DOI: 10.24996/ijs.2023.64.4.38

ISSN: 0067-2904

New Approach of Generating Ground-Truth for Local Surveillance Dataset Tested with Benchmark Background Subtraction Models

Maryam A. Yasir [1]*, Yossra H. Ali [2]
[1]Departement of Computer Science, College of Science, University of Baghdad, Baghdad, Iraq
[2]Departement of Computer Science, University of Technology, Baghdad, Iraq

Received: 7/5/2022 Accepted: 23/8/2022 Published: 30/4/2023

Abstract

Background subtraction is the dominant approach in the domain of moving object detection. Lots of research has been done to design or improve background subtraction models. However, there are a few well-known and state-of-the-art models that can be applied as a benchmark. Generally, these models are applied to different dataset benchmarks. Most of the time, choosing an appropriate dataset is challenging due to the lack of dataset availability and the tedious process of creating ground-truth frames for the sake of quantitative evaluation. Therefore, in this article, we collected local video scenes of a street and river taken by a stationary camera, focusing on dynamic background challenges. We presented a new technique for creating ground-truth frames using modeling, composing, tracking, and rendering each frame. Eventually, we applied three promising algorithms used in this domain: GMM, KNN, and ViBe, to our local dataset. Results obtained by quantitative evaluations revealed the effectiveness of our new technique for generating the ground-truth scenes to be benchmarked with the original scenes using a number of statistical metrics.

Keywords: Video surveillance, Background subtraction, Moving object detection, Ground-truth, Evaluation metrics

طريقة جديدة لتوليد مشاهد الحقيقة الاساسية لمجموعة بيانات المراقبة المحلية واختبارها باستعمال الطرق المعيارية في استخراج الخلفية

مريم عبد الرحيم ياسر [1]*, يسرى حسين علي [2]
[1] قسم علوم الحاسوب, كلية العلوم, جامعة بغداد , بغداد , العراق
[2] قسم علوم الحاسوب, الجامعة التكنلوجية , بغداد , العراق

الخلاصة

استخراج خلفية المشهد هو الأسلوب السائد في مجال اكتشاف الأجسام المتحركة. تم إجراء الكثير من الأبحاث لتصميم أو تحسين خوارزميات استخراج الخلفية من المشهد. ومع ذلك ، هناك عدد قليل من الخوارزميات المعروفة والحديثة التي تم تطبيقها كمعيار. بشكل عام ، يتم تطبيق هذه الخوارزميات مجموعة مختلفة من البيانات المعيارية. في معظم الأوقات ، يعد اختيار مجموعة البيانات المناسبة أمرًا صعبًا نظرًا لعدم

*Email: maryam.a@sc.uobaghdad.edu.iq

توفر مجموعات البيانات عوضاً عن صعوبة انشاء مشاهد الحقيقة الأساسية (Ground–truth) من أجل التقييم الكمي.
لذلك ، في هذه المقالة ، قمنا بجمع مشاهد فيديو محلية لشارع ونهر تم التقاطها بواسطة كاميرا ثابتة مع التركيز على تحدي الخلفية الديناميكية. لقد قدمنا تقنية جديدة لإنشاء إطارات حقيقة أساسية باستعمال النمذجة والانشاء والتتبع ومن ثم عرض كل إطار. في النهاية طبقنا ثلاث خوارزميات واعدة مستعملة في هذا المجال ، GMM و KNN و ViBe على مجموعة البيانات المحلية الخاصة. وقد كشفت النتائج التي تم الحصول عليها من خلال التقييمات الكمية فاعلية أسلوبنا الجديد في إنشاء مشاهد الحقيقة الاساسية (Ground–truth) ليتم قياسها مع المشاهد الأصلية باستعمال عدد من المقاييس الإحصائية.

1. Introduction

Detection of moving objects is the initial and essential workflow step of the video surveillance system, where further steps could be taken for classifying or recognizing the moving objects. Background subtraction techniques have been used to detect moving objects along with optical flow and temporal differencing techniques [1], although background subtraction is the most well-known and still the dominant technique among others due to its easiness and high performance when implemented in a vast scope of video surveillance environments [2].

Normally, background subtraction techniques deal with different challenges according to the various environments. For example, challenges can be dynamic backgrounds, illumination changes, camera jitter, low frame rates, or other challenges [3].

