Iraqi Journal of Science, 2024, Vol. xx, No. x, pp: xx
DOI: 10.24996/ijs.2024.65.7.46

ISSN: 0067-2904

Frame-Based Change Detection Using Histogram and Threshold to Separate Moving Objects from Dynamic Background

Hala A. Jasim , Loay E. George, Mohammed I. Abd-almajied

[1]Department of computer science, College of Science, University of Baghdad, Baghdad, Iraq
[2] University of Information Technology and Communication, Baghdad, Iraq
[3]Department of Remote Sensing & GIS, College of Science, University of Baghdad, Baghdad, Iraq

Received: 22/**6**/2023 Accepted: 5/11/2023 Published: xx

Abstract

Detecting and subtracting the Motion objects from backgrounds is one of the most important areas. The development of cameras and their widespread use in most areas of security, surveillance, and others made face this problem. The difficulty of this area is unstable in the classification of the pixels (foreground or background). This paper proposed a suggested background subtraction algorithm based on the histogram. The classification threshold is adaptively calculated according to many tests. The performance of the proposed algorithms was compared with state-of-the-art methods in complex dynamic scenes.

Keywords: Change Detection, Dynamic Background Separation, Moving Object Detection, Ground truth, proposed system.

اكتشاف التغيير القائم على الاطار بأستخدام الرسم البياني والعتبة لفصل الاجسام المتحركة عن الخلفية الديناميكية

هلا عبد السلام جاسم [1] ,لؤي ادور جورج[2], محمد أسماعيل عبد المجيد[3]

[1]قسم علوم الحاسبات, كلية العلوم, جامعة بغداد, بغداد, العراق.
[2]جامعة تكنلوجيا المعلومات والاتصالات, بغداد, العراق.
[3]قسم التحسس النائي ونظم المعلومات الجغرافية, كلية العلوم, جامعة بغداد, بغداد, العراق.

الخلاصة:

اكتشاف الاجسام المتحركة وطرحها من الخلفيات هيه واحدة من اهم الاهتمامات. تطور الكامرات وانتشار استخداماتها في المجالات الامنية, المراقبة و مجالات اخرى ادى الى مواجهه هكذا موضوع. يكمن صعوبة هكذا مجال كونه غير مستقر باتجاه تصنيف البكسلات (الامامية او الخلفية). في هذا البحث, خوارزمية مقترحة لطرح الخلفية تم اقتراحها بالاعتماد على الرسم البياني. تصنيف العتبة تم حسابها بشكل متكيف طبقا للعديد من الاختبارات. ان اداء الخوارزمية المقترحة تم مقارنتها مع احدث الاساليب ذات مشاهد لخلفيات ديناميكية معقدة.

*Email: Mohammed.ismael@sc.uobaghdad.edu.iq

1. Introduction

Motion Detection separates the image's background from the foreground or moves objects across the selected camera view or video sequences (1). The process of background detection from moving objects depends on the pixel movements or pixel-changing values. Using the pixel value to detect if the point is still in the same location or changing in the next frames (2). Detecting dynamic background depends on the same principle, but the pixel values change will be less than moving objects, so the threshold is used to detect the changing types (3).

When considering those video pixels as a random variable, a popular method used to determine between foreground or background pixels is by adopting a statistical method; a tunable threshold was established for the background model by using the video for checking the pixel that belongs to the foreground or background. Displaying video pixels as a random variable and determining whether a pixel belongs to the foreground or background is adopted by utilizing many methods like statistical methods, creating a background image within a video view setting, updating the minimum threshold of change, and determining whether the pixel considers as foreground or background one. Research has been done in this field. Reynolds, Douglas A. In 2009, the Gaussian mixture model (GMM), this method used multiple normal distributions to distinguish the pixel changes (4). Barnich and Droogenbroeck 2009 proposed the Visual Background Extractor (ViBe) technique; this method uses the distribution of the neighborhood pixel consistency on the first frame to produce a background model (5). Based on (ViBe, Hofmann, et al., 2012) proposed a pixel-based adaptive segmented (PBAS) and background image creation model. Then, the foreground or background decision depends on a predefined threshold (6). Lee et al. 2012 proposed a nonparametric modeling technique based on the kernel density function (KDE) to model the pixel distributions based on the distribution of existing pixel samples (7). Tianming Yu et al., 2019, propose a fuzzy approach (fuzzy histogram) for background subtraction. This approach is called fuzzy c-means clustering and the fuzzy nearness degree (FCFN. They used the distribution of the fuzzy nearness degree to calculate the segmentation threshold for each of the individual pixels (2).

