Hybrid Deep Learning Signature based Correlation Filter for Vehicle Tracking in Presence of Clutters and Occlusion

Shobha B.S, Deepu.R
Department of Computer Science and Engineering
ATME College of Engineering, Mysore, Mysore, India

***Abstract*—This vehicle tracking is an important task of smart traffic management. Tracking is very challenging in presence of occlusions, clutters, variation in real world lighting, scene conditions and camera vantage. Joint distribution of vehicle movement, clutter and occlusions introduces larger errors in particle tracking based approaches. This work proposes a hybrid tracker by adapting kernel and particle-based filter with aggregation signature and fusing the results of both to get the accurate estimation of target vehicle in video frames. Aggregation signature of object to be tracked is constructed using a probabilistic distribution function of lighting variation, clutters and occlusions with deep learning model in frequency domain. The work also proposed a fuzzy adaptive background modeling and subtraction algorithm to remove the backgrounds and clutters affecting the tracking performance. This hybrid tracker improves the tracking accuracy even in presence of larger disturbances in the environment. The proposed solution is able to track the objects with 3% higher precision compared to existing works even in presence of clutters.**

***Keywords*—Smart traffic management; background subtraction; vehicle detection; aggregation signature; hybrid tracker**

I. Introduction

The Video surveillance based smart traffic management is the cost-effective solution for traffic monitoring and control. With the convergence of image processing and artificial intelligence, various traffic related applications like vehicle classification, congestion detection, accident detection, over-speeding detection, tracking vehicles etc. can be realized for smart traffic management. Vehicle tracing is an important application in smart traffic management for various traffic violation detections and criminal forensics.

Traditional vehicle tracking algorithms predict the vehicle location in consecutive frames based on its position in the previous frames. Some of the well-known tracking algorithms in this category are Kalman filter [1], mean shift [2], particle filter [3] and tracking learning detection [4]. Many methods have been proposed extending these approaches. Edge feature-based extension to Kalman filter [5], SIFT based extension to mean shift filter [6], integrated mean shift with particle filter [7] and long-term TLC combining detection and tracking [8] are some of the works extending traditional tracking algorithms. The traditional algorithms and its extension have many challenges in tracking due to clutters, occlusion, speed of vehicle and lighting conditions [9]. In presence of these challenges, target vehicle localization becomes erroneous. The five important challenges considered in this work are lighting variations, object occlusion, fast movement, shape deformation and false positive. These challenges result in failure to detect and track the vehicles. Occlusion makes object detection difficult both in current frame and for updating future images. Fast movement introduces blurred effect in video frames. Clutters and occlusion jointly introduce either partial or complete deformation of the target object. False positives are introduced due to presence of multiple similar objects [10].

Based on functionality, the existing tracking methods can be divided into two classes: background modeling and foreground modeling [11]. Background modeling-based approaches uses the first frame or few initial frames to build a background model, and then compare the current frame with the background model to detect moving foreground objects, and finally review the background model. Most of the backgrounds modeling methods are supported on Bayesian filtering, Kalman and particle filters. Background modeling approaches suffer from illumination changes and dynamic background. Detailed survey on vehicle detection, recognition and tracking is provided [12].Foreground modeling approach uses vehicle shape properties to track the vehicle. Various features like local binary pattern (LBP), histogram, Scale invariant feature transform (SIFT) etc. are used for vehicle detection but in presence of occlusions, vehicle shape distortions create higher false positives in vehicle tracking. Thus the proposed solution addresses the problem of tracking in presence of lighting variations, object occlusion, fast movement, shape deformation and false positive which affects the precision of tracking the vehicles.

In our proposed work we combine the best of both background and foreground modeling for effective vehicle tracking. The proposed solution preprocesses the video using background modeling that is adaptive to eliminate the effect of shadows, illuminations and clutters in the foreground. A novel hybrid tracker combining particle filtering with aggregation signature-based object detection is proposed for accurate tracing of object in presence of clutters and occlusion. The proposed aggregation signature is a deep learning feature of probability distribution of object in presence of clutter and occlusions in the frequency domain.The regions in the image are scored combining the results of particle filters and aggregation signature tracking and the target vehicle is localized in the higher scored region.Following are the important contributions of this work.

1) Adaptive background modeling to eliminate shadows, illuminations and clutters.

2) A novel deep learning aggregation signature feature for occlusion and clutter distributed target object in frequency domain applying Quaternion Discrete Cosine Transform (QDCT) in frequency domain.

3) An integrated hybrid tracker combining particle filter with deep learning-based aggregation signature feature.

II. Literature Survey

The existing deep learning models for tracking of objects are detailed and the issues are presented in this section.

