Spatio-temporal Weighted Histogram based Mean Shift for Illumination Robust Target Tracking

Kalyani Deopujari
MES College of Engineering
Pune, India
kjdeopujari@gmail.com

Rajbabu Velmurugan
Indian Institute of Technology Bombay
Mumbai, India
rajbabu@ee.iitb.ac.in

Kanchan Tiwari
MES College of Engineering
Pune, India
kanchan.tiwari@mescoepune.org

ABSTRACT

This paper proposes a simple method to handle illumination variation in a video. The proposed method is based on generative mean shift tracker, which uses energy compaction property of discrete Cosine transform (DCT) to handle illumination variation within and across frames. The proposed method uses spatial and temporal DCT coefficient based approach to assign weights to target and candidate histograms in mean shift. The proposed weighing factor takes care of changes in illumination within a frame i.e., illumination change of the target with respect to background and also across the frames i.e., varying illumination between the consecutive time instances. The algorithm was tested using VOT2015 challenge dataset and also on sequences from OTB and CAVIAR datasets. The proposed method was also tested rigorously for illumination attribute. The qualitative and quantitative evaluation process of the proposed method was twofold. First, the tracker was compared with existing DCT coefficient based method and showed improved results. Secondly, the proposed algorithm was compared with other state of the art trackers. The results show that the proposed algorithm outperformed some state-of-the-art trackers while with others it showed comparable performance.

CCS Concepts

•Computing methodologies → Computer vision; Visual inspection; Vision for robotics; Tracking; •Applied computing → *Surveillance mechanisms;*

Keywords

Mean Shift; Discrete Cosine Transform; DC Coefficient; Illumination Variation

1. INTRODUCTION

Video tracking and analysis has wide range of applications such as human computer interaction, traffic management, surveillance, medical imaging, military applications, video editing, video communication and compression. Most tracking methods require an object detection mechanism either in every frame or when the object first appears in the video. Object detection involves locating objects in the frame of a video sequence. One of the main issues in robust visual tracking is to handle appearance change of the target object. Appearance of the target can change due to variation in illumination, occlusions, pose changes, background clutter and other factors. On the basis of appearance model used, tracking algorithms can be grouped as generative or discriminative trackers. Generative tracking represents the target object in a particular feature space and then searches for the best matching score within the image region. It does not require a large dataset for training. On the other hand discriminative tracking treats visual tracking as a binary classification problem, to define the boundary between a target image patch and the background. It requires a large dataset in order to achieve good performance. While numerous algorithms of both categories have been proposed with success, it remains a challenging task to develop a tracking algorithm that is both accurate and efficient. In order to address challenging factors in visual tracking, numerous features and models have been used till date to represent target objects. The appearance of an object changes drastically when illumination varies significantly. Thus, change in illumination is one of the major challenges to be dealt with in the context of appearance changes. In [10] authors have used locality sensitive histogram model to deal with variation in illumination. This locality sensitive histogram takes into account weighted contribution of each pixel of the image. This method fails because of its low computational speed for large image sizes. In [16] DCT based features are used to weigh the probability distribution functions used in mean shift. This method fails for long video sequences and drastic changes in illumination. A low complexity mean shift tracker is proposed in [4]. It fails when an object undergoes changes in illumination conditions. Extensive research addressing these issues of illumination exist in published literature [11], [18], [19] that refer to overcoming drawbacks of mean shift tracking using color as feature. Fragmented weighted mean shift tracking using color feature is proposed in [11] and [19]. However, they cannot handle sudden changes in light condition. Multi feature tracking algorithms were proposed in [20] to overcome dependency on color. These algorithms track various complementary features enabling robust tracking. Among the existing techniques, [15] can handle minor changes in illumination but fails with sudden changes in illumination. In [14] the au-

ICVGIP '16 Guwahati, Assam India

DOI: http://dx.doi.org/10.1145/3009977.3010059

thors proposed a face tracking technique to solve the problem of flash light by setting the DC coefficient to zero, but the method cannot handle low illumination and frame wise illumination changes. The proposed algorithm is based on modifying the DC coefficients based on spatial and temporal illumination changes.

2. MEAN SHIFT THEORY AND ILLUMINATION INVARIANCE

2.1 Mean Shift Algorithm

With advances in state-of-the-art tracking algorithms, mean shift still holds its importance because of its low complexity and fast convergence. Hence, it is preferred for real-time applications. Mean shift is a non-parametric method to find modes of the density functions. In other words, it finds out the densest region through an iterative process. The target and candidate histograms are as follows

$$\hat{q}_u = C \sum_{i=1}^{n} k\left(\|x_i^*\|^2\right)\delta[b(x_i^* - u)] \tag{1}$$

$$\hat{p}_u(y) = C_k \sum_{i=1}^{n} k\left(\left\|\frac{y - x_i}{h}\right\|^2\right)\delta[b(x_i) - u] \tag{2}$$

where $\hat{q}_u$ is the target color model in the first frame, $\hat{p}_u(y)$ is the candidate model for consecutive frames, C, C_k are normalization constants, y is the target center, x_b are set of data points, k represents Epanechnikov kernel, u represents color of the target model, and h represents bandwidth. The weights are calculated as

$$w_i = \sum_{u=1}^{m} \sqrt{\frac{\hat{q_u}}{\hat{p_u}(\hat{y_0})}}\delta[b(x_i) - u] \tag{3}$$