Sung In Cho et al. 2019 proposed a method to automatically identify the inconstant in the content of the moving image based on the histogram shape-based scene for frame rate up-conversion, in which a global and local scene change must be detected. The threshold value used in this method was suggested to be a fixed value to minimize the computation effect regardless of the image characteristics (8). Fakhri A. Khan et al. In 2020, a method for run-time threshold value defining the subtraction of the image's background was proposed by utilizing histogram peaks and valleys of the motion to show the moving objects' slow and high motion areas (9).

Mohamed M. Ismail et al. In 2020, a system of image-based fall detection. The system used the HOG feature to encode the visual property of the human body in the video frames. A classification model was acquired depending on the resulting feature vectors. The decisions of typical individual classifiers depended on the majority vote. Namely, the KNN, SVM algorithms, and Naïve Bayes were used as single learners (10). Ho Sub Lee and Sung In Cho, in 2022, propose a technique that utilizes a scene-change detection that analyzes the histograms method's luminance level for frame rate up-conversion (FRUC). This technique obtained each region's statistical average luminance value from the histogram generation (7). Kailai Sun et al. 2022 used the camera for the detection and estimation system; the developed idea was to filter the non-occupied frames and afford the switching value. An object entering or leaving a view was considered motion detection. The Fully Convolutional Head Detector

(FCHD) was utilized to discover indoor human heads. Lastly, defining the occupancy of numbers of information based on the previous results (11).

2. The Proposed Method

The algorithm of this research is to find the most redundant pixel value and consider it the fixed part. In the same way, less redundant values can be considered as moving objects and can be classified as dynamic background or moving objects. This algorithm is easy to understand and produces encouraging results. A flowchart of this algorithm is shown in Figure (1).

