

Tracking Objects on Detector Response using Extended Kalman Filter



Hemavathy R, Shobha G

Abstract: Video analytics plays a very important role in identification or detection and tracking of objects, this intern find application in many fields and domains. Novel learning methods or techniques built on Neural Networks requires larger dataset for training the results, the output obtained depends on how well the training is done. The proposed method of Weighted Cumulative Summation (WCS) is an approach based on background modelling to segment the moving objects. This method adapts and tunes the background variations instantaneously as the video frame arrives. The segmentation obtained is compared with other basic methods. The result obtained infers improvements in segmentation and in removal of ghost effect in the video. Extended Kalman Filter (EKF) is used to track the detector response. The responses of the detection from WCS are provided as input to EKF to track the moving object. The results are tabulated and represented in the form of graphs for analysis. The results are compared with three different video datasets and the results are noticeably good. The methods WCS can be used in the applications where data set is not available.

Keywords: Detector response, Kalman filters, moving objects, segmentation, tracking .

I. INTRODUCTION

Computer Vision is an expanding area of research that amalgamates the multifaceted problems of the dynamic environment with the powerful practices and techniques of computer vision. Under water videos or images disclose many interesting and useful information hidden beneath the ocean. To gain the knowledge and understanding of the underwater animal activities and its habitat, an automated detection and tracking system is necessary. The motivation behind this research is to correlate the knowledge of computer science with basic science, and to extend it for developing an automated system.

II. EXISTING BACKGROUND SUBTRACTION APPROACH

A. Introduction to segmentation

Deep Learning and Artificial neural network concepts are used for computer vision and image processing application

nowadays however these techniques uses a huge dataset for learning phase and the time taken for training is also more (time consuming). There are few applications and domains where huge data set or information is not available for research. So in such cases segmentation will be more efficient compared to the learning concepts.

B. Background Subtraction, Adaptive Background Learning (ABL) and Adaptive Selective Background Subtraction

Basic background (BG) modelling [1-3] is used to remove the background and to extract the foreground by performing subtraction. To adapt to the variations in the environment, the background (BG) is updated as video frames proceed. Frame difference method is the earliest and the simplest one used, which performs the difference between two consecutive frames to obtain the object displacement. To this result, threshold is applied and the binary segmented image is obtained. The absolute difference between two consecutive frames is given in equation 1. Where $cf(n)$, $cf(n-1)$ being the n^{th} , $n-1^{th}$ frame respectively.

$$bg = |cf_{(n)} - cf_{(n-1)}| \quad (1)$$

In Adaptive Background Learning (ABL) [1-3] approach, initially starting frame is considered as background denoted by bg as shown in equation 2. The bg is updated as the frames proceeds as shown in equation.3, this is termed as background modelling. Equation 3 contains a parameter alpha (α), cf current frame, i, j image coordinates, and α is equal to 0.05 a weightage parameter.

$$bg = cf \quad (2)$$

$$bg_{(i,j)} = (\alpha)cf_{(i,j)} + (1 - \alpha)bg_{(i,j)} \quad (3)$$

Background is updated by adding 5% of the current frame and 95% of the previous BG frame. Current frame is subtracted with the background first and then threshold is applied. A variation of ABL is Adaptive Selective Background Learning (ASBL) [3] which contains learning phase and updates the bg as in equation 3. After the learning phase is completed, the bg starts updating only for background pixel of the segmented image till the end of the video frame.

ASBL and ABL are unsupervised methods, it uses simple equations and execution time is less. For static background with conditioned environment these methods are sufficient. Hence, these facts are considered to improvise the existing methods. Even though there are many techniques which give more accuracy of segmentation, they require more execution time due to calculation of complex mathematical equations and in turn, require more memory storage.

Revised Manuscript Received on April 18, 2020.

* Correspondence Author

Hemavathy R*, Computer Science & Engineering, R. V. College of Engineering, Bengaluru, India. Email: hemavathy@rvce.edu.in

Shobha G, Computer Science & Engineering, R. V. College of Engineering, Bengaluru, India. Email: shobha@rvce.edu.in

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](http://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

An improvement over the earlier method may lead to new development or extension.