Using the target model and target candidate, mean shift vector estimates the target center in the next frame using

$$\hat{Y} = \frac{\sum_{i=1}^{n} x_i w_i g \left\|\frac{\hat{y}_0 - x_i}{h}\right\|^2}{\sum_{i=1}^{n} w_i g \left\|\frac{\hat{y}_0 - x_i}{h}\right\|^2} \tag{4}$$

Here w_i is obtained using (3) and is used to calculate the new center for the target as given in (4), where $g()$ is the negative of derivative of the kernel, and m is the number of histogram bins. The center of the kernel is then shifted from y_0 to a new center Y. This is repeated till the candidate model is close to the target model. Here, we use Bhattacharyya distance ρ for similarity measurement

$$\hat{\rho}(y) = \rho[\hat{p}(y), \hat{q}] = \sum_{u=1}^{m} \sqrt{\hat{p}_u(y), \hat{q}_u}. \tag{5}$$

2.2 Background Weighted Histogram and Illumination Invariance

In target tracking, the background information is often included in the detected target region. If the similarity between the target and the background is high, the tracking process will be poor. To reduce the interference of background, a representative model for background features is proposed in [5]. It selects discriminative features from the target region and the background region to be used as the weight for target model and target candidate. Here, histogram is used as a discriminative feature between foreground and the background to calculate weights. These weights can be used to define the target and candidate histograms during tracking and obtained as

$$v_u = \min\left(\frac{\hat{O}^*}{\hat{O}_u}, 1\right), \quad u = 1, 2, \ldots, m \tag{6}$$

where $\hat{O}^*$ is the background color histogram and $\hat{O}_u$ is the target color histogram, and v is the weight. Now let $\widehat{\lambda}$ be the weight for target histogram and $\widehat{\lambda}_p$ be weight for candidate histogram. Both $\widehat{\lambda}$ and $\widehat{\lambda}_p$ are derived from discriminative foreground and background features. Weight $\widehat{\lambda}$ is used to modify the target model and $\widehat{\lambda}_p$ to modify the target candidate. From (1) and (2), the modified target model and target candidate are obtained as.

$$q'_u = c_1 \sum_{i=1}^{n} k(\|x_i\|^2)\delta[\widehat{\lambda} b(x_i) - u], \text{ and} \tag{7}$$

$$p'_u(y\|) = c_2 \sum_{i=1}^{n} k\left(\left\|\frac{y - x_i}{h}\right\|^2\right)\delta[\widehat{\lambda}_p b(x_i) - u]. \tag{8}$$

These are used in (4) to calculate the target center. The weights calculated also determine the convergence of the tracking algorithm. Since we modify the target model and target candidate based on illumination, the weights get modified accordingly. Using (3) modified weight is obtained as

$$w'_i = \sum_{u=1}^{m} \delta[b(x_i) - u]\sqrt{\frac{q'_u}{p'_u(y)}} \tag{9}$$

and from equations (7),(8),(9) we have

$$w'_i = \sum_{u=1}^{m} \delta[b(x_i) - u]\sqrt{\frac{c_1 \sum_{i=1}^{n} k(\|x_i\|^2)\delta[\widehat{\lambda} b(x_i) - u]}{c_2 \sum_{i=1}^{n} k(\|\frac{y - x_i}{h}\|^2)\delta[\widehat{\lambda}_p b(x_i) - u]}} \tag{10}$$

Rewriting using normalization factors c_1 and c_2 in equation (10) we have

$$w'_i = \sum_{u=1}^{m} \delta[b(x_i) - u]\sqrt{\frac{c_1 c_2}{c_2 c_1}} \times \sqrt{\frac{\delta[\widehat{\lambda} b(x_i) - u]}{\delta[\widehat{\lambda}_p b(x_i) - u]}} \tag{11}$$

Simplifying (11) for particular bin u, we have

$$w'_i = \sqrt{\frac{\widehat{\lambda}}{\widehat{\lambda}_p}} \times w_i. \tag{12}$$

Eq. (12) clearly reflects the fact that the weight calculated using the modified target representation (using illumination) makes the algorithm illumination invariant.

3. PROPOSED METHODOLOGY

Figure 1 shows a block diagram for the proposed methodology. Video as a sequence of frames is given as an input to the tracker system. The target is represented as a rectangular bounding box defined by its center, height, and width. Target and background areas are defined in the first frame. Background area is taken two times the target area. In mean shift algorithm the target and candidate histograms are calculated as explained in Sec. 2. These histograms are weighted with a spatio-temporal weighing factor to compensate for changing intensity levels within and across frames. Thus, target is tracked in consecutive frames using mean shift with modified histograms resulting in an illumination robust tracking algorithm.

