DOI: 10.24996/ ijs.2017.58.3A.12

Iraqi Journal of Science
ISSN: 0067-2904

A Proposed Background Modeling Algorithm for Moving Object Detection Using Statistical Measures

Abdulamir A. Karim
Department of Computer Science, College of Science, University of Technology, Baghdad, Iraq.

Abstract

Extracting moving object from video sequence is one of the most important steps in the video-based analysis. Background subtraction is the most commonly used moving object detection methods in video, in which the extracted object will be feed to a higher-level process (i.e. object localization, object tracking).

The main requirement of background subtraction method is to construct a stationary background model and then to compare every new coming frame with it in order to detect the moving object.

Relied on the supposition that the background occurs with the higher appearance frequency, a proposed background reconstruction algorithm has been presented based on pixel intensity classification (PIC) approach. First, pixel intensity in a predetermined time period has been classified according to a proposed clustering method, second, pixels frequency of those clusters has been calculated, finally, the center of the cluster with the higher pixel frequency has been chosen as the background pixel intensity value.

The efficiency and effectiveness of the proposed algorithm has been confirmed through comparing its results with those of the most common traditional methods, besides , the results of the proposed algorithm in a number of testing environment which are traffic monitoring and pedestrian surveillance shows that the proposed algorithm can save space and economize computation time and give good accuracy.

Keywords: Background Modeling, Background Subtraction, Pixel Intensity Classification.

خوارزمية نمذجة خلفية مقترحة لكشف الأجسام المتحركة بأستخدام المقاييس الأحصائية

عبدالأميرعبدالله كريم
قسم علوم الحاسوب، كلية العلوم، الجامعة التكنلوجيه، بغداد، العراق.

الخلاصة

ان استخلاص الأجسام المتحركة من الفيديو تعد واحدة من اهم الخطوات في تحليل الفديو. تعتبر عملية طرح الخلفية من اكثر طرق الكشف عن الأجسام المتحركة شيوعا في الفديو، حيث يتم تغذية الجسم المستخلص الى عمليات ذات مستوى اعلى (اي تحديد موقع الجسم وتتبع الجسم).

المتطلب الرئيسي لطريقة طرح الخلفية هو بناء نموذج خلفية ثابت ومن ثم مقارنة كل اطار frameجديد بذلك النموذج من اجل الكشف عن الأجسام المتحركة.

اعتمادا على فرضية ان بكسل Pixels الخلفية تظهر بأعلى تردد ظهور ، تم اقتراح خوارزمية جديدة لأعادة بناء الخلفية بالأستناد الى نهج تصنيف كثافة البكسل اولا. تم تصنيف البكسل في فترة زمنية محددة

Email: ameer.aldelphi@yahoo.com

مسبقا وفقا لطريقة عنقدة clustering مقترحة ، ثانيا. تم حساب تردد البكسل لتلك العناقيد ، واخيرا. تم تحديد مركز العنقود ذا تردد البكسل الأعلى واعتماده كقيمة لبكسل الخلفية.

تم التحقق من فاعلية وكفاءة الخوارزمية المقترحة من خلال مقارنة نتائجها مع نتائج الطرق التقليدية الأكثر شيوعا. فضلا عن ذلك ، فأن نتائج الخوارزمية المقترحة في عدد من بيئة الأختبار (مراقبة حركة المرور و مراقبة المشاة) أظهرت ان الخوارزمية المقترحة يمكن ان توفر في المساحة الخزنية وان تقتصد في وقت الحساب وانها تعطي دقة جيدة.

1. Introduction

The extraction of the moving objects from video sequence is a fundamental step in video surveillance system (e.g. traffic and pedestrian monitoring and analysis) . Background subtraction is one of the most common approach used for moving object identification, specifically for video sequence captured from a static camera [1].

Although , using background subtraction methods is very popular in moving object detection and extraction from video, it needs to cope with a number of situations like change in illumination, dynamic (non static) background , and the shadow caused by moving objects [2].

A background subtraction is particularly suitable for applications like surveillance systems and video conferencing in which the background almost remain static during the monitoring or the conference period of time [3].

