An Overview of the Thermal Infrared Visual Object Tracking VOT-TIR2015 Challenge

Amanda Berg*†, Michael Felsberg*, Gustav Häger* and Jörgen Ahlberg*†
*Computer Vision Laboratory, Dept. of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden
Email: {amanda., jorgen.ahl, michael.fels}berg@liu.se, gustav.hager@liu.se

†Termisk Systemteknik AB, Diskettgatan 11B, SE-583 35 Linköping, Sweden
Email: {amanda., jorgen.ahl}berg@termisk.se

***Abstract*—The Thermal Infrared Visual Object Tracking (VOT-TIR2015) Challenge was organized in conjunction with ICCV2015. It was the first benchmark on short-term, single-target tracking in thermal infrared (TIR) sequences. The challenge aimed at comparing short-term single-object visual trackers that do not apply pre-learned models of object appearance. It was based on the VOT2013 Challenge, but introduced the following novelties: (i) the utilization of the LTIR (Linköping TIR) dataset, (ii) adaption of the VOT2013 attributes to thermal data, (iii) a similar evaluation to that of VOT2015. This paper provides an overview of the VOT-TIR2015 Challenge as well as the results of the 24 participating trackers.**

I. Introduction

Visual tracking is a challenging task that has attracted significant attention in the past two decades, e.g. [1], [2], [3]. The number of accepted motion or tracking papers in high profile conferences, such as ICCV, ECCV, and CVPR, has been consistently high ($\sim$40 papers annually), summing up to a few hundred relevant papers in the field. However, the lack of established performance evaluation methodology combined with this large number of publications makes it difficult to assess and understand the advancements made in the field. Several initiatives have attempted to establish a common ground in tracking performance evaluation, starting with PETS [4] and more recently with the Visual Object Tracking (VOT) challenges [5], [6], [7] and the Object Tracking Benchmark [8], [9].

In recent years, thermal cameras have improved in image quality and resolution while decreasing in both price and size. This development has opened up new application areas [10]. Historically, thermal cameras have delivered noisy images with low resolution, used mainly for tracking small objects (point targets) against colder backgrounds and have mainly been of interest for military purposes. Today, they are commonly used in various applications, e.g., cars and surveillance systems. Increasing image quality allows exploration of new application areas, often requiring methods for tracking of extended dynamic objects. Further, for some applications, the methods can not be restricted to stationary platforms. The main advantages of thermal cameras are their ability to see in total darkness, their robustness to illumination changes and shadow effects, and reduced privacy intrusion.

This paper gives an overview of the first thermal infrared (TIR), short-term tracking challenge, the Visual Object Tracking TIR (VOT-TIR2015) challenge, and the results obtained. The results were first presented in [11]. Like the VOT challenge, the VOT-TIR challenge considered single-camera, single-target, model-free, causal trackers, applied to short-term tracking. It was featured as a sub-challenge to VOT2015, organized in conjunction with ICCV2015. The challenge enabled participants not only to evaluate their results on visual data, but also to benchmark their trackers on thermal infrared sequences.

Available datasets for evaluation of tracking in thermal infrared had become outdated [12]. This caused researchers to evaluate their methods on proprietary datasets, which made it difficult to get an overview of advancement made in the field. Inconsistent performance measures across different papers contributed to this difficulty. The Visual Object Tracking challenge, provides an established evaluation methodology for data in the visible spectrum. The main idea of VOT-TIR2015 was to carry these ideas to the area of TIR data, based on a recently collected dataset [12].

A. Related work

A large number of benchmarks exist in the area of visual tracking, but far fewer for TIR tracking. Among visual spectrum (RGB) tracking, the most closely related investigations to the approach presented here is the VOT2015 challenge [7], as well as those of previous years [5], [6]. The online tracking benchmark (OTB) by Wu et al. [8], [9] contains 100 sequences and is a widely used tracking benchmark. In the OTB, trackers are compared using a precision score and a success score, without restarting a failed tracker. The precision score is the percentage of frames where the estimated bounding box is within some fixed distance to the ground truth, while the success score measures the area under the curve of number of frames where the overlap is greater than some fixed percentage. This area has been shown to be equivalent to the average overlap [13], [14]. For further discussion on OTB we refer to [8], [9] and for comparisons with the VOT evaluation to [15], [16].

