Efficient Diverse Ensemble for Discriminative Co-Tracking: Supplementary Material

Kourosh Meshgi, Shigeyuki Oba, Shin Ishii
Graduate School of Informatics, Kyoto University
606–8501 Yoshida-honmachi, Kyoto, Japan
{meshgi-k,oba,ishii}@sys.i.kyoto-u.ac.jp

1. Comparison with Existing Studies

From one hand, CMT [3] uses multiple-memory horizons for obtaining training data and QBT [4] uses a simple bagging of most recent "uncertain" data to update the classifier, but we construct artificial "diversity" data from the distribution of most recent samples.

On the other hand, MUSTer [2] uses a long-term key-point database to validate whereas TGPR uses long-term memory to regularize the result of the short-memory tracker. Both use fixed heuristics to override the overall result of the short-memory tracker, after the tracking. For clarification, the novelties of the study are: building ensemble on disputed data and maintain it by the online update, diversify ensemble members by generating plausible artificial data, and active switch between short-long memories to label samples, where the short-long memory fusion is performed during labeling, and the data is exchanged between two memories.

2. Elaboration on Proposed Idea

Methods such as ensemble tracking are well-known for label noise and breaking self-learning loop. In addition, co-tracking framework [5] breaks the self-learning loop by exchanging data between two parallel classifiers. We used both in a hierarchical fashion, and also utilized bagging as a part of ensemble model update, that promotes robustness against label noise. Furthermore, the batch update of the aux. classifier prevents the drift with a low amount of label noise and assists the ensemble to fight label noise in-turn.

On the other hand, since we used Gaussian sampling around the last target location, some samples (depending on the target type) are labeled positive, from which only the most similar one forms the tracker output, but others are used for retraining. For the initial frame, we perturbed the initial user bounding box of the target to generate initial training for both the ensemble and the aux. classifier.

The proposed diversification in each frame t, provides the ensemble member $\theta_t^{(c)}$ with a subset of training data. This makes the temporary ensemble $\mathcal{C}_t'$, trained on the obtained samples. Then the artificial data is from the sample's empirical distribution, but its label is selected in a way to challenge the ensemble's belief about those data. Once the model $\theta_t^{'(c)}$ is updated with generated "diversity" samples, the total accuracy of the ensemble on all current samples is measured. If the accuracy improved, the "diversity" samples are accepted, otherwise, new artificial samples are generated and the process repeats. By generating artificial data, the number of positive samples increases (samples are often negative $\rightarrow$ artificial data is often labeled positive), and since they are sampled from the data distribution (modeled by multivariate Gaussian), they are unlikely to be outliers.

3. Combining Long and Short Memories

Researchers have been combining long-term and short-term classifiers to realize robust tracking. TGPR categorized samples into auxiliary samples from early frames and update them slowly and carefully, and target samples from recent samples that are updated quickly and aggressively [1]. MEEM selects an snapshot of the classifier

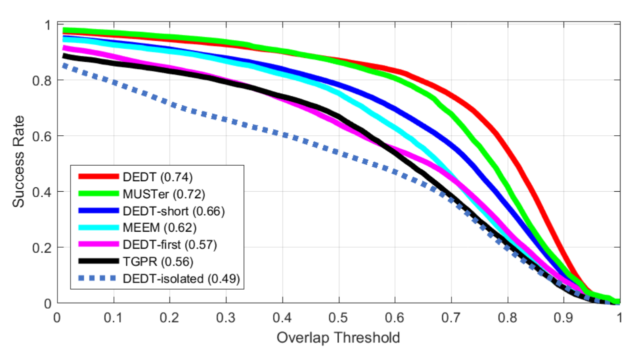


Figure 1. The effect of using long-term memory for auxiliary classifier of DEDT on the overall tracking results on OTB50 [6].

trained by the samples obtained from the beginning of the tracking to role-back inappropriate updates of the classifier [8]. MUSTer archives consistent key-points of the target in he long-term memory, and validates the tracking of the short-term tracker [2]. In our proposed tracker, however, an ensemble of short-memory classifiers invoke the long-term memory when deemed necessary and an active query mechanism governs this process to balance the use of long and short term memories. To see the performance of this scheme, we made DEDT-first that trains the auxiliary classifier on the first frame and do not update this classifier, DEDT-short that updates the auxiliary classifier on each frame, canceling its long-memory properties, and DEDT-isolated that isolate the ensemble from auxiliary classifier, and fuse their results in the end similar to [2]. Figure 1 shows that both of such strategies have inferior performance in our settings, which promotes the role of active query selection and loopy update of the auxiliary classifier.