3.1 Spatio-temporal Weighing Factor

Illumination variations are expected to be captured in the low frequency components. $DCT - II$ is used to transform a logarithmic image to frequency domain. The $M \times N$ 2-D DCT of frame f can be defined as in [8]. Proposed method is based on energy compaction property of DCT. By energy compaction property it means that most of the energy is concentrated in the 0th or DC coefficient of DCT. Hence, weighing factor for histogram is calculated in order to take care of illumination variation spatially as well as temporarily. By spatially we mean illumination changes within a frame, between target and the background. By temporarily we mean illumination from frame to frame. For calculating the weighing factor we consider ratio of DC coefficients of foreground and background in the previous as well as the present frame. Spatial illumination compensation ratio for previous and present frame are given by

$$\lambda_{t-1}^{spatial} = \frac{(\lambda_{dc}^{fr})_{t-1}}{(\lambda_{dc}^{bg})_{t-1}} \tag{13}$$

$$\lambda_{t}^{spatial} = \frac{(\lambda_{dc}^{fr})_{t}}{(\lambda_{dc}^{bg})_{t}} \tag{14}$$

$$\lambda_{temporal} = \frac{\lambda_{t-1}^{spatial}}{\lambda_{t}^{spatial}} \tag{15}$$

$\lambda_{t-1}^{spatial}$ and $\lambda_{t}^{spatial}$ are the ratios of DC coefficient of foreground and background of previous frame at $(t-1)^{th}$s and present frame at $(t)^{th}$s, respectively. The ratio of foreground to background DC coefficient takes care of spatial illumination variation. For example, for present frame at $(t)^{th}$s if illumination of foreground is increased with respect to background, the value of $(\lambda_{dc}^{fr})_t$ will be more with respect to $(\lambda_{dc}^{bg})_t$ thus increasing the ratio. Ratio decreases for the vice verse. The proposed weighing factor $\lambda_{temporal}$ is the ratio of these two spatial ratios at $(t)^{th}$ and $(t-1)^{th}$ frames. Now if illumination in frame is increasing from $(t-1)^{th}$ instance to a frame at $(t)^{th}$ instance, value of $\lambda_{t}^{spatial}$ increases with respect to $\lambda_{t-1}^{spatial}$, which in turn decreases $\lambda_{temporal}$ with respect to weighing factor of previous frame. When this reduced $\lambda_{temporal}$ is multiplied with candidate histogram $\hat{p}_u(y)$ in current frame at $(t)^{th}$ instance, it suppresses increased illumination effect on histogram feature. On the other hand, if illumination in frame is decreased from $(t-1)^{th}$ instance to a frame at $(t)^{th}$ instance, value of $\lambda_{t}^{spatial}$ decreases with respect to $\lambda_{t-1}^{spatial}$ which in turn increases $\lambda_{temporal}$ with respect to weighing factor of previous frame. When this reduced $\lambda_{temporal}$ is multiplied with candidate histogram $\hat{p}_u(y)$ in current frame at $(t)^{th}$ instance, it enhances decreased illumination effect on histogram feature.

4. EXPERIMENTAL RESULTS

The proposed tracker was evaluated on VOT2015 benchmark. It was also tested on videos that specifically include illumination variation attribute from other popular benchmarks like OTB and CAVIAR. In VOT evaluation, first the proposed tracker was evaluated with the baseline tracker given in [16]. Secondly, it was evaluated with nine other state-of-the-art trackers.

4.1 Evaluation in VOT2015 Challenge

The proposed algorithm was compared with baseline tracker from [16] along with nine other state-of-the-art trackers in VOT2015 challenge framework. The ten trackers included for evaluation process were EBT [24], Struck [9], IVT [17], DSST [6], OAB [1], MIL [2], SRDCF [7], MEEM [22], VOT2015 baseline (NCC) and baseline [16]. For fair evaluation of the proposed methodology, the tracker selection was done such that the comparison will be with popular and recent trackers. Comparison was done on 15 challenging videos with different attributes such as camera motion, size change, illumination variation, occlusion and motion change. Two weakly correlated performance measures (i) Accuracy and (ii) Robustness are used due to their high level of interpretability [12]. Accuracy measures the overlap of the bounding box predicted by the tracker with the ground truth. Robustness measures the number of times the tracker loses the target during tracking. VOT2015 implements an averaging scheme for participating trackers, i.e., considering a set of trackers in which four top-performing trackers perform equally well, the averaging scheme will assign them a rank of 2.5. It means that no tracker will be ranked as 1. Evaluation scheme used in the VOT2015 challenge is explained in [12]. In order to avoid label conflict with VOT baseline tracker, we address the baseline in [16] as 'wms'. In VOT analysis the proposed tracker performs significantly better than 'wms'. Figure 2 shows the experimental baseline results for the two trackers. Accuracy and Robustness rank of proposed tracker was 1 with respect to 'wms' rank of 1.73. Average overlap score for the proposed tracker was 0.55 against an overlap score 0.37 of for 'wms'. Average number of failures with the proposed tracker was 1.87 against 10.27 for 'wms'.

To analyse performance w.r.t. the different visual attributes, the two measures can be calculated only on the subset of frames in the dataset that contain a specific attribute (attribute subset). The trackers are ranked with respect to each measure. Figure 3 shows attribute specific ordering plot for the proposed and baseline tracker. It is a plot of average expected overlap vs rank. From [12] average expected overlap can be defined as averaging the average overlap on very large set of frames in long sequences. It can be seen in Fig. 3 that the proposed tracker outperforms the baseline tracker in all attributes, especially for illumination change. Because of its unique weighing parameters there is a drastic improvement in handling illumination variations. Better accuracy of features leads to good overlap score. It not only shows superior performance in illumination change attribute, but also excels significantly in other attributes like camera motion, occlusion handling, size change and motion change.