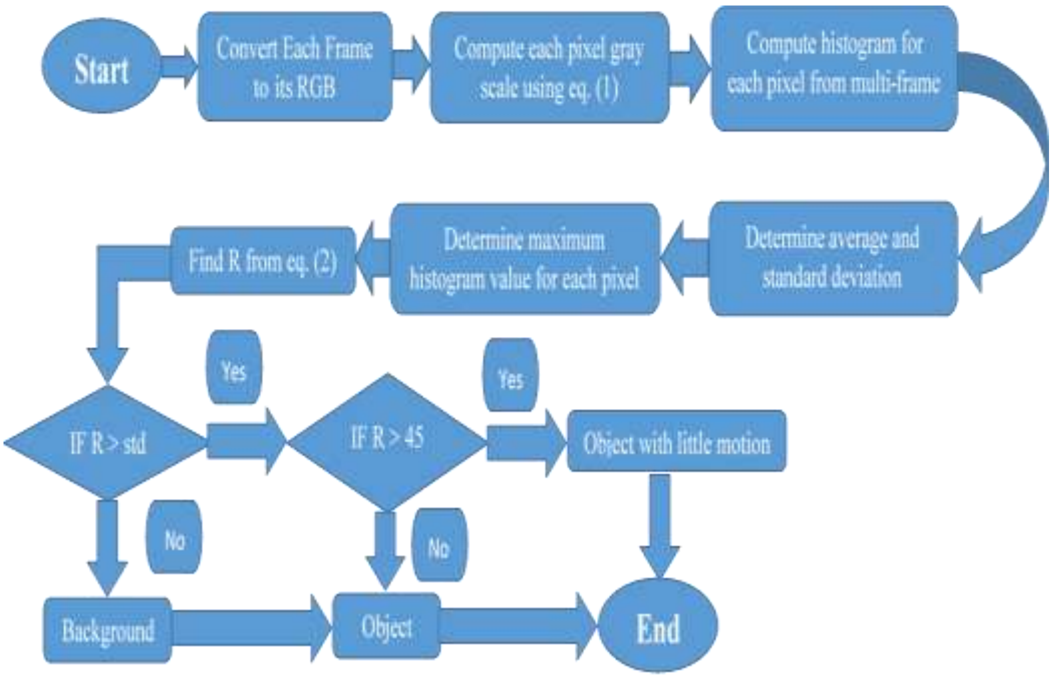


Figure 1: Flow-chart of the algorithm

2.1 Preprocessing

The Proposed System imports image frames; each input frame contains RGB layers (red, green, blue). The first step of work is converting the frame image from RGB color to gray-scaled space with only a single layer. This part of the work was to decrease the computation period and amount of memory required for processing each layer. To convert the image to Gray Scale, equation (1) was used (12):

$$P_{index}(\mathrm{X,Y}) = 0.51\ \mathrm{R} + 0.38\ \mathrm{G} + 0.1\ \mathrm{B} \tag{1}$$

Where

P_{index} is the index of pixel

R, G, and B are the red, green, and blue of pixel

2.2 Histogram

A histogram of all frames was calculated according to each pixel separately (13). The maximum value of each pixel was established in this histogram's output (14).

2.3 Standard Division

The proposed system suggested calculation of the Standard Division (STD) of the image and the average value of the processed frame, as seen in the following equations (15-17):

$$Std(X,Y) = Alpha \ * \sqrt[2]{\left(\frac{sm2}{no.Frames} \ - V \times V\right)} \quad (2)$$

Where

Alpha is a predefined value equal to 0.75

$$Sum2 = Sum2 + \ V \times i \quad (3)$$

$$V = hist(i) \times i \quad (4)$$

The average value is:

$$\text{Avg. Sum1} = \frac{\text{Sum1}}{\text{no. Frames}} \quad (5)$$

Where

$$Sum1 = Sum1 + V \quad (6)$$

Finally, an absolute difference exists between the Mod or final image and the frame images. Equation 7 will be used to determine this value (18):

$$R = |\ Frame(I,X,Y) - \ ModImg(X,Y)\ | \quad (7)$$

2.4 Point Separation

According to equation (7) previously mentioned, the proposed system will classify the point into three categories:

1. Objects
2. Dynamic Background.
3. Fixed Background.

For the first one, the object image was detected as a moving object; the other two types will be classified as background and ignored. The classification was done depending on the R-value and the predefined threshold value. If the std (X, Y) of the frame image is more significant than R and the suggested threshold, it will be classified as an object. If the std (X, Y) is less than R but more significant than the predefined threshold, it will be considered a dynamic background; otherwise, it will be considered a fixed background. Table (1) shows the effect of the threshold on the detection of objects and backgrounds of the resulting image.

Table 1: Threshold effect on detected objects and background

Image	**Threshold**	**Image**	**Threshold**
	30		35
	40		45
	50		55
	60		65
	70		75
	80		

If the threshold value is between 30 and 60, the sensitivity of the detection of motion will be increased, while if it is between 65 and 80, the detected motion will be less sensitive, as shown in Table (1). The objects that were detected in the dynamic background motion will be ignored. The proposed system's best threshold value is 80 (19).