Through the decades, many background subtraction models have been developed and introduced to tackle the background subtraction challenges. These models are classified into different categories by several surveys and review articles [4][5]. ***Basic models*** are the simplest models; they depend on a threshold to decide whether it's a foreground or background pixel, and it is more convenient to models with a single background distribution [6]. Mean models [7], median models [8], and histogram analysis models [9] are examples of the basic models. ***Filter models*** expect the background to depend on the intensity or orientation of the previous pixels [10], Wiener filter [11], Tchebychev filter [12], Correntropy filter [13], optical flow [14] and Kalman filter [15], which are examples of the single processing models (filter models). ***Mathematical models*** consist of two classes: statistical parametric models and statistical non-parametrical models. Gaussian Mixture Model (GMM) [16] is an example of the parametrical statistical models [17], while Visual Background extractor (ViBe) [18], Substance Sensitivity Segmenter (SuBSENSE), kernel density estimation (KDE) [19] , and fuzzy models [20] are examples of the non-parametrical statistical models[4].

Clustering models depend on the color intensity of the pixel to recognize whether a pixel belongs to the background or the foreground clusters. K-means [21], Codebook [22], and background reconstruction [23] are examples of clustering models.
Furthermore, ***machine learning models*** are the state-of-the-art models that encompass various techniques like support vector machines (SVM) [24], robust subspace tracking [25], reconstructive and discriminative subspace learning techniques [26][27], deep learning neural networks [28][29] and convolutional neural networks (CCN) [30], which have been broadly embraced due to the massive development of hardware processing power [31]. Fusion of several models and strategies is another approach that has been adopted in many articles for better performance. Real-time semantic background subtraction (RT-BSB) [32] is an example.

The background subtraction technique mainly consists of pipelined stages; four main stages; and two secondary stages. The main stages are: background initialization, background modeling, background maintenance, and background detection. The secondary stages are: the pre-processing and post-processing stages [33]. Generally, background subtraction models are applied to a benchmark dataset. Choosing a suitable dataset is quite challenging due to the lack of available datasets that provide a specific challenge to be tested or provide the ground-truth frames for quantitative evaluations. As a result, only a few datasets can be considered benchmarks. Change Direction Network (CDnet) 2012 and CDnet 2014 [3] are the most well-known benchmark datasets used in this domain for providing a variety of challenges along with supporting the ground-truth frames. Therefore, in this article, we shed light on how to collect a local dataset with dynamic background challenges using a stationary camera and propose a new technique for creating ground-truth frames for quantitative evaluation. We applied the most well-known and benchmarked background subtraction algorithms to our local dataset for qualitative and quantitative assessment. In the remainder of this paper, section 2 is devoted to elaborating on background subtraction models, while section 3 demonstrates the details of the local original scenes along with the new approach for ground-truth generation. This paper provides experimental results with both quantitative and qualitative evaluations, illustrated in tables and figures respectively in section 4. Eventually, section 5 presents the conclusion.

2. Background Subtraction Models

Hundreds of models in the domain of background subtraction for detecting a moving object have been developed over decades, but only a few models are considered a benchmark for delivering high performance and are thus used in well-known computer vision and image processing libraries and frameworks. The following are three eminent background subtraction algorithms that were applied in this article.

2.1 Gaussian Mixture Model (GMM)

The GMM was first introduced in 1999 by Stauffer and Grimson [16] as a breakthrough approach in the field of background modeling where each pixel is modeled as a mixture of Gaussian using a real-time approximation for updating the model [34]. A GMM is a statistical parametrical and unsupervised model using three main parameters: mean vectors, covariance matrices, and mixture weights from all component densities[35]. GMM is famous for its robust performance with respect to gradual illumination changes and dynamic background challenges. However, the GMM suffers from slow recovery and poor performance with unexpected lighting changes and background abnormal motions[36]. In addition, the efficiency of GMM is prone to being reduced due to the parametrical nature of the algorithm, like in selecting inaccurate parameters or the time consumed in selecting these parameters [37].

Many studies have been performed to enhance the GMM against different background subtraction challenges, like in Boosted Gaussian Mixture Model (BGMM), where the performance is boosted using a color space classifier and dynamic learning for updating the background model [38]. Improved Gaussian background modeling (GBM) is also an example of an enhanced model using wavelet denoising applied to a foreground object. This model performs better with respect to shadow and illumination change challenges [39]. GMM has been used frequently in the field of foreground detection and is still applied in many optimized versions[40]. The following are the general steps of the GMM algorithm:

Algorithm 1: GMM algorithm

Input: Initialize the **mean** μc, the **covariance matrix** ∑c and the **mixing coefficients** πc by some random values. (or other values)

Output: foreground mask of the moving object in the video scene.
1. Compute the **ϒc** values for all **k** // k= number of clusters
2. Recalculate all the parameters (μ**c**, ∑**c** *and* **πc**) using the current **ϒc** values.
3. Compute log-likelihood function.
4. Set some convergence criterion.
5. If the log-likelihood value converges to some value

(or if all the parameters converge to some values)
then stop,
else return to Step 2.