In Figure 4 sequence specific AR and Ranking plots are shown. For representation purpose results of 4 sequences out of 15 have been displayed. In each of the 15 videos proposed tracker surpassed baseline in accuracy and robustness thus securing first rank. For sequence ball1 accuracy rank of msdct is 1 with overlap score of 0.80 (0.69 for wms) and robustness rank is also 1 with 0 failures. For sequence handball1 accuracy rank of msdct is 1 with overlap score of 0.60

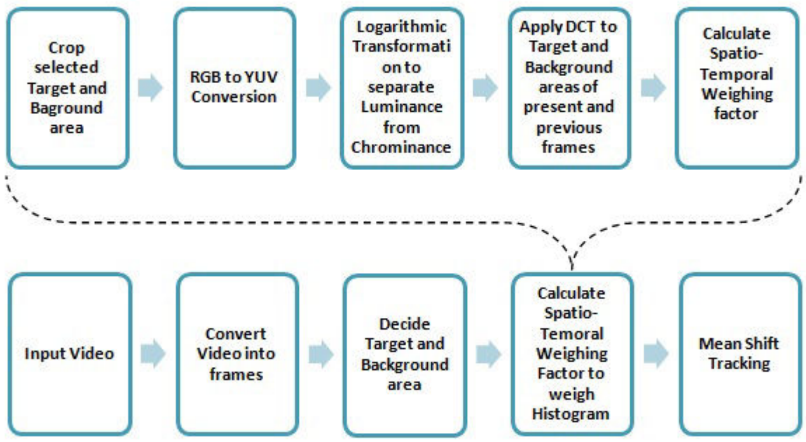
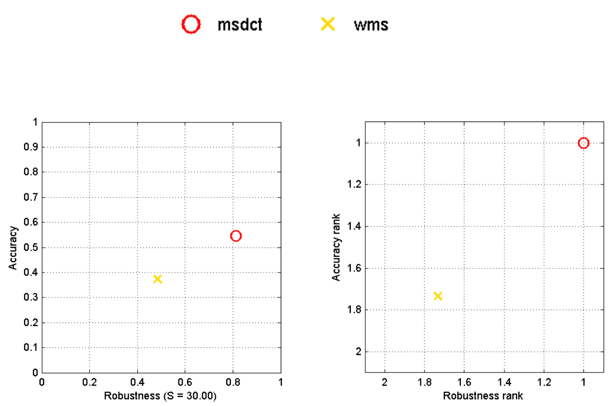


Figure 1: Block diagram of the proposed methodology.



(a) (b)

Figure 2: (a) AR and (b) Ranking plot for proposed (msdct) and baseline (wms) tracker. Accuracy and Robustness rank for the proposed tracker is 1 with average overlap score of 0.55 (against 0.37 for wms) and average number of failures are 1.87 (against 10.27 for 'wms').

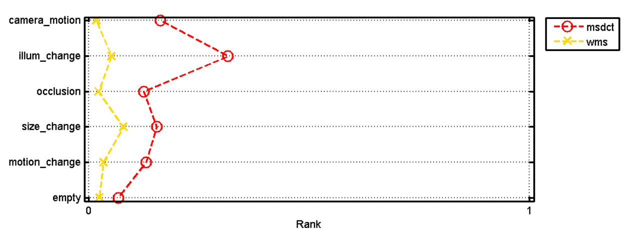


Figure 3: Attribute specific ordering plot of average expected overlap for proposed tracker and baseline 'wms' tracker.

(0.32 for wms) and robustness rank is also 1 with 5 failures (8 failures for wms). For sequence iceskater1 accuracy rank of msdct is 1 with overlap score of 0.47 (0.20 for wms) and robustness rank is also 1 with 0 failures (61 failures for wms). For sequence motocross2 accuracy rank of msdct is 1 with overlap score of 0.62 (0.27 for wms) and robustness rank is also 1 with 1 failure (2 failures for wms).

4.1.1 *Evaluation of proposed tracker with state of the art trackers*

Proposed tracker was compared with 10 trackers on 15 benchmark videos as an attempt for fair evaluation. Table 1 gives AR ranking for all trackers. Proposed method ranked first (1.00) in accuracy followed by MEEM (1.33) and SRDCF (1.40). On the other hand, in robustness it ranked fourth (3.67) after MEEM (3.53). It showed better performance than other trackers such as EBT, IVT, baseline, Struck, DSST, OAB and MIL. Figure 6 show plots for experimental evaluation of proposed tracker with 10 other trackers. In both AR and ranking plots proposed msdct can be seen surpassing the performance of other 10 trackers, quite significantly. But because of the low robustness values no tracker could show outstanding performance in AR plot. The proposed tracker was followed by MEEM, Struck, SRDCF and EBT in baseline performances. Figure 7 shows attribute specific ordering plot for all 10 trackers. It is a plot of average expected overlap vs. rank. For all the 5 attributes the proposed tracker shows comparable results with other popular state-of-the-art trackers. Since the proposed tracker does not consider scaling it cannot be seen performing best in the category, though shows comparable performance with trackers that handle scale change. Figure 5 shows sequence

Trackers	Accuracy	Robustness
msdct (Proposed)	1.00	3.67
wms [16]	5.60	8.13
SRDCF [7]	1.40	3.27
EBT [24]	1.53	1.60
IVT [17]	3.87	7.53
baseline (VOT)	2.67	5.20
Struck [9]	1.80	3.27
DSST [6]	2.87	6.67
OAB [1]	3.67	7.80
MIL [2]	3.47	6.20
MEEM [22]	1.33	3.53

Table 1: Ranking of msdct with respect to 10 state-of-the-art trackers. Color code 'red' corresponds to rank 1. Color code 'blue' and 'green' corresponds to rank 2 and rank 3, respectively.