The main idea of background subtraction is simply to subtract the current frame (image) from the reference frame (still background) , hence, after subtraction, only the moving or new object are left [1]. Most of the existing techniques does not consider the construction of background model (i.e. still background image has been acquired from the real scene before any object move in), on the other hand , robust techniques try to construct a background model and use it to reconstruct a background image [2].

Obtaining a reference image (still background) from a dynamic scene is a very difficult task , because it required acquiring a new background whenever the scene change, hence, reconstruction (modeling) of background is at the core of majority of background subtraction technique [4, 5]. Hence, key problem of background subtraction methods is how to acquire a precise background (reference) image under the condition of non-stationary background or moving object in the scene [6].

The reconstructed background image (reference image) ought to be a representation of the scene without any moving objects or new incoming objects and must be updated regularly so that it can react quickly to changes in luminance or motion changes [4].

In this work , a new background modeling (reconstruction) algorithm relied on pixel intensity classification has been proposed, firstly , pixels intensity are stored and then classified according to their intensity value, secondly, cluster center and pixel frequency for each cluster has been calculated, finally, cluster with the maximum pixels intensity frequency has been chosen as the candidate background intensity value.

Results show that the proposed background reconstruction algorithm can efficiently and effectively construct a stationary background in dynamic scene so that it can adapt its behavior according to variation in background, which can later use for moving vehicle and pedestrians localization and detection in video. The proposed algorithm rely on constructing robust (stable) background model with small memory requirement and fast enough so that it can accommodate real time requirement.

2. Background modeling techniques

Background modeling (estimation) is the core of any background subtraction method. Below is a brief description of some of the most popular conventional background modeling techniques.

2.1 median modeling: 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 [7, 8]. The main drawback of median method is that its result is not predicted i.e. any extreme pixels intensity value from the past, can be effected on the estimated background model [4], besides it is over sensitive to changes in scene lighting and to extraneous events [9]. Also it needs comparable storage , due to the buffer size that median filter required [8].

2.2 Time Average Background Image (TABI): TABI is one of the most common statistical model used for background reconstruction. This method is simple, it store the incoming **N** frames pixel intensity values in a buffer , then the mean (average) of those pixels in that buffer is calculated (using equation 1 below), and will be used to represent the model (estimated background) pixel at location (x, y). This procedure will be repeated for all frame locations (x, y).

$$\mu\,(x, y) = \frac{1}{n}\sum_{i=1}^{n} f(x, y) \qquad \ldots\ldots\ldots\ldots \quad (1)$$

TABI method cannot cope with situations in which scene contains many moving objects, especially if those objects moves slowly, so that resulted background may contains foreground objects blending with background images [5, 10].

2.3. **Mixture of Gaussians (MoG)**: in this technique a probability density function (pdf) has been maintained for each pixel [7]. In MoG we have to modeled the background at each pixel location (x, y) independently, by fitting a Gaussian probability density function (pdf) over a preceding **N** frames. The Gaussian pdf parameters (mean and standard deviation) at each pixel location (x, y) which has been stored in a buffer , will be calculated , and used to represent the background model [2]. Any incoming frame pixel intensity value at location(x, y) will be compared against its corresponding parameter of Gaussian pdf , if they don't fit with those parameters , that's mean the pixel belongs to foreground (moving object) , otherwise it belongs to the background which must be adjusted accordingly [2, 10]. MoG parameter can be calculated using equation (1) above and equation (2) below :

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} \left(f(x, y) - \mu\,(x, y) \right)^2 \qquad \ldots\ldots \ (2)$$

where μ , σ are the mean and standard deviation for n pixels intensity values stored in a buffer of size n.

The MoG method can manipulate slow changes in illumination efficiently ,and it can deal with long term changes in the scene , but it is very sensitive to sudden changes in global illumination. On the other hand, MoG method is computationally expensive, due to algorithm complexity, and its parameters (μ , σ) should be tuned carefully [7, 8]

2.4. Pixel Intensity Classification (PIC) : the basic assumption that PIC rely on is that background pixel intensity has been occurred in the image sequence with the higher frequency, and the pixel intensity value with the higher frequency is chosen as the expected background (model) pixel intensity value [1, 7], hence, in order to classify pixel intensity values, a clustering technique is required which must be computationally economical and with low memory space requirement. To cluster a pixel intensity values the following basic steps has to be performed [1].