Many deep learning models have been proposed for tracking of vehicles. These deep learning models are used separately or in combination with particle filter for tracking the vehicles. Zhang et al. [14] proposed a method to track vehicles using recurrent convolutional neural network. Recurrent network trained with reinforcement learning predicts the future locations in the next frames. Learning is done based on continuous inter frame correlation. The approach does not consider the presence of occlusions in learning process. Shobha B S et al. [13] proposed a traffic adaptive multiple vehicle classification using deep learning which tries to address the problem of vehicle categorisation in presence of clutters and occlusion.Multiple online trackers are fused for vehicle tracking by Leang et al. [15]. Set of trackers are combined in different stages of tracking. But in presence of occlusions, the localization error is amplified. Chen et al. [16] addressed the problem of occlusions in tracking but their approach was template based and computational complexity was very high. Hua et al. [17] used deep learning to localize the patches of interest. But the accuracy of localization decreases with decrease in the visibility of patch. Speed was prediction based on localization of patches. But with error in localization patches, the speed prediction also becomes erroneous. A two-stage model to track vehicles is proposed by Liu et al. [18]. In the first stage, recurrent convolutional neural network is adapted to detect object and in the second stage, Kalman filtering is done to track vehicles. In presence of clutters and occlusions, first stage fails and due to this tracking also fails. Integrated YOLOv3 with ORB tracking is used for localization of vehicles by Song et al. [19]. Feature points needed for ORB cannot be extracted in presence of occlusions and thus this approach fails in presence of occlusions. Xu et al. [20] used faster RCNN to detect objects. Features collected from target object were classified by RCNN to detect object. But the overhead is very high in this approach and in presence of large variances in features due to occlusion, the detection fails. A particle-based filtering based on motion of particles is proposed by Fang et al. [21] to track vehicles. But the model needs to be trained for all poses of vehicle. A fully convolutional region-based detector was used totrack vehicles by Dai et al. [22]. The method is faster compared to R-CNN but it is not resilient to even a small shape distortion. A correlation filter using the activations from convolutional layer of CNN was proposed Danelljan et al. [23]. Though method was able to detect vehicles better than hand crafter features, it is very sensitive to occlusions. Ma et al. [24] model the hierarchy of convolutional layers as a pyramid representation of image and use these multiple levels of abstraction for visual tracking. The correlation between the filters at each layer and the visual appearance is adaptively learnt and the response of each layer to detect the target is analysed. The features from the detected layer are then used for object localization. In presence of occlusions, the detection layer response is varying and thus training becomes cumbersome.. Various filters extending linear correlation filter for object tracking have been proposed. Works by Bolme et al. [27], Henriques et al. (2015)and Wang et al. [30] are some of the noteworthy extensions. Bolme et al. [27] extended the linear filter with sum of square error filtering. Though this extension made the approach robust against lighting, scale, pose and non-rigid transformations, the filter performance degrades in presence of clutters and occlusions.A dual correlation filter extending the linear correlation filter was proposed by Henriques et al. [26]for tracking the vehicles. The computational complexity increases exponentially with increase in the number of objects in the frames and the approach is not resilient to occlusions and clutters.A discriminant correlation filter (DCF) was proposed by Wang et al. [25]. It is an end to end light weight network architecture using Siamese network and back propagation trained to provide a heat map of object location in the frames. In presence of occlusion, the false positive is higher in this approach.Yang et al. [28] proposed a detection-based tracking framework using YOLO deep learning model. A light weight feature extraction for object using YOLO is combined along with filter template matching to localize the object. In presence of occlusions, the features have higher spatial difference compared to the region to be tracked in this method. This reduces the accuracy of tracking.

III. Hybrid Tracker

The architecture of the proposed hybrid technique to track vehicles is given in Fig.1 Background is subtracted from the current frames using a fuzzy adaptive background model. The target vehicle to be tracked is selected by marking a rectangular bounding region. Aggregation signature is learnt for the target vehicle. Kernel correlation filter (KCF) and Particle correlation filter (PCF) are adapted to use the aggregation signature and provide the response. The response of the KCF and PCF filter are fused with weighted Newton fusion function to get the predicted position of the target vehicle in any frame. The details of each of the process are detailed in following subsections.