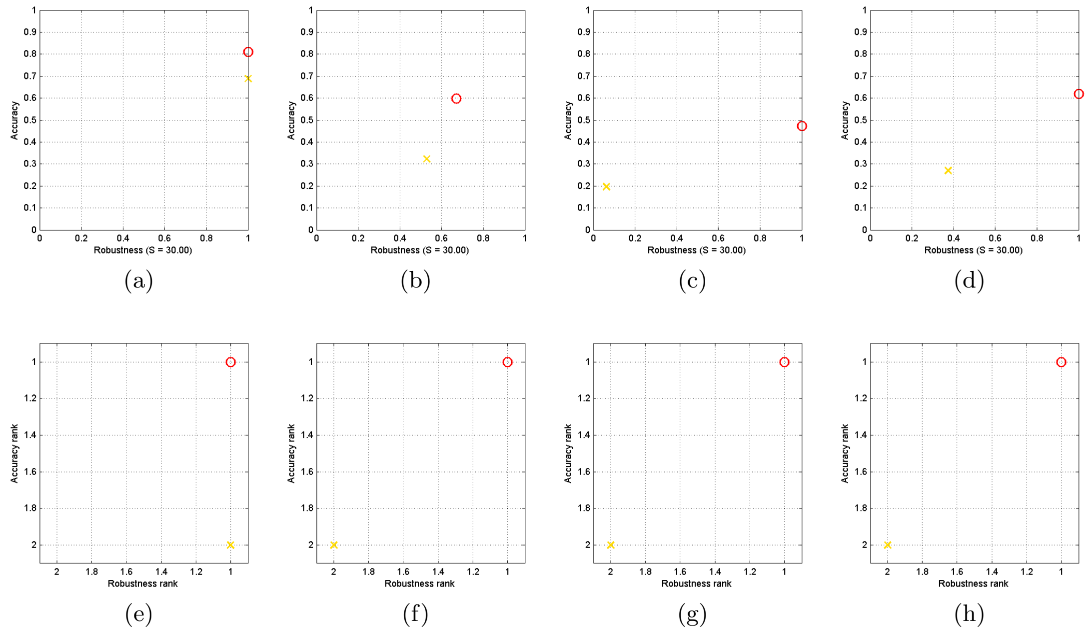


Figure 4: Sequence specific AR and Ranking plots for proposed tracker comparison with baseline (wms) are shown from (a)-(h). First row from (a)-(d) are AR plots for sequences *ball1*, *handball1*, *iceskater1*, *motocross2*, respectively. Second row from (e)-(h) are Ranking plots for sequences *ball1*, *handball1*, *iceskater1*, *motocross2*, respectively. Proposed msdct outperformed baseline (wms) in all the videos.

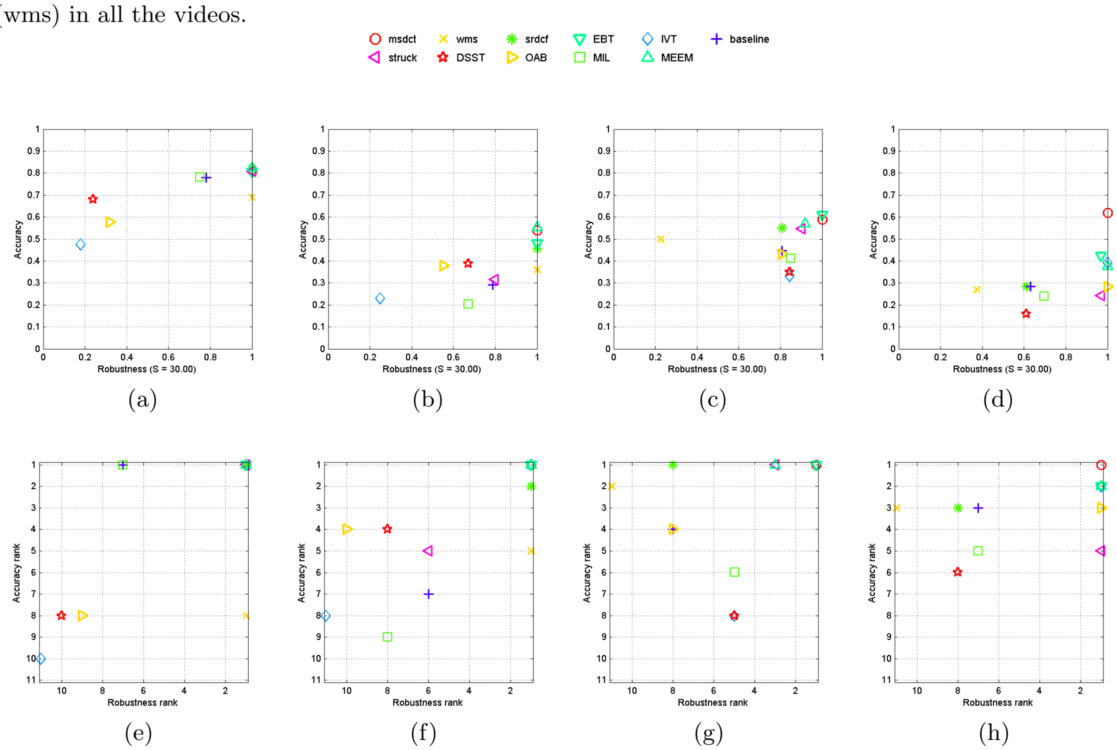
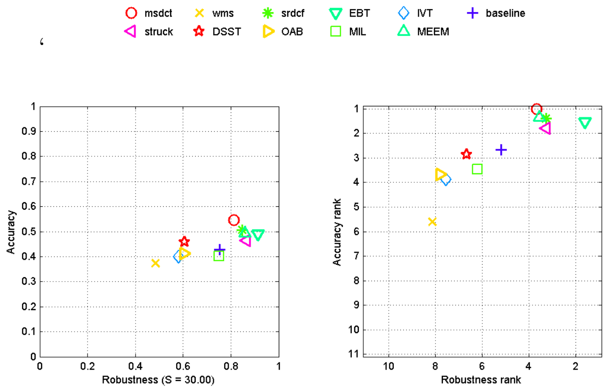


