Extracting Background Model in Video Surveillance By Using Hybrid Techniques

Yossra Hussain Ali, Ph.D (Asst. Prof.)* **Mohammed Rajah Mohammed***

Abstract: One of the most important methods of video surveillance is background subtraction and one of its main steps is modeling of initial background image. In this paper the initial background image is formed by using hybrid techniques (i.e. k-mean and histogram). When the proposed method is employed, the outcomes proposes that the correct color value of highest frequency in the greater cluster leads to modeling of the initial background image, which in turn results in more accurate image of background . This accuracy is measured through PSNR (Peak Signal to Noise Ratio) metric. Furthermore, the current study proposed a comparative study with other four methods (mean, median, histogram and k-mean) through the use of PSNR metric.

Keywords: k-mean, histogram, initial background, PSNR.

* Computer Science Department, University of Technology, Baghdad, Iraq.

1. Introduction

Image processing, computer vision and pattern identification are very important areas of study, concerned with the automatic analyzing of images and sequences of images, with a wide range of applications like remote sensing, medical diagnosing, human-computer communication or video compression, etc. Benefitting from the emerging of these areas, robotized video-based reconnaissance has emerged as a claim inquiries about point which has picked up a ton of consideration in the late years.

In the area of video surveillance, change detection has been regularly utilized as a part of request to foreground items from the back-ground. Foreground items articles are the objects of automated surveillance system. The divided foreground items are afterwards related between frames with a specific end goal to play out a scene investigation and identify occasions of premium. Those basic elements of an automated video-based surveillance system are illustrated in Fig. 1[1].

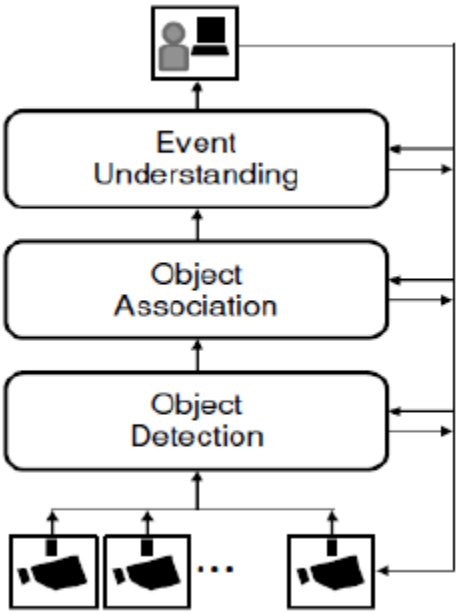

Figure (1) A diagram of A General Video-Based Surveillance System

In this manner, the segmenting the foreground items via the detection of the variations based on a back-ground model is typically referred to as back-ground subtracting. The main issues facing the back-ground subtraction are [1, 2]:

- Sudden variations in lighting illumination that typically happen indoors via switching artificial illumination sources on and off, and also outdoors due to unstable weather conditions like when clouds suddenly hide the sun.
- Smooth variations in lighting that typically happen outdoors throughout various daytimes and have an effect on the appearance of the items in the viewed image.
- Dynamical back-ground, they're the portions of the back-ground that exhibit various appearances due to the fact that they contain a type of moving items like waving trees, rippling water, escalators, etc. that aren't of much interest to the interpretation of the scene.
- Shadows typically caused by items in motion and cause a complication in the accurate item segmentation (stationary items that belong to the back-ground cast shadows as well; however, those aren't quite problematic to the process of back-ground subtraction due to the fact that they're constantly casted at the same place -or at slow moving locations in outdoors cases based on the position of the sun - and may be more easily included in the back-ground model).
- Bootstrapping that is necessary due to the overall training unfeasibility of a back-ground model with an entirely void scene.
- Camouflage, resulted from items whose appearance is hard to identify from the appearance of the back-ground.

2. Theoretical Background

This section proved an overview on the brainer work of the current study which employed in the proposed method and to compare with the proposed method to show the efficiency and effectiveness of the proposed method by using the (PSNR). The next section discusses the above concept in more details:

2.1 K. Mean method

This approach (Mac Queen, 1967) is one of the simplest un-supervised learning algorithms which answer the well-known clustering issue technique commonly used which is simple and a fast method. It is easy to implement and has small number of iterations. K-means is another clustering. The stages of the method are the following [3]:

Step1. Select the number k of clusters, arbitrarily or according to some heuristics.

Step2. Produce k number of clusters and choose the center of each cluster.