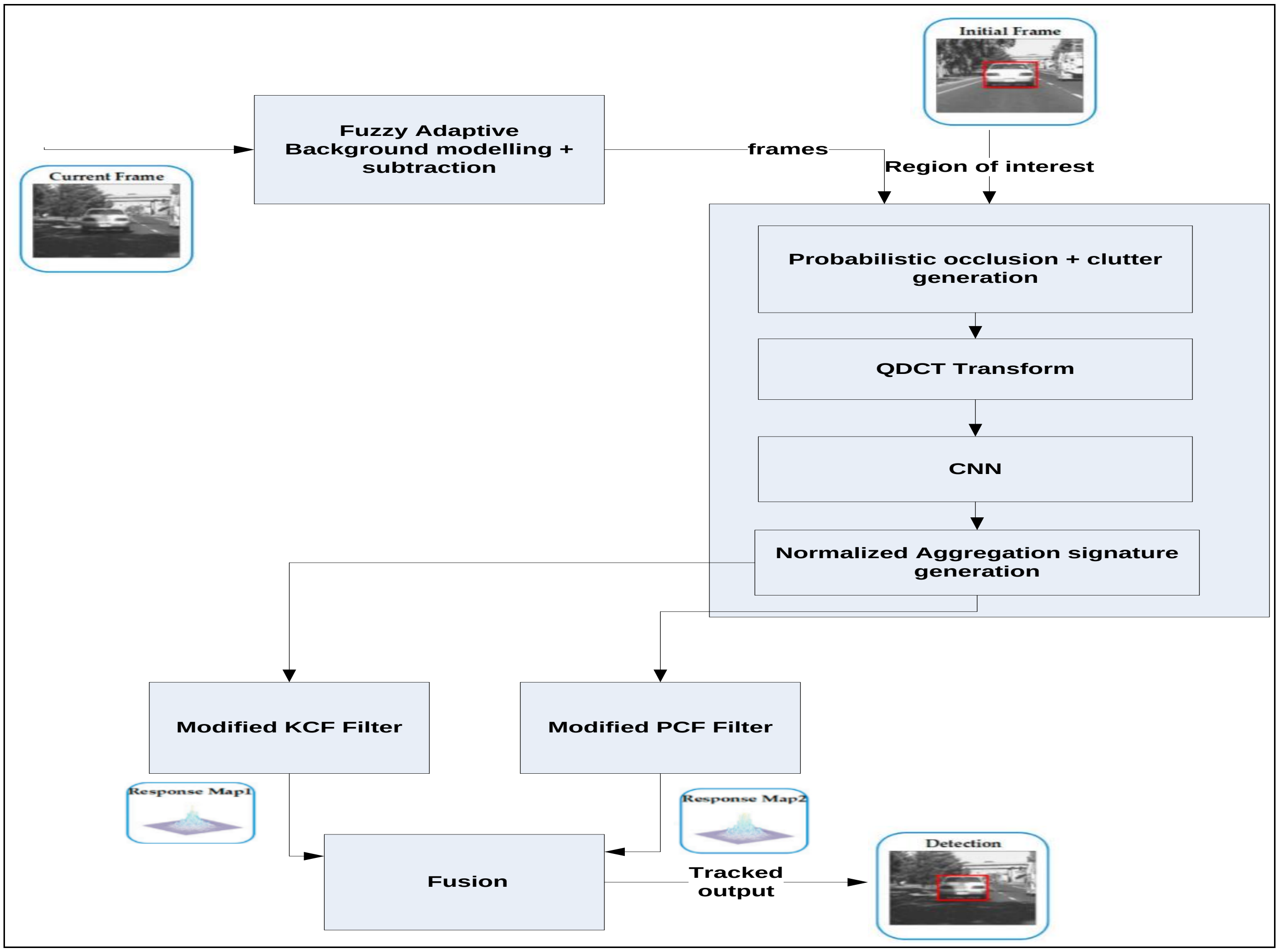

Fig. 1. Hybrid Tracking Architecture.

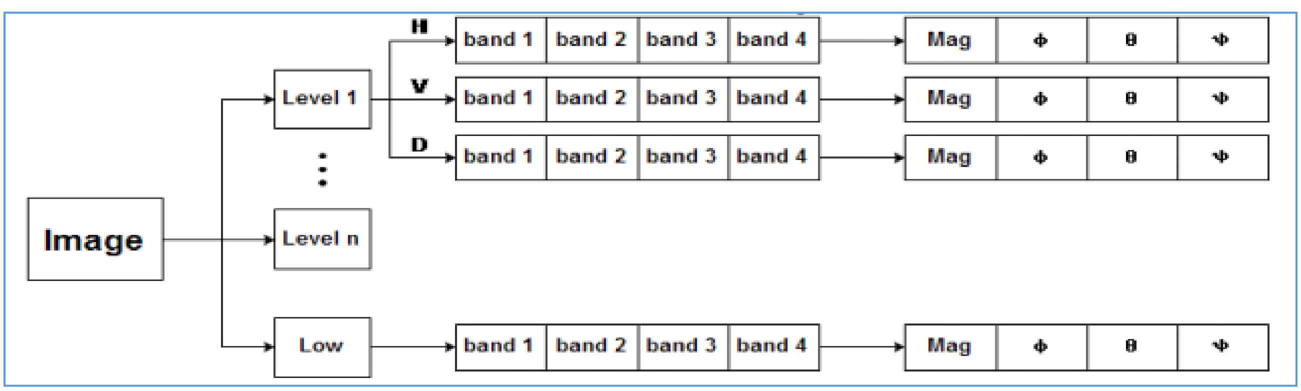

Fig. 2. QDCT Coeffiecints.

A. Fuzzy Adaptive Background Modelling

A fuzzy adaptive background model is constructed from few initial frames and it is updated on every successive frames. The background for every frame is subtracted from this model. The model is designed to eliminate the impact of shadows, illuminations and clutters in the foreground. The initial model is constructed based on first k frames as

$$p_k(x) = \tilde{p}_{k-1}(x) + \frac{1}{G_k\sqrt{2\pi\sigma^2}} exp\,(\frac{-1}{2}(\frac{x - x_k}{\sigma})^2)$$

where

$$\tilde{p}_k(x) = \frac{p_k(x)}{\sum_0^N p_k(x)} \quad (1)$$

Where, x_k is the pixel value observed at the k^{th} frame. The value of $\tilde{p}_k$ is calculated for N frames. The parameter G_k is the learning rate whose value varies from 0 to 1. This value is adapted based on minimization of energy value between the current background and previous background. The energy (U) of a background is calculated in spatial and temporal level as

$$U = \sum_i E_T(i) + \sum_i E_s(i) \quad (2)$$

Where $E_T(i)$ is the temporal energy of a pixel i and $E_s(i)$ is the spatial energy of a pixel i.

$$E_s(i) = \sum_{j \in N_i} Sim(i,i) \quad (3)$$

Where N_i is the eight pixel neighborhood of i and Sim is the similarity function between pixel i and j.

$$Sim(i,i) = \begin{cases} 1, if\ |intensity_t(i) - intensity_t(j)| < D \\ -1, otherwise \end{cases} \quad (4)$$

The temporal energy of pixel i is calculated as

$$E_t(i) = \sum_{j \in N_i} TSim(i,i) \quad (5)$$

$$TSim(i,i) = \begin{cases} 1, if\ |intensity_{t+1}(i) - intensity_t(j)| < D \\ -1, otherwise \end{cases}$$

Every time before selecting the best value ofG_k , the best value from 0 to 1 is selected which leads to a small difference to U.

B. Aggregation Signature

Most of existing trackers fail to localize the target vehicle in presence of occlusions. In this work an aggregation signature for target vehicle is created accommodating the clutters.Aggregation signature of image patch is formed from a set of probable noise occluded image patches using frequency domain deep learning model. On the image patch of the target vehicle, occlusion patches of various probabilistic distributions are added. QDCT is applied to the Noised image patches of target vehicles to get the low and high frequency components. QDCT for an image $f(x,y)$ is calculated as.

$$(x,y) = A_n^q f(x,y) + \sum_{s=1}^{n} [D_{s,1}^q f(x,y) + D_{s,2}^q f(x,y) + D_{s,3}^q f(x,y)] \quad (6)$$

Where $A_n^q f(x,y)$and$D_{s,1}^q f(x,y)$ are low frequency and high frequency band of the image respectively.

A low frequency part and n groups of high frequency parts are obtained after QDCT is applied on the image.

The frequency of the coefficients are as shown in Fig.2.