III. PROPOSED BACKGROUND MODELLING

The proposed technique is improvised version of Adaptive Background Learning model (ABL) and Adaptive Selective Background Learning (ASBL) model [99]. The proposed technique utilizes Cumulative weight Summation concepts. BG is modelled by a series of weightage on cumulative sum. This is a deterministic model. ABL and ASBL discern that the BG updated accords more weightage to the previous *bg* modelled BG, than the current frame, that results in ghosting effects in the segmented image. The ghosting effect formed in segmented image, is due to the presence of traces of previous frame while creating the BG. In the proposed method, ghosting effects and False Positive (FP) is considerably reduced.

Background (BG) modelling phase: Modelling BG is very important as frame differencing will not segment the moving object properly. Once BG is modelled, the segmentation will be more accurate and any variations in the video can be updated. Three previous intermediate backgrounds are *bg1*, *bg2*, *bg3* which are initialized to the starting frame Each time the background is updated by 10% of current frame *cf*, 20% of *bg1*, 30% of *bg2* and 40% of *bg3*.

By considering the set of *bgs* as explained, traces of the previous frame in the background is reduced. After the background is updated, the absolute difference between the current frame and the background is performed to obtain the result.

$$bg = W_1 cf_{(n)} + W_2 \sum_{k=1}^{n-1} f_{(k)} + W_3 \sum_{j=0}^{n-2} f_{(j)} + W_4 \sum_{i=0}^{n-3} f_{(i)} \quad (4)$$

$$bg_1 = \sum_{k=1}^{n-1} f_{(k)} \quad (5)$$

$$bg_2 = \sum_{j=0}^{n-2} f_{(j)} \quad (6)$$

$$bg_3 = \sum_{i=0}^{n-3} f_{(i)} \quad (7)$$

$$bg = W_1 cf_{(n)} + W_2 bg_1 + W_3 bg_2 + W_4 bg_3 \quad (8)$$

Updating phase: The updating phase continues updating the background and adapts the variation of the preceding or successive frames in a video. Each time the calculated *bg* is used in the equation 8 for next iteration and swapping of consecutive *bgs* are shown in equation 9, 10, 11 below.

$$bg_3 = bg_2 \quad (9)$$

$$bg_2 = bg_1 \quad (10)$$

$$bg_1 = bg \quad (11)$$

The swapping of statements updates the *bg* values and iterates through the last frame in the learning phase.

Adaptive selective updating phase: After the learning phase is completed, in the adaptive selective phase, the modelled BG is updated, only for the pixels considered as part of background in the segmented binary image. By updating only the BG of the output frame, the time of execution is significantly reduced. This proposed approach is applied on the videos with static background having fewer variations.

Steps in adaptive selective phase are as follows:

Step 1: Apply background subtraction and segment the object

Step 2: Check the pixel intensity of the segmented frame, if pixel value is 0 go to step 4 else go to step 5.

Step 4: Update BG model for pixel value 0.

Step 5: Loop through step 2 to 4 until the end of the frame.

This phase runs throughout the end of the video, since it updates only the pixel intensity value 0, the time of execution is less. This approach can be applied for static conditioned videos with less variation. Since background modelling is based on three previous background estimations, the traces of the previous frames will be reduced noticeably in the results, as depicted later in the result analysis.

