Non-negative Matrix Factorization based Illumination Robust Meanshift Tracking

Gargi Phadke, Rajbabu Velmurugan and Shubham Dawande
Indian Institute of Technology Bombay
Mumbai, India 400076
email: {gargiphadke, rajbabu }@ee.iitb.ac.in
shubhamdawande.sd@gmail.com

***Abstract*—Object tracking is a critical task in surveillance and activity analysis. One main issue in tracking is illumination variation. We propose a method which is robust to illumination by incorporating a feature that is less variant to illumination. The proposed feature is a reflectance histogram obtained using sparsity constrained non-negative matrix factorization (NMFsc). Using NMFsc, illumination and reflectance components are separated in each frame of the input video. The obtained reflectance histogram is used as feature vector in a meanshift framework for target tracking. We also use an adaptive target model to handle target appearance changes. Experimental results using standard benchmark videos show that the proposed scheme can lead to better tracking in challenging illumination scenarios, when compared to several existing algorithms.**

I. Introduction

One of the main challenges in robust visual tracking is the change in appearance of the target. In this paper, we focus on tracking when there are changes in light condition and propose a modification to the meanshift (MS) algorithm. The MS algorithm and subsequent variants are extensively discussed from a tracking perspective in [1]. A robust and low complexity MS algorithm is proposed in [2]. The MS algorithm fails if the feature considered changes with illumination. Multiple features such as color model, motion model, and target model updates have been used to overcome the failure in MS tracking when illumination condition changes [3], [4], [5]. In [6], [7] spatio-temporal oriented energy histogram is used to handle changes in light condition. In most of these methods the target model changes with illumination leading to failure of these methods. In [8] illumination changes within a frame is handled, but not across the frame. The visual tracking decomposition method (VTD) [9], tracking, learning, detection (TLD) tracker [10], real-time compressive tracker (CT) [11], and distribution field tracker (DFT) [12] are popular methods to handle different challenges in video tracking. These methods also handle illumination changes.

As is well known, images are characterized by illumination and reflectance components [13]. The illumination component is dependent on the light source and the reflectance depends on the object. Most MS algorithms [2] depend on intensity features, which changes with illumination or light condition. In [13], homomorphic filtering is suggested as an approach to separate illuminance and reflectance. In [14] homomorphic filtering along with wavelet transforms is used for face recognition under illumination variation. Another method to separate illuminance and reflectance is using non-negative matrix factorisation (NMF) with sparseness constraints [15]. In [16] authors proposed NMF for face recognition under illumination changes. In [17] NMF is used for video tracking using particle filtering. But there the use of NMF is used to obtain a set of basis to represent the target and use it in the context of target tracking.

In this paper, we propose an efficient illumination robust tracking algorithm. Tracking is typically based on pixel information and for successful tracking target pixel values should be invariant to illumination. We propose to use the target reflectance histogram, obtained using a NMF approach [15], as the feature. An outline of the proposed method along with sparseness constrained NMF for feature extraction is explained in Sec. II. The meanshift (MS) algorithm using the proposed feature is presented in Sec. III. In Sec. IV, the successful working of the proposed method is demonstrated through visual and quantitative results on several standard tracking benchmark videos and VOT2015 challenge dataset with varying illumination conditions and different attributes.

II. NMFsc for feature extraction and tracking

In the proposed method, a reflectance histogram obtained using NMF is used to represent the target and candidate models. This modified target feature is used as MS vector. An adaptive mechanism to update the target model is also proposed to handle target appearance changes. The NMF algorithm is described next.

A. Non-negative matrix factorization with sparseness constraint (NMFsc)

Non-negative matrix factorization (NMF) is an approach to obtain parts-based, linear representations of non-negative data. For a given non-negative data $\mathbf{V} \in \mathbb{R}^{N \times T}$, NMF factorization with M basis components is given as a product of non-negative matrices $\mathbf{W} \in \mathbb{R}^{N \times M}$ and $\mathbf{H} \in \mathbb{R}^{M \times T}$ as in