2.2 K-Nearest neighbor (KNN)

KNN is one of the well-known non-parametrical supervised algorithms where K denotes the number of neighbors to be considered in voting for the class detection. It was first introduced in the 1970s as a statistical model. Since then, KNN has been used in classification and regression problems [41]. The KNN is influenced by the concept of similar things being close to each other. In this context, the algorithm finds the closest k samples (neighbors) to the query using distance and then determines the query label with the same class label as the closest sample. The following are the general steps of the KNN algorithm:

Algorithm 2: KNN algorithm

Input: unknown example(query) **x**, the dataset **S** and the distance **d**
Output: foreground mask of the moving object in the video scene.
1. for(x',l') ϵ **S** do
   compute the distance d(x',x)

end for

2. Sort the |**S**| distance by increasing order.
3. Count the number of occurrences of each class lj among the K nearest neighbor.
4. Assign to **x** the most frequent class.

2.3 Visual Background extractor (ViBe)

ViBe was first introduced by Barnich and Droogenbroeck in 2009 as a background extraction algorithm, which uses a novel technique that considers the effect of a value in a multicolor space to be restricted to a local neighborhood [18]. In ViBe, classification is done by comparing a value to its closest values in the set of samples instead of using the probability density function (pdf) to update the background model for obtaining better results [41]. The ViBe model has been improved over time in various works, such as [42]. The following are the general steps of the ViBe algorithm [43]:

Algorithm 3: ViBe algorithm

Input: Given samples **N**
The values of a sample **x v(x)**
The distance threshold **R**
Classification threshold **λ**
Subsampling rate **t**
Output: foreground mask of the moving object in the video scene.
1. **for** each pixel **x do**
2. **while** neighbors < λ **and** index < N **do**
3. Calculate Euclidean distance between v(x) and v(i)
4. **if** distance < R **then**
neighbors= neighbors + 1
end if
5. index=index + 1
end while
6. **if** neighbors ≥ λ **then**
7. Store that pixel ϵ background
8. Update current pixel background model with probability 1/t
9. Update neighboring pixel background model with probability 1/t
else
10. Store that pixel ϵ foreground
end if
end for

3. Local Video Scenes

In addition to the global benchmark dataset used by researchers around the world, in this article, we attempt to collect local scenes to be tested by benchmark background subtraction models. The utilization of a locally gathered dataset along with the global benchmark datasets is of importance as it adds the value of testing algorithms on local scenarios captured using locally widespread technology, as the use of Closed-Circuit Television (CCTV) cameras has increased significantly with the rise of security demands. For this purpose, we captured three videos with different dynamic background challenges and distinct moving objects to be detected. The video scenes were taken by a common stationary CCTV camera with a frame rate of 30 fps. Furthermore, we used a new technique to generate the ground-truth frames for each input frame. Table 1 contains details about the local dataset videos.

3.1 Original Video Scenes

In this section, we present the details with sample frames from each original video captured as follows:

3.1.1 Street Scenes

In the street scenes, two different moving vehicles are captured in both right and left directions. The dynamic background challenge is the moving palm frond in the upper right corner of the scene frame. The following Figure 1 depicts frame samples of street videos. The scenes (a) and (c) are before the emergence of the target moving vehicle, while (b) and (d) are the scenes with the target moving vehicle.

(a) **(b)** **(c)** **(d)**

Figure 1: Frame samples before and after the emergence of the target moving vehicle

3.1.2 Tigris River scene

The Tigris river scene is captured from the riverbank, capturing the river and the moving ferry. In this scenario, the moving water surface is the dynamic background challenge. Figure 2 depicts the frame samples, where (a) is the original scene before the emergence of the moving ferry, and (b) is the frame containing the target moving ferry.

(a) **(b)**

Figure 2: Frame samples before and after the emergence of the target moving ferry

Table 1: Dynamic background videos from local Dataset

Videos	Dynamic challenge description	Total frames	Region of interest frames
Van	Moving palm frond in the upper right corner	346	188 - 313
Car	Moving palm frond in the upper right corner	491	254 - 383
Ferry	Moving water surface	3811	235 - 3412

3.2 Ground-truth scenes

In this section, we present the steps of generating the correspondent ground-truth frames for each original frame in our video scenes.

3.2.1 Modeling

Generally, 3D modeling is the process of creating edges, polygons, and vertices for an object through representing the mathematical coordinates of the object's shell in a virtual three-dimensional space using specific applications [44]. Furthermore, there are various approaches to generating a 3D model: the manual approach, in which the designer is responsible for creating the model from scratch; the procedural approach, in which the designer follows an algorithmic path to build the 3D model; and scanning, in which the designer scans the real object to create the 3D model [45]. In this work, three moving objects were modeled manually using the Autodesk 3DS Max application. The following Figure 3 depicts the modeling perspectives for the three moving objects (van, car, and ferry).