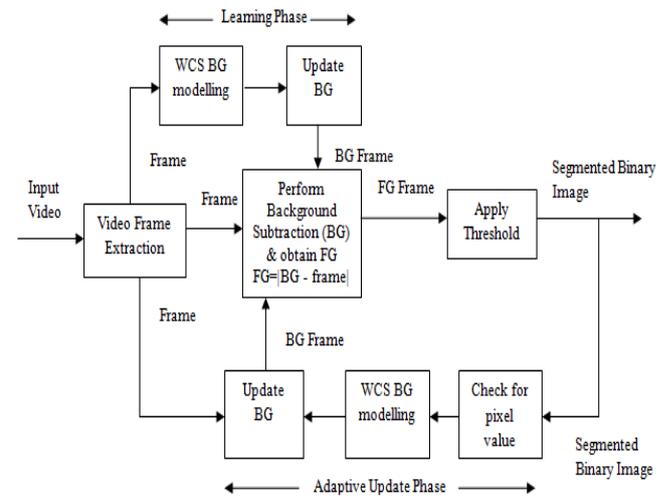


Figure 1 Modules of Weighted Cumulative Summation (WCS) method

Figure 1 explains the WCS background modelling and the operations performed as discussed. Results of the proposed work are documented and analyzed in the succeeding sections. The output of this module is used as detector response and tracked using Extended Kalman Filter.

IV. TRACKING BY DETECTOR RESPONSE

The responses of the detection are used to track the moving object. Since EKF [4,5] evaluates both the predictors and corrector equations to identify the changes in the video frames. This algorithm uses a series of measurements observed over time, with random variations, estimates unknown variables that are more accurate. The videos considered in this research work are dynamic in nature and the time taken to track will be more as evaluation and updating phases take time for execution. Tracking can be made more efficient in terms of execution time if the variations and noise under dynamic conditions are eliminated considerably. It reduces computation time considerably since only pixels belonging to the foreground objects need to be dealt with.

At the time *k*, there is information on the forecast X^f_k with the covariance P^f_k , the measurement Z_k with the covariance R_k . K_k is the unbiased Kalman gain.

Data Assimilation Step/Corrector:

$$K_k = P_k^f (J_h^T(X_k^f) (J_h(X_k^f) P_k^f J_h^T(X_k^f) + R_k))^{-1} \quad (12)$$

$$X_k^a \approx X_k^f + K_k (Z_k - h(X_k^f)) \quad (13)$$

$$P_k = (I - K_k J_h(X_k^f)) P_k^f \quad (14)$$

It is a continuous cycle of prediction and update. Now k becomes $k + 1$ and loops till the end. Moving object is tracked using forecast step which forecasts the location of the object in the present frame and rectifies the location based on measurement noise. EKF tracks the objects throughout the video frames[4,5].

The random variables V_k represent the measurement noise, W_k represent model noise or measures uncertainties in the model, V_k and W_k are uncorrelated. X_k is the state vector, Z_k observational vector, $f(\cdot)$, $h(\cdot)$ represent non linear vector function.

$$\begin{aligned} X_k &= f(X_{k-1}) + W_{k-1} \\ Z_k &= h(X_k) + V_k \end{aligned} \quad (15)$$

In the initial stage, the only information available is mean μ_0 and covariance P_0 . The initial optimal estimate X_0^a and P_0 error covariance are given in the equation 16

Initialization:

$$X_0^a = \mu_0 \text{ with error covariance } P_0 \quad (16)$$

The forecast value of X_k^f is as in equation 17. Forecast error covariance P_k^f is in equations 17 and 18. J_f the Jacobian of $f(\cdot)$. Q_{k-1} , is the process noise covariance.

Model Forecast Step/Predictor:

$$X_k^f \approx f(X_{k-1}^a) \quad (17)$$

$$P_k^f = J_f(X_{k-1}^a) P_{k-1} J_f^T(X_{k-1}^a) + Q_{k-1} \quad (18)$$

At the time k , there is information on the forecast X_k^f with the covariance P_k^f , the measurement Z_k with the covariance R_k . K_k is the unbiased Kalman gain.

Data Assimilation Step/Corrector:

$$K_k = P_k^f (J_h^T(X_k^f) (J_h(X_k^f) P_k^f J_h^T(X_k^f) + R_k))^{-1} \quad (19)$$

$$X_k^a \approx X_k^f + K_k (Z_k - h(X_k^f)) \quad (20)$$

$$P_k = (I - K_k J_h(X_k^f)) P_k^f \quad (21)$$

It is a continuous cycle of prediction and update. Now k becomes $k + 1$ and loops till the end. Moving object is tracked using forecast step which forecasts the location of the object in the present frame and rectifies the location based on measurement noise. EKF tracks the objects throughout the video frames.