$$\mathbf{V} = \mathbf{W}\mathbf{H} \tag{1}$$

978-1-5090-5356-8/17/$31.00 © 2017 IEEE

The non-negativity of all three matrices makes NMF useful for describing decomposition with a potential physical meaning. Here each column of $\mathbf{H}$ contains the coefficient vector $\mathbf{h}_t$ corresponding to the measurement vector $\mathbf{v}_t$ and $\mathbf{W}$ is the matrix containing the basis vectors $\mathbf{w}_m$. Given a data matrix $\mathbf{V}$, depending on the choice of distance measure considered, the matrices $\mathbf{W}$ and $\mathbf{H}$ can be obtained to minimize the error in the factorization. A standard distance measure is the euclidean distance, and the error is minimized to obtain the factors. As proposed in [18], multiplicative update rules can be used to obtain $\mathbf{W}$ and $\mathbf{H}$. One such approach to obtain a meaningful decomposition is to enforce sparseness on either the $\mathbf{W}$ or $\mathbf{H}$ depending on the application, as proposed in [19]. Accordingly the modified cost function to be minimized is

$$E = \|\mathbf{V} - \mathbf{W}\mathbf{H}\|_F = \sum_{i,j} (\mathbf{V}_{i,j} - (\mathbf{W}\mathbf{H})_{i,j})^2 \tag{2}$$

$$\text{s.t.,sparseness}(\mathbf{w}_i) = S_w, \forall i, \text{sparseness}(\mathbf{h}_i) = S_h, \forall i \tag{3}$$

where S_w and S_h are the desired sparseness of $\mathbf{W}$ and $\mathbf{H}$, respectively. An algorithm to obtain $\mathbf{W}$ and $\mathbf{H}$ for given sparsity constraints was proposed in [19], and we refer to this as the NMFsc algorithm. Here sparseness is measured by the l_1 norm of the vector.

B. Reflectance histogram using NMFsc

In the basic image formation model, image intensity of a pixel at a location $[x, y]$ in an image $f[x, y]$ is assumed to be the product of reflectance component $r[x, y]$ and illumination component $l[x, y]$ [13]. Consider $f[x, y]$ as the current frame in a video with reflectance $r[x, y]$ and illumination as $l[x, y]$. Then the image formation model can be given as,

$$f[x, y] = l[x, y]r[x, y] \tag{4}$$

Taking logarithmic transform of the image intensity and considering the image pixels as a vector, we have

$$log(\mathbf{f}) = log(\mathbf{l}) + log(\mathbf{r}) = \mathbf{l}_l + \mathbf{r}_l \tag{5}$$

The reflectance component in log space ($\mathbf{r}_l$) can be further represented as a weighted linear combination of R reflectance components $r = \sum_{i=1}^{R} \mathbf{j}_i \gamma_i$ where the vectors $\mathbf{j}$ are approximately independent reflectance components and γ are the weighting coefficients [15]. The advantage of this representation is that the measurement data is modeled in terms of additive components only. This image model can be combined with sparseness constrained NMF (NMFsc) representation to estimate the illumination and reflectance components as proposed in [15]. An image is assumed to contain one illumination component and one reflectance component ($R = 1$), i.e., one basis vector for each component. With $M = 2$ and $T = 1$, (1) can be written as,

$$\mathbf{V} = \sum_{i=1}^{M} \mathbf{w}_i h_{i1} = \mathbf{w}_1 h_{11} + \mathbf{w}_2 h_{21} = \mathbf{V}_l + \mathbf{V}_r \tag{6}$$

Comparing equations (5) and (6), we have the illumination component $\mathbf{V}_l$ and reflectance component $\mathbf{V}_r$ using this factorization. Here, $\mathbf{w}_1$, h_{11} and $\mathbf{w}_2$, h_{21} are the basis vectors and weights for illumination and reflectance, respectively. The input image is in the RGB plane and the input $\mathbf{V}$ is generated by concatenating the three color planes into one vector. Negative log of this normalized vector is a input matrix for NMFsc decomposition [15]. NMFsc can be used to solve for the illumination and reflectance components as it allows the sparseness for each basis vector to be controlled individually. NMFsc can be use to obtain two basis vectors, one with sparseness constraint of the illumination component (S_h) set close to 0, and another with the sparseness constraint of the reflectance (S_w) set close to 1. Thus, illumination and reflectance part of an image can be separated as shown in Fig. 1. It also shows that the reflectance component estimated using NMFsc remains invariant to change in light condition.