Since monitoring systems focus the cameras on a fixed and specific place, in this proposed paper, the threshold value should be set as high in continuously moving places like trees or water. For example, the value of the seas and trees should be chosen 80 or more so that the large moving objects can be identified while the small moving objects, like fallen leaves or the movement of the waves, can be ignored. In certain areas, such as the one presented in the image illustrated in Table 1, which contains large moving objects such as cars and small moving objects such as humans, the threshold value must be chosen at a lower value for the upper part of the view and higher value in the lower part, to avoid identifying the movement of the wave. Two cameras and systems should be used to move objects to identify them within the cameras' scopes.

4. Performance Metrix

To evaluate the performance of the proposed system, each pixel will be compared with the famous GroundTruth Method (GroundTruth, which an expert gives, and the system result, which was the outcome of the suggested system in this article). The result will be one of the following four classes (20).

1. True Positive (TP): The first class (TP) refers to the case of the two systems correctly predicted moving objects; both systems return Yes.
2. True Negatives (TN): The second class (TN) this class refers to the point that was classified correctly in both systems as Fixed background; both systems return No in this case.
3. False Positives (FP): the third class refers to the case where the base system classified the points as Fixed (No), and the proposed system classified the same points as Moving objects (Yes).
4. False Negatives (FN): The last class refers to the cases that were classified in the base system as Moving Objects (Yes), and the proposed system classified them as Fixed points (No).

4.1 Accuracy

Accuracy is the ratio of correctly predicted observations (TP and TN) to the total observations; it is an intuitive performance measure of the system result (2).

$$Accuracy = \frac{(TP + TN)}{(TP + FP + FN + TN)} \tag{8}$$

4.2 Precision

Precision represents the ratio of correctly predicted positive classified points (TP) to the total predicted positive classified points in the proposed system (21).

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

4.3 Recall (Sensitivity)

The recall is the rate between the correctly predicted positive classified point (TP) and all classified points in the based system. The resulting value was considered good if above 0.5 (2).

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

4.4 F1 score

The F1 Score represents the relation between the weighted average of Precision and Recall (1). Therefore, this score considers false positives and negatives (2, 8).

$$F1_{Score} = 2 * \frac{(Recall * Precision)}{(Recall + Precision)} \tag{11}$$

5. Data Set

The dataset used for the proposed system was published on http://changedetection.net/ website. It contained a video. The website has two types of datasets (2012, 2014). The dataset used in this paper was in 2014, which includes the following defiance: dynamic background, camera jitter, intermittent object motion, shadows, and thermal signatures. The proposed system focuses on the dynamic background field and uses the videos in Table (2) to test its performance. Figure (3) shows a single image of (a) the boat dataset and (b) the canoe dataset.

Table 2: The Used DataSet (22)

Dataset Type	Video Name	No. of Frame
Dynamic Background	Boats	7999
	Canoe	1189

(a) **(b)**

Figure 3: single image of (a) boat dataset and (b) canoe dataset.

6. Result and Discussion

Table (3) shows the results of two video sequences mentioned earlier. To check the power and weaknesses of the proposed system, the obtained results were compared with the Ground Truth algorithm. The system's accuracy was outstanding, and the precision always gives a good result as it must be greater than 0.5; recall and F1_score were also good.

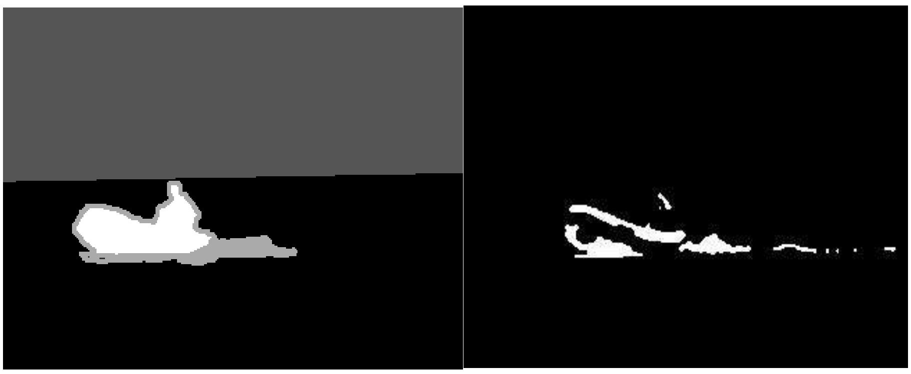
Table 3: Test Result of the Two Datasets

