INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

Window Based Min-Max Feature Extraction for Visual Object Tracking

Rubeena Banu[1]*, **M H Sidram**[2]

1 Research Scholar, Department of Electrical and Electronics Engineering, Sri Jayachamarajendra College of Engineering, VTU University, Mysuru-570006, Karnataka, India
2 Associate Professor, Department of Electrical and Electronics Engineering, Sri Jayachamarajendra College of Engineering, VTU University, Mysuru-570006, Karnataka, India

OPEN ACCESS

Received: 06-07-2022

Accepted: 12-09-2022

Published: 18-10-2022

Citation: Banu R, Sidram MH (2022) Window Based Min-Max Feature Extraction for Visual Object Tracking. Indian Journal of Science and Technology 15(40): 2047-2055. https://doi.org/10.17485/IJST/v15i40.1395

* **Corresponding author**.

rubyms2011@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Background/Objectives: Visual object tracking is considered as difficult problem and becomes more challenging because of the various environmental conditions. In order to achieve the efficient result in the area of visual tracking, feature extraction will play the important role. This study demonstrates the Min-max feature extraction method to improve the tracker robustness in the area of visual tracking. **Methods**: The proposed min-max features along with existing Histogram of Oriented Gradient (HOG), and Convolution Neural Network (CNN) features are given to the Spatial Temporal Regularized Correlation Filter (STRCF) to find the new position of the target and it is successfully solved through Alternative Direction Method of Multipliers (ADMM). By employing both Spatial and temporal regularization methods, without much compromise in the efficiency, the boundary effect is handled. The min-max feature will extract the object's window-based features as foreground and background. The foreground consists of higher color values than the background. As compared to the Color Names (CN) proposed min-max feature method gives accurate features to identify the objects in a video. In order to present the performance, the method is tested on the OTB dataset image sequences and compared with the state-of-the-art tracker and achieved the promising results for all of tested videos. **Findings**: Our method, using Min-max feature, gives Mean OP and FPS 61.44% &18.27 respectively, which shows improvement in the tracking accuracy along with the computational speed as compared to CN feature.

Keywords: MinMax feature; CNN; Histogram of Oriented Gradient; Occlusion; Visual tracking; Video surveillance

1 Introduction

The visual object tracking is a rudimentary issue in the area of computer vision, analytic of video and multimedia fields. The purpose of tracking process is to identify the new position of the target, where first frame of the video sequences is annotated with a bounding box. The object tracking may be of any type of arbitrary object; therefore, it is employed in a different practical condition. It has a lot of applications along with video surveillance, unmanned vehicle driving, analysis of driving sport video, action identification, editing the video and others. Due to different circumstances like low contrast, changes in the illumination, full or partial occlusion, variation in the appearance, fast motion etc., visual tracking is observed as a difficult problem and becomes more challenging. In order to deal with these kinds of circumstances enormous tracking methods have been developed and accomplished notable results which denoted as performance of tracking and robustness. Since, an object and background are dynamic in nature among video sequences, the primary role of object tracking is to bring off effective tracking output in a systematic manner(1–4).

Most of the proposed methods of tracking are normally categorized as generative or discriminative. Generative methods normally utilize appearance features to model the target and later employ this model to determine the foremost candidate features as the output of tracking. Discriminative methods use a classifier that attempts to discriminate the target from the background by updating its positive and negative samples.