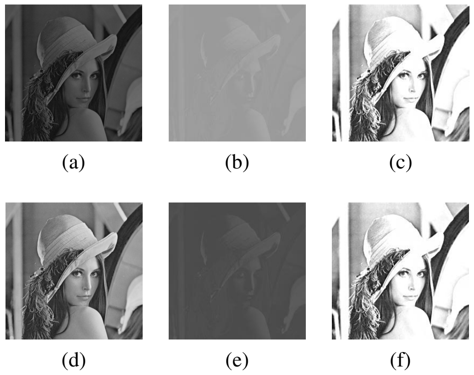


Fig. 1: Two differently illuminated *Lena* images. Separation of illumination and reflectance using $S_w = 0.3$ (for reflectance) in NMFsc. Row 1 and Row 2 show components for poorly and well illuminated images, respectively. The reflectance components in (c) and (f) obtained using NMFsc are similar, though (a) and (d) have varying illumination condition.

In videos, with change in light condition the illumination component changes but the reflectance of the object remains same. So we propose to use the reflectance component as the target histogram, which can be estimated using the NMFsc algorithm. To see this we provide the variation in the target histogram with respect to the original target in the reference frame using reflectance histograms for a sample video(*David* [20]). The similarity of histograms is obtained using the earth mover distance (EMD) [21]. The EMD between histograms $\mathbf{hist}_x$ and $\mathbf{hist}_y$ with N bins is given by,

$$EMD = \sum_{i=0}^{N-1} |\mathbf{hist}_x(i) - \mathbf{hist}_y(i)| \tag{7}$$

where $\mathbf{hist}_x(i)$ indicates the value of the i-th bin in $\mathbf{hist}_x$. Lesser EMD indicates similar histogram. We compare the variation in intensity histogram, histogram obtained using the homomorphic wavelet transform method [14], and the proposed reflectance histogram obtained using the NMFsc approach in Fig. 2. It can be seen that the reflectance histogram

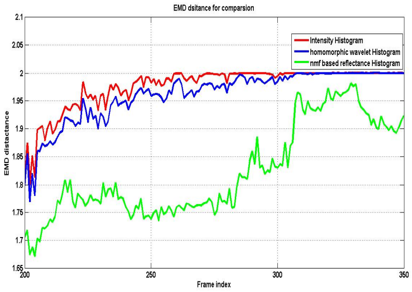


Fig. 2: Comparison of intensity histogram, histogram obtained using homomorphic wavelet transform and proposed reflectance histogram, using EMD distance on *David* video [22]. It can be seen that the reflectance histogram is relatively less variant to illumination.

is relatively less variant, though the illumination in the video changes with time.

III. Meanshift based tracking

In meanshift tracking [2], target tracking is achieved by choosing a target histogram from the region-of-interest (ROI) corresponding to the object to be tracked in the reference frame. This target histogram is compared with a candidate histogram to obtain the meanshift vector, which gives the target position. Since this is a standard algorithm, we only provide the relevant modifications to the original algorithm. The target and candidate histograms are

$$q_u = c_1 \sum_{i=1}^{n} k(\|X_i\|^2)\delta[b(X_i) - u], \text{ and} \tag{8}$$

$$p_u(Y) = c_2 \sum_{i=1}^{n} k(\|\frac{Y - X_i}{B}\|^2)\delta[b(X_i) - u]. \tag{9}$$

where c_1, c_2 are normalization constants, Y is the target center, $(X_i)_{i=1,2,\ldots,n}$ are pixel co-ordinates of the target model, k represents the Epancechnikov kernel, B is bandwidth, b is bins of the reflectance component $\mathbf{V}_r$ (as against the intensity histogram in meanshift algorithms). Using the target histogram, candidate histogram, and meanshift vector [2], the center for the target in the next frame is given as

$$Y = \frac{\sum_{i=1}^{n} X_i z_i g(\frac{Y_0 - X_i}{B})^2}{\sum_{i=1}^{n} z_i g(\frac{Y_0 - X_i}{B})^2}, \tag{10}$$

Here g is the negative of derivative of kernel k, and z is the weight calculated from the target and candidate histograms, which are used to calculate the new center of the target. The center of the kernel is then shifted from Y_0 to a new center point Y. This is repeated till the candidate model is close to target model. Here we use Bhattacharyya distance for similarity measurement. Target model and candidate histograms are obtained from the reflectance component of each frame, which is less variant to illumination. Hence, the proposed meanshift vector is robust to illumination variation.