Tracking by detector response approach is used in this project to exploit the benefits of the proposed segmentation algorithm. The tracking works in two folds:

1. Segment or detect the object in the video using the WCS method.
2. Track the detector response using EKF [4,5].

Decrease in the time of tracking could be noticed. The time taken in estimation and updating is considerably reduced, due to the elimination of variation by the detector response. The modules used in Tracking by Detector contains three blocks as shown in Figure 2, the frame extractor, extracts each frame from a video and provides input frame to the segmentation process.

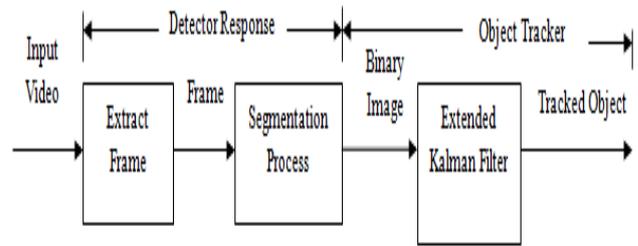


Figure 2 Modules involved in the tracking process of the multiple object

The output of this process is a binary frame with reduction of noise, which is connected to EKF module. The EKF module tracks the segmented frame using prediction and correction. Since the detector response is clearer in dynamic environment the performance of the tracking task is improved.

V. RESULT ANALYSIS

The tracking of segmented objects under static and dynamic environment is performed by EKF. The execution time is considerably reduced since only the frames related to objects in motion are segmented by reducing the effects of unwanted variations and movements. The performance of Tracking by Detector response is checked on both the types of video datasets and the time of execution is tabulated and depicted in the graph. The research video dataset is downloaded in the Fish4Knowledge website [6].

Figure 3 illustrates tracking on three different algorithms Table 1 provides the time in milliseconds per frame (ms/f) and Figures 4 and 5 depict the performance of tracking by detector response.

Table 1 Tracking time on ABL, ASBL and WCS method in milliseconds/frame

Frame Number	EKF with ABL(ms/frame)	EKF with ASBL(ms/frame)	EKF with WCS Method(ms/frame)
46	0.10233	0.0204827	0.0112224
48	0.0714094	0.0122343	0.00539931
50	0.167334	0.0125604	0.00492658
52	0.118584	0.00957482	0.00454803
54	0.106422	0.0175757	0.00681147
57	0.0976663	0.0181666	0.00415318
60	0.114058	0.0101762	0.00417733
63	0.497748	0.0199073	0.00359471
66	0.252712	0.0177665	0.00442184
70	0.16477	0.0165421	0.00429143
Average	0.16930337	0.015498662	0.005354628

The average time taken is about 0.01549ms/f in ASBL and 0.005354 ms/f in the proposed method; it is around 2.9 times faster than the ASBL. Hence, increase in the speed of tracking. Figures 4 and 5 give the graphs showing the performance improvement.

Video1: In static background [6]

Frame Number	Input Frame	EKF with ABL	EKF with ASBL	EKF with WCS Method
46				
48				
50				
52				
54				
57				
60				
63				
66				
70				