Step3. Give every pixel of the image to the clusters which diminish the distance between the pixel and the center of the cluster (Distance is the squared or absolute difference between a pixel and a center).

Step4. Re-calculate the cluster center via averaging all cluster pixels.

Step5. Re-do steps 2 and 3 till a convergence is reached (for instance, the center of the cluster stays un-changed).

In case of background modeling which represents every one of the pixels in the frame with a cluster set. The background initialization is achieved offline. The clusters are arranged with respect to the possibility that they construct the back-ground and are trained to deal with back-ground and illumination changes. New pixels undergo matching with the mapping set of clusters and are classified depending on whether the matching cluster is considered to be part of the back-ground [4].

2.2 Histogram Method

This approach is one of methods that used to modeling initial background image, where introduces an algorithm to extract background images from traffic video streams. The algorithm analyzes each pixel's color values in a series of frames captured during a particular time period and then uses the mode of the series as the correct color value for the background image[5].

2.3 Median Method

Median is one of the most common techniques used for background modeling. Based on the supposition that the background pixel must stay in the buffer more than half of the number of frames accommodated in that buffer, the estimated background is assigned to be the median pixel at each pixel location (x,y) of the preceding N frame in that buffer [6].

2.4 Mean Method

The simplest approach involves calculating an average background frame whilst no moving objects are present. Subsequently, when objects enter the scene, they will cause the current frame to diverge from the background frame and their presence can be easily detected by thresholding the difference between them[7] .

2.5 PSNR Metric

Peak Signal to Noise Ratio is one of the measures utilized for quantifying the quality of the image. The parameter of Peak Signal to Noise Ratio is typically considered as a standard level of similarity between reconstructed and the original images. Bigger value of the Peak Signal to Noise Ratio will result in a more optimal image quality [8] . The arithmetic representation of the Peak Signal to Noise Ratio is depicted in the following formula:

$$\mathbf{PSNR} = \mathbf{20}\ \mathbf{log_{10}}\left(\frac{\mathbf{MAXf}}{\sqrt{\mathbf{MSE}}}\right) \qquad (\mathbf{1})$$

Whereas the MSE (Mean Squared Error) is calculated by the following equation:

$$\boldsymbol{MSE} = \frac{\mathbf{1}}{\boldsymbol{m} * \boldsymbol{n}} \sum_{\mathbf{0}}^{\boldsymbol{m}-\mathbf{1}} \sum_{\mathbf{0}}^{\boldsymbol{n}-\mathbf{1}} \left|\left|\boldsymbol{f}(\boldsymbol{i},\boldsymbol{j}) - \boldsymbol{g}(\boldsymbol{i},\boldsymbol{j})\right|\right|^{\mathbf{2}} \qquad (\mathbf{2})$$

Where

f : represents the matrix data of our original image

g: represents the matrix data of our degraded image in question

m :represents the numbers of rows of pixels of the images and I represents the index of that row

n: represents the number of columns of pixels of the image and j represents the index of that column

MAXf : is the maximum signal value that exists in our original "known to be good" image

3. Related Work

Due to the fact that a prior knowledge of a scene's back-ground isn't typically available, the key for any back-ground subtraction approach is the way of learning and modeling it. There are many of the methods and techniques developed in the field of video surveillance. Here the most important research that touched on the background modeling scheme is presents.

W. Wang, et al. [9] the typically, fastest and the most memory efficient back-ground modeling is known as the running average approach. In this approach, back-ground extraction is performed with mathematical averaging on train series. Following the back-ground extracting, back-ground could change throughout the detecting of moving items. The variations in lighting are a reason of back-ground variations. Due to scene lighting variations and a number of other reasons, back-ground image has to be altered in every frame. Therefore, the approach of the running average is an Infinite Impulse Response (IIR) filter system. However, due to low computational complexity and high memory compactness, this approach is utilized in real-time systems **.**

Butler et al. [4] proposed an approach which represents every one of the pixels in the frame with a cluster set. The background initialization is achieved offline. The clusters are arranged with respect to the possibility that they construct the back-ground and are trained to deal with back-ground and illumination changes. New pixels undergo matching with the mapping set of clusters and are categorized depending on whether the matching cluster is considered to be part of the back-ground. To improve the robustness.

J. Zheng et al. [5] introduces an innovative approach for the extraction of back-ground images from traffic video sequences. The approach performs an analysis of every pixel's color value in a frame sequence recorded within a specific period of time and afterwards utilizes the mode

of the series as the proper value of a color for the back-ground image. For handling little disturbances of back-ground color during a short period of time, a function is utilized for the aggregation of neighboring values of colors into a single bin for mode computation. Via this function, the approach is of higher robustness and of lower disturbance impact resulting due to environmental causes. The algorithm doesn't need sorting and therefore is easy to apply and fast to obtain back-ground images.