A. *Target model update*

A target model is initialized at the beginning of the tracking. However, during tracking there may be variations in the background. If the original target model is used without updating, the tracking accuracy will be reduced because the current background may be different from the background used in the original target model. Therefore, it is necessary to update the target model for robust tracking. Here we propose a simple target model update method. First, target candidate from the current frame is calculated. Then the Bhattacharyya similarity between the target candidate and the target model is computed. If it is smaller than a threshold, implying that there are considerable changes in the background, then it indicates that the candidate is different from the original target model. For accurate tracking, we update the target model as in

$$q_{u,n} = \begin{cases} p_{u,n-1}, & \text{if } \rho < Th_n \\ q_{u,n-1}, & \text{otherwise} \end{cases} \tag{11}$$

where n indicates the current frame index for target model. To make the threshold adaptive, we use the average of Bhattacharyya coefficients of past ten frames as given in

$$Th_n = \frac{\sum_{i=n-10}^{n} \rho(i)}{10} \tag{12}$$

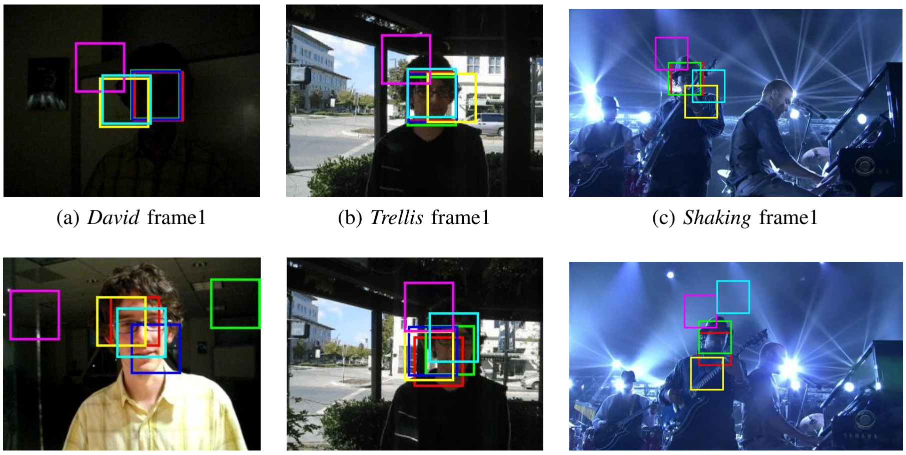
where ρ indicates Bhattacharyya coefficient. If the Bhattacharyya coefficient of current frame is less than average value of Bhattacharyya coefficient of past ten frames, it indicates that the current frame target area is different from all the previous frames. The average value of Bhattacharyya coefficient of past ten frames indicates similarity among the ten frames.

IV. Results and discussion

In this section, we provide results to demonstrate the effectiveness of the proposed tracking method. We have implemented the proposed method and evaluated on several videos. Here results using video sequences from the PETS2009 [23], Caviar [24], AVSS2007 [20], the online benchmark (OTB) [22] and VOT2015 challenge [25] are provided. Evaluation is done using performance measures from OTB based evaluation [22] and VOT2015 challenge [25]. In OTB illumination varying videos with illumination variation across frames, illumination variation within the target area, and sudden illumination change (flash-light effect) are considered for evaluation. Figure 3 shows output frames of three test videos indicating that the proposed method works well under different illumination variation conditions.

A. *Quantitative performance evaluation*

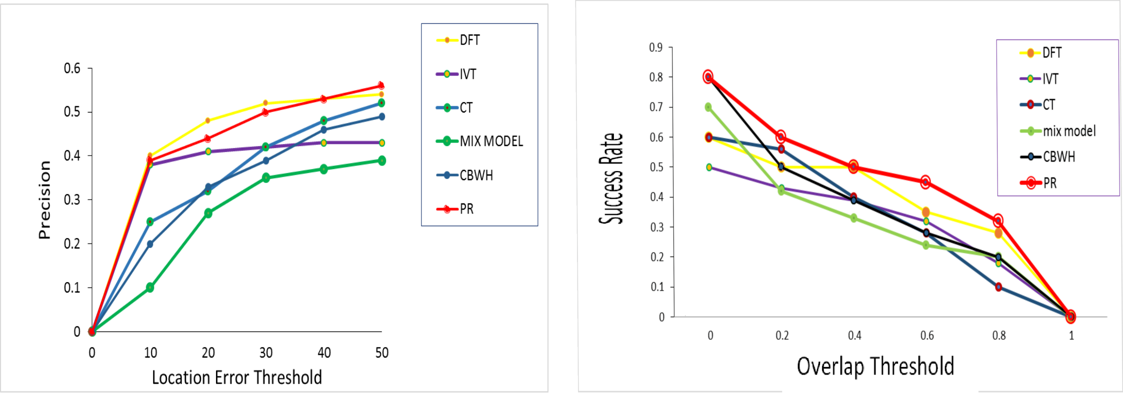
The proposed algorithm is compared with existing methods that are popular in video tracking such as corrected background weighted histogram (CBWH) algorithm [26], joint histogram of color and texture based video tracking (MM) [3] real-time compressive tracker (CT) [11], DCT weight updated histogram using mean shift tracking (IWHMS) [8], the distribution field tracker (DFT) [12], and incremental learning



