Background Subtraction Method for Object Detection and Tracking

Satrughan Kumar and Jigyendra Sen Yadav

Abstract Video object extraction and its tracking is one of the fundamental tasks of computer vision that require a close observation on video content analysis. However, these tasks become sophisticated due to spatial and temporal changes in the video background. In this work, we have proposed a background subtraction algorithm that efficiently localizes the object in the scene. In the next stage, a regional level process is integrated by calculating the Shannon energy and entropy to correctly examine the nonstationary pixels in the frames. In order to extract the object efficiently, the background model is updated to the dynamics changes that reduces the false negative pixels on foreground. Further, an adaptive Kalman filter is integrated to track the object in consecutive frames. Qualitative and quantitative analysis on some experimental videos shows that the method is superior to some existing background subtraction methods used in tracking.

Keywords Background subtraction · Shannon entropy · Energy · Object tracking

1 Introduction

Video object segmentation has been a motivating area of computer vision system in the last decade. It solves the varieties of problem in target localization, tracking, and action analysis [1–3]. However, the varying nature of video background due to rippling water, waving tree, quasi-stationary motion, and its changing appearance due to bad resolution or illumination make the object segmentation and extraction tasks more exigent. In [3], Hu et al. categorized the moving object detection methods into three basic classes that are frame difference [4], optical flow [5], and

Satrughan Kumar (✉) · J.S. Yadav
Department of Electronics and Communication, MANIT,
Bhopal 462003, India
e-mail: satrughankumar@gmail.com

J.S. Yadav
e-mail: jsyadav74@gmail.com

R. Singh and S. Choudhury (eds.), *Proceeding of International Conference on Intelligent Communication, Control and Devices*, Advances in Intelligent Systems and Computing 479, DOI 10.1007/978-981-10-1708-7_125

background subtraction [1–3]. The temporal difference method uses the pixel-wise difference between the two or three consecutive frames in the video to localize the moving object especially in sudden or gradual illumination changes, but the object extraction fails when the object becomes stationary in the scene. It is also affected due to serious hole, ghost, and aperture distortion. The optical is computationally complex and its smoothness constraint limited to few pixels movement in the successive frame.

In this concern, we proposed a background subtraction algorithm that has the capability to provide the sufficient sample size to tracking module without the any prior assumption and suitable under static camera arrangement. The article is structured as follows: Sect. 2 explains some methods related to object extraction and tracking. In Sect. 3, the proposed algorithm is explained. Experimental results are shown in Sect. 4, while the concluding remarks are given in Sect. 5.

2 Related Work

Some of the existing background subtraction methods are reviewed in this section. Many tracking frameworks use these models to trace the object in the scene.

In [2], Manzanera and Richefeu proposed Σ-Δ method (SDE and utilized the difference image and time variance to calculate the foreground pixels. The method is suitable for real-time application but produces insufficient accuracy in case of multiple objects in a scene. In [6], a Gaussian mixture model (GMM) method handles a single pixel using at least three Gaussian components that are updated adaptively over a time in the consecutive frames. Although GMM can work well under gradual illumination conditions and local background motion, but it has greater time complexity. A statistical method proposed in [7], does not update the background pixels that makes it less useful under changing illumination. In [8], Jing et al. integrate spatial-temporal processing to get the moving pixels, but the update of background model is done according to traditional schemes. In [9], a recursive filter is used in updating the background that solely depends on the learning parameters. The dependency on the learning parameters may cause either trails or delay to update the background model. In order to extract the object, a frame difference method is proposed in [10]. However, the frame difference method stops the extraction procedure when the target becomes stationary. In [11], Yao and Ling proposed an improved version of GMM, but it fails to detect the object near camouflage region.

Previous studies reveal that the sample size of object either is lost or buried under the noise under complex condition. Even though some methods are applicable in real-time scenario, but most of them suffer from either ghost effect or aperture distortion. Methods depend on fast learning parameter causes trails behind the object and reduce the accuracy, while methods having low learning rate, do not update the background accurately. Therefore, a regional level processing may be beneficial to update only the changing background pixels and for detecting the

actual non-stationary pixels. In this paper, we integrate the regional level processing by evaluating the block wise entropy and energy that provided the actual moving pixels on the foreground.