Sinha et al. [10]. New video segmentation performs a decomposing of image frames into back-ground and foreground. Where, a mix of simplified mean-shift filter and K-Means clustering are utilized to model the back-ground. Where the mean shifting filter is utilized for computing the local peaks (number of clusters) in the back-ground at every one of the pixel positions with the use of a pre-determined number of training image frames. The mean shifting filter utilizes each one of the pixel values of the video frames as a starting point and converges to the local peak (densest point). As soon as the convergence is done for every pixel, a cluster of average values is performed with the use of the K-Means clustering method. Every one of the input values from testing frame undergoes comparison with those clusters via calculating the distances of the clusters. In the case where the smallest distance is bigger than the threshold then the pixel is considered as a foreground pixel, in the opposite case, it's considered as a back-ground pixel. The testing proved the fact that there has been approximately 28% gain in time when simplified calculation for mean shifting has been utilized.

4. Software And Hardware Requirement

The suggested method has been tested on video stream that was taken from ATON Dataset testing images from its source web-site [11]. That utilized in video surveillance evaluating the resulted images of this data set are 320 x240 pixel. This data set includes multiple video sequence of in-door or outdoor sequence in video monitoring area. as summarized in table (1) .

Where test-bed consisted of processor Intel® core ™ i3-3120M CPU @ 2.50GH_Z and installed memory: 4 GB (2.58 usable) and system type 32-bit operating system. All algorithms were implemented in c#.

Table (1): Videos Used In The Comparative Evaluation.

						
Frame	***number***	1800	440	2227	887	300
	Size	320X240	320X240	320X240	320X240	320X240
Scenes	***Type***	Indoor	Outdoor	Outdoor	Indoor	Indoor
	Noise	Medium	Medium	Medium	Low	Medium
Objects	***Type***	People	Vehicles	Vehicles	people/ other	People
	Size	Variable	Large	Small	Medium	Medium

5. Proposed Approach

The aims of this paper is to improve the already existing video surveillance system where based on used hybrid methods, a modification process was carried out its accuracy feature. The proposed method finds the histogram of lager cluster. Then put the value of highest histogram in corresponding location in initial background. This process was performed in three stages:

Stage one : non-linear buffering: this stage involved input of the video stream in order to receive an output of array of image (frame), it was achieve through loading video and checking the frame number. If the frame number was even then two frame would be neglected and stored one frame in array, on the other hand, if the frame number was odd then one frame would be neglected and stored one frame in array, as show in algorithm (1).

The output will be a buffer of 100 frame and stored in an array_frame, while the frame number was reached to 250, which leads to obtain a solution to find a 1st background structure when over a half of the training buffer contains foreground items.

This operation is called (non-linear buffer). The algorithm (1) explain steps of extraction frame and stored them in array.

Algorithm (1) Non-Linear Buffer
Input : *Training Video (Width =320, Height=240)* *: N : Number Of Frame In Buffer*
Output : *Array Of Frame*
Begin ***Step 1:*** *T= 0 , I=0 // T : Time , I: Index Of Frame* ***Step 2:*** *Repeat* *Capture Frame (T) From Video* *Set Frame To Array Of Frame* *T++* *If (T Is Even)* *T++* *End If* *T ++* *I++* *Until (I < N)* ***End***

Stage two: vector of pixel: to obtain vector of pixel form the array of frame that has been obtain from stage 1. This paper applying the algorithm (2) the vector construction algorithm that extraction an array of pixel (100) value at each pixel location (i, j) and send this array _pixel to stage 3 to apply the k-mean .

Algorithm (2) vector construction algorithm
Input : *array_frame*
Output : *array_pixel*
Begin ***Step 1:*** *Chose 100 frames from the first 250 Frames, by using algorithm(1) for non-linear buffer .* ***Step 2 : For i=****0 to frame _width* ***For j =****0 to frame_height* ***For k=*** *0 to number_frames* *get pixel of indexing(i,j) from frame (k) and set it in array_pixel[k].* ***End for K*** ***End For j*** ***End for i*** ***End***

Stage three: Image Formation: To explain modeling of initial background image the proposed method employed the following steps: Firstly, k-mean method has been applied on (array_pixel) and selected the lager cluster that used in next steps, as in algorithm (3) k-mean algorithm.