Over the last few years, the results of the tracking have been improved based on correlation filter, due to its constructive decision strategy and search mechanism. But, this method will go weak when it goes through the video with severe distracters from undesired boundary effect. To solve this issue numerous methods for tracking related to correlation filter have been developed using both handcrafted and deep features. Zhu, Xue-Feng, et al(5) proposed attribute aware discriminative correlation filter to extract the important information from different channel and used HOG, CN and intensity channel for feature extraction. Yang, Haoran, et al.(6) presented spatio-temporal context and Kalman filtering (STC-KF) to collect important context information of the target. Euclidean distance is considered to identify the location of the target in the process of tracking. Fang, Sheng, et al.(7) developed confidence based multi feature correlation filtering. This method selects HOG features then multi-feature fusion technique is adopted to enhance the tracking algorithm robustness and also model rollback mechanism is used to lessen the effect of model corruption. Zhu, Hong, et al.(8) presented hybrid cascade filter to combine both deep feature and handcrafted features to improve the robustness and accuracy in visual tracking. Xu, Tianyang, et al.(9) introduced an approach called group feature selection method for discriminative correlation filter (GFS-DCF) to choose feature across both channel and spatial dimension and significantly achieved the performance of tracking. Huang et al.(10) combined CN features and the HOG features to learn the filters, and to mitigate the computational cost principal component analysis method is utilized. Kumar et al.(11) used color histogram, LBP and pyramid of the histogram of gradients features for the object's appearance and presented multi-clue particle filter for visual tracking. Along with this numerous method have proposed based on correlation filter using multiple features to train the filter(12) (13). The normally utilized handcrafted features are, scale invariant feature transform (SIFT), color names (CNs), the local binary pattern (LBP) and the histogram of oriented gradients (HOG). These features will give the color and shape knowledge of the object for tracking(14–16).

Although correlation filter has been attained the significant performance in visual tracking accuracy, still the robustness of the tracker will reduce by using single feature or by adopting weak feature extraction method. From the analysis of survey, we noticed that effective feature extraction method is required to achieve the great performance of the tracker. So, we proposed window-based min-max feature extraction method. Our method introduces novel features, which selects the window-based feature of the object. The advantages of this method are that, it gives the foreground features of the target more accurately than the background.

The main motivation of this research work is to handle the effective severance of the intended object from the respective background. For two decades, various algorithms are implemented based on pixel level, edge level and region level, but very less number of works had been done on probabilistic based methods. Hence, this work proposed a min-max features-based algorithm to effective separation of intended objects from the background by combining the pixel level along with probabilistic concept. The max-min features study the pixel intensity level by comparing the maximum value and minimum value of fixed region to identify the foreground pixels. The importance of the proposed min-max features outcome helps to extract the high-quality properties of histogram of gradient values to represent the interested object and these HOG values more contributes for effective tracking of that object even in the complex background nature. The prime contributions of the work are summarized as below:

- This research article proposed Min-Max feature algorithm by combining the pixel level operation and probabilistic concept to effective severance of the interested object in the video.
- HOG features are extracted from the outcome of Min-Max feature algorithm to identical representation of the interested object in the video.
- CNN architecture is modified with STRCF framework is implemented for object tracking system.
- A constructive alternating-direction method-of-multipliers (ADMM)-based algorithm is modified using Lagrangian multiplier concept for minimizing the issues (optimization).

The remaining section of this paper is presented as follows: In section 2 Mathematical explanations are described. Results and discussions are carried out in section 3. Finally, the Section 4 concludes the overall contribution of the research work.

2 Methodology

Over the last few years, the results of the tracking have been improved based on correlation filter, due to its constructive decision strategy and search mechanism. But, this method will go weak when it goes through the video with severe distracters from undesired boundary effect. To solve this issue numerous methods for tracking related to correlation filter have been developed using both handcrafted and deep features. In this work, a window-based min-max feature extraction method has been developed with STR correlation filter framework. The min-max feature extraction method is evolved based on the concept used in the paper[17]. Along with this, HOG and CNN based feature extraction methods have been utilized. These different features are combined with STRCF to achieve a more precise and accurate results.

2.1 Min-Max feature extraction

RGB color information play a vital role to distinguish foreground from the respective background. In an image, foreground will have higher color values compared to the background. The method consists of three-color sub-band images i.e. R, G and B. In order to find out high color values in these sub bands min-max criteria has been used. This min-max method will select the maximum and minimum values from each pixel of R, G and B sub band. Then, the third value from this sub band is compared with above obtained maximum and minimum values in order to detect the nearest value. If the third obtained value is nearest to the maximum value, then it is replaced with max value. If the value is nearest to minimum value, then it is replaced with min value. The basic idea of this method is that, compared to its background, target pixel has high intensity value in any one of three sub-bands. The step-by-step procedure of min-max feature extraction method is detailed below:

Step1: Maximum and Minimum of R, G, B is selected from the original image using below condition

$$A(x,\ y) = max(r(x,\ y),\ max(g(x,\ y),\ b(x,\ y)))$$

$$B(x,\ y) = min(r(x,\ y),\ min(g(x,\ y),\ b(x,\ y))) \tag{1}$$

Step2: Other than maximum and minimum i.e. Third value is determined for each pixel

$$C\ (x,\ y) = r\ (x,\ y) \tag{2}$$

Step3: This step finds whether the third value is near to maximum or minimum value

$$D\ (x,\ y) = A\ (x,\ y) - C\ (x,\ y))$$

$$E\ (x,\ y) = C\ (x,\ y) - B\ (x,\ y)) \tag{3}$$

If $E < D$ then it forms maximum group value or else it forms minimum group value.

Step4: Find the maximum value of max group value and assign it to the original image size and do the same for minimum value as well.

$$image1(x,\ y) = max(A(x,\ y),\ C(x,\ y))$$

$$image2(x,\ y) = min(B(x,\ y),\ C(x,\ y)) \tag{4}$$

The obtained value is gray scale value. In order to enhance the pixel value which increases the gap between target and background, the sliding window concept i.e. 3x3 window is applied for this obtained gray scale value.

Step5: The maximum and minimum value will be selected from the 3x3 window

$$F = max(max(wind));$$

$$G = min(min(wind)); \tag{5}$$

Then, obtained max and min value are compared with each value in the window. If value is near to max value it forms maximum value group or else minimum value group as explained in step 3. After the first iteration, matrix will be updated and this will be given as input to the next iteration. Same procedure is followed throughout the image. The resultant values of min-max feature algorithm are used to extract the prominent characteristics of histogram of gradient values to identical representation of the objects.

2.2 Histogram of Oriented Gradient

The Histogram of Oriented Gradient (HOG) values are obtained from the final output of min-max feature method to indicate the objects separately in the given input. HOG is a dense feature extraction method for images, which extracts features for all location in an image. The distributions of directions of gradients are utilized as features. Gradients i.e. x and y derivatives of an image are helpful since the magnitude of gradients is high near edges and corners because corner and edges consist much information about shape of an object than the flat regions. To compute the HOG descriptor, first vertical and horizontal gradients are calculated, which is done by simply filtering the image with the Sobel derivative kernels. The gradient consists magnitude and direction for every pixel. The gradient of the three channels is estimated for the color images. The magnitude of gradient at each pixel is the maximum of the magnitude of gradients of the three channels and the angle corresponding to the maximum gradient. The HOG descriptor of the frame is normally observed by plotting the 9×1 normalized histogram in the 8×8cells.

2.3 Convolution Neural Network (CNN)

The prominent features of histogram oriented gradient values are identified from the local level to higher level by training the CNN architecture for tracking only intended object in the video. CNN model is used in each image related issue. Compared to predecessors, the advantage of CNN is that it detects the significant features without any human guidance. For example, if we give any image of car or cat it will learn distinctive features by itself for every class. CNN model is computationally effective. It utilizes convolution, pooling performances and executes parameter sharing. The CNN model is globally appealing as it run on any device. The architecture of CNN is shown in Figure 1. It executes a series pooling and convolution operations, followed by number of fully connected (FC) layers. If the multi-class classification is performed, then output is softmax. CNN model is very strong and effective. Here, feature extraction will be done automatically in order to accomplish superhuman accuracy. Finally, these prominent features are employed on spatial temporal regularized correlation filter for the purpose of visual tracking of specific object in the video.

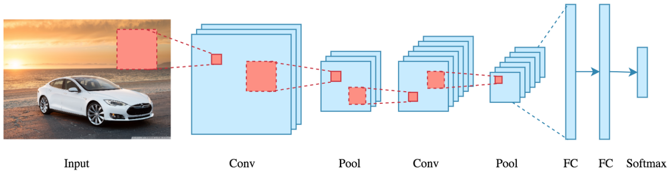

Fig 1. CNN models architecture

2.4 Spatial Temporal Regularized Correlation Filter (STRCF) Visual Tracking

The trained CNN model provides the efficient key features for spatial temporal regularized correlation filter[18] to track the entire region of the intended object in the video. This technique uses both spatial and temporal information to track accurate region of the interested object. In the online classification, algorithm will forecast its label when new instance occurs on every round. Later, according to the newly instance label pair, the classifier has been updated. Meanwhile, in order to model the updated classifier equivalent to earlier one, the learning algorithm must be passive. Also, this learning algorithm should be aggressive to make sure the new instance is classified precisely. Therefore, Passive aggressive (PA) algorithm has been suggested[19]. The regularization process of spatial and temporal information is explained as follows.