Figure 5: Sequence specific AR and Ranking plots for the proposed tracker in comparison with state-of-art trackers are shown from (a)-(h). First row from (a)-(d) are AR plots for sequences *ball1*, *handball1*, *iceskater1*, *motocross2* respectively. Second row from (e)-(h) are Ranking plots for sequences *ball1*, *handball1*, *iceskater1*, *motocross2* respectively.Proposed msdct outperformed in all the videos.

wise performance of the trackers. In majority of the 15 benchmark videos the proposed tracker showed improved results while in others it showed comparable results with top state of art trackers like MEEM, EBT, SRDCF, Struck. It outperformed the other trackers including baseline tracker for VOT. Analysis in VOT 2015 benchmark confirms that the proposed tracker not only proved robust to variation in illumination but it also excelled when compared with 10



(a) (b)

Figure 6: (a) AR and (b) Ranking plot for all the 10 trackers. Proposed method ranked first (1.00) in accuracy. In robustness it ranked fourth (3.67).

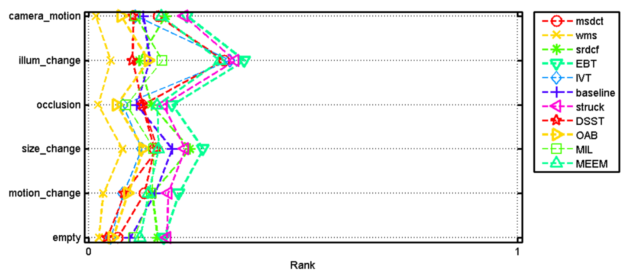


Figure 7: Attribute specific ordering plot of average expected overlap for 10 trackers.

popular state-of-the-art trackers in VOT 2015 challenge. Visual analysis of different trackers can be shown in Figure 8. All the sequences are from VOT database.

4.2 Evaluation on OTB and CAVIAR dataset

The quantitative evaluation of the proposed tracker on sequences specific to illumination changes from OTB [21] and CAVIAR dataset was performed. The trackers included in the evaluation were IVT [17], CT [23], Mix model [3] and VTD [13]. The criteria used was root mean square error (RMSE) which is the Euclidean distance between center of the target in tracked frame $\{x_i, y_i\}$ and ground truth center $\{centerx_i, centery_i$ in each frame i=1,2,..,N_f, where N_f is the total number of frames. RMSE for both the coordinates is given by

$$RMSE_x = \sqrt{\frac{1}{N_f}\sum_{i=1}^{N_f}(centerx_i - x_i)^2} \tag{16}$$

$$RMSE_y = \sqrt{\frac{1}{N_f}\sum_{i=1}^{N_f}(centery_i - y_i)^2} \tag{17}$$

$$RMSE_{xy} = \frac{RMSE_x + RMSE_y}{2} \tag{18}$$

and $RMSE_{xy}$ is the performance parameter. Figure 9 shows the plots of RMSE for videos *david*, *trellis*, *frame* and *OneLeaveShop2cor*, which have significant illumination variation throughout the sequences. In all four sequences the proposed method has minimum Euclidean distance values when compared with other four trackers, indicating that the proposed method has better robustness for illumination variation than existing methods included in the study.

5. CONCLUSION

In this work, we proposed a simple but effective tracking method for handling illumination variations. The method was compared using the VOT2015 benchmark with trackers such as EBT, MEEM, MIL, IVT, OAB, DSST, baseline (VOT), SRDCF, Struck and wms. We considered 15 video sequences with challenging illumination variations and other attributes like camera motion, size change, occlusion, shape change and motion change. When compared with baseline trackers the proposed algorithm showed exceptional robustness towards illumination changes. It ranked first in accuracy followed by MEEM and SRDCF for both baseline and overall evaluation on 15 videos. On the other hand, in Robustness it stood fourth. It showed better performance than EBT, IVT, baseline, Struck, DSST, OAB and MIL. In the second stage of evaluation, the proposed tracker was tested on illumination specific videos from OTB and CAVIAR with other set of trackers. RMSE plot for all the videos showed better performance of the proposed method. Limitation of the proposed method is that it is not adaptive to handle scale change or orientation of the target. We plan to address these issues in future work.