3 Proposed Method

The section explains the proposed method into two stages. The first stage or phase describes the object extraction phase using the background subtraction technique, while the next stage works on tracking the trajectory of object using an adaptive Kalman filter. We have utilized the gray scale videos for the experimental set up, which are recorded under static camera arrangement.

3.1 Background Subtraction and Object Extraction

Initially, some 'K' frames are taken to generate the reference background model using the modified moving approach. These initial frames consist of no foreground object. The reference background $B_r(x, y)$ is given as follows:

$$B_r(x, y) = I_0(x, y) + \frac{(I_t(x, y) - I_0(x, y))}{K} \tag{1}$$

where $I_0(x, y)$ and $I_t(x, y)$ are the first and current frame of the video. The 'x' and 'y' are the height and width of the frame.

Further, it creates a difference image by subtracting reference background from the current frame. The moving pixels in the difference image are filtered out by selecting the proper threshold function. Finally, it updates the reference background model over a time to adapt the temporal variation due to environmental changes. The difference image $D_t(x, y)$ is computed as

$$D_t(x, y) = \left| I_t(x, y) - B_t^R(x, y) \right| \tag{2}$$

However, the state of pixels in $D_t(x, y)$ may be affected due to the dynamic or illumination changes in the background, which may lead to the appearance of irrelevant pixels on the foreground. In this concern, block wise Shannon entropy and Shannon energy are evaluated to examine the actual moving pixels in the initial motion field. The moving pixels that belong to the constant intensity area have low entropy and energy as compared to the fluctuating intensity region. The entropy and energy depict the information content present in the video. As seen, the lower gray level distributed region has higher energy. Taking these assumptions, the gray levels inside the block of $D_t(x, y)$ having size $c \times c$ are taken to evaluate the

probability density function 's'. The value of 'c' is taken as 8. The 'R' are gray levels inside the block. The Shannon entropy 'E_t' and energy 'EN_t' are computed as follows:

$$E_t = -\sum_{R=R_{\min}}^{R_{\max}} s \log_2(s) \tag{3}$$

$$EN_t = -\sum_{R=R_{\min}}^{R_{\max}} s^2 \log_2(s^2) \tag{4}$$

Based on the ratio of energy to entropy, the approximate moving region is defined and the background update is done.

$$B_t(x,y) = \left\{ \begin{array}{ll} B_{t-1}(x,y) + \text{signum}(I_t(x,y) - I_{t-1}(x,y)), & \text{if } \left(\frac{EN_t}{E} < \tau_1\right) \\ B_{t-1}(x,y), & \text{else} \end{array} \right\} \tag{5}$$

The corresponding motion mask is evaluated as follows:

$$M_t(x,y) = \left\{ \begin{array}{ll} 1 & \text{if } (D_t(x,y) > \tau_2 \text{ or } \frac{EN_t}{E_t} < \tau_1 \\ 0 & \text{otherwise} \end{array} \right\} \tag{6}$$

where 'τ_1' and 'τ_2' are user defined thresholds. Signum is a mathematical function.

3.2 Object Tracking Using Kalman Filtering

The Kalman filter has the capability to estimate tracking positions using the minimum sample size of the detected object. The adaptive Kalman filtering method proposed in [10] is integrated for tracking in the object extraction module. As seen, the tracking may be affected due to unconstrained measurement and local disturbance in the background. These difficulties may be overcome using the accurate predicted sate. The Kalman filter utilizes a state model that requires current input and previous output to estimate the next location in the successive frames.

The matrices that belongs state $\hat{T}(t)$ and measurement $m(t)$ model are defined as

$$\hat{T}(t) = B\hat{T}(t-1) + p(t) \tag{7}$$

$$m(t) = H(t)\hat{T}(t) + q(t) \tag{8}$$

where 'B' stands for the state transition matrix and $H(t)$ refers to measurement matrix used in the estimation procedure. The Gaussian noise $p(t)$ and $q(t)$ having the zero mean may arise in the system model due to unconstrained measurement. The white noise is assuming in this experiment.