To reduce the dimension of the coefficients, average fusion is done on low frequency sub bands. Fusing of High frequency sub bands based on maximum value of energy of coefficients. Fusing the low frequency bands is carried by average of the coefficients pair wise between the Low frequency coefficients of two patch images. Selecting the maximum value of coefficient between the pair wise high frequency sub band is the fusion rule for fusing the high frequency sub bands.

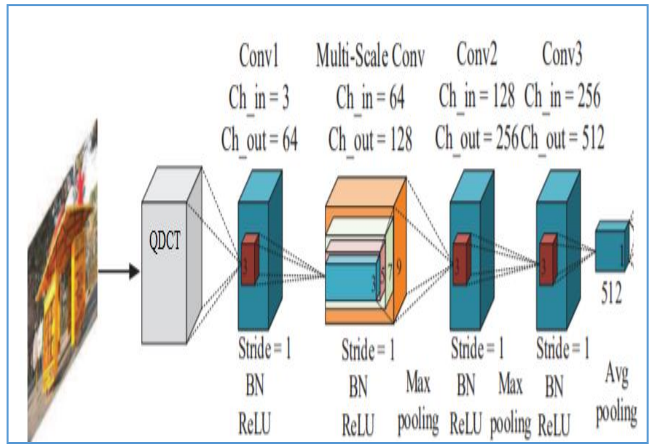

Fig. 3. Deep Learning Features Extraction.

The QDCT coefficients are given as input to a frequency domain convolutional neural network shown in Fig.3. The coefficients pass through a sequence of ReLU and max pooling layer and a final average pooling layer to provide an output of 1× 512 dimension feature vector. The CNN configuration used for feature extraction is given in Table I. An aggregation signature is constructed from the feature vectors belonging to same image patch as below:

- A unit random vector of dimension d (d<512) is generated $\{r_0, r_1, \ldots r_d\}$. A Gaussian function with mean 0 and variance 1 is used to sample each element. The d vector is put together into a matrix D of dimension 512 × d. This is generated on time at time of collecting the video as input for tracking.
- An inner product between the feature vectors v and the matrix D is done to get vector $u = D^T v$.
- For every vector u, following transformation function tf is applied produce the transformed feature vector $\bar{u}$
- $tf(u) = \begin{cases} 1\ r.u \geq 0 \\ 0, r.u < 0 \end{cases}$
- $\bar{u} = \{\mathrm{tf_{r1}(u), tf_{r2}(u), \ldots . tf_{rd}(u)}\}$

The feature vectors belonging to same image patch is now represented as bit stream of length d called as aggregation signature of the target image patch.

The benefits of converting the features of same patch to binary bit stream of aggregation signature have two benefits of: compressed form and reduced time complexity for matching the aggregation distance between aggregation signatures. The value of d can be selected by making a compromise between the matching time and accuracy of matching. The time complexity for construction of aggregation signature is $O(n \times 512 * d)$ where n is the number of feature vectors.

TABLE I. CNN CONFIGURATION FOR FEATURE EXTRACTION

Layer	Input	Filter	Stride	Padding	Out
MaxPool	32×32×128	2×2	2	0	16×16×128
Conv2	16×16×128	3×3	1	1	16×16×256
MaxPool	16×16×256	2×2	2	0	8×8×256
Conv3	8×8×256	3×3	1	1	8×8×512
AvgPool	8×8×512	8×8	1	0	1×1×512
FC	512×1	-	-	-	2×1

C. Hybrid Tracking

The hybrid trackers performs vehicle tracking by fusing the results of modified KCF filter and modified PCF filter. Though KCF filter [24] is known for its effectiveness and robustness, it cannot handle the appearance variations caused due to occlusions. This is because it does not make decisionsbased on tracking result but utilizes the maximum response in the output response graph to localize the target and update the template. Occlusion results in tracking deviation in KCF. The modified KCF filter adds a re-detection module to the standard KCF filter. The re-detection is activated by calculating the average peak correlation energy(AC) as in [28] and comparing it historical mean.

$$\boldsymbol{AC} = \frac{|\boldsymbol{F_{max}} - \boldsymbol{F_{min}}|^2}{\boldsymbol{mean}(\sum_{w,,h}(\boldsymbol{F_{w,h}} - \boldsymbol{F_{min}})^2} \tag{7}$$

Target re-detection is initiated when the AC is greater than historical mean. Re-detection is done by combining sliding window and template matching procedure.Template matching is done by calculating the hamming distance between the aggregation signature between the target and the candidate regions. Template matching on entire area of image takes significant amount of time, to avoid the probable area for matching is selected using Gray model GM (1,1) approach proposed in [29]. For each of the candidate regions identified by GM(1,1), template matching with aggregation signature is done with sliding window approach to get the hamming distance matching scores for each of the candidate regions as.

$$R1 = \{< C_1, h_{s1} >, < C_2, h_{s2} >, \ldots . < C_n, h_{sn} >\} \tag{8}$$

Particle filters apply Monte Carlo simulation for visual tracking [30]. A particle filter can be effective only if it can capture all the variations in state space. But this introduces a higher computation overhead. Many approaches [31-34] have been proposed to increase the capability of particle filters.But these models don't accommodate the appearance variations introduced due to occlusions. A correlation particle filter model is proposed in [35] which solves the problem of occlusions using a mixture and correlation filter. We modify the correlation particle filter algorithm proposed in [35] at the step 5 by removing the prediction step.

Algorithm 1: The proposed correlation particle filter tracking algorithm.

Input : Image Sequences and Initialization.
Output: Tracking Results $\mathbf{s}_t$ $\forall t$.
1 **for** *each frame* **do**
2 Generate particles using the transition model $p(\mathbf{s}_t|\mathbf{s}_{t-1})$ and re-sample them.
3 Shift particles with a mixture correlation filter $\mathbf{s}_t^i \rightarrow \mathcal{S}_{mcf}(\mathbf{s}_t^i)$.
4 Update particle importance weights
5 Predict target object state
6 **end**

Fig. 4. Correlation PCF [35].