Figure 3 EKF applied on ABL, ASBL and on WCS background modelling

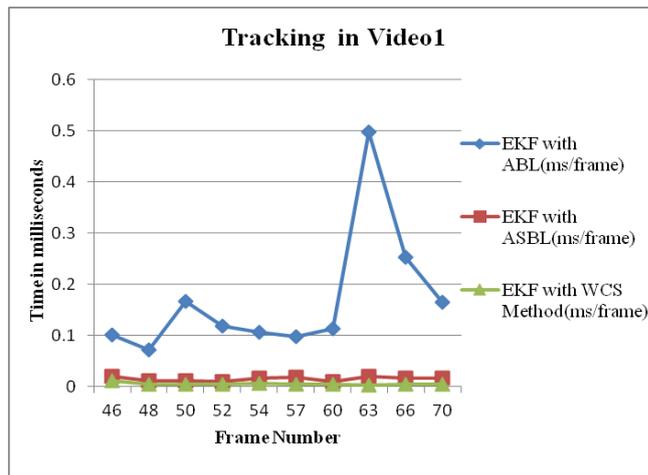


Figure 4 Tracking time on ABL, ASBL and WCS methods

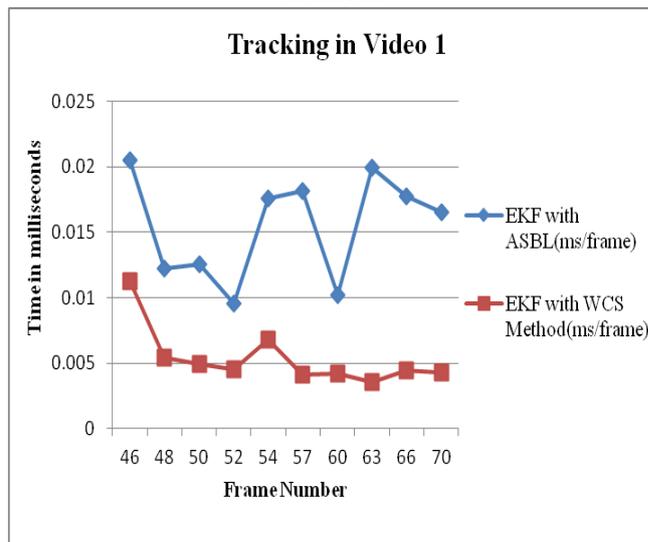


Figure 5 Tracking time of ASBL and WCS method
Video 2: In static background

Frame Number	Input Frame	EKF with ABL	EKF with ASBL	EKF with WCS Method
13				
41				
64				

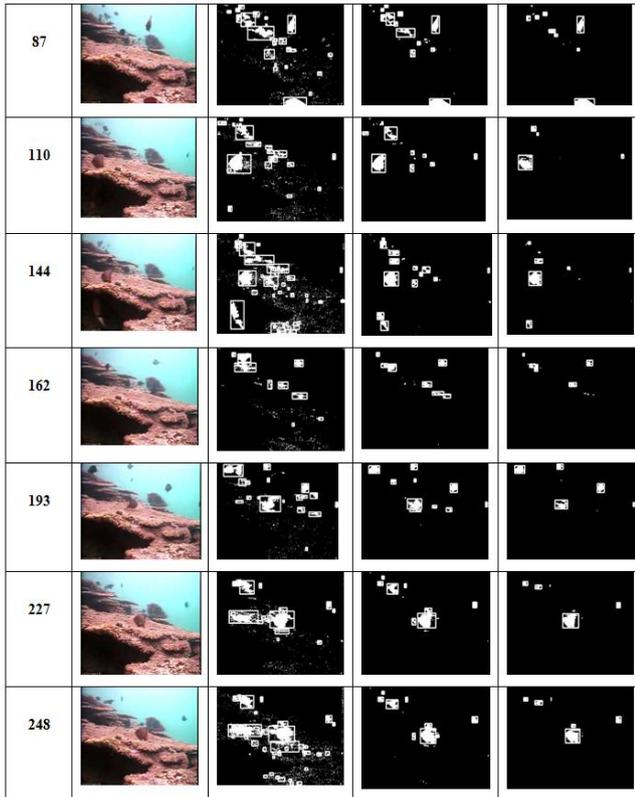


Figure 4 EKF applied on ABL, ASBL and on WCS proposed method

Figure 4 shows the tracking of the objects. Considerable reduction in the tracking time is seen as per measurements that are tabulated in Table 2

Table 2 Shows tracking time of ABL, ASBL and WCS method in milliseconds/frame is tabulated

Frame Number	EKF with ABL(ms/frame)	EKF with ASBL(ms/frame)	EKF with WCS Method(ms/frame)
13	0.166169	0.0190506	0.00616486
41	0.603448	0.017627	0.00525381
64	0.285301	0.0141681	0.00505155
87	0.670032	0.0168639	0.00497971
110	0.582878	0.0101043	0.00426125
144	0.394604	0.00500405	0.00273316
162	0.118588	0.00541078	0.00332725
193	0.134022	0.00511533	0.00353554
227	0.14002	0.0063321	0.00262509
248	0.133276	0.00605075	0.00245846
Average	0.3228338	0.010572691	0.004039068