The series of workshops on Performance Evaluation of Tracking and Surveillance (PETS) [4] have organized thermal infrared challenges on two occasions. The first has taken place in 2005 and the second in 2015, where the challenge was detection, multi-camera/long-term tracking and behavior (threat) analysis. In contrast to VOT-TIR, the challenges concerned multiple research areas while VOT-TIR focused on the problem of short-term tracking only. The lack of further related work within the area of thermal infrared tracking challenges motivates the VOT-TIR initiative.

B. The VOT-TIR2015 challenge

The VOT-TIR2015 challenge targeted a specific set of trackers. All participating trackers were required to be: (i) Causal – sequence frames had to be processed in sequential order; (ii) Short-term – trackers were not required to handle reinitialization; (iii) Model-free – pre-built models of object appearances were not allowed.

Performance of participating trackers was automatically measured using the VOT2014 evaluation kit [6]. The toolkit performs a standardized experiment and stores resulting bounding boxes. If the tracker fails, it is re-initialized. Participants were required to integrate their trackers into the toolkit. Tracking results were analyzed using the VOT2015 evaluation methodology [7].

Participants were expected to submit a single set of results per tracker as well as binaries for result verification. A different set of parameters does not constitute a new tracker. Tracker parameters set by the participant is required to be equal for all test sequences. Detection (by the tracker) of a specific test sequence in order to set hand-tuned parameters was not permitted. However, the tracker itself was allowed to internally change parameters using, e.g., the bounding box size. Further details regarding participation rules are available from the challenge homepage[1].

Differences from the visual spectrum challenge: Compared to the visual equivalent, VOT2015 [7], there are some differences in annotation as well as acquisition and evaluation procedure. The annotated bounding boxes were not allowed to rotate. Further, due to the limited amount of freely available thermal infrared datasets and sequences, sequence selection could not be done as in VOT2015. A new dataset, LTIR (the Linköping Thermal IR dataset) [12], was created for this purpose. Seven different sources were asked to contribute with data and the provided data that contained sufficiently challenging tracking events were included in the dataset. A more detailed description can be found in Section II.

The VOT-TIR2015 challenge applied the same evaluation methodology as VOT2015 [7], except for the practical difference evaluation. This evaluation requires multiple annotations, which were not (yet) available for the LTIR dataset.

C. Outline

The dataset used in the VOT-TIR2015 challenge is described in Section II. Section III briefly summarizes the performance measures and evaluation methodology used in the challenge. Analysis and results are presented in Section IV and, finally, conclusions are drawn in Section V.

[1]http://www.votchallenge.net/vot2015/participation.html

II. The VOT-TIR2015 dataset

The dataset used in VOT-TIR2015 was LTIR, the Linköping Thermal IR dataset [12]. Sequences included in the dataset were collected from seven different sources using eight different types of sensors. The included sequences originate from industry, universities, a research institute and an EU FP7 project. Resolutions range from 320×240 to 1920×480 pixels and the average sequence length is 563 frames. Some sequences in the LTIR dataset are available with both 8- and 16-bit pixel values, however, for this challenge, only 8-bit sequences were used. The main reason for this restriction is that several of the submitted methods cannot deal with 16-bit data. There are sequences from indoor and outdoor environments, and the outdoor sequences were recorded in different weather conditions. Example frames from four sequences are shown in Fig. 1.

All benchmark annotations were in accordance with the VOT2013 annotation process [5] and were done manually. One object within each sequence is annotated in each frame with a bounding box that encloses the object throughout the sequence. The bounding box was allowed to vary in size but not to rotate. In addition to the bounding box annotations, global attributes were per-sequence annotated and local attributes per-frame annotated.

a) Global attributes: The per-sequence global attributes from VOT had to be adapted to the properties of TIR in order to be useful. Below, the global attributes have been arranged according to similarity to VOT-attributes.