The Spatial-temporal regularized CF model

$$\min_{r}\frac{1}{2}\|\sum_{m=1}^{M}x_t^m * r^m - y\|^2 + \frac{1}{2}\sum_{m=1}^{M}\|w * r^m\|^2 + \frac{\mu}{2}\|r - r_{t-1}\|^2 \tag{6}$$

Here, $\sum_{m=1}^{M}\|w * r^m\|^2$ represents spatial regularizer, r_{t-1}indicates CF employed in (t-1)$^{\text{th}}$frame. Where, $\|r - r_{t-1}\|^2$ represents temporal regularization and μ is the parameter of regularization.

STRCF is considered as online learning of linear regression and the samples in STRCF arrive at batch level at every round. So, STRCF will acquire quality of PA by balancing the trade-off between passive and aggressive model learning. Hence, in case of huge changes in the appearance this leads to robust model. Hence, correlation filtered spatial and temporal information effectively tracks the interested object even in the complex background and low resolution nature.

2.5 Optimization Algorithm

The objective function of the model in equation 1 is convex. The ADMM algorithm [(20)] is used to solve the minimization issue. Hence, introduced an auxiliary variable by restricting f=n and γ is step size parameter and then established Lagrangian form of Equation (6) is composed as:

$$L(w,n,z) = \frac{1}{2}\|\sum_{m=1}^{M}x_t^m * f^m - y\|^2 + \frac{1}{2}\sum_{m=1}^{D}\|w \cdot n^m\|^2$$

$$+\sum_{m=1}^{M}(f^m - n^m)^T z^m + \frac{\gamma}{2}\sum_{m=1}^{M}\|f^m - n^m\|^2 + \frac{\mu}{2}\|f - f_{t-1}\|^2 \tag{7}$$

Where, z denoted as Lagrangian multiplier and μ is penalty factor. By mentioning $l = \frac{1}{\gamma}z$ above Equation7 can be represented as:

$$L(w,n,l) = \frac{1}{2}\|\sum_{m=1}^{M}x_t^m * f^m - y\|^2 + \frac{1}{2}\sum_{m=1}^{D}\|w \cdot n^m\|^2$$

$$+\frac{\gamma}{2}\sum_{m=1}^{M}\|f^m - n^m + l^m\|^2 + \frac{\mu}{2}\|f - f_{t-1}\|^2 \tag{8}$$

The optimization issue is divided into several sub-problems as given below

$$f^{i+1} = \underset{f}{argmin}\|\sum_{m=1}^{M}x_t^m * f^m - y\|^2 + \gamma|f - n + l\|^2 + \mu\|f - f_{t-1}\|^2 \tag{9}$$

$$n^{i+1} = \underset{g}{argmin}\sum_{m=1}^{M}\|w \cdot n^m\|^2 + \gamma|f - n + l\|^2 \tag{10}$$

$$l^{(i+1)} = l^{(i)} + f^{i+1} - n^{(i+1)} \tag{11}$$

Solution for each sub-problem in details is given below

Solving f: By employing the Parseval's theorem, the Equation 9 is reformulated in the Fourier domain as

$$\underset{f}{argmin}\|\sum_{m=1}^{M}\widehat{x_t}^m * \widehat{f}^m - \widehat{y}\|^2 + \gamma|\widehat{f} - \widehat{n} + \widehat{l}\|^2 + \mu\|\widehat{f} - \widehat{f}_{t-1}\|^2 \tag{12}$$