The average time taken is about 0.01549ms/f in ASBL and 0.005354 ms/f in the proposed method; it is about 2.7 times quicker than the ASBL. Hence, increase in the speed of tracking. Figures 5 and 6 give the graph showing the performance improvements

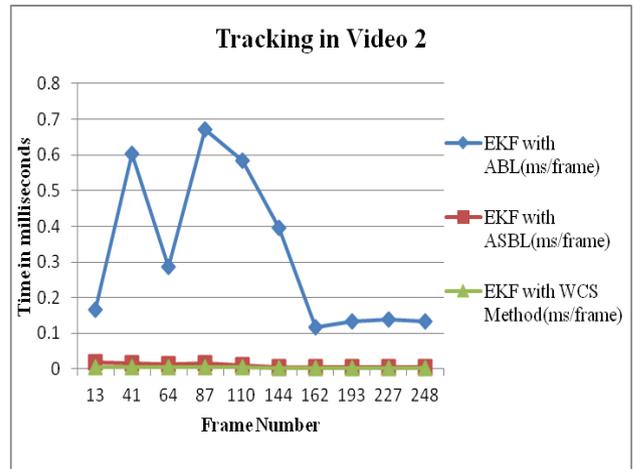


Figure 5 Tracking time on ABL, ASBL and WCS method

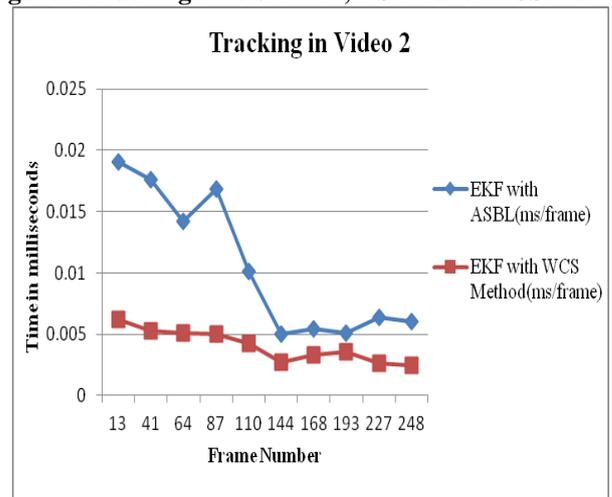
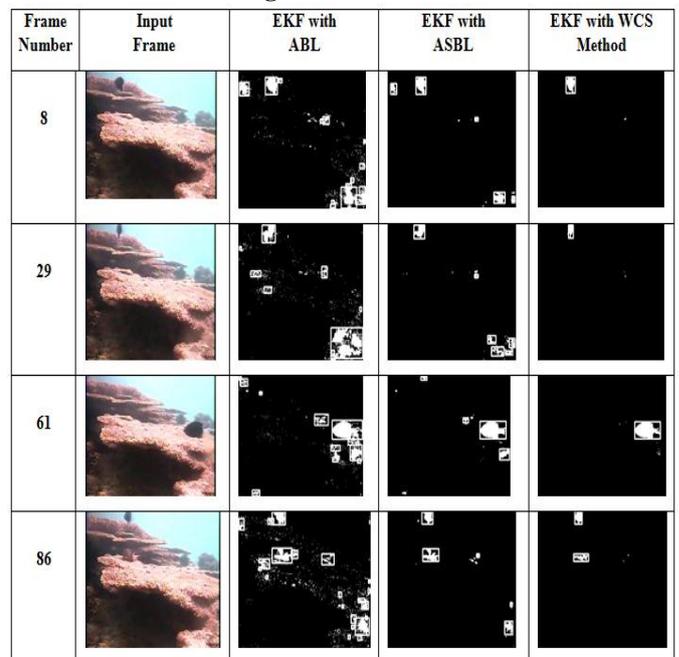