1. **N** previous frames are selected from the video sequence.
2. For any pixel location (x, y), the **N** pixels intensity values which belong to N frames have to be clustered using a clustering techniques (e.g. K-mean clustering).
3. Number of pixels intensity (frequency) for each cluster is calculated.
4. The cluster with the higher frequency is selected as the candidate cluster.
5. The center of the candidate cluster is calculated and it is chosen to represent the estimated background pixel intensity.
6. Step 1-5 are repeated for all pixel locations (x, y).

3. Proposed background modeling method

3.1 Preview

Many background subtraction approaches assume a picture of the scene taken in absent of moving object as a background image (reference frame) ,this approach is not efficient , because the frequent changes in the scene required capturing a new picture whenever the scene change. The second approach, which has been adopted in this work , is to estimate the background image using a background modeling technique (e.g. median filter). Between several background subtraction approaches, the crucial different is that how does the background modeled.

The proposed modeling algorithm is considered as a Pixel Intensity Classification (PIC) method. The key problem in modeling a background by PIC is the selection of the clustering technique, and how does this clustering technique goes through. Once the background is modeled, a comparison between the pixels of the inputted frame and the pixels of the background model is done, and the inputted pixel which largely deviated from the model pixel is considered as a foreground (object) pixel.

The proposed algorithm assume that the video under consideration is a grey scale video , since adding color in background modeling leads to increase the complexity of the model estimation process, besides it does not add any positive effect to this process.

3.2. adaptive mean value

In the proposed algorithm two mean value has been adopted which are differ from the conventional statistical mean value, those are :

3.2.1 Adaptive Arithmetic Mean

The classical formula used to calculate the average value of the (**n)** pixels value located in (x,y) coordinate is as illustrated in equation (1), this formulas required recalculating the average value whenever a new frame is entered in the scene, which in turn reduce the speed of calculation.

In this work an adapted formula for calculating the average value has been considered , this formula does not required recalculating the average value whenever a new frame is entered to the scene (i.e. it doesn't required recalculating the summation of all the pixel values) ,hence, the new pixel value is taken in consideration without resuming all the preceding pixel values. The adopted formula for adaptive mean value as shown in equation (3) :

$$m_{new} (x, y) = \frac{1}{n+1} \quad (f_{new} (x,y) + n * m_{old} (x, y)) \qquad \ldots\ldots\ldots (3)$$

where **n** is the number of pixel under consideration.

f(x, y) pixels intensity value at location x, y.

According to the above formula , calculating m_{new} required only adding the value of the new pixel to the value of **$n * m_{old}$ (x, y)**

3.2.2 Weighted Mean

Consider two samples S_1, S_2 value with mean values m_1, m_2 and with sample size n_1, n_2 (i.e. pixel frequency for each sample). When we want to merge those two samples S_1, S_2 , a new mean has to be calculated for the new merged sample, this new mean must take the frequency of each sample into consideration (i.e. frequency is considered as a weight) and this what we called the weighted mean . in this work the below formulas is adopted for calculating the weighted mean.

$$m_{weighted} = \frac{m1.n1.+m2.n2}{n1+n2} \qquad \ldots\ldots\ldots (4)$$

3.3 background modeling algorithm

The below steps present a detailed description of the proposed background modeling main algorithm which in turn classify the pixel intensity values by calling intensity clustering algorithm.

Algorithm 1: background modeling algorithm
Input : gray scale video **Output** : the background model (reference Frame)
Step 1: read **N** frame sequence from the input video sequentially. Step2: maintain an independent clustering model for each pixel (x,y) 2.1 classify pixel intensity value at location (x,y) for **N** frames based on online clustering using algorithm 2. Step 3: if the different between two cluster centers is less than a selected threshold; $\lvert Ci - Cj \rvert < T$ Combine those two clusters so that : 3.1 use the weighted mean (equation 4) to calculate the average value of the new cluster (combined cluster) $C_{new} = \frac{C\,i.ni + Cj.nj)}{ni+nj}$ 3.2 new cluster frequency is equal to the summation of the two merged cluster frequency : $n_{new} = n_i + nj$ step 4: searching for the cluster with maximum frequency . step 5: select the background intensity value : a. adopt cluster average of the cluster with the highest frequency as a background pixel value at location (x, y) Step 6 : repeat step 2-5 so that the entire pixels of N frames will be classified.