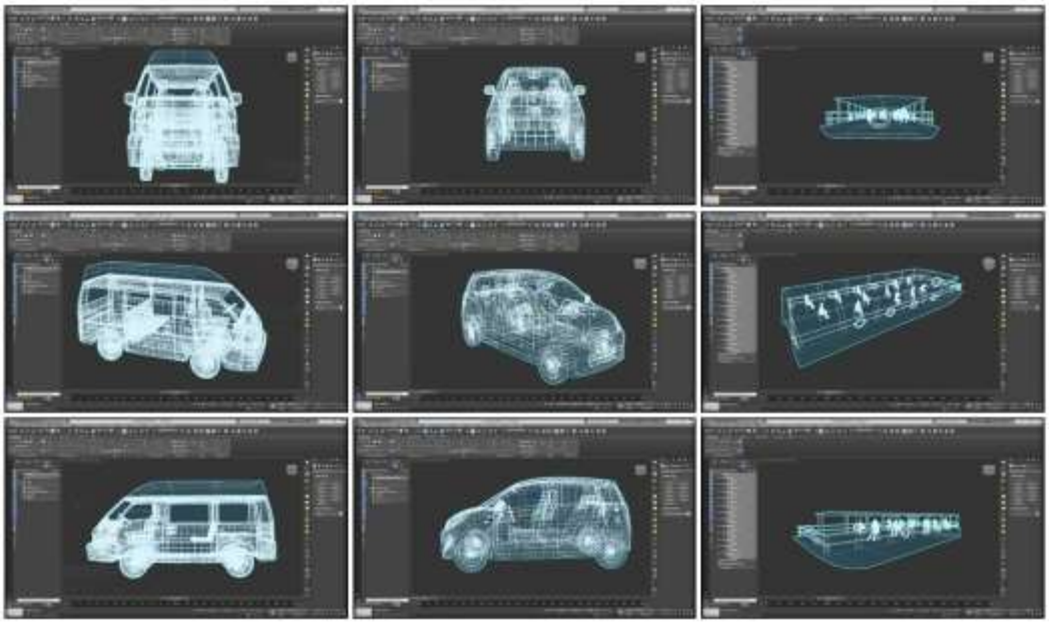


Figure 3: Modeling perspectives

3.2.2 Composing

In general, this step involves choosing the suitable camera lenses, controlling the camera angle, posing and blocking if the object is static. It could also involve some specific composition techniques [46]. In this work, we used the real dimensions of the objects and the real camera distance with lenses similar to the CCTV camera lens. The following Figure 4 depicts the composition screenshots.

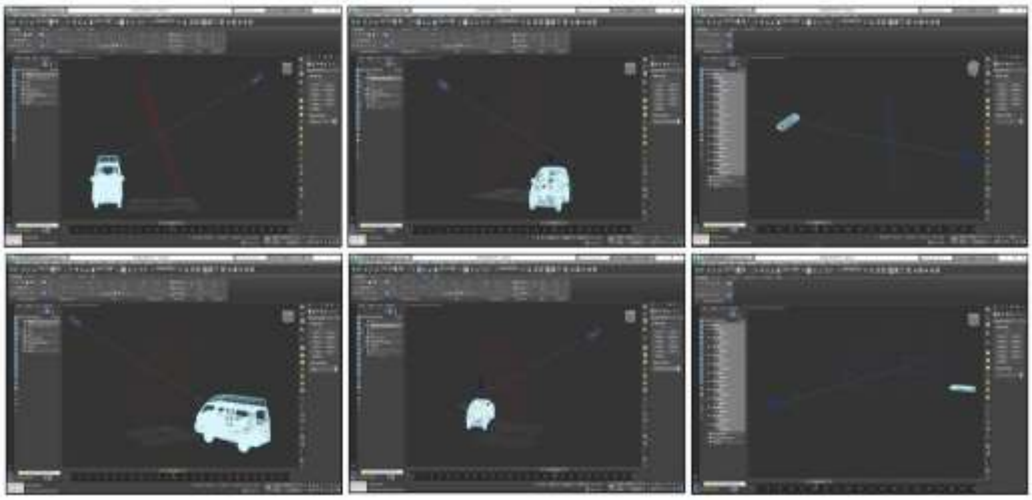


Figure 4: Composing the camera to the objects

3.2.3 Tracking

Animating the moving object is the approach we followed to imitate and track the movement of our objects in the original scene. Figure 5 shows the screenshots of the tracking of the moving objects.

(a) **(b)** **(c)**

Figure 5: Tracking (a) van (b) car (c) ferry

3.2.4 Rendering

Rendering is the final step where the designer generates a moving object based on the 3D scene. In this work, the negative mode is used to create the scene. Figure 6 shows the rendering process.

(a) **(b)** **(c)**

Figure 6: Rendering on (a) van (b) car and (c) ferry

4. Experiment and evaluation

In this article, we compare the foreground detection using the forementioned background subtraction models (GMM, KNN, and ViBe). The models are tested using our local dataset, considering the dynamic background challenge.