Instead of step 5 of correlation PCF shown in Fig.4, we calculate the hamming distance scores for the results achieved in step 4 between aggregation signature of target and particle regions and return the result. The modified algorithm is given below.

Algorithm: Modified PCF
Input: frame ,agg_sig
Output: Candidate regions
1.Collect the current frame
2Generate particles using the transition model
p$(s_t|s_{t-1})$ and resample them
3shift particles with a KCF correlation filter
$s_t^i \rightarrow S_{mcf}(s_t^i)$
4.$R2 = \{\}$
5for each particle region r
6R2= {R2, (r, Hamming(r, agg_sig))}
7end
8return

From the results of modified KCF $(R1)$ and modified PCF(R2) , the final target is found by fusion. Fusion is done by selecting the region with least hamming distance score in R1 and R2 and calculating the overlap of these regions. The overlap provides the target region.KCF filter is through the procedure of re-detection.The entire process flow of hybrid tracker is given in Fig. 5.

TABLE II. COMPARISON FOR DIFFERENT DISTURBANCES

Parameters/Solutions	KCF tracker	Parallel correlation filter	Combined detector-tracker		MUSTER	Proposed hybrid filter
Illumination variation						
Precision	0.548	0.846	0.907		0.862	0.961
SR	41.4	81	84		86	97
Background clutter						
Precision	0.366	0.748	0.83		0.87	0.962
SR	45.22	76	82.12		89.6	91.23
Occlusions						
Precision	0.635	0.743	0.855		0.881	0.97
SR	40.54	75	82.12		89.1	91.96

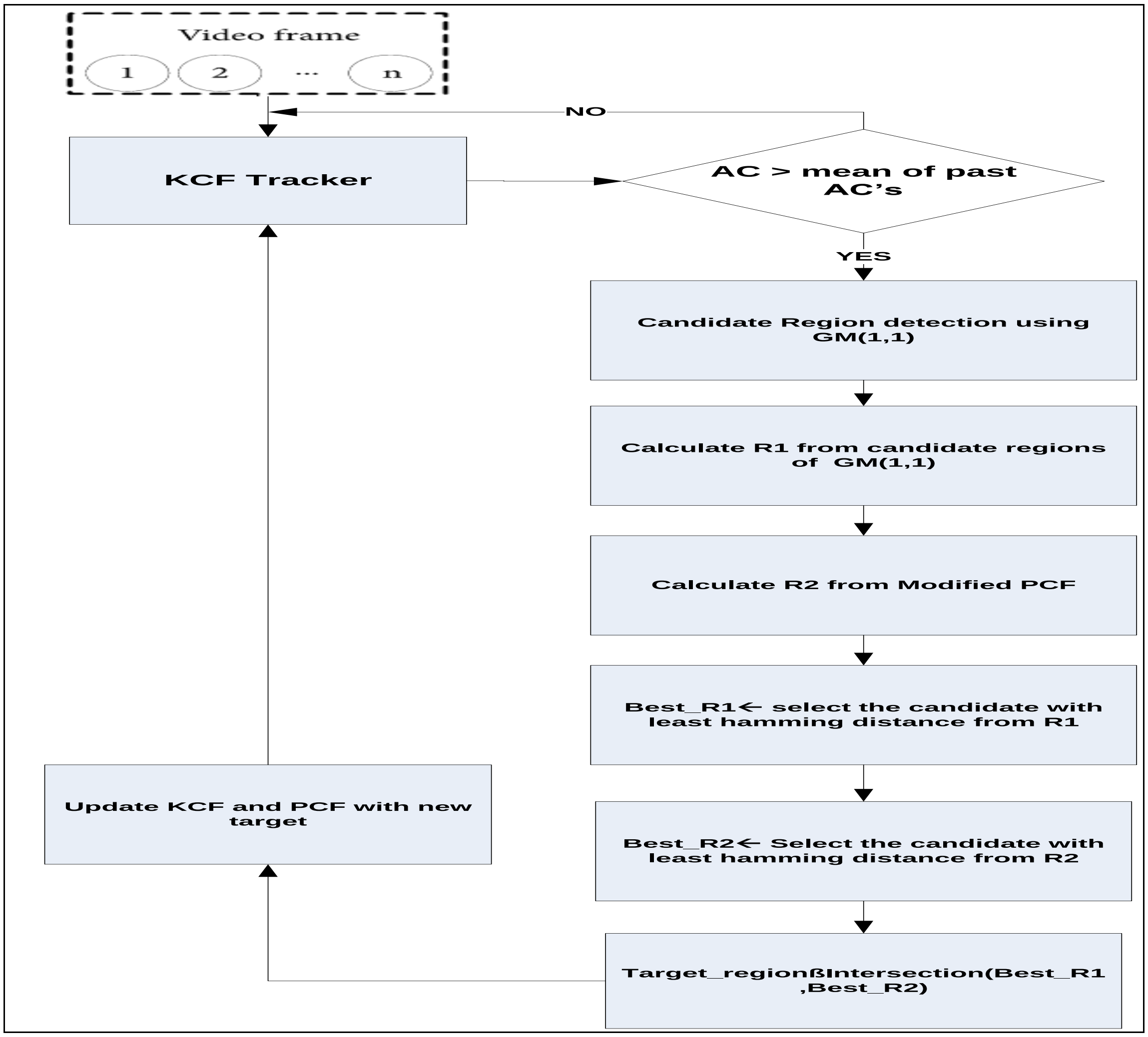

Fig. 5. HybridTracker Process Flow.