6. REFERENCES

[1] B. Babenko, M.-H. Yang, and S. Belongie. Visual tracking with online multiple instance learning. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 983–990. IEEE, 2009.

[2] B. Babenko, M.-H. Yang, and S. Belongie. Robust object tracking with online multiple instance learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8):1619–1632, 2011.

[3] F. Bousetouane, L. Dib, and H. Snoussi. Improved mean shift integrating texture and color features for robust real time object tracking. *The Visual Computer*, 29(3):155–170, 2013.

[4] D. Comaniciu, V. Ramesh, and P. Meer. Real-time tracking of non-rigid objects using mean shift. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 142–149. IEEE, 2000.

[5] D. Comaniciu, V. Ramesh, and P. Meer. Kernel-based object tracking. *IEEE Transactions on pattern analysis and machine intelligence*, 25(5):564–577, 2003.

[6] M. Danelljan, G. Häger, F. Khan, and M. Felsberg. Accurate scale estimation for robust visual tracking. In *British Machine Vision Conference, Nottingham, September 1-5, 2014*. BMVA Press, 2014.

[7] M. Danelljan, G. Hager, F. Shahbaz Khan, and M. Felsberg. Learning spatially regularized correlation filters for visual tracking. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4310–4318, 2015.

[8] Gonzalez and Woods. *Digital Image processing.* Prentice hall, fifth edition, 2005.

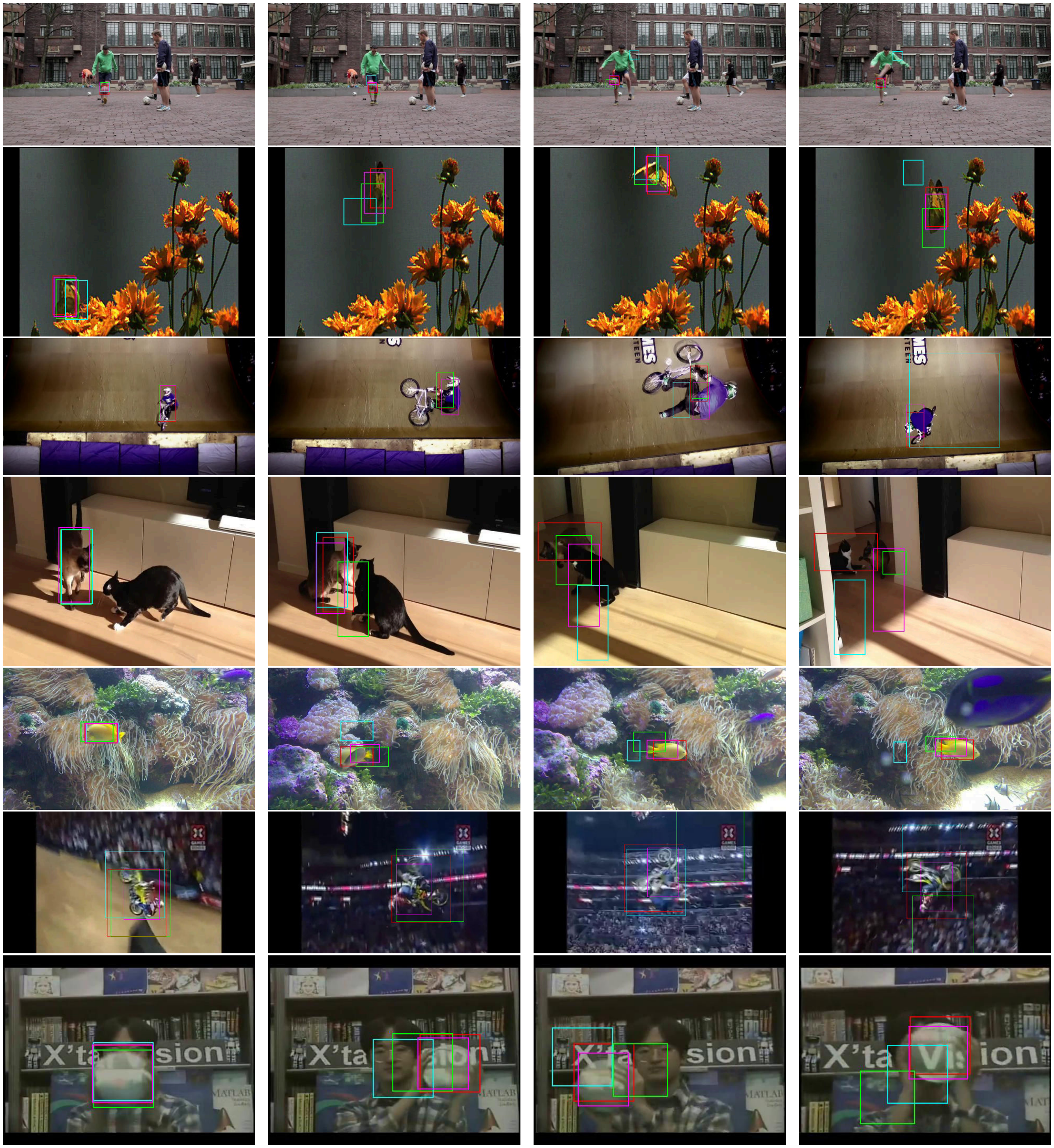


Figure 8: Visual analysis of proposed msdct with respect to baseline wms, OAB and EBT on sequences (row wise) ball1, butterfly, bmx, fernando, fish3, motocross2 and sphere. All the sequences are from VOT database.