- Attributes different from VOT: *Dynamics change* and *temperature change* have been introduced instead of *illumination change* and *object color change*. Not all cameras provide the full 16-bit range, instead, an adaptively changing 8-bit dynamics are sometimes used. *Dynamics change* indicates whether the dynamics is fixed during the sequence or not. *Temperature change* refers to changes in the thermal signature of the object during the sequence
- Attributes similar to VOT: In TIR, *Blur* indicates blur due to motion, high humidity, rain or water on the lens.
- Attributes equal to VOT: *Camera motion*, *object motion*, *background clutter*, *size change*, *aspect ratio change*, *object deformation*, and *scene complexity*.

b) Local attributes: The local, per-frame annotated attributes are: *motion change*, *camera motion*, *dynamics change*, *occlusion*, and *size change*. The attributes are used in the evaluation process to weigh tracking results. They can also be used to evaluate the performance of the method on frames with specific attributes.

III. Performance measures and evaluation methodology

The performance measures as well as evaluation methodology for VOT-TIR2015 are equal to the ones for

(a) running rhino (b) quadrocopter (c) crowd (d) street

Fig. 1: Snapshots from four sequences included in the LTIR dataset. The annotated bounding box is marked in yellow.

VOT2015, except for the practical difference evaluation. Therefore, only a brief summary is given here, further details can be found in [7].

Similar to the VOT2015 challenge, the two weakly correlated performance measures, accuracy and robustness, are used due to their high level of interpretability [13], [14]. The accuracy measurement measures the overlap between the predicted bounding box and the ground truth while the robustness measurement measures how many times the tracker fails. If a tracker is considered to have failed, it is re-initialized five frames later. Overlap calculations, re-initialization, definition of a failure and the rank-based evaluation methodology is further explained in [7].

IV. Analysis and Results

A. The VOT2015 experiments

In our evaluation, and in contrast to VOT2014 [6], we considered the baseline experiment only. We did not consider the region noise experiment for three reasons: First, the results of previous experiments hardly differed [6]. Second, the experiments need significantly more time. Finally, the reproducibility of results would have required to store the seed, which has not been foreseen in the evaluation kit.

B. Submitted trackers

In total, 24 trackers were included in the VOT-TIR2015 challenge. Among them, 20 trackers were submitted and 4 trackers were added by the VOT Committee (3 novel and 1 baseline trackers). The committee have used the accompanying binaries/source code for result verification. For the baseline trackers, the default parameters were selected, or, when not available, were set to reasonable values. All entries are briefly described below and references to the original papers are given where available. More detailed descriptions are given in the Appendix of [11].

Twenty trackers participated in both the VOT2015- and VOT-TIR2015 challenge while 4 trackers were only entered in the VOT-TIR2015 challenge.[2]

One tracker, EBT [17], uses object proposals [18] for object position generation or scoring. Several trackers are based on Mean Shift tracker extensions [19], ASMS [20], PKLTF [21], SumShift [22], and its derivative DTracker. CMIL is based on online boosting and sPST [23] is based on tracking-by-detection learning. A number of trackers can be classified as part-based trackers. These were LDP, G2T, AOGTracker, MCCT, and FoT [24]. A number of trackers come from a class of holistic models that apply regression-based learning for target localization. Out of these, one is based on structured SVM learning, Struck[3]. Several regression-based trackers use correlation filters [26], [27] as visual models. Some correlation filer based trackers maintain a single model for tracking, i.e., NSAMF, OACF, SRDCFir [28], sKCF, STC [29], MKCF+, CCFP, and several trackers apply multiple templates to model appearance variation, i.e., SME, and KCFv2. One tracker, ABCD, applies a global, generative model exploiting channel representations. Finally, the VOT Committee added a baseline tracker, the HotSpot tracker, to the set of submitted trackers. Tracking by detecting hot areas is still state-of-the-art in many TIR applications, e.g. pedestrian detection [30]. The HotSpot tracker detects objects by pixel intensity thresholding and tracks detections using a Kalman filter with a Global Nearest Neighbour approach to the association problem.

C. Results

The results are summarized in sequence pooled and attribute normalized AR rank and AR raw plots in Figure 2. The sequence pooled AR rank plot is obtained by concatenating the results from all sequences and creating a single rank list, while the attribute normalized AR rank plot is created by ranking the trackers over each attribute and averaging the rank lists. Similarly the AR raw plots were constructed. The raw values for the sequence pooled results are also given in Table I.

The following trackers appear either very accurate or very robust among the top performing trackers (closest to the upper right corner of rank plots): SME, MCCT, sPST, SRDCFir, ABCD, and AOG. In contrast to VOT2014, where methods based on correlation filters were largely dominating [6], top performers in VOT-TIR2015 belong to several different classes.