here, $\hat{f}$ represents discrete Fourier Transform (DFT) of filter f. It can be observed that across all M channel, j-th element of $\hat{y}$depends on j-th element of $\hat{f}$ filter and$\hat{x}$. Assume $B_j(f) \in R^M$ the vector contains j-th elements of f along all M channels. The Equation 12 is decomposed in to following PQ sub problems,

$$\underset{B_j(\widehat{f})}{arg\,min}\|B_j(\widehat{x_t})^T B_j(\widehat{f}) - \widehat{y}_j\|^2 + \mu\|B_j(\widehat{f}) - B_j(\widehat{f}_{t-1})\|^2 + \gamma\|B_j(\widehat{f}) - B_j(\widehat{n}) + B_j(\widehat{l}\|^2 \tag{13}$$

By considering above equation to be zero will get solution for $B_j(\hat{f})$

$$B_j(\widehat{f}) = (B_j(\widehat{x_t})B_j(\widehat{x_t})^T + (\mu+\gamma)I)^{-1}q \tag{14}$$

The equation 14 is solved using Sherman Morrsion formula[17], and got

$$B_j(\widehat{f}) = \frac{1}{\mu+\gamma}(I - \frac{(B_j(\widehat{x_t})B_j(\widehat{x_t})^T}{\mu+\gamma+B_j(\widehat{x_t})^T(B_j(\widehat{x_t})})q \tag{15}$$

The above Equation 15 consists multiply-odd operation so it can be calculated effectively. Solving n: From the Equation 10 every element in the n can be computed separately. Therefore, closedform of solution for n is calculated by

$$n = (W^TW + \gamma I)^{-1}(\gamma f + \gamma l) \tag{16}$$

Here, W denotes the MPQ × MPQ diagonal matrix is integrated with the diagonal matrices M. Updating parameter**:** The γ step size parameter is updated is given in Equation 17,

$$\gamma^{(i+1)} = min(\gamma^{max}, \rho\gamma^{(i)}) \tag{17}$$

Here, maximum value of $\gamma$is indicated by γ^{max} andscale factor is represented by ρ.

The proposed method contains the heuristic complex mathematical analysis for extracting and tracking the object in the video. Hence Alternating Direction Method of Multipliers (ADMM) is employed to reduce the convex optimization problem by dividing them in to local sub problems to identify the efficient solution for global problem. Where, ADMM is a dual decomposition technique and combines the Lagrangian concept for constraint optimization. The convergence of proposed algorithm is based on the convexity of finiteness of f. The Lagrangian concept in ADMM restricts the step size parameter to get the accurate tracking results by minimizing issues.

3 Results and Discussion

This section gives the experiments, to examine the performance of proposed method in visualtracking. To analyze the performance, the model is compared with state-of-art-of the trackerSTRCF by utilizing OTB50 and OTB100 dataset. Initially, region of interest in an image is cropped as square region. Then the features called Min-max, HOG and CNN is extracted for that region. To lessen the boundary discontinuity, the obtained features are then weighted by cosine window.The parameter is used same as STRCF $\gamma^{(0)} = 10, \gamma^{(max)} = 100$, and scale factor ρ =1.2. In this work, the method is implemented using MATLAB 2018a, on a PC with an Intel i5, 10300H CPU processor (2.50GHz), and NVIDIA GTX 1650 GPU. The benchmark dataset OTB-2015 is a popular tracking dataset[21]. It has 100 fully annotated video sequences along with 11 various attributes i.e. occlusion, scale variation, motion blur illumination variation etc.

3.1 Proposed method compared with tracker based on features

We compared the proposed method with the STRCF based on feature extraction. The tracker has been evaluated on One Pass Evaluation (OPE), where mean overlap precision (mean OP) metric is used by computing the bounding box overlaps exceeding 0.5 in a sequence is shown in below Equation 18.

$$OP \ = \ mean \ (success \ number \ of \ overlap) \ / \ size \ (groundTruthboxes, \ 1) \tag{18}$$

The running speed was computed in frames per second (FPS) from the obtained overall result. Table 1 shows the comparison results of mean OP and FPS on the OTB 2015 dataset. The method using min-max feature gives mean OP and FPS scores 61.44 % and 18.27, as state-of-the-art method using CN shows 60.42% and 16.77 respectively. From the result it has been observe that, using min-max feature our method gives significantly better result compare to the color names feature. By employing min-max it extracts the foreground feature more accurately than the background in order to track the object and also improves the computational speed.