The outcome of this step was clustering (100) value of pixel at location (x, y) in three cluster and return the largest cluster that have the maximum number of members.

Algorithm (3) k-mean algorithm
Input : *vector of data(array_pixel[100])*
Output : *greater cluster*
Begin ***Step1****: Select the number k of clusters, arbitrarily.* ***Step2:*** *Produce k number of clusters and choose the center of each cluster.* ***Step3****: Give every pixel of the image to the clusters which diminish the distance between the pixel and the center of the cluster* ***.*** ***Step4:*** *Re-calculate the cluster center via averaging all cluster pixels.* ***Step5:****Re-do steps 2 and 3 till a convergence is reached (for instance, the center of the cluster stays un-changed).* ***Step6:*** *find the greater cluster // that have the maximum member .* ***End***

Secondly, find the value of highest histogram of largest cluster and put this value in initial background in corresponding location p(x,y) for each pixel .

To do this step the proposed method applied the histogram on large cluster, as in algorithm (4) histogram algorithm.

Algorithm (4) histogram algorithm
Input : *vector of data(Data[n])*
Output : *vector of frequency(Frequency[n])*
Begin *for i=0 to N* *Item = Data[i].* *Frequency_Conut=0.* *for j=0 to N* *if(Item = Data[j]) then Frequency_Conut= Frequency_Conut +1.* *End For j* *Frequency[i]=Frequency_Conut* *End for i* ***End***

The proposed method finds the histogram of lager cluster. Then put the color value of highest histogram in corresponding location in initial background image. The aim of this step is to construction initial background with more accuracy, as show in algorithm (5) Image Formation.

Algorithm (5) Image Formation
Input : *victor array_ pixel of [100]*
Output : *modeling pixel*
Begin ***Step 1:*** *apply k-mean on victor array_pixel[100]* ***Step2 :*** *apply histogram on cluster with largest member* ***Step 3 :*** *modeling pixel = pixel value with max frequency* ***End***

For clarifying the objective of this step it is necessary to recognizing that when using only the histogram method in some cases the foreground is the highest frequency. In this case the initial background is construction in a shape contain a part of the foreground. Whereas using only the k-mean method give incorrect value because it is adding and dividing.

While using the proposed method, the result will be correct color value of highest frequency in the greater cluster, therefore modeling of the initial background will be more accuracy and that can be proved through PSNR metric.

6. Results

The implementation of the proposed method is evaluated against the background modeling algorithms. To test the background modeling performance of the five methods mentioned earlier the (PSNR) metric employed in this paper.

The proposed method has been done in two steps: firstly used **non-linear buffer** to extraction video frame and store them in array of frame that used later in next step. The aim of this step was to find a 1st background structure when over a half of the training contains sequence foreground items. Figure (2) show the differences between used sequences buffering frame and the buffer by used non-linear buffer that proposed in this paper.

(a)

(b)

Figure (2): Initial background (A): Buffer Using First 100 Frame, (B): Buffer Using Proposed Method.

Secondly , to make sure that the hybrid method performance was more sufficient than other methods the comparison was done with grand truth and other four methods (running average, median, histogram and k-mean), as shown in figure (3) where the first column is grand truth background, the second column is the background obtained by using the running average, the third column is the background obtained by using the median, the fourth column is the background obtained by using the

Histogram, the fifth column is the background obtained by using the k-mean, and the sixth column is the background obtained by using the proposed method.

