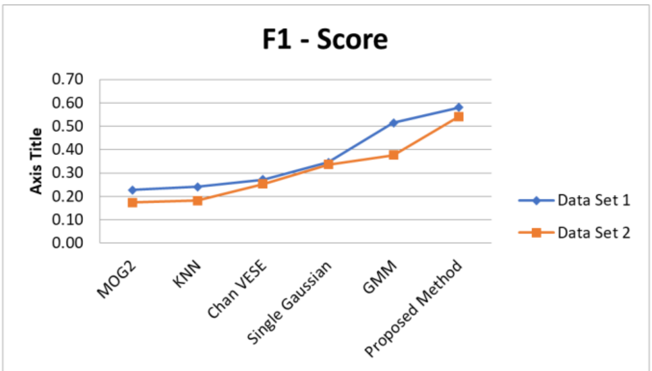

Fig 8- Comparison of F1-Score value over different segmentation techniques

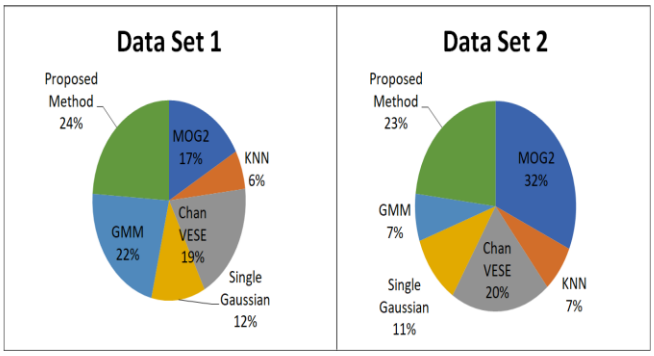

Fig 9- Comparison of SSIM over different segmentation techniques

7 Conclusion and Future Work

This article offers a comprehensive examination of five frequently used foreground detection algorithms that strive to achieve illumination invariance. Furthermore, we have presented an innovative approach for contemplation. Our primary objective was to evaluate the efficacy of these algorithms in detecting foreground objects under varying illumination circumstances, including both sudden and gradual changes. In order to objectively evaluate their performance, we employed various metrics including recall, precision, accuracy, structural similarity index (SSIM), and F1-score.

The results of our investigation provide clear evidence that the recently presented algorithm exhibits superior performance compared to all five alternative algorithms in all aspects. The experimental findings indicate that the Gaussian Mixture Model (GMM) exhibits effective adaptability to gradual shifts in illumination. Additionally, our suggested technique provides superior performance in scenarios characterized by rapid illumination changes, while still displaying satisfactory performance in cases involving gradual illumination variations. In addition, the technique we present demonstrates enhanced computational efficiency in comparison to the Gaussian Mixture Model (GMM), making it very suitable for real-time applications.

In subsequent investigations, given the increasing need for real-time applications across diverse domains such as autonomous vehicles, robotics, and medical imaging, future research endeavors may concentrate on customizing image segmentation methodologies to effectively address the unique demands of these applications. This task may entail the enhancement of segmentation algorithms to achieve efficient processing in contexts with limited resources and strict latency requirements.

The combination of different sensor inputs, such as RGB, depth, or infrared data, has the potential to enhance the precision and resilience of real-time picture segmentation algorithms. Further research might be conducted to explore fusion methodologies and assessment criteria that are tailored to the task of multi-modal segmentation.

Declarations

Conflict of Interest

On behalf of all the authors, corresponding author declares that there is no conflict of interest involved in conducting the study.

Funding

Authors declare that no funding was received for this research work.

References

[1] Piccardi M 2004 Background subtraction techniques: a review. *IEEE IntConfSyst Man Cybern* 4:3099–3104

[2] Benezeth Y, Jodoin Pierre-Marc, Emile Bruno, Laurent Helene and Rosenberger Christophe 2012 Comparative study of background subtraction algorithms. *J Electron Imaging* 19(3):12

[3] Ahmed Sumaya H, El-SayedKhaled M and ElhabianShireen Y 2008 Moving object detection in spatialdomain using background removal techniques-state-of-art. *Recent Pat ComputSci* 1(1):32–54

[4] Bouwmans Thierry 2014 Traditional and recent approaches in background modeling for foreground detection: an overview. *ComputSci Rev* 11(12):31–66

[5] Sen-ching SC and Chandrika K 2004 Robust techniques for background subtraction in urban traffic video. *Proceedings of the SPIE 5308: 881-892*

[6] Li L, Huang W, Irene YH Gu and Qi Tian 2003, November Foreground object detection from videos containing complex background. *ACM International Conference on Multimedia*, pp: 2–10