4. Discussion

The proposed tracker tackled some of the important topics in tracking community: noisy labels, sparse positive samples, and model drift due to self-learning loop.

To alleviate label noise and breaking self-learning loop, methods such as ensemble tracking has been established in the literature. In addition, co-tracking framework [5] breaks the self-learning loop by exchanging data between two parallel classifiers. We used both of them in a hierarchical fashion, and also we utilized bagging as a part of ensemble model update, that promotes robustness against label noise. Furthermore, the batch update of the auxiliary classifier prevents the drift with a small amount of label noise and serves as the helper of the ensemble to fight label noise in-turn.

On the other hand, since we used Gaussian sampling around the last target location, some of the samples (depending on the target type) are labeled positive, from which only the most similar one is considered as the tracker output, but the others are used for positive samples in the retraining. For the initial frame, following a popular routine, we perturbed the initial user-annotated bounding box of the target to generate initial training for both the ensemble and the aux. classifier.

The proposed diversification mechanism, in each frame t, provides the ensemble member $\theta_t^{(c)}$ with a subset of the training data (which we ensured to have enough positive data in the implementation). This makes the temporary ensemble $\mathcal{C}'_t$, trained on the obtained samples. Then the artificial data is generated using the same distribution of the samples, but its label is selected in a way to challenge the belief of the ensemble about such data. Once the model $\theta_t^{'(c)}$ is updated with this generated "diversity" samples, the total accuracy of the ensemble on all current samples is measured. If the accuracy was improved, the "diversity" samples are accepted, otherwise, new artificial samples are generated and the same routine repeats. By generating artificial data, the number of positive samples increases (samples are usually negative $\rightarrow$ artificial data is usually labeled positive), and since they are sampled from the data distribution (modeled by multivariate Gaussian here), these samples are unlikely to be outliers.

Detailed success plots of comparisons against state-of-the-art trackers on OTB50 and OTB100 datasets are provided in Figures 2 and 3 respectively.

References

[1] J. Gao, H. Ling, W. Hu, and J. Xing. Transfer learning based visual tracking with gaussian processes regression. In *ECCV'14*, pages 188–203. Springer, 2014. 2

[2] Z. Hong, Z. Chen, C. Wang, X. Mei, D. Prokhorov, and D. Tao. Multi-store tracker (muster): a cognitive psychology inspired approach to object tracking. In *CVPR'15*. 1, 2

[3] K. Meshgi, S. Oba, and S. Ishii. Active discriminative tracking using collective memory. In *MVA'17*. 1

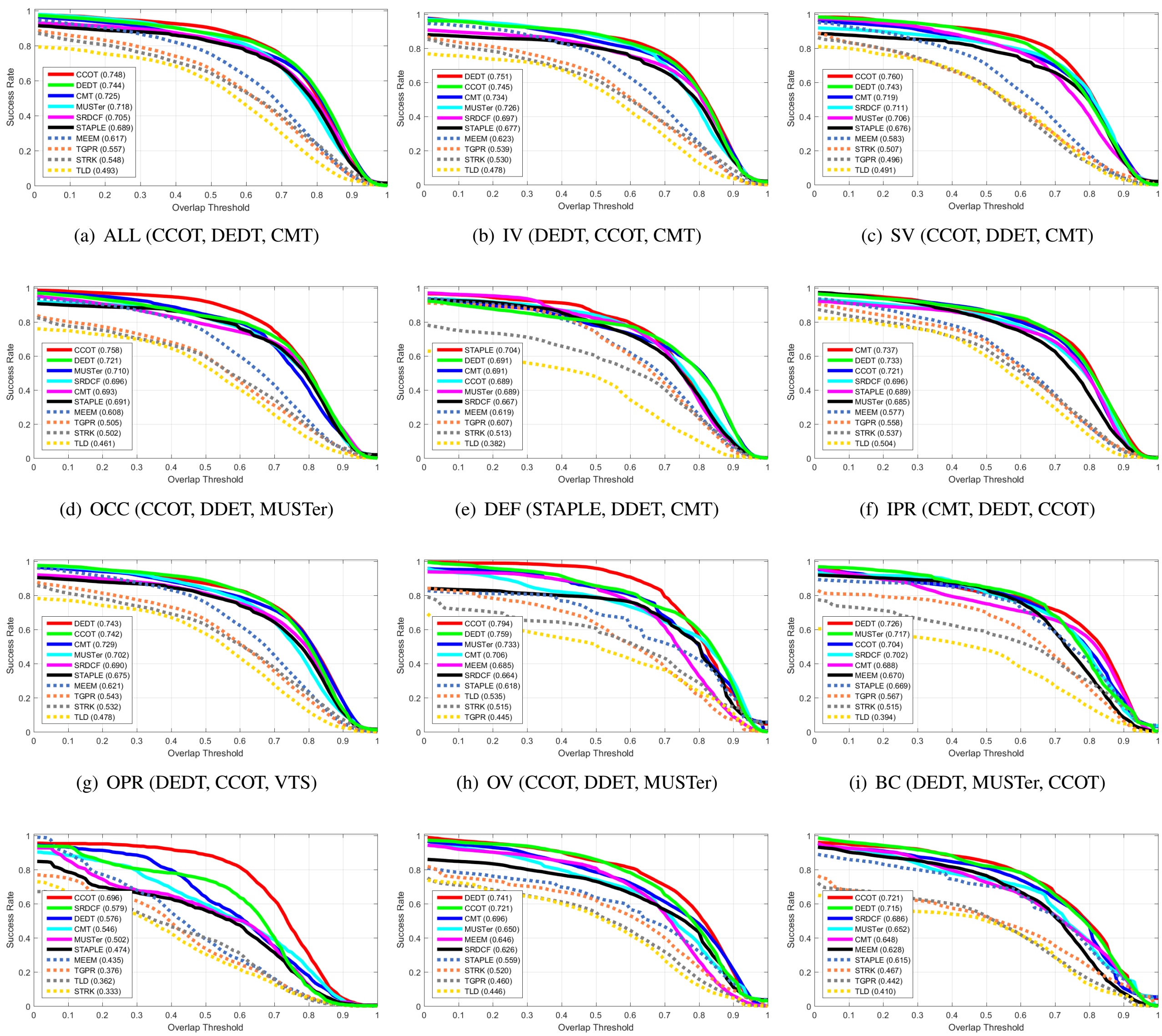
[4] K. Meshgi, S. Oba, and S. Ishii. Robust discriminative tracking via query-by-committee. In *AVSS'16*, 2016. 1