Table 1. Comparison results of STRCF and proposed method

Sl.No	Video Sequences	STRCF using CN		Proposed Method using Min-max feature	
		Mean OP in %	FPS	Mean OP in %	FPS
1	Basketball	36.97	19.18	77.09	20.1
2	Bolt	68.05	18.16	67.38	19.7
3	Board	78.03	6.57	80.68	8.95
4	Biker	37.65	30.4	36.92	31.9
5	Bird	19.15	27.02	5	27.5
6	BlurBoday	68.74	13.14	69.63	13.7
7	Blurcar2	87.16	15.1	86.78	15.5
8	Blurface	85.89	15.86	86.13	16
9	Blurowl	87.14	17.04	83.36	18.4
10	Box	74.32	16.84	74.66	17.1
11	Carscale	69.82	21.62	70.23	22.8
12	Cardark	84	30.49	84.99	32.8
13	Couple	67.92	18.29	70.37	20.9
14	Crowds	69.97	23.63	72.3	25.7
15	Car24	88.09	17.29	88.96	18.3
16	Coke	55.01	16.56	56.6	16.75
17	David	24.51	16.37	24.63	17.6
18	Deer	79.34	17.85	79.27	19.4
19	Diving	21.94	18.62	22.32	21
20	Doll	81.79	10.32	81.73	10.85
21	Dragonbaby	50.61	17.14	50.65	19.4
22	Faceocc	75.52	7.59	76.72	7.68
23	Gym	41.03	13.12	45.57	13.73
24	Girl	70.18	16.77	66.87	17.8
25	Human2	74	7.31	75.84	7.52
26	Human3	61.8	15.49	61.44	16.4
27	Human5	78.12	15.78	79.27	16.6
28	Human6	79.26	25.65	79.48	26.4
29	Human9	53.03	18.74	63.87	19.6
30	Ironman	10	18.11	2	19
31	Jump	9	13.95	19	15.4
32	Liquor	83.56	12.93	83.78	13.6
33	Lemming	74.45	10.76	75.74	16.26
34	Matrix	21.47	19.43	10.14	20.8
35	MountainBike	64.02	11.88	65.5	14.08
36	Motorrolling	9.37	14.27	9.81	14.4
37	Panda	39.41	25.9	47.01	27.8
38	RedTeam	75.09	28.68	74.9	31
39	Shaking	73.84	10.26	74.54	12.9
40	Singer1	86.25	9.19	86.77	9.83
41	Singer2	79.71	12.39	80.35	12.3
42	Skating1	42.72	10.62	41	14.9
43	Skiing	7	20	5	29.9
44	Soccer	50.99	9.46	49.26	12.6
45	Sufer	67.62	21.29	68.98	23.5
46	Tiger2	52.3	17.86	53.77	18.7
47	Trellis	75.78	18.95	75.26	19.7
48	Woman	77.42	11.06	77.67	11.2
49	Walking	72.5	17.36	73.19	18.3
50	Walking2	79.95	16.69	79.73	17.4

Figure 2 shows the qualitative evaluation on video sequence (such as Basketball, Couple, Crowds, Human9, Dog, Panda, Surfer, Tiger2, Jump, Gym, Board and Rubik) for both the methods using the same frames from the video sequences. From the Figure 2 we can analyze that by using min-max feature extraction method, the model will give better bounding box over the target compared to the state of-the-art method. The occlusion is handled by updating correlation filter passively to retain it nearer to the previous one. Due to limited number of pages only few results are tabulated and displayed. Figure 3 and Figure 4 show the comparison of the Mean OP and FPS plot with the STRCF tracker on OTB dataset.