[7] Wahyono A and Filonenko Jo KH 2016 Unattended object identification for intelligent surveillance systems using sequence of dual background difference. *IEEE Trans IndInf* 12(6):2247–2255

[8] Wang K, Liu Y, Gou C and Wang FY 2016 A multi-view learning approach to foreground detection for traffic surveillance applications. *IEEE Trans VehTechnol* 65(6):4144–4158

[9] Cheng FC, Huang SC and Ruan SJ 2011 Illumination-sensitive background modeling approach for accurate moving object detection. *IEEE Trans Broadcast 57(4):794–801*

[10] Barnich O and Van Droogenbroeck M 2011 ViBe: a universal background subtraction algorithm for video sequences. *IEEE Trans Image Process* 20(6):1709–1724

[11] Manzanera A and Richefeu JC 2004, December a robust and computationally efficient motion detection algorithm based on background estimation. *Indian Conference on Computer Vision, Graphics and Image Processing*, pp: 46–51

[12] Lou J, Yang H, Hu W and Tan T 2002, January An illumination invariant change detection algorithm. *Asian Conference on Computer Vision*, pp: 13–18

[13] Holtzhausen PJ, Crnojevic V and Herbst BM 2015 An illumination invariant framework for real-time foreground detection. *J Real Time Image Process* 10(2):423–433

[14] Elharrouss O, AbbadA,Moujahid D and Tairi H 2018 Moving object detection zone using a block-based background model. *IET Comput Vision* 12(1):86–94

[15] Kim W and Kim Y 2016 Background subtraction using illumination-invariant structural complexity. *IEEE Signal Process Letters* 23(5):634–638

[16] Mahmoudpour S and Robust M Kim 2016 Foreground detection in sudden illumination change, *Electron. Letters*, vol.52, no.6, pp : 441–443

[17] Cheng FC, Huang SC and Ruan SJ 2011 Illumination-sensitive background modeling approach for accurate moving object detection, *IEEE Trans. Broadcasting*, 57(4):794–801

[18] Jin Ran, Han Xiaozhen and Yu Tongrui 2021 A Real-Time Image Semantic Segmentation Method Based on Multilabel Classification *Hindawi Mathematical Problems in Engineering* vol 2021, 13 pages .

[19] Zhou Tianfei , Porikli Fatih , Crandall David , Gool Luc Van and Wang Wenguan 2022 A Survey on Deep Learning Technique for Video Segmentation *IEEE Transactions On Pattern Analysis and Machine Intelligence* Volume: 45, pp: 7099 – 7122

[20] Robinson Andreas, JäremoLawin Felix, Danelljan Martin, Khan FahadShahbaz and Felsberg Michael 2020 Learning Fast and Robust Target Models for Video Object Segmentation *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*

[21] Kim Hyungjoon, Lee Jae Ho and Lee Suan 2023 A Hybrid Image Segmentation Method for Accurate Measurement of Urban Environments *Electronics*, 12(8), 1845.

[22] Karthikeyan Panjappagounder Rajamanickam and Periyasamy Sakthivel 2019 Entropy Based Illumination-Invariant Foreground Detection *IEICE TRANS. INF. & SYST.*, VOL.E102–D, NO.7 pp:1437-1437

[23] Holtzhausen P J , Crnojevic V and Herbst B M, 2015 An illumination invariant framework for real-time foreground detection *Journal of Real-Time Image Processing* volume 10, pp: 423–433

[24] Karthikeyan P R, Sakthivel P and Karthik T S, 2018 Comparative study of illumination-invariant foreground detection *The Journal of Supercomputing* volume 76, pp:2289–2301

[25] Hu Yuan-Ting, Huang Jia-Bin, and Schwing Alexander G. 2018 VideoMatch: Matching based

Video Object Segmentation *Computer Vision ECCV* , pp: 56–73

[26] Wang Kunfeng, Liu Yuqiang, Gou Chao, and Wang Fei-Yue 2016 A Multi-View Learning Approach to Foreground Detection for Traffic Surveillance Applications *IEEE Transactions on Vehicular Technology* 65(6) :4144 – 4158

[27] Chen Haiyan, Zhang Huaqing and Zhen Xiajun 2022 A hybrid active contour image segmentation model with robust to initial contour position *Multimedia Tools and Applications*, 82, 10833

[28] Wenjuan Ma and Feng Xu Retraction 2021 Underwater image segmentation based on computer vision and research on recognition algorithm *Arabian Journal of Geosciences*, 14:1-11

[29] Cigaroudy Ladan Sharafyan and Aghazadeh Nasser 2016 A multiphase segmentation method based on binary segmentation method for Gaussian noisy image *Signal, Image and Video Processing*, 11, pp:825–831