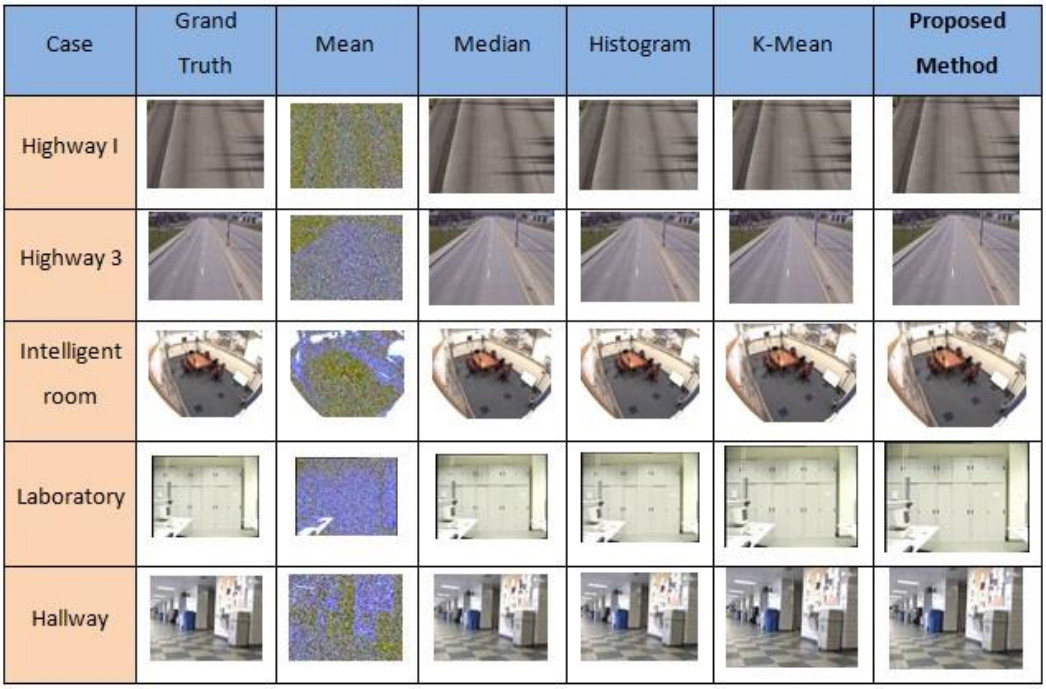

Figure (3) Qualitative Comparative of Different Methods

The quantitative comparison of different methods was done by using PSNR metric. Illustrated in table (2).

Table (2): Result of (PSNR) Ratio Implemented Per Five Methods .

Case	Mean	Median	Histogram	k-mean	Proposed
high way 1	14.40	33.36	33.87	33.76	34.20
high way 2	13.63	27.05	26.93	27.66	27.92
high way 3	13.90	42.49	42.18	41.48	42.72
Intelligent room	12.48	35.78	35.47	35.83	35.88
Laboratory	10.87	39.44	39.37	39.06	39.57
hall way	12.05	32.30	30.92	32.59	33.33

Figure (4) shows the performance of each method across all sequences. Every line illustrates the efficiency of an approach that ranges from the original sequences (10% PSNR) to sequences (45% PSNR).

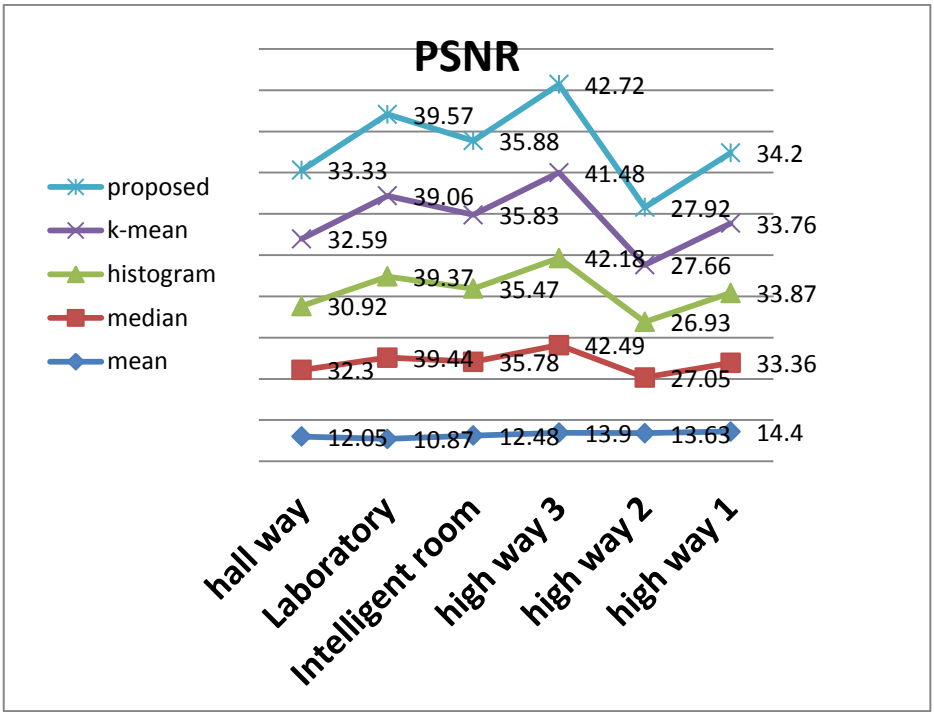

Figure (4): The Performance of Five Methods Across All Sequences

From the data collected shown in table (2) and the illustrated in figure (4) , this current study comes to the conclusion that, the method proposed has proven more efficient and accurate in comparison to the methods mention earlier in this paper . In other words, 6 case studies were proposed by means of 4 methods (mean, median, histogram and k-mean) as illustrated in table (2). The resulting PSNR metric in column (proposed) came out higher than the other methods which indicate high level of efficiency. For further clarification figure (4) illustrated the performance level of the five methods.