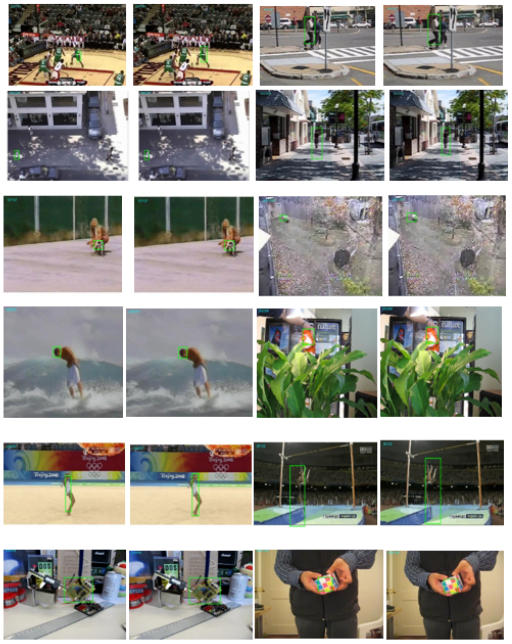

Fig 2. Qualitative evaluation on video sequences (Basketball, Couple, Crowds, Human9, Dog, Panda, Surfer, Tiger2, Jump, Gym, Board and Rubik). Figure shows the comparison results of STRCF and Proposed method respectively using same frames of the videos.

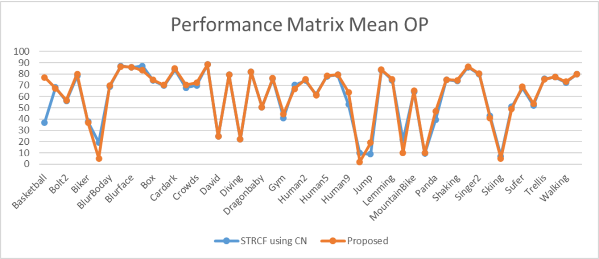

Fig 3. Comparison plot of performance matrix Mean OP

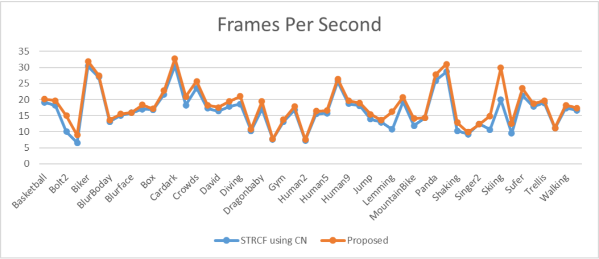

Fig 4. Comparison of performance matrix FPS

4 Conclusion

This study newly introduces a min-max feature extraction method to extract the window-based feature; wherein the minimum and maximum value of each pixel of R, G and B will be modelled to extract the target feature to achieve the better tracking. Along with this, HOG and CNN feature extraction methods are combined. The STRCF model is adopted which is more robust in case of any large variation in the appearance and ADMM is used to solve optimization problem. The proposed model when compared with STRCF based on features, has accomplished favorably against state-of-the-art tracker by increasing computational speed 18.27 and Mean OP 61.44%. The model was performed on OTB 2015 benchmark dataset and the output shows that min-max feature extraction gives superior result over color name feature used in the STRCF in terms of mean op and FPS. Here, we worked on single object tracking and used single dataset. In future, we will focus on merging both CN and Min-max features, in order to track both single and multiple object using multiple dataset for better performance.