he main idea behind algorithm 2 (adaptive mean) is to apply a proposed clustering technique on a buffer of N pixels, the algorithm read the buffer pixels sequentially, and compare each incoming pixel at location (x,y) (i.e. $f_{new}(x,y)$ with all the preceding cluster averages). The new pixel will be assigned to the closest cluster average and that cluster average will be recalculated using equation(3), besides the cluster frequency will be incremented by 1, finally a number of cluster centers and its corresponding frequencies will be obtained.

Algorithm 2: adaptive mean algorithm

Input : a buffer contains **N** pixels
Output : a predefined cluster centers of the buffer pixels
a predefined cluster frequency of the buffer pixels

Step 1: parameter initialization.
set MAX to maximum number of clusters determined by the user.
set all cluster average $\mathbf{m_c(x,y)}$ to zero // { $m_1(x,y)=0$, $m_2(x,y)=0$, …. , $m_{max}(x,y) = 0$ }
T= [3 5] // set threshold value to the range from 3 to 5 according to the video illumination changes
C= 1 // set cluster no. to 1
$m_C(x,y) = f_1(x,y)$ // set cluster 1 average to the first frame pixel value.
$n_C(x,y) = 1$ // set cluster 1 frequency to 1
Step2 : calculate cluster average and frequency
N = 2 // set the number of clusters under consideration to 2
Repeat steps (2.1, 2.2 , 2.3) until N= MAX
2.1 read pixel value from buffer sequentially // read $f_{new}(x,y)$
2.2 for C = 1 to N do // C is the cluster number
If $| f_{new}(x,y) - m_C(x,y) | < T$ // apply equation (3) to calculate cluster average.
$m_C (x, y) = \frac{1}{nc(x,y) +1} (f_{new} (x,y) + n_C(x,y) * m_C (x, y))$
$n_C(x,y) = n_C(x,y) + 1$ // increment cluster frequency by 1
2.3 N = N + 1 // increment number of clusters under consideration by 1
Step 3: return all cluster averages { $m_1(x,y)$, $m_2(x,y)$, ... , $m_{max}(x,y$}
return all cluster frequency { $n_1(x,y)$, $n_2(x,y)$, ... , $n_{max}(x,y)$ }
end.

4. Experimental results and comparisons

The proposed algorithm is implemented in C# programming languages, under visual studio 2015 environment with 2.13 GHz CPU and 4 GB RAM .

In order to establish the efficiency of the proposed algorithm, it has been tested on two types of videos, vehicles passing on highway and pedestrian video sequence. Figure-1 below show a number of frames of each video, and the estimated background using the proposed algorithm.

a. Frame no. 107 b. frame no. 130 c. estimated background

d. Frame no. 224 e. Frame no. 320 f. Estimated background

g. Frame no. h. Frame no. i. Estimated background

Figure 1- (a-f) vehicles passing on highway video frames and its reconstructed background.(g-i) pedestrian video frames and its reconstructed background.

To validate the proposed algorithm, results obtained by it are compared with those of traditional background subtraction methods (Median, TABI, PCI) and three fidelity criteria which are (MSE, SNR, PSNR) are used to validate the accuracy of the reconstructed background. Table-1 below show the obtained results.

Table 1- The results of The Evaluation Criteria After Applying The Traditional Methods (Median, TABI, PCI) and the Proposed Algorithms on Pedestrian Video.

method \ criteria	RMSE	SNR	PSNR
TABI	15.492	7.462	24.326
Median	7.093	18.341	31.113
PCI	1.747	73.673	43.276
OUR's	1.293	99.397	45.881

In term of accuracy, fidelity criteria in Table-1 shows that the proposed method gives the best estimated background, also PCI method gives a good estimated background among the other traditional methods, while TABI was the worst.
From computation time point of view, the time required to obtain **one** background pixel from a buffer of size 25, 50, 100 pixels is calculated for the four background modeling techniques, Table- 2 below illustrate the time required to estimate **one** BG pixels from a buffer of different size.