The filter predicts the next state $\hat{T}^+(t)$ by incorporating the prior estimate of state $\hat{T}^-(t)$. The prior state used for the actual measurement in the system.

$$\hat{T}^+(t) = \hat{T}^-(t) + K(t)(m(t) - H(t)\hat{T}^-(t)) \tag{9}$$

'$K(t)$' stands for the Kalman gain and is expressed as

$$K^+(t) = \hat{P}^-(t)H(t)^T(H(t)\hat{P}^-(t)H(t)^T + R(t))^{-1} \tag{10}$$

The Kalman gain includes a prior error covariance matrix $\hat{P}^-(t)$ and $\hat{P}^+(t)$.

The aim is to estimate correct state using Eq. (9) through correcting the Kalman gain using Eq. (10). As seen, higher would be the Kalman gain, it will reduce the measurement error. The final aim is to get a posterior covariance matrix using Eq. (11). The previous posterior estimate is utilized to compute a new prior estimate in order to correct the measurement.

$$\hat{P}^+(t) = (I - K(t)H(t))\hat{P}^-(t) \tag{11}$$

4 Experimental Analysis

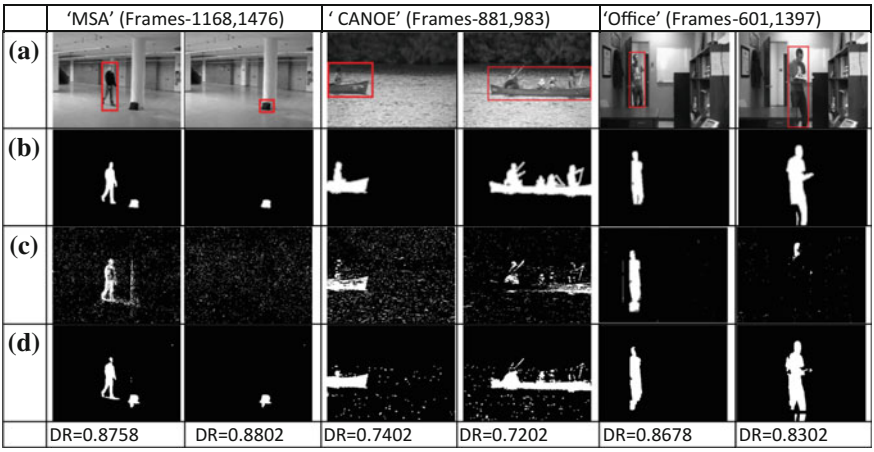
In this section, we have shown the visual and quantitative performance on some experimental video sequences. In this regard, we consider 'MSA,' 'CANOE,' and OFFICE sequences that consist some deviations in the background. All the experiment results are simulated on MATLAB 7.1 on a desktop with configuration 3.2 GHz Intel CPU, 2 GB RAM.

In Fig. 1, first row shows the sampled video frames. The first row of Fig. 1 also shows tracking results through the proposed method, while the last row shows the segmented results of this method. As one can observe, the proposed method classifies accurately between the foreground and background in both static and dynamic background conditions. Moreover no ghost effect, over-segmentation error and aperture distortion are seen on the foreground mask.

The parameters Similarity, F1, and Detection rate are the quantitative metrics. These metrics explain the output image with respect to its ground truths. These metrics depend on '*tn'*, '*fp'*, '*fn,'* and *'tp',* which are true negative, false positive, false negative, and true positive pixels, respectively. The '*fp'* and '*fn*' are the mistakenly detected foreground and background pixels, respectively. The '*tp*' and '*tn*' are accurately detected foreground and background pixels, respectively.