The robustness ranks with respect to the visual attributes are shown in Figure 3. The top three trackers with respect to the different visual attributes are mostly SRDCFir, sPST, and MCCT. A significant exception is camera motion, where SME and EBT come second and third.

[2]Here, we consider SRDCF and SRDCFir being the same, despite the fact that SRDCFir uses a slightly different feature vector.

[3]The implementation used here is a recent improvement of [25].

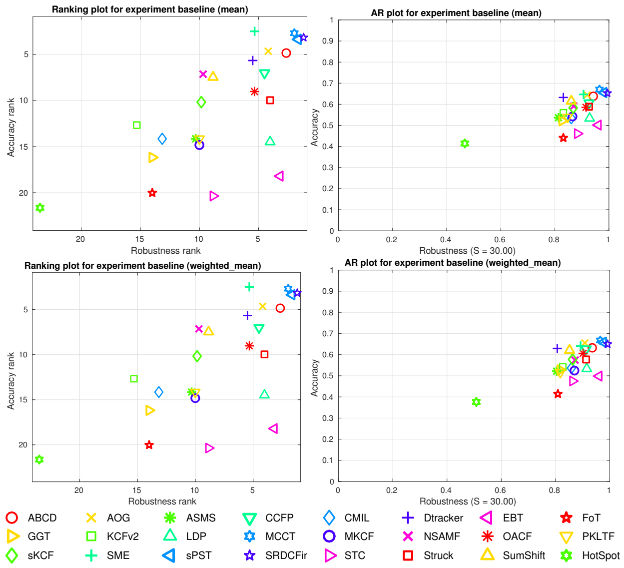

Fig. 2: The AR rank plots and AR raw plots generated by sequence pooling (upper) and by attribute normalization (below).

Tracker	A	R	$\hat{\Phi}$	Speed	Impl.
SRDCFir	0.65	**0.58**	**0.70**	3.17	M C
sPST	*0.66*	*2.18*	*0.64*	0.61	M C
MCCT	**0.67**	3.34	0.55	15.05	M C
EBT	0.50	3.50	0.43	1.08	M C
CCFP	0.63	8.55	0.36	1.03	M C
ABCD	0.63	5.81	0.34	6.88	M
Struck	0.58	8.48	0.30	2.90	C
SME	0.64	9.97	0.30	6.67	M C
LDP	0.53	8.33	0.29	6.96	M C
NSAMF	0.57	12.63	0.28	10.69	M
OACF	0.61	9.57	0.28	3.22	M C
AOG	0.65	8.76	0.27	1.27	*binary*
sKCF	0.58	13.90	0.27	**255.13**	C
CMIL	0.54	14.04	0.25	5.31	C
MKCF+	0.52	12.61	0.24	1.60	M C
KCFv2	0.54	17.81	0.23	14.78	M
STC	0.48	13.85	0.23	29.92	M
SumShift	0.62	15.67	0.19	19.78	C
G2T	0.53	18.59	0.18	0.39	M C
FoT	0.41	19.40	0.17	131.57	C
PKLTF	0.52	19.30	0.16	23.65	C
Dtracker	0.63	19.69	0.16	11.55	C
ASMS	0.52	20.03	0.14	*163.42*	C
HotSpot	0.38	62.27	0.04	5.98	M

TABLE I: The table shows raw accuracy and the average number of failures, expected average overlap, tracking speed (in EFO), and implementation details (M is Matlab, C is C or C++).

The latter turns also out to rank well in the overall criterion *expected average overlap*, see Figure 4. The expected average overlap curve is given by the average bounding-box-overlap averaged over a set of sequences of certain length, plotted over the sequence length N_s [7]. These curves confirm previous statements on the three top performing methods MCCT, sPST, and SRDCFir, where the latter gives the best overall performance. The fact that EBT is ranked fourth underpins the importance of robustness for the expected average overlap.