[5] F. Tang, S. Brennan, Q. Zhao, and H. Tao. Co-tracking using semi-supervised support vector machines. In *ICCV'07*. 1, 2

[6] Y. Wu, J. Lim, and M.-H. Yang. Online object tracking: A benchmark. In *CVPR'13*, pages 2411–2418. IEEE, 2013. 1, 3

[7] Y. Wu, J. Lim, and M.-H. Yang. Object tracking benchmark. *PAMI*, 37(9):1834–1848, 2015. 4

[8] J. Zhang, S. Ma, and S. Sclaroff. Meem: Robust tracking via multiple experts using entropy minimization. In *ECCV'14*. 2



(a) ALL (CCOT, DEDT, CMT)

(b) IV (DEDT, CCOT, CMT)

(c) SV (CCOT, DDET, CMT)

(d) OCC (CCOT, DDET, MUSTer)

(e) DEF (STAPLE, DDET, CMT)

(f) IPR (CMT, DEDT, CCOT)

(g) OPR (DEDT, CCOT, VTS)

(h) OV (CCOT, DDET, MUSTer)

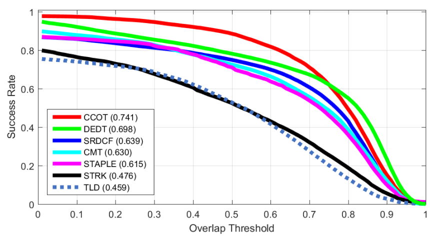
(i) BC (DEDT, MUSTer, CCOT)

(j) LR (CCOT, SRDCF, DEDT)

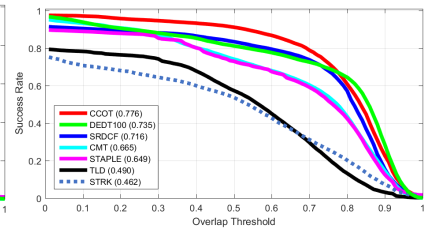
(k) FM (DEDT, CCOT, CMT)

(l) MB (CCOT, DEDT, SRDCF)

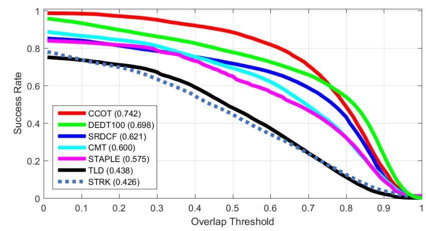
Figure 2. Quantitative evaluation of trackers under different visual tracking challenges (Top three performing trackers are listed in the order of their AUC values). The DEDT is plotted against other state-of-the-art algorithms. DEDT outperformed other trackers (except in overall and DEF (Fig. 2(e)) category) when dealing with different tracking challenges of OTB50 [6] at all of the subcategories. It is shown in 2(a) that DEDT, clearly has a better overall performance compared to other trackers.