(a) *David* frame1 (b) *Trellis* frame1 (c) *Shaking* frame1

(d) *David* frame2 (e) *Trellis* frame2 (f) *Shaking* frame 2

Fig. 3: Visual analysis using the proposed method for videos with various illumination changes. Bounding boxes in yellow, green, blue, magenta, skyblue, red correspond to estimates using CT, Mixmodel, DFT, CBWH, IVT, and the proposed method, respectively.



(a) Precision plot (b) Success rate plot

Fig. 4: (a) Precision plot and (b) Success rate plot from one-pass evaluation (OPE) using videos that were tagged as illumination being the challenging aspect in [22]. The proposed method (PR) performs favorably against state-of-the-art algorithms.

based tracker (IVT) [27]. We provide comparison results with these methods as suggested in [22]. The videos in the benchmark dataset [22] are annotated with attributes, which describe the challenges that a tracker will have in each video (illumination changes, occlusions, etc.). Figure 4 shows the overall comparative results in terms of precision and successful tracking of the proposed method with existing methods using videos that were tagged as illumination being the challenging aspect. It shows that the proposed method (PR) performs better.

Another database for video tracker evaluation is the VOT2015 challenge. The proposed algorithm is compared with ten other state-of-art trackers in the VOT2015 challenge framework. The ten trackers included for evaluation are PTZ-MOSSE, OAB, BDF, CMT, CMIL [25], IVT [27], ACMS [28], ACT [29], MIL [30], MEEM [31]. Comparison was done on videos annotated with different attributes like camera motion, size change, illumination variation, occlusion and motion change. Figure 5 shows proposed method's superior performance in illumination change attribute. It also excels significantly in other attributes like occlusion handling and motion change. Figure 6 shows sequence specific performance of the trackers in terms of robustness and accuracy. In majority of videos the proposed tracker shows improved results.

V. Conclusion

In this paper, we proposed a simple but effective framework for tracking under illumination changes. We proposed a reflectance histogram feature, obtained using sparsity constrained NMF (NMFsc) algorithm, which is invariant to illumination. We also used an adaptive target model update method which is invariant to background changes in videos. We considered several video sequences with challenging illumination variations and compared with existing popular methods using OTB and VOT 2015 challenge. The results clearly show that for changes in illumination the proposed method performs well compared to other methods.

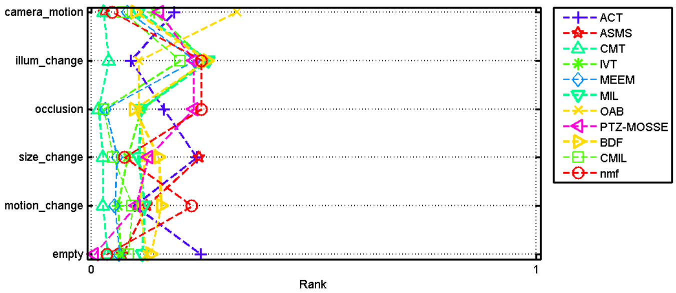


Fig. 5: Comparison of the proposed tracker (nmf) with other existing methods using visual attributes such as size changes, illumination changes, motion, camera motion, and occlusion. The y-axis label *empty* denotes frames that do not correspond to any of the five attributes.

Acknowledgment

The authors would like to acknowledge the support provided by National Center of Excellence in Technology for Internal Security (NCETIS) (an initiative by IIT Bombay and DEitY).

References

[1] J.Yilmaz and O. Shah, "Object tracking: A survey," *ACM Comput. Surv.*, Dec. 2006.