Figure 6 Tracking time of ASBL and WCS method Video 3: In static background



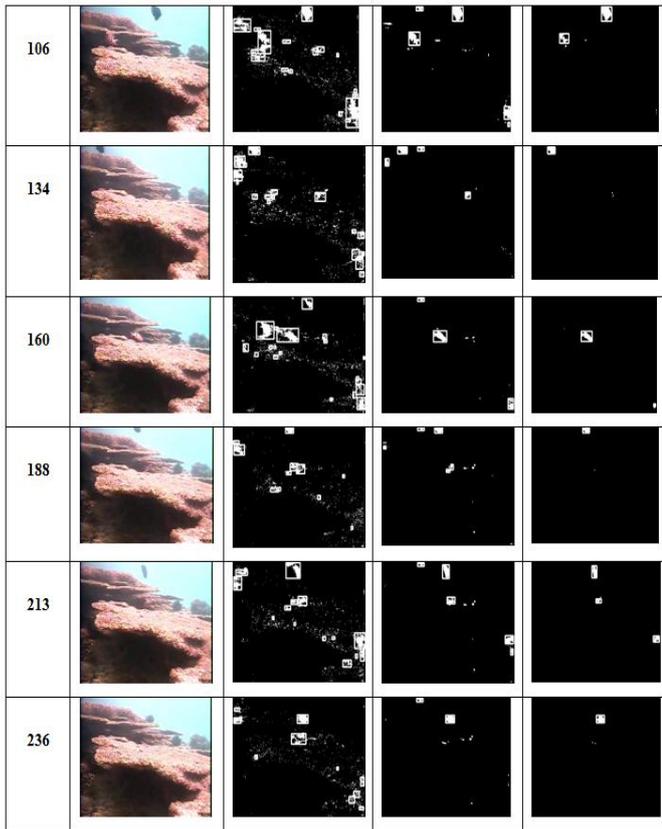


Figure 7 EKF applied on ABL, ASBL and WCS method

Table 3 Tracking time of ABL, ASBL and WCS on video 3

Frame Number	EKF with ABL(ms/frame)	EKF with ASBL(ms/frame)	EKF with WCS Method(ms/frame)
8	0.706176	0.0223235	0.004255279
29	0.456046	0.0103669	0.00208715
61	0.494423	0.00637738	0.00159208
86	0.232031	0.00567703	0.00226043
106	0.287844	0.01971720	0.00301149
134	0.443724	0.01833880	0.00278086
160	0.303676	0.00610026	0.00270962
188	0.248175	0.00903688	0.0044786
213	0.25864	0.00711214	0.00410548
236	0.303795	0.01446640	0.00198935
Average	0.373453	0.011951649	0.0029270339