	Frame No.	Accuracy	Precision	Recall	F1_Score
Boats	1965	94.7891	0.71651	0.70020	0.70826
	1966	94.6198	0.72167	0.69361	0.70736
	1970	95.4297	0.73590	0.77089	0.75299
	1980	94.4948	0.74131	0.73583	0.73856
	1990	93.6562	*0.74013*	0.71437	0.72702
	2020	93.3672	0.69332	0.61056	0.64931
Canoe	845	98.6927	0.82899	0.52102	0.63988
	860	99.1068	0.73507	0.84702	0.78708
	870	97.5846	0.60279	0.77204	0.67699
	880	97.4453	0.64565	0.84374	0.73152
	890	96.7812	0.59996	0.88082	0.71375
	899	95.8294	0.58463	0.89426	0.70703

Ground-truth data discarded the motions of far objects, while the proposed system detected all motions like people's movements in the background or car movements; comparing the resulting frame with the ground truth may affect the result of the proposed Syston. Figure (4) below shows the difference between the proposed system and ground truth for the same frames of the dataset. Figure (4a) is the result of the ground truth where the background is in the upper part of the figure. Figure (4b) reveals the result of the proposed system. The average result of each dataset of the Table (3) is listed in Table (4). The accuracy of the proposed system was a good result for the Boats and Canoe objects. The accuracy in detecting Canoe was better than that of Boats because of the difference in background motions and shadow, while the precision is suitable for boat images with the background motion compared to the canoe dataset.

Table 4: Average Result for each Dataset

Dataset	*Accuracy*	*Precision*	*Recall*	*F1_Score*
Boats	94.3928	0.7248	0.7042	0.7139
Canoe	97.5734	0.6662	0.7932	0.7094



(a) **(b)**

Figure 4: difference between the (a) ground truth and (b) proposed system.

7. Conclusion

Predicting changing or moving objects in a camera's field of view is significant in computer vision and video processing. This work proposed a new frame-change detection algorithm that relies on a histogram extraction for each frame pixel. The system is initially trained on a set of frames of images, and the most frequented pixel value was selected to be the value of this pixel in the resulting image. The proposed algorithm calculated the STD for each point in an individual frame and classified these points according to STD and the predefined threshold value. The threshold value depends on the camera's view, and the number of frames also affects the algorithm's performance. Additionally, as a final step, the resulting images were filtered from noise using the Median filter to enhance the algorithm performance. The results reveal that the average F1 score of this work was up to 0.7, and the algorithm's accuracy was between 94% and 97% for different datasets.