[9] S. Hare, S. Golodetz, A. Saffari, V. Vineet, M.-M. Cheng, S. Hicks, and P. Torr. Struck: Structured output tracking with kernels. 2015.

[10] S. He, Q. Yang, R. W. Lau, J. Wang, and M.-H. Yang. Visual tracking via locality sensitive histograms. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2427–2434, 2013.

[11] J. Jeyakar, R. V. Babu, and K. Ramakrishnan. Robust object tracking with background-weighted local kernels. *Computer Vision and Image Understanding*, 112(3):296–309, 2008.

[12] M. Kristan, J. Matas, A. Leonardis, M. Felsberg, L. Cehovin, G. Fernandez, T. Vojir, G. Hager, G. Nebehay, and R. Pflugfelder. The visual object tracking VOT2015 challenge results. In *The IEEE*

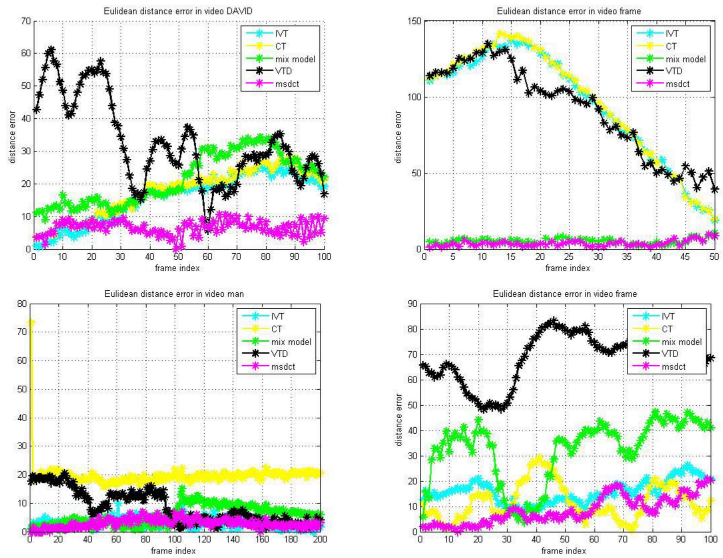


Figure 9: Comparison of the proposed tracker using OTB and CAVIAR sequences with other state-of-the-art trackers.

International Conference on Computer Vision (ICCV) Workshops, December 2015.

[13] J. Kwon and K. M. Lee. Visual tracking decomposition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1269–1276. IEEE, 2010.

[14] M. La Cascia, S. Sclaroff, and V. Athitsos. Fast, reliable head tracking under varying illumination: An approach based on registration of texture-mapped 3d models. *IEEE Transactions on pattern analysis and machine intelligence*, 22(4):322–336, 2000.

[15] J. Ning, L. Zhang, D. Zhang, and C. Wu. Robust mean-shift tracking with corrected background-weighted histogram. *IET Computer Vision*, 6(1):62–69, 2012.

[16] G. Phadke et al. Improved weighted histogram for illumination invariant mean-shift tracking. In *Proceedings of the 2014 Indian Conference on Computer Vision Graphics and Image Processing*, page 83. ACM, 2014.

[17] D. A. Ross, J. Lim, R.-S. Lin, and M.-H. Yang. Incremental learning for robust visual tracking. *International Journal of Computer Vision*, 77(1-3):125–141, 2008.

[18] J. Vermaak, P. Pérez, M. Gangnet, and A. Blake. Towards improved observation models for visual tracking: selective adaptation. In *European Conference on Computer Vision*, pages 645–660. Springer, 2002.

[19] F. Wang, S. Yu, and J. Yang. Robust and efficient fragments-based tracking using mean shift. *AEU-international Journal of Electronics and Communications*, 64(7):614–623, 2010.

[20] J. Wang and Y. Yagi. Integrating color and shape-texture features for adaptive real-time object tracking. *IEEE Transactions on Image Processing*, 17(2):235–240, 2008.

[21] Y. Wu, J. Lim, and M.-H. Yang. Online object tracking: A benchmark. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2013.

[22] J. Zhang, S. Ma, and S. Sclaroff. MEEM: Robust tracking via multiple experts using entropy minimization. In *European Conference on Computer Vision*, pages 188–203. Springer, 2014.

[23] K. Zhang, L. Zhang, and M.-H. Yang. Real-time compressive tracking. In *European Conference on Computer Vision*, pages 864–877. Springer, 2012.

[24] G. Zhu, F. Porikli, and H. Li. Tracking randomly moving objects on edge box proposals. *arXiv:1507.08085*, 2015.