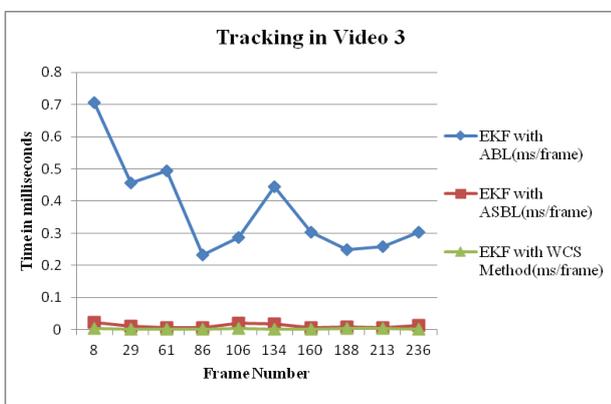


Figure 8 Tracking time on ABL, ASBL and WCS method

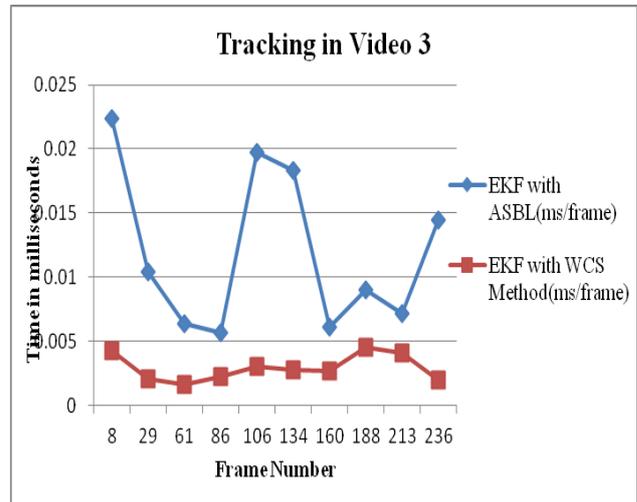


Figure 9 Tracking time for ASBL and WCS method

As observed from the Table [3] the average time taken is about 0.011951649 ms/f in ASBL and 0.0029270339 ms/f in the proposed method, it is about 4 times faster than the ASBL. Hence increase in the speed of tracking. Figure 8 and 9 gives the graph showing the performance improvement.

From the above observations, it can be concluded that the time taken for tracking is reduced by 3.2 times as compared with ASBL.

VI. CONCLUSION

In the method the popularly used EKF is applied to observe the performance of the proposed segmentation methods. There is a very good improvement in the speed of tracking and reduction in execution time. In Weighted Cumulative Summation approach (WCS) the following points are observed, WCS method uses unsupervised approach as the WCS background modelling uses simple arithmetic equations, the time of execution is considerably less as compared to other methods using stochastic or statistical models with complex equations. WCS method provides noticeable reduction in the ghosting effects caused due to the presences of traces of previous frames in the background modelling. There is improvement in precision, on an average it is about 1.7 times higher and PWC is found to be 2.4 times lesser when compared to ASBL. Count of the total number of objects tracked is less in the proposed method when compared to the traditional method as it reduces the false positive pixels noticeably. Even the time taken is significantly reduced in the proposed WCS method; it is about 4.3 times faster than the ASBL method.

REFERENCES

1. Stauffer Chris and W. Eric L. Grimson, "Adaptive background mixture models for real-time tracking", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2, 1999.
2. Zivkovic, Zoran. "Improved adaptive Gaussian mixture model for background subtraction." IEEE 17th International Conference on Pattern Recognition (ICPR), Vol. 2, 2004, pp. 28-31.
3. Andrews Sobral, "BGSLibrary: An OpenCV C++ Background Subtraction Library", Ninth Workshop de Visao Computacional (WVC), Rio de Janeiro, Brazil, Jun 2013, pp. 2-10.

4. Ribeiro, Maria Isabel. "Kalman and extended kalman filters: Concept, derivation and properties", Institute for Systems and Robotics Feb 2004: 43
5. Terejanu, Gabriel A., "Extended Kalman Filter tutorial", online Disponible:<http://users.ices.utexas.edu/~terejanu/files/tutorialEKF.pdf>, 2008.
6. Fishforknowledge website <http://www.inf.ed.ac.uk/index.html>

AUTHORS PROFILE



Dr. Hemavathy R.B.E, M.Tech. Ph.D (VTU), Associate Professor, Department of CSE, R. V. College of Engineering, Bangalore, She has 18 years of teaching experience. Research on Computer Vision , Machine Learning. She has guided many UG, PG students. Worked on research and consultancy projects. Authored more than 15 journals, ISTE, IEEE member.



Dr. Shobha G, Professor, Department of CSE, R. V. College of Engineering, Bangalore, She has 25 years of teaching experience. She has guided UG, PG and 7 PhD students. She has received Career Award for young teachers by All India Council for Technical Institution for the year 2008-09, IBM Shared University Research Award in 2019, she received Accenture Woman award for best teacher. She was Principal Investigator for research projects under NRB, ARB, DRDO, AICTE, Mission on Food Processing, Govt. of Karnataka. Mentoring consultancy projects at CISCO Sys, IBM, CITRIX. She has authored around 70 Journals. ISTE, IEEE , CSI member.