Apart from tracking accuracy, robustness, and expected

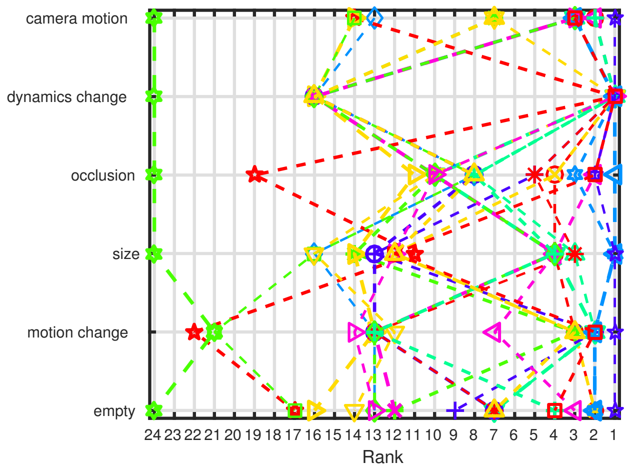

Fig. 3: Robustness plots with respect to the visual attributes. See Figure 2 for legend.

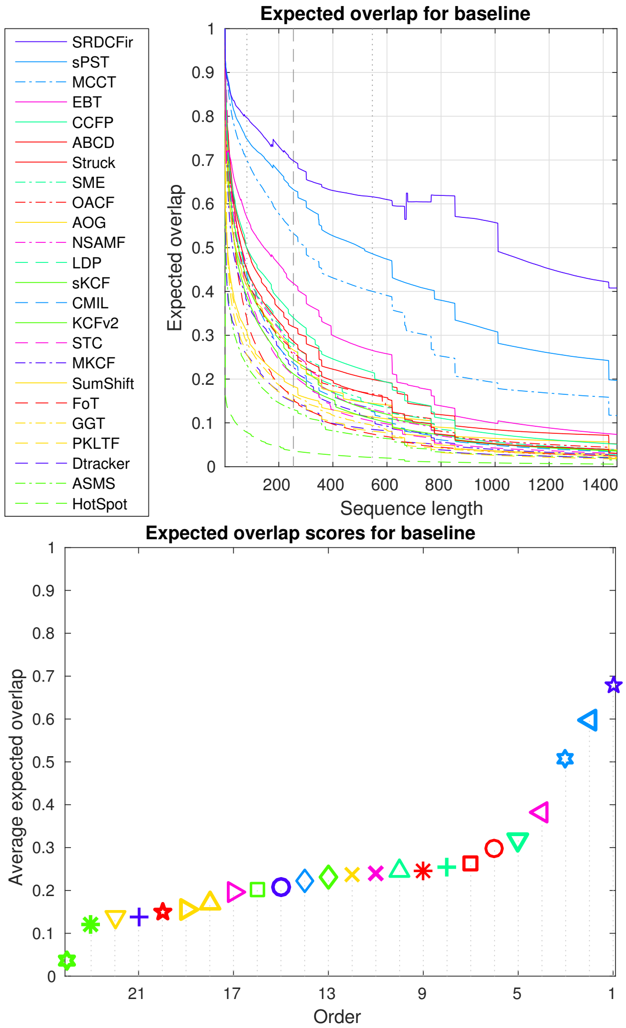

Fig. 4: Expected average overlap curve (above) and expected average overlap graph (below) with trackers ranked from right to left. The right-most tracker is the top-performing according to the VOT2015 expected average overlap values. See Figure 2 for legend. The vertical lines in the upper plot show the range of typical sequence lengths.

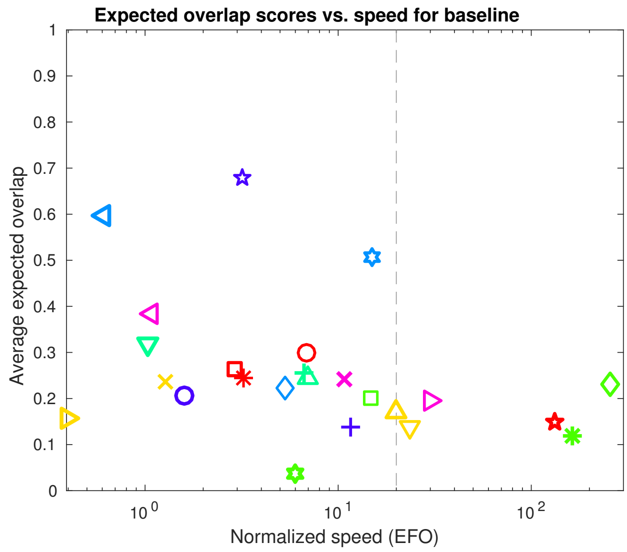

Fig. 5: Expected average overlap scores w.r.t. the tracking speed in EFO units. The dashed vertical line denotes the estimated real-time performance threshold of 20 EFO units. See Figure 2 for legend.