References

[1] D. Szolgay, J. Benois-Pineau, R. Mégret, Y. Gaëstel, J.-F. J. P. A. Dartigues, and Applications, "Detection of moving foreground objects in videos with strong camera motion," vol. 14, pp. 311-328, 2011.
[2] T. Yu, J. Yang, and W. J. I. A. Lu, "Dynamic background subtraction using histograms based on fuzzy c-means clustering and fuzzy nearness degree," vol. 7, pp. 14671-14679, 2019.
[3] M. Harville, "A framework for high-level feedback to adaptive, per-pixel, mixture-of-gaussian background models," in *Computer Vision—ECCV 2002: 7th European Conference on Computer Vision Copenhagen, Denmark, May 28–31, 2002 Proceedings, Part III 7*, 2002, pp. 543-560: Springer.
[4] D. A. J. E. o. b. Reynolds, "Gaussian mixture models," vol. 741, no. 659-663, 2009.
[5] 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: IEEE.
[6] M. Hofmann, P. Tiefenbacher, and G. Rigoll, "Background segmentation with feedback: The pixel-based adaptive segmenter," in *2012 IEEE Computer Society Conference on computer vision and pattern recognition workshops*, 2012, pp. 38-43: IEEE.
[7] H. S. Lee and S. I. J. I. A. Cho, "Luminance Level of Histogram-Based Scene-Change Detection for Frame Rate Up-Conversion," vol. 10, pp. 15968-15977, 2022.
[8] S. I. Cho and S.-J. J. I. A. Kang, "Histogram shape-based scene-change detection algorithm," vol. 7, pp. 27662-27667, 2019.
[9] F. A. Khan, M. Nawaz, M. Imran, A. U. Rahman, and F. J. M. S. Qayum, "Foreground detection using motion histogram threshold algorithm in high-resolution large datasets," vol. 27, pp. 667-678, 2021.
[10] M. Ismail, A. Ben, and O. J. J. C. S. Bchir, "Fall detection using the histogram of oriented gradients and decision-based fusion," vol. 16, pp. 257-265, 2020.
[11] K. Sun, Q. Zhao, Z. Zhang, X. J. E. Hu, and Buildings, "Indoor occupancy measurement by the fusion of motion detection and static estimation," vol. 254, p. 111593, 2022.
[12] L. E. George and E. A. Al-Hilo, "Fractal Color Image Compression by Adaptive Zero-Mean Method," in *2009 International Conference on Computer Technology and Development*, 2009, vol. 1, pp. 525-529: IEEE.
[13] I. A. Saad, L. E. J. I. J. o. E. T. George, and T. i. C. Science, "Robust and fast iris localization using contrast stretching and leading edge detection," vol. 3, no. 2, pp. 61-67, 2014.
[14] A. Ibrahim, L. E. George, and E. K. J. I. J. o. S. Hassan, "Color image compression system by using block categorization based on spatial details and DCT followed by improved entropy encoder," pp. 3127-3140, 2020.
[15] M. I. Abd-ALmajied, L. E. George, and R. S. J. I. J. o. S. Hameed, "Algorithm Development for Full Gaps of Landsat 7 Images," pp. 2635-2647, 2023.
[16] M. I. ABD-ALMAJIED, L. E. GEORGE, K. M. J. J. o. T. ABOOD, and A. I. Technology, "A PROPOSED METHOD OF SELECTING PAIRS OF SIFT POINTS," vol. 97, no. 3, 2019.
[17] J. Ali, A. R. Ahmad, L. E. George, C. S. Der, and S. J. I. J. o. C. S. I. Aziz, "Red blood cell recognition using geometrical features," vol. 10, no. 1, p. 90, 2013.

[18] N. N. Dawoud, B. B. Samir, and J. J. I. J. o. C. A. Janier, "Fast template matching method based optimized sum of absolute difference algorithm for face localization," vol. 18, no. 8, pp. 0975-8887, 2011.

[19] P. Hasanzadeh, A. Shahmirzaie, and A. Rezaie, "Motion detection using differential histogram equalization," in *Proceedings of the Fifth IEEE International Symposium on Signal Processing and Information Technology, 2005.*, 2005, pp. 186-189: IEEE.

[20] C. Goutte and E. Gaussier, "A probabilistic interpretation of precision, recall and F-score, with implication for evaluation," in *Advances in Information Retrieval: 27th European Conference on IR Research, ECIR 2005, Santiago de Compostela, Spain, March 21-23, 2005. Proceedings 27*, 2005, pp. 345-359: Springer.

[21] H. Hofbauer, F. Alonso-Fernandez, P. Wild, J. Bigun, and A. Uhl, "A ground truth for iris segmentation," in *2014 22nd international conference on pattern recognition*, 2014, pp. 527-532: IEEE.

[22] N. Goyette, P.-M. Jodoin, F. Porikli, J. Konrad, and P. Ishwar, "Change detection. net: A new change detection benchmark dataset," in *2012 IEEE Computer Society conference on computer vision and pattern recognition workshops*, 2012, pp. 1-8: IEEE.