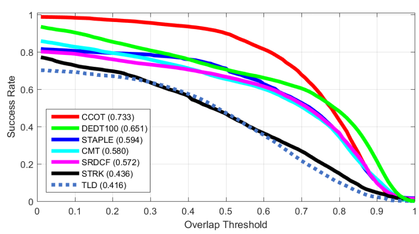
(a) ALL (CCOT, DEDT, SRDCF)



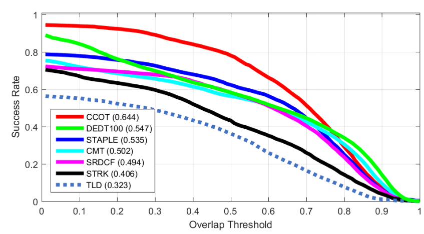
(b) IV (CCOT, DEDT, SRDCF)



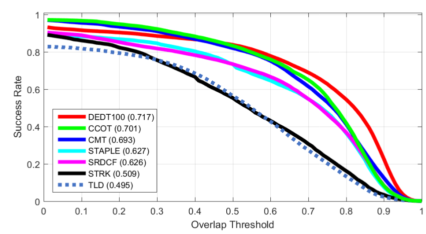
(c) SV (CCOT, DEDT, SRDCF)



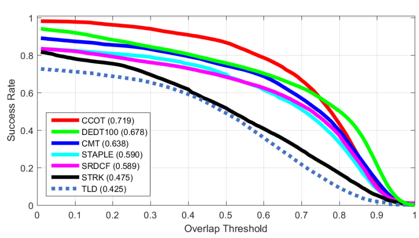
(d) OCC (CCOT, DEDT, STAPLE)



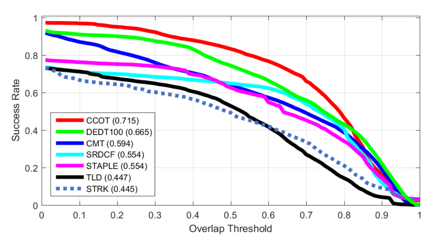
(e) DEF (CCOT, DEDT, STAPLE)



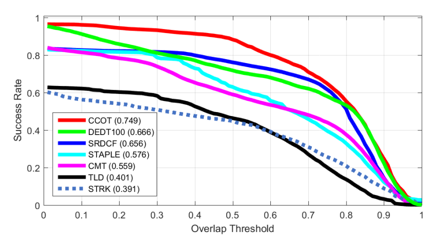
(f) IPR (CCOT, DEDT, CMT)



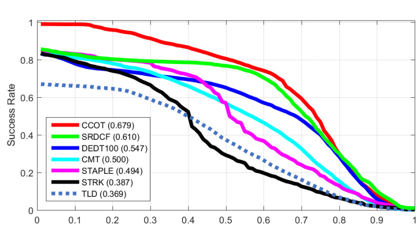
(g) OPR (CCOT, DEDT, CMT)



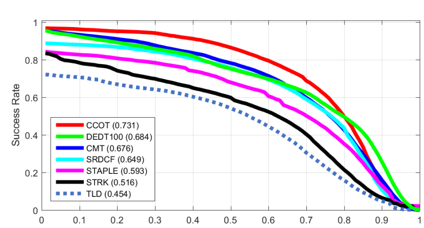
(h) OV (CCOT, DEDT, CMT)



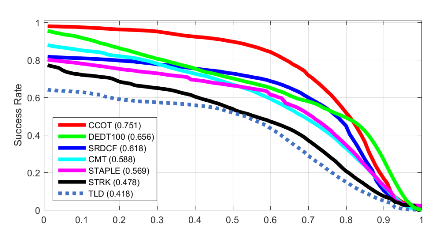
(i) BC (CCOT, DEDT, SRDCF)



(j) LR (CCOT, SRDCF, DEDT)



(k) FM (CCOT, DEDT, CMT)



(l) MB (CCOT, DEDT, SRDCF)

Figure 3. Quantitative evaluation of trackers under different visual tracking challenges (Top three performing trackers are listed in the order of their AUC values). The DEDT is plotted against other state-of-the-art algorithms. Except CCOT, DEDT outperformed other trackers (except in the LR 3(j) category) when dealing with different tracking challenges of OTB100 [7] at all of the subcategories. It is shown in 2(a) that CCOT and DEDT, clearly has an edge comparing to other trackers, while CCOT employs deep features and DEDT uses HOG.