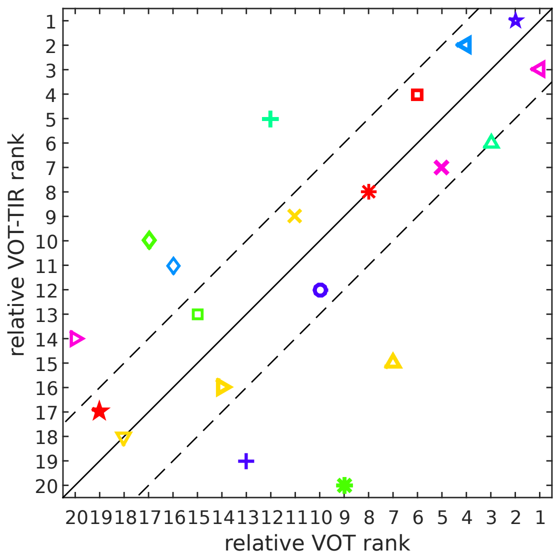

Fig. 6: Comparison of relative ranking of 20 trackers in VOT and VOT-TIR. See Figure 2 for legend.

average overlap at N_s frames, the tracking speed is also crucial in many realistic tracking applications. We therefore visualize the expected overlap score with respect to the tracking speed measured in EFO units in Figure 5. To put EFO units into perspective, a C++ implementation of a NCC tracker provided in the toolkit runs with average 140 frames per second on a laptop with an Intel Core i5-2557M processor, which equals to approximately 160 EFO units.

The vertical dashed line in Figure 5 indicates the real-time speed (equivalent to approximately 20fps). Among the three top-performing trackers, MCCT comes closest to real-time performance. The top-performing tracker in terms of expected overlap among the trackers that exceed the real-time threshold is at the same time the overall fastest tracker, sKCF.

D. *TIR-specific analysis and results*

A particular interesting question in context of VOT-TIR is the effect of the differences between RGB sequences and TIR sequences on the ranking of the trackers. For this purpose, the joint ranking for VOT and VOT-TIR of the 20 common trackers[2] is shown in Figure 6. The only VOT-TIR trackers that have not been run on VOT are MCCT, CCFP, ABCD, and the HotSpot detector.

The dashed lines are the margin of a rank-change by more than three positions. Any change of rank within this margin is considered insignificant and only 7 trackers change their rank by more than three positions. The most dramatic change occurs for ASMS, which ranks 23 in VOT-TIR, but 20 (out of more than 60) in VOT, corresponding to rank 9 within the set of 20 trackers. Other trackers that perform significantly worse are SumShift, and DTracker.

On the other hand, SME, sKCF, STC, and CMIL perform significantly better on VOT-TIR than on VOT according to the relative ranking. Similar as for the overall performance, it is difficult to identify a systematic correlation between improvement and type of tracking methods. Tracking methods that do not use color are likely to perform better on TIR sequences than color-based methods, such as ASMS, SumShift, and DTracker. Also the size of targets differ between VOT (larger) and VOT-TIR (smaller). It is also believed that the tuning of input features is more important to maintain good performance on VOT-TIR, e.g. SRDCFir introduces additional features beyond HOG and works better on TIR sequences than SRDCF with features as used in VOT2015.

V. Conclusion

The VOT-TIR challenge received 20 submissions and compared in total 24 trackers, which we consider a good success and the results presumably give a good guidance to future research within TIR tracking. Best overall performance has been achieved by SRDCFir, closely followed by sPST and MCCT. However, further analysis of the results will be required in order to draw deeper conclusions.

For future challenges, the dataset needs to be extended to become larger and more challenging. Annotation and evaluation need to be adapted to the current VOT standard: multiple annotations and rotating bounding boxes. Also challenges with mixed sequences (RGB and TIR) might be interesting to perform.

Acknowledgment

This work was supported in part by the following research programs and projects: M. Felsberg and G. Häger were supported by the Swedish Foundation for Strategic Research through the project CUAS and the Swedish Research Council trough the project EMC^2. J. Ahlberg and A. Berg were supported by the European Union 7th Framework Programme under grant agreement 312784 (P5) and the Swedish Research Council through the contract D0570301. Some experiments where run on GPUs donated by NVIDIA.