The parameters Detection Rate, Similarity, and F1 are computed as

$$\text{Detection Rate} = \text{tp}/(\text{tp} + \text{fn}) \tag{12}$$



DR*=Detection rate

Fig. 1 **a** Sample frames with tracking results **b** Ground truth **c** Results using GMM method **d** Foreground output using proposed method

$$\text{Similarity} = \text{tp}/(\text{tp} + \text{fp} + \text{fn}) \tag{13}$$

$$\text{F1} = 2 \times \text{Precision} \times \text{Recall}/(\text{Precision} + \text{Recall}) \tag{14}$$

where precision and recall are the irrelevant and relevant true positive pixels, respectively. Figure 1 presents the detection rate on sampled frames through this proposed method. Table 1 shows the comparison between GMM and proposed method. Here, GMM [6] initially have good detection rate, but its performance is tainted on subsequent frames when object either becomes stationary or moves near camouflage region. However, the detection rate through our method is far better than GMM for each video sequence. The average similarity and F1 obtained through this method is up to 60 % greater than GMM for the MSA sequence. However, in dynamic background of CANOE sequence, it is approximately 21 % greater than GMM method. The visual inspection and quantitative analysis exemplify that the method provide enough cues to meet the requirement of video surveillance.

Table 1 Performance comparison between proposed method and GMM method

Sequences	Evaluation	Proposed method	GMM
MSA	Similarity	0.8551	0.2726
	F1	0.9219	0.3001
Canoe	Similarity	0.6834	0.4420
	F1	0.8119	0.6370
Office	Similarity	0.8248	0.2721
	F1	0.9040	0.4279

5 Conclusion

In this work, a background-updating scheme is reported to update the dynamic background pixels, which in turn provided a sufficient sample size to the tracking module and enhanced the results. Based on the ratio of Shannon energy to entropy, the relevant moving pixels are accessed on the foreground. The proposed method efficiently reduces the over-segmentation error, aperture distortion, and ghost effect. The accurately segmented object on the foreground may provide important cues to many postprocessing applications. One can focus to extend this work for multiple object detection and tracking in unconstrained videos. Experimental results on some challenging video sequences show that the method outperforms the other state-of-the-art background subtraction methods used in tracking.

References

1. R. Cucchiara, C. Grana, M. Piccardi, A. Prati: Detecting moving objects, ghosts, and shadows in video streams. IEEE Transactions on Pattern Analysis and Machine Intelligence 25, 1337–1342 (2003).
2. Antoine Manzanera, Julien C. Richefeu: A new motion detection algorithm based on Σ-Δ background estimation. Pattern Recognition Letters 28, 320–328 (2006).
3. Weiming Hu, Tieniu Tan, Liang Wang, S. Maybank: A Survey on Visual Surveillance of Object Motion and Behaviors. IEEE Transactions on Systems Man and Cybernetics 34, 334–352 (2004).
4. Mandellos N. A., Keramitsoglou I., Kiranoudis C. T.: A background subtraction algorithm for detecting and tracking vehicles. Expert Systems with Applications 38, 1619–1631 (2011).
5. Lucas B. D., Kanade T.: An iterative image registration technique with an application to stereo vision. IJCAI 81, 674–679 (1981).
6. Stauffer C., Grimson W. E. L.: Learning patterns of activity using real-time tracking. IEEE Transactions on Pattern Analysis and Machine Intelli. *22*, 747–757 (2000).
7. M. Oral, U. Deniz: Center of mass model: A novel approach to background modeling for segmentation of moving objects. Image and Vision Computing 25, 1365–1376 (2007).
8. Jing G., Siong C. E., Rajan D.: Foreground motion detection by difference-based spatial temporal entropy image. IEEE Conference 2004 379–382 (2004).
9. Fu Z., Han, Y.: Centroid weighted Kalman filter for visual object tracking. Journal of Measurement 45, 650–655 (2012).
10. Weng S. K., Kuo C. M., Tu S. K.: Video object tracking using adaptive Kalman filter. Journal of Visual Communication and Image Representation 17, 1190–1208 (2006).
11. Yao, Li, Ling, M.: An Improved Mixture-of-Gaussians Background Model with Frame Difference and Blob Tracking in Video Stream. The Scientific World Journal 2014, 1–9 (2014).