IV. Results

The performance of the proposed hybrid tracker is tested against 10 videos from OTB 100 dataset [27]. OTB dataset is the visual tracking benchmark widely used for testing the performance of object tracking algorithms. The dataset has 100 video sequences in different categories of objects, humans, vehicles etc.The video sequences have wide distribution of clutter, illumination variations and occlusions. The samples for testing are selected with presence of background clutter, illumination variation and occlusions. The performance is compared against vehicle tracking algorithms of KCF tracker [24], Parallel correlation filter [37], combined detector-tracker [26] and Multistore Tracker (MUSTER) [36]. The performance is measured in terms of precision and success rate. Precision is calculated in terms of central location error (CLE). It is calculated as the Euclidean distance with respect to the central location of tracked objects and manually labelled ground truth. Precision is computed as

$$Precision = \frac{\prod_{N} CLE<Th}{N} \tag{9}$$

Where N is the total number of frames and the Th value is set as 20 in the experiment as tracking performance evaluated for every 20 frames. Keeping a higher value, may lose the minor changes and reducing to lower value, increases computational complexity. So an optimum value of 20 was selected trial and error in the experimental setup.

Success rate is computed as

$$SR = \frac{area(R_t \cap R_{gt})}{area(R_t \cup R_{gt})} \tag{10}$$

R_t is the bounding box object tracked and R_{gt} is the bounding box corresponding to the ground truth. The performance is also compared in terms of two parameters of success rate and precision for varied degrees of occlusions/clutters.

The average precision and success rate for videos are measured for various disturbances and the result is given in Table II. Both precision and success rate is higher compared to all the existing works. Comparison of precision and success rate across different solutions is shown in Fig.6 and Fig.7 respectively.

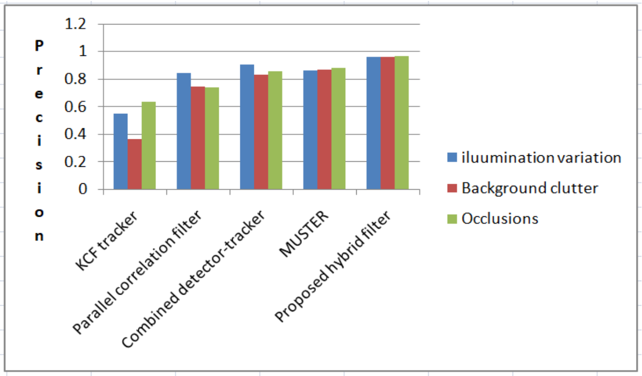

Fig. 6. Comparison of Precision Across Solutions.

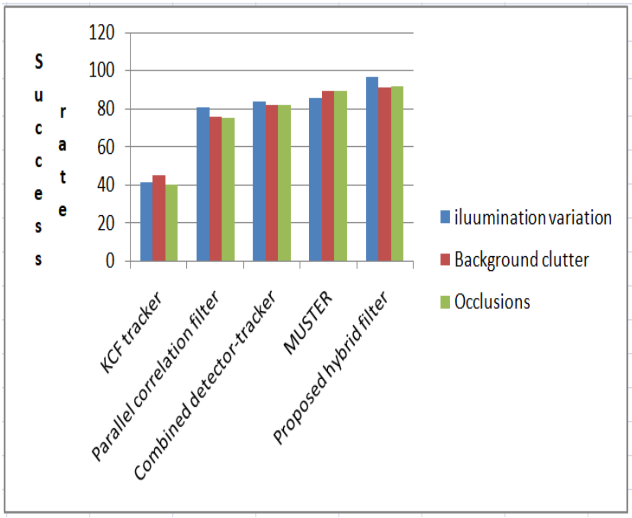

Fig. 7. Comparison of Success Rate Across Solutions.

The precision and success rate has increased in proposed solution due to use of hybrid trackers which combines the best of modified KCF and modified PCF. Also the influence of lighting variations, clutters, occlusion have been reduced due to use of novel background modeling with adaptive gain in the proposed solution.

The precision is measured for different error thresholds and the result is given in Fig. 8.

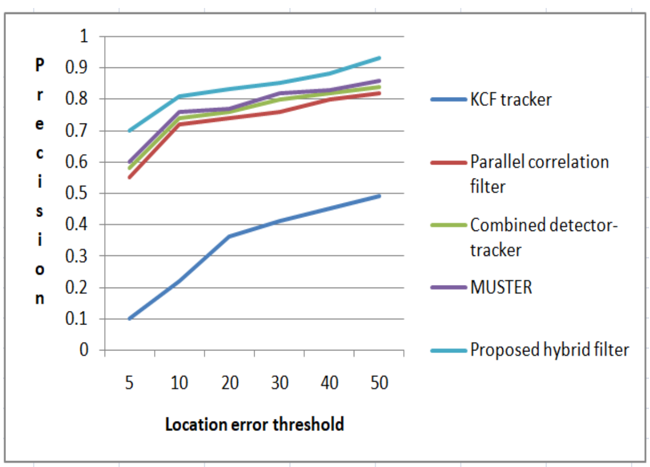

Fig. 8. Precision vs Location Error Threshold.

As location error threshold increases, the precision increases. Proposed hybrid tracker achieves highest precision among all other solutions even when location error threshold increases.Even if 50% of object is visible, the proposed solution is able to localize the vehicle accurately due to use of aggregation signature rather than exact matching. Aggregation signature is constructed with consideration for different probabilities of occlusions in the proposed solution and this helps to achieve better localization.

The success rate is measured for different overlap threshold and the result is given in Fig. 9.

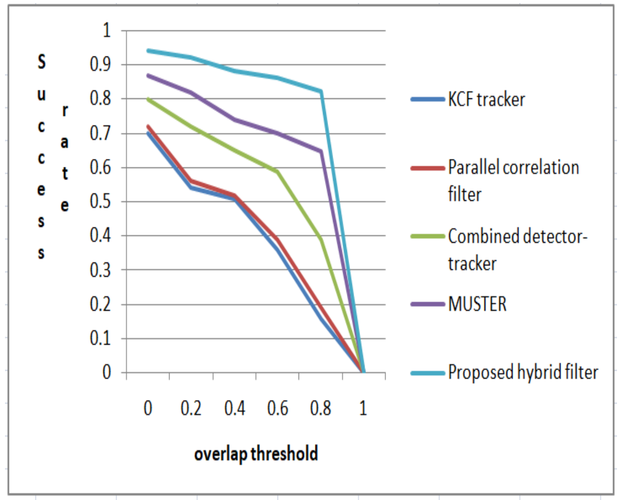

Fig. 9. Success Rate vs Overlap Threshold.