In these experiments, we are comparing the background subtraction image with the correspondent ground-truth image to evaluate the performance of each method with respect to quantitative evaluation metrics at the pixel level, and the background subtraction method classifies the pixels into background or foreground. Seven metrics are used for the performance evaluation, as illustrated in the following equations [47].

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall(sensitivity) = \frac{TP}{TP + FN} \tag{3}$$

$$F - Measures(F1) = \frac{2 * Precision * Recall}{Precision + Recall} \tag{4}$$

$$FPR = \frac{FP}{FP + TN} \tag{5}$$

$$FNR = \frac{FN}{TP + FN} \tag{6}$$

$$Error\ rate(PWC) = \frac{FP + FN}{TP + FN + TN + FP} * 100 \tag{7}$$

Where, **TP** is the number of foreground pixels correctly classified, **TN** is the number of background pixels correctly classified, **FP** is the number of background pixels incorrectly classified as foreground pixels, and **FN** is the number of foreground pixels incorrectly classified as background pixels.

Accuracy indicates the correct classification for a pixel, whether it is a foreground or a background pixel**.** ***Precision*** indicates the proportion of truly detected foreground pixels to the number of all pixels classified as foreground pixels; ***recall*** indicates the number of pixels that are correctly classified as foreground of all the foreground pixels; and the ***F-measure*** is the harmonic mean of recall and precision. On the other hand, we have the metrics: the False Positive Rate (***FPR)*** is the number of background pixels that are misclassified as foreground pixels, the False Negative Rate (***FNR)*** is the number of foreground pixels that are misclassified as background pixels, and percentage weight loss (***PWC)*** indicates the error rate, which is the percentage of misclassified pixels to the original pixels [4].

Normally, we measure relevance by recall and precision. A low recall is an indication of over segmentation of the foreground objects, where a low precision is an indication of under segmentation of the foreground objects. High F-measures are an indication of a robust background subtraction algorithm, and the lower the FPR, FNR, and PWC, the better the performance.

In Tables 2, 3, and 4, the analytical metrics results of applying the GMM, KNN, and ViBe models respectively to our local dataset videos are illustrated. The highest F1 is highlighted in bold. Table 5 illustrates the average performance metrics of GMM, KNN, and ViBe on the local dataset.

Table 2: Performance metrics of GMM on local dataset

Video	Accuracy	Precession	Recall	F1	FPR	FNR	PWC
Van	0.985	0.2482	0.6351	**0.6445**	0.0092	0.3649	1.4952
Car	0.9575	0.1494	0.7303	0.6309	0.0414	0.2697	4.252
Ferry	0.9942	0.2299	0.3861	0.3088	0.0039	0.6139	0.584

Table 3: Performance metrics of KNN on local dataset

Video	Accuracy	Precession	Recall	F1	FPR	FNR	PWC
Van	0.9774	0.2619	0.5862	0.6294	0.0157	0.4138	2.2598
Car	0.9830	0.1941	0.6436	**0.6555**	0.0152	0.3564	1.7045
Ferry	0.9937	0.3241	0.8654	0.5255	0.006	0.1346	0.6330

Table 4: Performance metrics of ViBe on local dataset

Video	Accuracy	Precession	Recall	F1	FPR	FNR	PWC
Van	0.9853	0.2795	0.4164	**0.5383**	0.0033	0.5836	1.4717
Car	0.9935	0.2806	0.3962	0.5182	0.0015	0.6038	0.6508
Ferry	0.9958	0.4102	0.3346	0.376	0.0016	0.6654	0.4202