[2] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," in *IEEE Proc. CVPR*, 2000.

[3] J.Ning, L. Zhang, and D. Zhang, "Robust object tracking using joint color-texture histogram," *Pattern recognition and artificial intelligence*, pp. 1245–1263, 2009.

[4] A. Lehuger, L. Patrick, and P. Patrick, "An adaptive mixture color model for robust visual tracking," in *IEEE Proc. ICIP*, 2006, pp. 573–576.

[5] L. Hong, Y. Ze, Z. Hongbin, Z.Yuexian, and L. Zhang, "Robust human tracking based on multi-cue integration and mean-shift," in *Elsevier Science Inc. Pattern Recogn. Lett.*, 2009.

[6] Cannons, J. Kevin, and W. Richard, "Spatiotemporal oriented energy features for visual tracking." in *Springer ACCV*, ser. Lecture Notes in Computer Science, 2007.

[7] D. Suryanto, H. Kim, and K. SungJea, "Spatial color histogram based center voting method for subsequent object tracking and segmentation," in *Image and Vision Computing*, 2011.

[8] G. Phadke and R. velmurugan, "Improved weighted histogram for illumination invariant mean-shift tracking," in *ICVGIP*, 2014.

[9] J. Kwon and K. Lee, "Visual tracking decomposition," in *IEEE Proc. CVPR*, 2010.

[10] Z.Kalal, J.Matas, and K.Mikolajczyk, "P-n learning: Bootstrapping binary classifiers by structural constraints," in *IEEE Proc. CVPR*, 2010.

[11] K.Zhang, L. Zhang, and Y.Ming-Hsuan, "Real-time compressive tracking," in *Springer Proc. ECCV 2012*, 2012.

[12] L. Sevilla-Lara and E. Miller, "Distribution fields for tracking," in *IEEE Proc. CVPR*, 2012.

[13] R. C. Gonzalez and R. E. Woods, *Digital Image processing*, 5th ed. Prentice hall, 2005.

[14] H. Han, S. Shan, X. Chen, and W. Gao, "Illumination transfer using homomorphic wavelet filtering and its application to light-insensitive face recognition," in *8th IEEE International Conference on Automatic Face and Gesture Recognition*, 2008.

[15] L. Shi, B. Funt, and W. Xiong, "Illumination estimation via nonnegative matrix factorization," *Journal of Electronic Imaging*, 2012.

[16] I. Buciu and I. Nafornita, "Non-negative matrix factorization methods for face recognition under extreme lighting variations," in *Intl. Symp. Signals, Circuits and Systems (ISSCS)*, 2009.

[17] Y. Wu, B. Shen, and H. Ling, "Visual tracking via online nonnegative matrix factorization," *IEEE Trans. Circuits and Systems for Video Technology*, 2014.

[18] D. D. Lee and H. S. Seung, "Learning the parts of objects by nonnegative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.

[19] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *The Journal of Machine Learning Research*, 2004.

[20] "AVSS 2007," http://www.eecs.qmul.ac.uk/ãndrea/avss2007_d.html, accessed: 5 Jan 2016.

[21] Y.Rubner, C. Tomasi, and L. Guibas, "The earth mover's distance as a metric for image retrieval," in *International Journal of Computer Vision*, 2000.

[22] Y. Wu, J. Lim, and M. Yang, "Online object tracking: A benchmark," in *IEEE conf. on CVPR*, 2013.

[23] "PETS 2009," http://www.cvg.reading.ac.uk/PETS2009, accessed: 14 May 2016.

[24] "CAVIAR data," http://homepages.inf.ed.ac.uk/rbf/CAVIAR/, accessed: 5 Jan 2016.

[25] http://www.votchallenge.net/index.html.

[26] J. Ning, L.Zhang, D. Zhang, and C.Wu, "Robust mean-shift tracking with corrected background-weighted histogram," in *IET Computer Vision*, Jan. 2012.