Even in presence of higher percentage of overlap, the proposed tracker is able to localize the object with higher success rate due to use of aggregation signature which already accommodates overlaps.

Video with a synthetic shape in varied percentage of occlusion is created. The performance of precision and success rate is measured with this synthetic data and the result is given in Fig.10 and Fig. 11.

Even in presence of 50% occlusion, the proposed hybrid tracker is able to achieve success rate of 0.82 and precision of 84%.

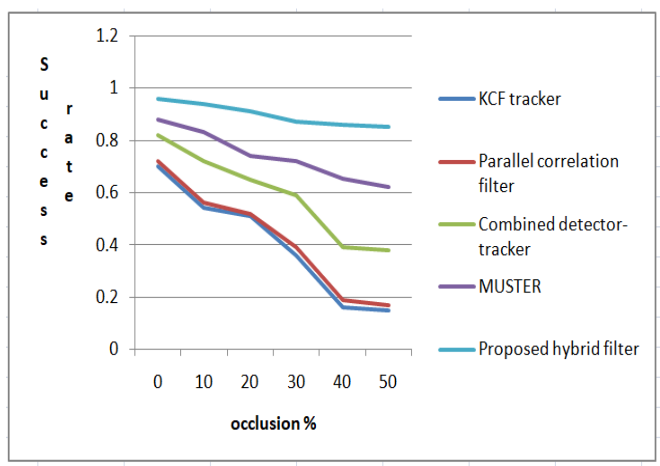

Fig. 10. Success Rate vs Occlusion %.

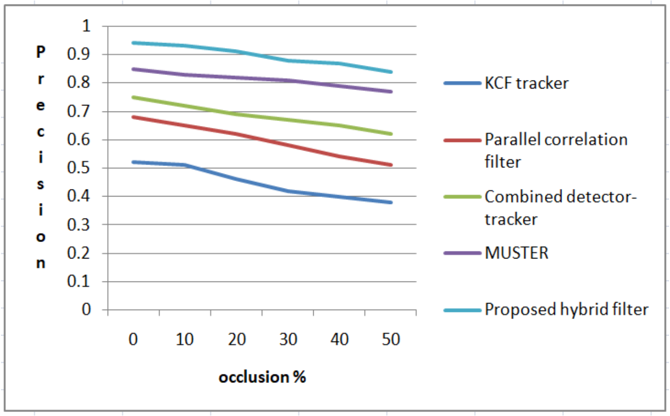

Fig. 11. Precision vs Occlusion %.

The success rate decreases as the occlusion percentage increases, but the success rate is higher in proposed solution compared to other solutions. The average success rate in proposed solution is at least 3% higher compared to others. Usage of aggregation signature-based matching within the KCF and PCF has increased the success rate in the proposed solution. The average precision in the proposed solution is at least 3% higher compared to other solutions/ Use of deep learning features with operating on image in frequency domain has increased the precision in the proposed solution.

A. Discussion

Through performance results, the proposed tracker is found to effectively localize the target vehicle even in presence of various degrees of occlusion and overlap. Localization accuracy has improved due to use of aggregation signature in the proposed solution. The aggregation signature is constructed considering different probabilities of occlusions and it is binary signature. Matching is realized using hamming distance. Due to this, vehicle localization accuracy has increased. Most of the existing works used in comparison used deep learning features and comparison of the features to localize the object. But in presence of occlusion, the distance match crossed the threshold and they failed to localize the object. Due to error in localization, their tracking performance too reduced. Tracking performance improved in proposed solution due to two reasons: (i) avoidance of unnecessary background objects through adaptive gain background subtraction and(ii) hybrid tracker to localize combining both kernel and particle filtering.This helped to localize object even in presence of higher overlap.

V. Conclusion

A hybrid tracking technique combining deep learning-based aggregation signature features with modified KCF & PCF is proposed in this work. Aggregation signature of target vehicle is created in frequency domain using QDCT and dimension reduced using deep learning. The proposed solution is able to achieve about 82% success rate and 86% precision in target even at 50% occlusion rate. In addition, the solution performs well against disturbances of occlusion, clutter and illumination variations.Improving the solution to work for even higher occlusion rate and various lighting conditions is in scope of future work.

References

[1] Yin, S. (2019). Estimation of rotor position in brushless direct current motor by memory attenuated extended Kalman filter, European Journal of Electrical Engineering, 21(1): 35-42.

[2] Ha, D.B., Zhu, G.X., Zhao, G.Z. (2010). Vehicle tracking method based on corner feature and mean-shift. Computer Engineering, 36(06): 196-197.

[3] Cao, X., Gao, C., Lan, J., Yuan, Y., Yan, P. (2014). Ego motion guided particle filter for vehicle tracking in airborne videos. Neurocomputing, 124: 168-177.

[4] Oh, J., Min, J., Choi, E. (2010). Development of a vehicle image-tracking system based on a long-distance detection algorithm. Canadian Journal of Civil Engineering, 37(11).

[5] Choi, I.H., Pak, J.M., Ahn, C.K., Mo, Y.H., Lim, M.T., Song, M.K. (2015). New preceding vehicle tracking algorithm based on optimal unbiased finite memory filter. Measurement, 73: 262-274.

[6] Li, Z., He, S., Hashem, M. (2015). Robust object tracking via multi-feature adaptive fusion based on stability: contrast analysis. The Visual Computer, 31(10): 1319-1337.