REFERENCES

[1] D. M. Gavrila, "The visual analysis of human movement: A survey," *Comp. Vis. Image Understanding*, vol. 73, no. 1, pp. 82–98, 1999.

[2] X. Li, W. Hu, C. Shen, Z. Zhang, A. R. Dick, and A. Van den Hengel, "A survey of appearance models in visual object tracking," *arXiv:1303.4803 [cs.CV]*, 2013.

[3] T. B. Moeslund, A. Hilton, and V. Kruger, "A survey of advances in vision-based human motion capture and analysis," *Comp. Vis. Image Understanding*, vol. 103, no. 2-3, pp. 90–126, November 2006.

[4] D. P. Young and J. M. Ferryman, "Pets metrics: On-line performance evaluation service," in *ICCCN '05 Proceedings of the 14th International Conference on Computer Communications and Networks*, 2005, pp. 317–324.

[5] M. Kristan, R. Pflugfelder, A. Leonardis, J. Matas, F. Porikli, L. Čehovin, G. Nebehay, G. Fernandez, T. Vojir, A. Gatt, A. Khajenezhad, A. Salahledin, A. Soltani-Farani, A. Zarezade, A. Petrosino, A. Milton, B. Bozorgtabar, B. Li, C. S. Chan, C. Heng, D. Ward, D. Kearney, D. Monekosso, H. C. Karaimer, H. R. Rabiee, J. Zhu, J. Gao, J. Xiao, J. Zhang, J. Xing, K. Huang, K. Lebeda, L. Cao, M. E. Maresca, M. K. Lim, M. E. Helw, M. Felsberg, P. Remagnino, R. Bowden, R. Goecke, R. Stolkin, S. Y. Lim, S. Maher, S. Poullot, S. Wong, S. Satoh, W. Chen, W. Hu, X. Zhang, Y. Li, and Z. Niu, "The Visual Object Tracking VOT2013 challenge results," in *ICCV Workshops*, 2013, pp. 98–111.

[6] M. Kristan, R. P. Pflugfelder, A. Leonardis, J. Matas, L. Cehovin, G. Nebehay, T. Vojir, G. Fernandez, A. Lukezi, A. Dimitriev, A. Petrosino, A. Saffari, B. Li, B. Han, C. Heng, C. Garcia, D. Pangersic, G. Häger, F. S. Khan, F. Oven, H. Possegger, H. Bischof, H. Nam, J. Zhu, J. Li, J. Y. Choi, J.-W. Choi, J. F. Henriques, J. van de Weijer, J. Batista, K. Lebeda, K. Ofjall, K. M. Yi, L. Qin, L. Wen, M. E. Maresca, M. Danelljan, M. Felsberg, M.-M. Cheng, P. Torr, Q. Huang, R. Bowden, S. Hare, S. YueYing Lim, S. Hong, S. Liao, S. Hadfield, S. Z. Li, S. Duffner, S. Golodetz, T. Mauthner, V. Vineet, W. Lin, Y. Li, Y. Qi, Z. Lei, and Z. Niu, "The Visual Object Tracking VOT2014 Challenge Results," in *Computer Vision - ECCV 2014 Workshops*, ser. Lecture Notes in Computer Science, vol. 8926. Springer, 2014, pp. 191–217.

[7] M. Kristan, J. Matas, A. Leonardis, M. Felsberg, L. Čehovin, G. Fernández, T. Vojíř, G. Nebehay, R. Pflugfelder, and G. Häger, "The visual object tracking vot2015 challenge results," in *ICCV workshop on VOT2015 Visual Object Tracking Challenge*, 2015.

[8] Y. Wu, J. Lim, and M. H. Yang, "Online object tracking: A benchmark," in *Comp. Vis. Patt. Recognition*, 2013.

[9] Y. Wu, J. Lim, and M. Yang, "Object tracking benchmark," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1834–1848, 2014.

[10] R. Gade and T. B. Moeslund, "Thermal cameras and applications: A survey." *Machine Vision & Applications*, vol. 25(1), 2014.