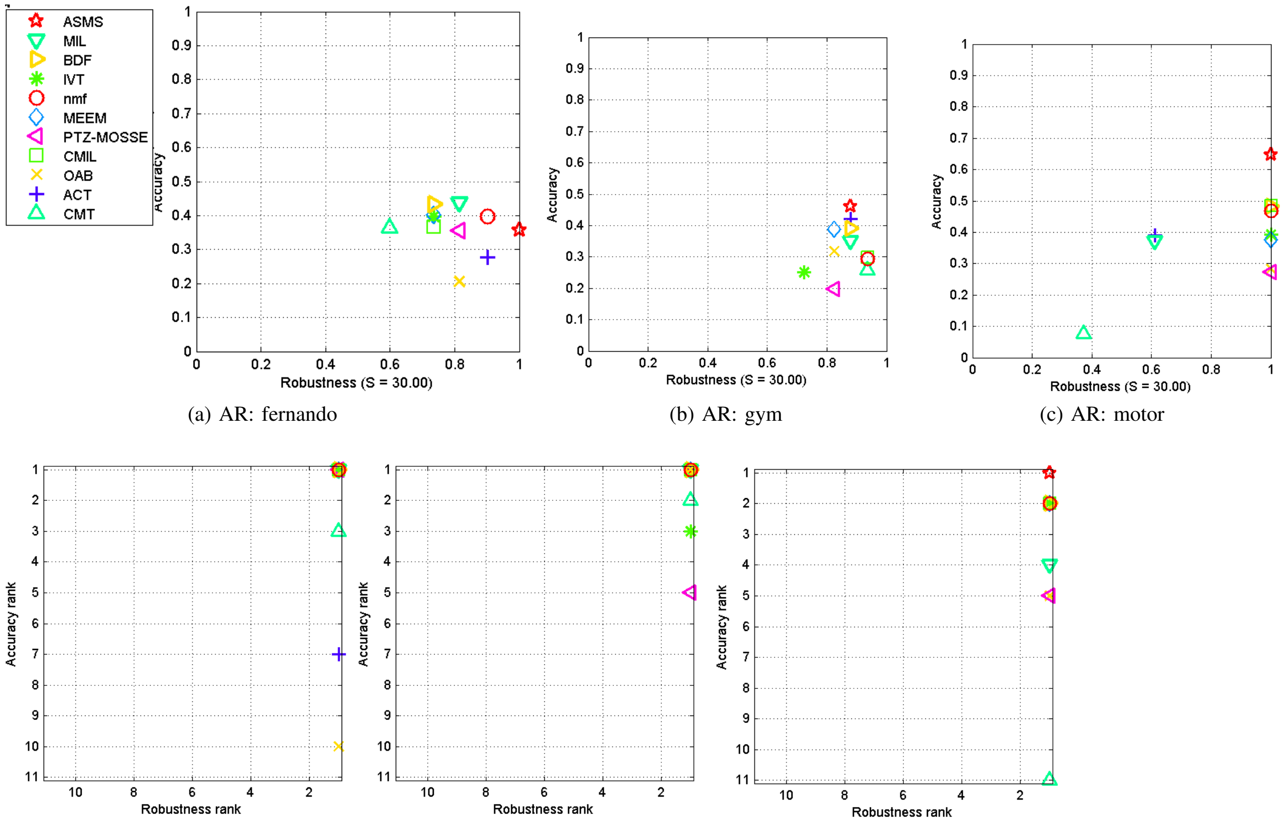
[27] D.Ross, Lim, and Ming-Hsuan, "Incremental learning for robust visual tracking," in *Springer J. Computer Vision*, 2008.

[28] T. Vojir, J. Noskova, and J. Matas, "Robust scale-adaptive mean-shift for tracking," in *Scandinavian Conference on Image Analysis*. Springer, 2013, pp. 652–663.

[29] M. Felsberg, "Enhanced distribution field tracking using channel representations," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2013, pp. 121–128.

[30] B. Babenko, M.-H. Yang, and S. Belongie, "Robust object tracking with online multiple instance learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1619–1632, 2011.

[31] J. Zhang, S. Ma, and S. Sclaroff, "MEEM: robust tracking via multiple experts using entropy minimization," in *Proc. of the European Conference on Computer Vision (ECCV)*, 2014.



(a) AR: fernando

(b) AR: gym

(c) AR: motor

(d) Rank: fernando

(e) Rank: gym

(f) Rank: motor

Fig. 6: Sequence specific AR and Ranking plots for the proposed tracker (nmf) in comparison with state-of-art trackers. First row from (a)-(c) are AR plots for sequences fernando, gymnastics4, motocross2,respectively from the VOT2015 dataset. Second row from (d)-(f) are Ranking plots for sequences fernando, gymnastics4, motocross2 respectively.