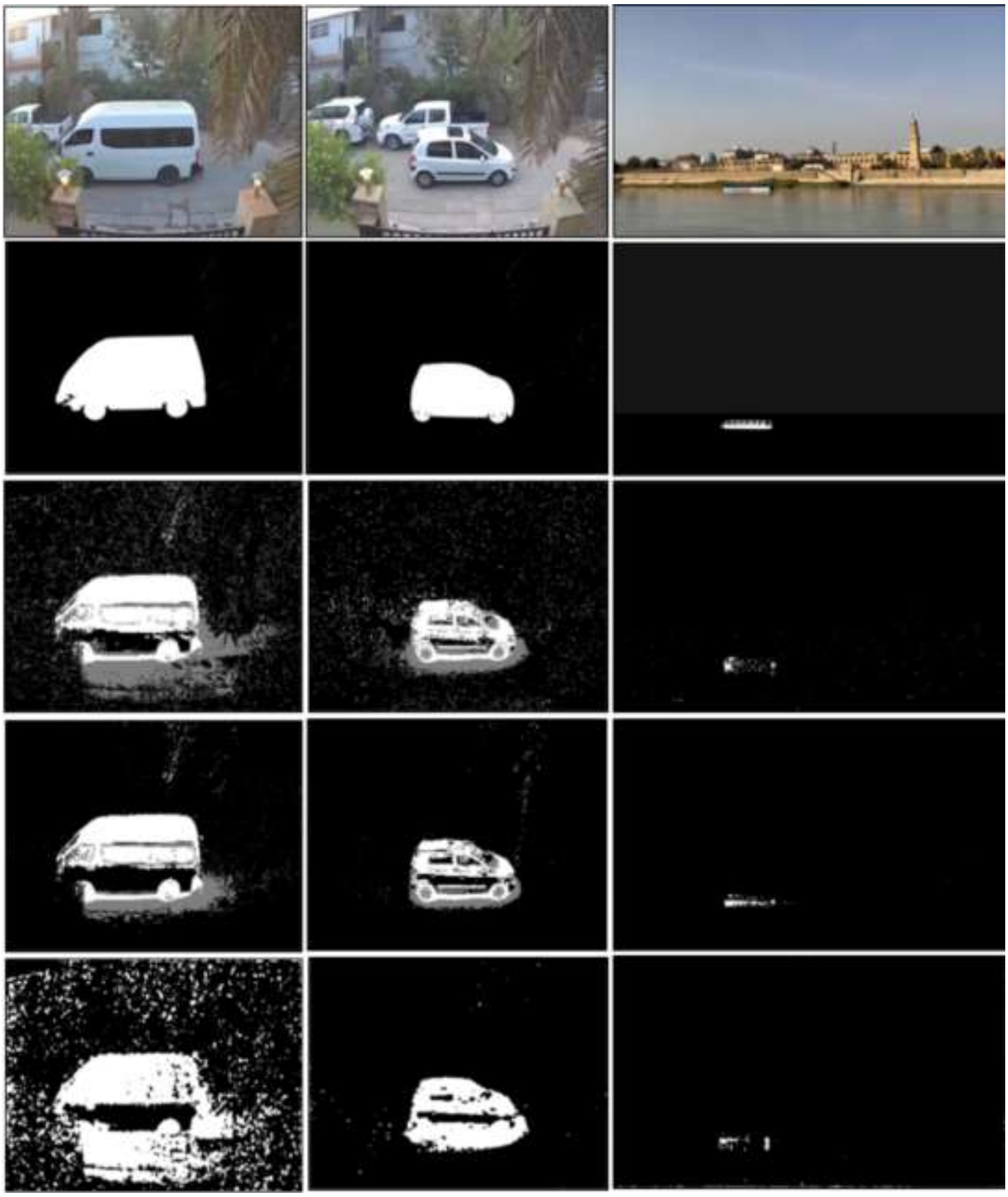
Table 5: Average performance metrics of GMM, KNN and ViBe on local dataset

Model	Accuracy	Precession	Recall	F1	FPR	FNR	PWC
GMM	0.979	0.209	0.584	0.528	0.018	0.416	2.110
KNN	0.985	0.260	0.698	**0.603**	0.012	0.302	1.532
ViBe	0.992	0.323	0.382	0.478	0.002	0.618	0.848

Overall, the forementioned comparison results of each model on each video from the local datasets have been made, specifically on the F-measure metric results. This is due to the fact that it is the harmonic mean of recall and precision, and the F-measure is the most important metric to be considered for evaluating the overall robustness of the background subtraction model.

The overall results of all models show that the best performance in F-measure is achieved more frequently on the car and van videos. This generally indicates that models behave similarly to the environmental video challenges even though each model produces its own performance results.
What can be noticed from Table 5 is that KNN on average outperforms the other two models when applied to the local dataset.

Moreover, Figure 7 depicts the visual comparison of the foreground results of applying the three models to the local dataset. Where (a) is the original scene, (b) is the ground-truth created for this dataset, (c) is the foreground mask created by the GMM model, (d) is the foreground mask created by the KNN model, and (e) is the foreground mask created by the ViBe model.



(a) (b) (c)

Figure 7: Comparison of foreground detection results (a) original scenes (b) ground-truths (c) GMM subtraction (d) KNN subtraction (e) ViBe subtraction

5. Conclusion

In this article, we recorded three local videos with dynamic background challenges in an attempt to prepare a local dataset. We proposed a new technique for creating ground truth for our local dataset by utilizing the concept of 3D modeling. A ground-truth is created for each original frame to be employed for quantitative evaluation. Benchmark algorithms (GMM, KNN, and ViBe) were applied to the local dataset and both quantitative and qualitative assessments are presented. Qualitative evaluations are illustrated in the screenshot figures depicting the visual assessment of the subtracted mask resulted from each algorithm. On the other hand, different evaluation metrics have been employed for quantitative evaluation. Our results showed the efficiency of the proposed ground-truth generation technique in creating suitable input for benchmark algorithms, thus allowing the developers of practical computer vision software targeting the local environment to test their solutions on local scenes.