[11] M. Felsberg, A. Berg, G. Häger, J. Ahlberg, M. Kristan, A. Leonardis, J. Matas, G. Fernandez, and L. Cehovin, "The thermal infrared visual object tracking vot-tir2015 challenge results," in *ICCV workshop on VOT2015 Visual Object Tracking Challenge*, 2015, p. accepted.

[12] A. Berg, J. Ahlberg, and M. Felsberg, "A thermal object tracking benchmark," in *12th IEEE International Conference on Advanced Video- and Signal-based Surveillance, Karlsruhe, Germany, August 25-28 2015*. IEEE, 2015.

[13] L. Čehovin, M. Kristan, and A. Leonardis, "Is my new tracker really better than yours?" *WACV 2014: IEEE Winter Conference on Applications of Computer Vision*, 2014.

[14] L. Čehovin, A. Leonardis, and M. Kristan. (2013) Visual object tracking performance measures revisited. arXiv:1502.05803 [cs.CV]. arXiv.org. [Online]. Available: http://arxiv.org/abs/1502.05803

[15] M. Kristan, R. Pflugfelder, A. Leonardis, J. Matas, F. Porikli, L. Cehovin, G. Nebehay, G. Fernandez, and T. Vojir, "The vot2013 challenge: overview and additional results," in *Computer Vision Winter Workshop*, 2014.

[16] M. Kristan, J. Matas, A. Leonardis, T. Vojir, R. Pflugfelder, G. Fernandez, G. Nebehay, F. Porikli, and L. Cehovin, "A novel performance evaluation methodology for single-target trackers," *arXiv:1503.01313*, 2015.

[17] G. Zhu, F. Porikli, and H. Li, "Tracking randomly moving objects on edge box proposals," in *CoRR*, 2015.

[18] C. L. Zitnick and P. Dollar, "Edge boxes: Locating object proposals from edges," in *Proc. European Conf. Computer Vision*, 2014, pp. 391–405.

[19] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 5, pp. 564–577, 2003.

[20] T. Vojir, J. Noskova, and J. Matas, "Robust scale-adaptive mean-shift for tracking," *Pattern Recognition Letters*, vol. 49, no. 0, pp. 250 – 258, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167865514001056

[21] A. González, R. Martín-Nieto, J. Bescós, and J. M. Martínez, "Single object long-term tracker for smart control of a PTZ camera," in *International Conference on Distributed Smart Cameras*, 2014, pp. 121–126.

[22] J.-Y. Lee and W. Yu, "Visual tracking by partition-based histogram backprojection and maximum support criteria," in *Proceedings of the IEEE International Conference on Robotics and Biomimetic (ROBIO)*, 2011.

[23] Y. Hua, K. Alahari, and C. Schmid, "Online object tracking with proposal selection," in *Int. Conf. Computer Vision*, 2015.

[24] T. Vojir and J. Matas, "The enhanced flock of trackers," in *Registration and Recognition in Images and Videos*, ser. Studies in Computational Intelligence, R. Cipolla, S. Battiato, and G. M. Farinella, Eds. Springer Berlin Heidelberg: Springer Berlin Heidelberg, January 2014, vol. 532, pp. 113–136.

[25] S. Hare, A. Saffari, and P. H. S. Torr, "Struck: Structured output tracking with kernels," in *Int. Conf. Computer Vision*, D. N. Metaxas, L. Quan, A. Sanfeliu, and L. J. V. Gool, Eds. IEEE, 2011, pp. 263–270.

[26] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2010.

[27] J. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 583–596, 2015.

[28] M. Danelljan, G. Häger, F. S. Khan, and M. Felsberg, "Learning spatially regularized correlation filters for visual tracking," in *Int. Conf. Computer Vision*, 2015.

[29] K. Zhang, L. Zhang, Q. Liu, D. Zhang, and M.-H. Yang, "Fast visual tracking via dense spatio-temporal context learning," in *Proc. European Conf. Computer Vision*, 2014, pp. 127–141.

[30] J.-E. Källhammer, D. Eriksson, G. Granlund, M. Felsberg, A. Moe, B. Johansson, J. Wiklund, and P.-E. Forssén, "Near Zone Pedestrian Detection using a Low-Resolution FIR Sensor," in *Intelligent Vehicles Symposium, 2007 IEEE*, ser. Intelligent Vehicles Symposium. Istanbul, Turkey: IEEE, 2007.