7. Conclusion

In this paper the initial background image is formed by using hybrid techniques (i.e. K-mean and histogram). When the proposed method is applied, the outcomes proposes that the correct color value of highest frequency in the greater cluster leads to modeling of the initial background image , which in turn results in more accurate image of the background. This current study comes to the conclusion that, the method proposed has proven more efficient and accurate in comparison to the methods mention earlier in this paper.

References

[1] S. Kaur , " Background subtraction in Video Surveillance" , M.Sc. thesis, Computer Science , University of Windsor , Windsor, Ontario, Canada , 2017 .

[2] Thierry Bouwmans, Fatih Porikli , Benjamin Höferlin , Antoine V acavant ." Background Modeling and Foreground Detection for Video Surveillance". International Standard Book Number-13: 978-1-4822-0538-1 (eBook - PDF) , LLC , 2015.

[3] Jipkate, B. R. and Gohokar,V. A Comparative Analysis of Fuzzy C-Means Clustering and K Means Clustering Algorithms. IJCER, 2 (3) :737-739 , 2012.

[4] D. Butler, V. Bove, and S. Shridharan, "Real time adaptive foreground/background segmentation". EURASIP, pp. 2292–2304, 2005.

[5] J. Zheng, Y. Wang, N. Nihan, and E. Hallenbeck. "Extracting roadway background image: A mode based approach". Journal of Transportation Research Report,1944:82–88, 2006.

[6] Abdulamir A. Karim , "A Proposed Background Modeling Algorithm for Moving Object Detection Using Statistical Measures " , Iraqi Journal of Science, Vol. 58, No.3A, pp: 1282-1289 , 2017.

[7] T. Horprasert, D. Harwood, and L. S. Davis, “A robust background subtraction and shadow detection” in Proc. 4th Asian Conference on Computer Vision (ACCV ’00), Taipei, Taiwan, January 2000.

[8] Albertus J. S., Lukito E. N., Gede B. S., Risanuri H. "Compression Ratio and Peak Signal to Noise Ratio in Grayscale Image Compression using Wavelet". IJCST Vol. 2 , Issue 2, June 2011 .

[9] W. Wang, D. Chen, W. Gao, J. Yang, "Modeling background from compressed video", Second Joint IEEE International Workshop on Video Surveillance and Performance Evaluation of Tracking and Surveillance, in conjunction with the Tenth IEEE International Conference on Computer Vision. Beijing, China, October 2005.

[10] S. Sinha, M. Mareboyana. “Video Segmentation into Background and Foreground Using Simplified Mean Shift Filter and K-Means Clustering ". ASEE Zone I Conference, April 3-5, University of Bridgeport, Bridgeport, CT, USA, 2014.
[11] ATON Dataset: http://cvrr.ucsd.edu/aton/shadow/index.html.

استخراج نموذج الخلفية في نظم مراقبة الفيديو باستخدام تقنيات هجينه

أ. م .د يسرى حسين علي* الباحث: محمد راجح محمد*

المستخلص: واحدة من أهم أساليب المراقبة بالفيديو هو الطرح الخلفية واحدة من خطواتها الرئيسية هو النموذجة من صورة الخلفية الأولية .في هذه الورقة يتم تكوين صورة الخلفية الأولية باستخدام تقنيات هجينة (بمعنى K-mean Histogram) . وعند استخدام الطريقة المقترحة، فأن النتائج أن قيمة اللون الصحيحة لأعلى تردد في الكتلة الأكبر تؤدي إلى نمذجه صورة الخلفية الأولية، الأمر الذي يفرض بدوره إلى صورة أكثر دقة للخلفية. وتقاس هذه الدقة من خلال قياس نسبة الإشارة إلى نسبة الضوضاء. (PSNR) وعلاوة على ذلك، اقترحت الدراسة الحالية دراسة مقارنة مع أربع طرق أخرى (mean، median، k-mean histogram,) من خلال استخدام مقياس (PSNR).

الكلمات المفتاحية: k-mean ، الرسم البياني ، الخلفية الأولية ، PSNR.

*الجامعة التكنولوجية - قسم علوم الحاسوب