[7] Wu, S.Y., Liao, G.S., Yang, Z.W., Li, C.C. (2010). Improved track-before-detect algorithm based on particle filter. Control and Decision, 25(12): 1843-1847.

[8] Kalal, Z., Mikolajczyk, K., Matas, J. (2011). Tracking-learning-detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(7): 1409-1422.

[9] Wang, Y. (2020). Moving vehicle detection and tracking based on video sequences. Traitement du Signal, Vol. 37, No. 2, pp. 325-331B. Abidi, A. Jilbab, and E. H. Mohamed, “An energy efficiency routing protocol for wireless body area networks,” J. Med. Eng. Technol., vol. 42, no. 4, pp. 290–297, 2018.

[10] Shrestha, Sandesh. (2019). Vehicle Tracking Using Video Surveillance. 10.5772/intechopen.89405.

[11] Xu, Yong & Dong, Jixiang& Zhang, Bob & Xu, Daoyun. (2016). Background modeling methods in video analysis: A review and comparative evaluation. CAAI Transactions on Intelligence Technology.

[12] Shobha B S and R. Deepu, "A Review on Video Based Vehicle Detection, Recognition and Tracking," 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS), 2018, pp. 183-186,doi: 10.1109/CSITSS.2018.8768743.

[13] Shobha B. S and Deepu. R, “Traffic Adaptive Deep Learning based Fine Grained Vehicle Categorization in Cluttered Traffic Videos” International Journal of Advanced Computer Science and Applications (IJACSA), 12(9), 2021. http://dx.doi.org/10.14569/IJACSA.2021.0120911.

[14] D. Zhang, H. Maei, X. Wang, and Y.-F. Wang. Deep reinforcement learning for visual object tracking in videos. arXiv preprint arXiv:1701.08936, 2017.

[15] I. Leang, S. Herbin, B. Girard, et al., On-line fusion of trackers for single-object tracking [J]. Pattern Recognition 74, 459–473 (2018).

[16] Z. Chen, X. You, B. Zhong, et al., Dynamically modulated mask sparse tracking [J]. IEEE transactions on cybernetics 47(11), 3706–3718 (2016).

[17] Hua S , Kapoor M, Anastasiu D C. Vehicle Tracking and Speed Estimation from Traffic Videos[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, 2018: 153-160.

[18] Liu S , Wang S , Shi W , et al. Vehicle tracking by detection in UAV aerial video[J]. SCIENCE CHINA Information Sciences, 2019, 62(2): 24101.

[19] H.S. Song, Y. Li, J. Yang, Vehicle object tracking method based on highway scenario [J]. Computer Systems & Applications 28(6), 82–88 (2019).

[20] Xu Y , Wang J. A unified neural network for object detection, multiple object tracking and vehicle re-identification[J]. arXiv preprint arXiv:1907.03465, 2019.

[21] Fang Y , Wang C , Yao W , et al. On-road vehicle tracking using part-based particle Filter[J]. IEEE Transactions on Intelligent Transportation Systems, 2019:1-15.

[22] Dai J, Li Y, He K, et al. R-fcn: Object detection via region-based fully convolutional networks[C]//Advances in neural information processing systems. 2016: 379-387.

[23] Danelljan M, Hager G, Shahbaz Khan F, et al. Convolutional features for correlation filter based visual tracking[C]//Proceedings of the IEEE International Conference on Computer Vision Workshops. 2015: 58-66.

[24] Ma C, Huang J B, Yang X, et al. Hierarchical convolutional features for visual tracking[C]//Proceedings of the IEEE international conference on computer vision. 2015: 3074-308.

[25] Wang Q, Gao J, Xing J, et al. Dcfnet: Discriminant correlation filters network for visual tracking[J]. arXiv preprint arXiv:1704.04057, 2017.

[26] J.F. Henriques, R. Caseiro, P. Martins, et al., High-speed tracking with kernelized correlation filters[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence 37(3), 583–596 (2015).

[27] D. S. Bolme, J. R. Beveridge, B. A. Draper and Y. M. Lui, Visual object tracking using adaptive correlation filters, 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, 2010, pp. 2544-2550.

[28] Yang, B., Tang, M., Chen, S. et al. A vehicle tracking algorithm combining detector and tracker. J Image Video Proc. 2020, 17 (2020).

[29] https://paperswithcode.com/dataset/otb.

[30] M. Wang, Y. Liu, Z. Huang. Large Margin Object Tracking with Circulant Feature Maps, IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2017,1:4800-4808.

[31] Q. Yu, J. Lyu, L. Jiang, and L. Li, “Traffic anomaly detection algorithm for wireless sensor networks based on improved exploitation of the GM(1,1) model,” International Journal of Distributed Sensor Networks, vol. 12, no. 7, 2016.

[32] M. Isard, A. Blake, Condensation - conditional density propagation for visual tracking, International Journal of Computer Vision 29 (1) (1998) 5–28.

[33] C.Yang, R. Duraiswami, L. Davis, Fast multiple object tracking via a hierarchical particle filter, in: Proceedings of the IEEE International Conference on Computer Vision, 2005.

[34] Z. Khan, T. Balch, F. Dellaert, A rao-blackwellized particle filter for eigentracking, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2004.

[35] S. K. Zhou, R. Chellappa, B. Moghaddam, Visual tracking and recognition using appearance-adaptive models in particle filters, IEEE Transactions on Image Processing 11 (1) (2004) 1491–1506. 3.

[36] A. Doucet, N. De Freitas, N. Gordon, Sequential monte carlo methods in practice, in: Springer-Verlag, New York, 2001.

[37] T. Zhang, S. Liu, C. Xu, B. Liu and M. Yang, "Correlation Particle Filter for Visual Tracking," in IEEE Transactions on Image Processing, vol. 27, no. 6, pp. 2676-2687, June 2018.