Table 2-The Results Of The time (in seconds) required to estimate **one** BG pixels from a buffer of different size 25, 50, 100 respectively when Applying The Traditional Methods (Median, TABI, PCI).

method criteria	25 frame	50 frame	100 frame
TABI	0.0003	0.0007	0.0011
Median	0.0007	0.0012	0.0217
PCI	0.0077	0.0121	0.0259
OUR's	0.0004	0.0009	0.0013

In term of computation time, results from Table-2 shows that TABI method is the faster in term of computation time, while the proposed method come in the second place, while PCI give the worst result.
Note that Table-2 above shows the time required to estimate **one** background pixels from a buffer of different size. In order to speed up the work when estimating the pixels of the whole background frame (e.g. with frame size of 160 × 120 the total number of buffers is 19,200) multi-threading technique which C# support has been used . Multi-threading speed up the calculation time and make it approximate real time requirement.

5. Conclusion

In this work, a proposed background modeling algorithm has been presented. The algorithm is to estimate the background pixels based on Pixel Intensity Classification (PIC) approach. The proposed algorithm can save space and economize computation time. Saving space has been achieved through dispense using buffer, because median, TABI and traditional PIC methods required a buffer space for each pixel, with buffer size equal to the number of frames required to construct the background pixel. While the economic in computation time come from that this method does not required sorting the pixels like median, and does not required to recalculate the average whenever a new pixel is taken into consideration like TABI, also it does not required to perform a classical clustering technique (e.g. K-mean) on the pixel in buffer like traditional PIC.

When comparing the proposed algorithm with traditional BG modeling techniques (Median, TABI, PIC) , results shows that the proposed algorithms is better than other technique in term of accuracy of the reconstructed background. The proposed algorithm lead to better background subtraction , in other words, an accurate background can be constructed which can be used later to detect moving objects and give accurate shape.

Mention

The traditional BG modeling techniques (Median, TABI, PIC) which result has been used to be compared with the proposed algorithm, was programmed by the author, using C# programming language, and Pentium CORE i3 personal computer, hence, the obtained results may be little different according to the computer specifications , language type and the manner in which those methods has been programmed.

References

1. Xiao, M., Han, C. and Kang, X. **2006**. A Background Reconstruction for Dynamic Scenes. 9th International Conference on Information Fusion.
2. Subudhi, B. N., Ghosh, S. and Ghosh, A. **2013**. Change Detection for Moving Object Segmentation with Robust Background Construct Under Wronskian Framework. *SPRINGER*.

3. Patil, P. A. and Deshpande, P. A. **2015**. Moving Object Extraction Based on Background Reconstruction. *International Journal of Innovative Research in Computer and Communication Engineering*, **3**(4).

4. Mohanty, A. and Shantaiya, S. **2015**. A Survey on Moving Object Detection using Background Method in Video. Proceedings on National Conference on Knowledge, Innovation in Technology and Engineering (NCKITE).

5. Hou, Z. and Han, C. **2004**. A Background Reconstruction Algorithm Based on Pixel Intensity Classification in Remote Video Surveillance System. Preceding of the Seventh International Conference on Information Fusion , Fairborn. United State, pp: 754-759.

6. Devi, R. B., Chanu Y. J. and Singh, K. M. **2016**. A survey on different Background Subtraction Method for Moving Object Detection. *International Journal For Research in Emerging Science and Technology*, **3**(10).

7. Mei, X. and Lei, Z. **2008**. Using Modified Basic Sequential Clustering for Background Reconstruction. *Information Technology Journal*, **7**(7), PP: 1037-1042.

8. Cheung, S. and Kamath, C. **2004**. *Robust Techniques for Background Subtraction in Urban Traffic Video.* IS&T/SPIE's Symposium on Electronic Imaging San Jose, CA, United States.

9. Mervin, K. and Sarala, B. **2015**. Moving Object Recognition Using Background Subtraction and Image Processing Techniques. *International Journal of Scientific Engineering and Applied Science (IJSEAS)*, **1**(8): 274-282.

10. Niranjil K. A. and Sureshkumar C. **2015**. Background Subtraction in Dynamic Environment Based on Modified Adaptive GMM with TDD for Moving Object Detection. *The Korean Institute of Electrical Engineers*, **10**(1), PP: 372-378.