References

[1] A. A. Karim, “Construction of a Robust Background Model for Moving Object Detection in Video Sequence”, Iraqi Journal of Science, vol. 59, no. 2B, pp. 969–979, May 2018.

[2] D. Zeng, M. Zhu, and A. Kuijper, “Combining background subtraction algorithms with convolutional neural network,” *J. Electron. Imaging*, vol. 28, no. 1, p. 13011, 2019.

[3] Y. Wang, et al., “CDnet 2014: An expanded change detection benchmark dataset,” in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 387–394, 2014.

[4] M. Yasir and Y. Hussain Ali, "Review on Real Time Background Extraction: Models, Applications, Environments, Challenges and Evaluation Approaches," International Journal of Online and Biomedical Engineering (iJOE), vol 17, 2021. doi: 10.3991/ijoe.v17i02.18013.

[5] T. Bouwmans and B. Garcia-Garcia, “Background subtraction in real applications: Challenges, current models and future directions,” *arXiv Prepr. arXiv1901.03577*, 2020.

[6] S. Lee and D. Kim, “Background subtraction using the factored 3-way restricted Boltzmann machines,” *arXiv Prepr. arXiv1802.01522*, 2018.

[7] P. W. Power and J. A. Schoonees, “Understanding background mixture models for foreground segmentation,” in *Proceedings image and vision computing New Zealand*, vol. 2002, pp. 10–11, 2002.

[8] M. Hadiuzzaman et al., “Pixel-based heterogeneous traffic measurement considering shadow and illumination variation,” *Signal, Image Video Process.*, vol. 11, no. 7, pp. 1245–1252, 2017.

[9] Z. Zivkovic and F. Van Der Heijden, “Efficient adaptive density estimation per image pixel for the task of background subtraction,” *Pattern Recognit. Lett.*, vol. 27, no. 7, pp. 773–780, 2006.

[10] K. Goyal and J. Singhai, “Review of background subtraction methods using Gaussian mixture model for video surveillance systems,” *Artif. Intell. Rev.*, vol. 50, no. 2, pp. 241–259, 2018.

[11] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, “Wallflower: Principles and practice of background maintenance,” in *Proceedings of the seventh IEEE international conference on computer vision*, vol. 1, pp. 255–261, 1999.

[12] R. Chang, T. Gandhi, and M. M. Trivedi, “Vision modules for a multi-sensory bridge monitoring approach,” in *Proceedings. The 7th International IEEE Conference on Intelligent Transportation Systems (IEEE Cat. No. 04TH8749)*, pp. 971–976, 2004.

[13] G. T. Cinar and J. C. Príncipe, "Adaptive background estimation using an information theoretic cost for hidden state estimation," The 2011 International Joint Conference on Neural Networks, pp. 489-494, 2011. doi: 10.1109/IJCNN.2011.6033261.

[14] X. Yu, X. Chen, and H. Zhang, “Accurate motion detection in dynamic scenes based on ego-motion estimation and optical flow segmentation combined method,” in *2011 Symposium on Photonics and Optoelectronics (SOPO)*, 2011, pp. 1–4.

[15] F. Lei and X. Zhao, “Adaptive background estimation of underwater using Kalman-Filtering,” in *2010 3rd International Congress on Image and Signal Processing*, 2010, vol. 1, pp. 64–67.

[16] C. Stauffer and W. E. L. Grimson, “Adaptive background mixture models for real-time tracking,” in *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, 1999, vol. 2, pp. 246–252.

[17] P.-L. St-Charles, G.-A. Bilodeau, and R. Bergevin, “Subsense: A universal change detection method with local adaptive sensitivity,” *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 359–373, 2014.

[18] O. Barnich and M. Van Droogenbroeck, “ViBe: a powerful random technique to estimate the background in video sequences,” in *2009 IEEE international conference on acoustics, speech and signal processing*, 2009, pp. 945–948.

[19] M. Hua, Y. Li, and Y. Luo, “Robust Background Modeling with Kernel Density Estimation.,” *Int. J. Online Eng.*, vol. 11, no. 8, 2015.

[20] F. El Baf, T. Bouwmans, and B. Vachon, “Type-2 fuzzy mixture of Gaussians model: application to background modeling,” in *International Symposium on Visual Computing*, 2008, pp. 772–781.