References

1) Lu X, Ma C, Ni B, Yang X. Adaptive Region Proposal With Channel Regularization for Robust Object Tracking. *IEEE Transactions on Circuits and Systems for Video Technology*. 2021;31(4):1268–1282.
2) Ondrasovic M, Tarabek P. Siamese Visual Object Tracking: A Survey. *IEEE Access*. 2021;9:110149–110172.
3) Kumar A, Walia GS, Sharma K. Recent trends in multicue based visual tracking: A review. *Expert Systems with Applications*. 2020;162:113711–113711.
4) Kumar A, Walia GS, Sharma K. A novel approach for multi-cue feature fusion for robust object tracking. *Applied Intelligence*. 2020;50(10):3201–3218.
5) Zhu XF, Wu XJ, Xu T, Feng ZH, Kittler J. Robust Visual Object Tracking Via Adaptive Attribute-Aware Discriminative Correlation Filters. *IEEE Transactions on Multimedia*. 2022;24:301–312.
6) Yang H, Wang J, Miao Y, Yang Y, Zhao Z, Wang Z, et al. Combining Spatio-Temporal Context and Kalman Filtering for Visual Tracking. *Mathematics*. 2019;7(11):1059–1059. Available from: https://doi.org/10.3390/math7111059.
7) Fang S, Ma Y, Li Z, Zhang B. A visual tracking algorithm via confidence-based multi-feature correlation filtering. *Multimedia Tools and Applications*. 2021;80(16):23963–23982. Available from: https://doi.org/10.1007/s11042-021-10804-4.
8) Zhu H, Han Y, Wang Y, Yuan G. Hybrid Cascade Filter With Complementary Features for Visual Tracking. *IEEE Signal Processing Letters*. 2021;28:86–90. Available from: https://doi.org/10.1109/LSP.2020.3039933.
9) Xu T, Feng ZH, Wu XJ, Kittler J. Joint group feature selection and discriminative filter learning for robust visual object tracking. *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019;p. 7950–7960. Available from: https://doi.org/10.48550/arXiv.1907.13242.
10) Huang Y, Zhao Z, Wu B, Mei Z, Cui Z, Gao G. Visual object tracking with discriminative correlation filtering and hybrid color feature. *Multimedia Tools and Applications*. 2019;78(24):34725–34744. Available from: https://doi.org/10.1007/s11042-019-07901-w.
11) Kumar A, Walia GS, Sharma K. Real-time visual tracking via multi-cue based adaptive particle filter framework. *Multimedia Tools and Applications*. 2020;79(29-30):20639–20663. Available from: https://doi.org/10.1007/s11042-020-08655-6.
12) Wang Y, Luo X, Ding L, Wu J, Fu S. Robust visual tracking via a hybrid correlation filter. *Multimedia Tools and Applications*. 2019;78(22):31633–31648. Available from: https://doi.org/10.1007/s11042-019-07851-3.
13) Hu G, Dixit C, Qi G. Discriminative Shape Feature Pooling in Deep Neural Networks. *Journal of Imaging*. 2022;8(5):118–118. Available from: https://doi.org/10.3390/jimaging8050118.
14) Liu M, Ma J, Zheng Q, Liu Y, Shi G. 3D Object Detection Based on Attention and Multi-Scale Feature Fusion. *Sensors*. 2022;22(10):3935–3935. Available from: https://doi.org/10.3390/s22103935.
15) Zhao J, Ji S, Cai Z, Zeng Y, Wang Y. Moving Object Detection and Tracking by Event Frame from Neuromorphic Vision Sensors. *Biomimetics*. 2022;7(1):31–31. Available from: https://doi.org/10.3390/biomimetics7010031.
16) Qi G, Zhang Y, Wang K, Mazur N, Liu Y, Malaviya D. Small Object Detection Method Based on Adaptive Spatial Parallel Convolution and Fast Multi-Scale Fusion. *Remote Sensing*. 2022;14(2):420–420. Available from: https://doi.org/10.3390/rs14020420.
17) Basavaraju H, Aradhya VM, Guru D. A novel arbitrary-oriented multilingual text detection in images/video. *Information and decision sciences*. 2018;701:519–529. Available from: https://doi.org/10.1007/978-981-10-7563-6_54.
18) Li F, Tian C, Zuo W, Zhang L, Yang MH. Learning spatial-temporal regularized correlation filters for visual tracking. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018;p. 4904–4913. Available from: https://doi.org/10.48550/arXiv.1803.08679.
19) Crammer K, Dekel O, Keshet J, Shalev-Shwartz S, Singer Y. Online passive aggressive algorithms. 2006. Available from: https://www.jmlr.org/papers/volume7/crammer06a/crammer06a.pdf.
20) Petersen KB, Pedersen MS. The matrix cookbook. *Technical University of Denmark*. 2008;7(15):510–510. Available from: https://ece.uwaterloo.ca/~ece602/MISC/matrixcookbook.pdf.
21) Wu MY, Lim J, Yang MH. Object tracking benchmark. 2015;37:1834–1848. Available from: https://doi.org/10.1109/TPAMI.2014.2388226.