[21] D. Butler, S. Sridharan, and V. M. J. Bove, “Real-time adaptive background segmentation,” in

2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03)., 2003, vol. 3, pp. III--349.
[22] L. Wang and C. Pan, "Effective multi-resolution background subtraction," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 909–912.
[23] M. Xiao, C. Han, and X. Kang, "A background reconstruction for dynamic scenes," in *2006 9th International Conference on Information Fusion*, 2006, pp. 1–7.
[24] I. Junejo, A. Bhutta, and H. Foroosh, "Dynamic scene modeling for object detection using single-class SVM," in *Proc. of IEEE International Conference on Image Processing (ICIP)*, 2010, vol. 1, pp. 1541–1544.
[25] N. Vaswani, T. Bouwmans, S. Javed, and P. Narayanamurthy, "Robust subspace learning: Robust PCA, robust subspace tracking, and robust subspace recovery," *IEEE Signal Process. Mag.*, vol. 35, no. 4, pp. 32–55, 2018.
[26] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *J. ACM*, vol. 58, no. 3, pp. 1–37, 2011.
[27] D. Farcas, C. Marghes, and T. Bouwmans, "Background subtraction via incremental maximum margin criterion: a discriminative subspace approach," *Mach. Vis. Appl.*, vol. 23, no. 6, pp. 1083–1101, 2012.
[28] T. Minematsu, A. Shimada, H. Uchiyama, and R. Taniguchi, "Analytics of deep neural network-based background subtraction," *J. Imaging*, vol. 4, no. 6, p. 78, 2018.
[29] L. Maddalena and A. Petrosino, "Self-organizing background subtraction using color and depth data," *Multimed. Tools Appl.*, vol. 78, no. 9, pp. 11927–11948, 2019.
[30] M. Braham and M. Van Droogenbroeck, "Deep background subtraction with scene-specific convolutional neural networks," in *2016 international conference on systems, signals and image processing (IWSSIP)*, 2016, pp. 1–4.
[31] Bouwmans, Thierry et al. "Deep neural network concepts for background subtraction:A systematic review and comparative evaluation," Neural networks : the official journal of the International Neural Network Society vol. 117 (2019): 8-66. doi:10.1016/j.neunet.2019.04.024
[32] A. Cioppa, M. Van Droogenbroeck, and M. Braham, "Real-time semantic background subtraction," in *2020 IEEE International Conference on Image Processing (ICIP)*, 2020, pp. 3214–3218.
[33] Y. H. . Ali and M. R. Mohammed, "Background modeling in video surveillance by using parallel computing", Iraqi Journal of Science, vol. 58, no. 3B, pp. 1516–1522, Dec. 2021.
[34] C. Lin, B. Yan, and W. Tan, "Foreground detection in surveillance video with fully convolutional semantic network," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 4118–4122.
[35] D. A. Reynolds, "Gaussian Mixture Models.," *Encycl. biometrics*, vol. 741, pp. 659–663, 2009.
[36] T. Bouwmans, F. El Baf, and B. Vachon, "Statistical background modeling for foreground detection: A survey," in *Handbook of pattern recognition and computer vision*, World Scientific, 2010, pp. 181–199.
[37] B. White and M. Shah, "Automatically tuning background subtraction parameters using particle swarm optimization," in *2007 IEEE International Conference on Multimedia and Expo*, 2007, pp. 1826–1829.
[38] I. Martins, P. Carvalho, L. Corte-Real, and J. L. Alba-Castro, "BMOG: boosted Gaussian mixture model with controlled complexity," in *Iberian Conference on Pattern Recognition and Image Analysis*, 2017, pp. 50–57.
[39] X. Lu et al., "Improved background subtraction method for detecting moving objects based on GMM," *IEEJ Trans. Electr. Electron. Eng.*, vol. 13, no. 11, pp. 1540–1550, 2018.
[40] M. A. Yasir and Y. H. Ali, "Comparative Analysis of GMM, KNN, and ViBe Background Subtraction Algorithms Applied in Dynamic Background Scenes of Video Surveillance System," *Eng. Technol. J.*, vol. 40, no. 4, pp. 617–626, 2022.
[41] 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, 2010.
[42] L. F. Huang et al., "Block background subtraction method based on ViBe," in *Applied Mechanics and Materials*, 2014, vol. 556, pp. 3549–3552.

[43] X. Zhou et al., “Improving video segmentation by fusing depth cues and the visual background extractor (ViBe) algorithm,” *Sensors*, vol. 17, no. 5, p. 1177, 2017.
[44] “3D modeling,” *Siemens Digital Industries Software* .
[45] Carmel, “How to 3D scan with a phone: Here are our best tips,” *Sculpteo*. Sculpteo, 2022, accessed on 22-Aug-2022 at https://www.sculpteo.com/en/3d-learning-hub/best-articles-about-3d-printing/3d-scan-smartphone/.
[46] “3D photography,” *3D Modeling Resources*. 2019.
[47] M. A. Yasir and Y. H. Ali, “Dynamic Background Subtraction in Video Surveillance Using Color-Histogram and Fuzzy C-Means Algorithm with CosineSimilarity,” *Int. J. Online Biomed. Eng.*, vol. 18, no. 09, 2022.