Robust Tracking with Weighted Online Structured Learning: Appendix

Rui Yao[1,2], Qinfeng Shi[2], Chunhua Shen[2], Yanning Zhang[1], and Anton van den Hengel[2]

[1] School of Computer Science, Northwestern Polytechnical University, China
[2] School of Computer Science, The University of Adelaide, Australia

This appendix contains two parts. In Section 1, we present the full proof of Proposition 1, Theorem 1 and Corollary 1, which are appeared in main body of paper. We show more experimental results in Section 2. Furthermore, we report all visual results in demonstration videos[3].

1 Proofs

Lemma 1 (Reservoir Mean Preservation). *For any function $g : \mathcal{Z} \to \mathbb{R}$ at any iteration t, and if $R_t \subset \mathcal{Z}$ with weights $(\omega_1, \omega_2, \cdots, \omega_t)$, where $\sum_{i=1}^{t} \omega_i = 1$ and $\omega_i \geq 0$ is a weighted reservoir in [1], the following holds*

$$\mathbb{E}_{R_t}\Big[\frac{1}{|R_t|}\sum_{\mathbf{z}\in R_t} g(\mathbf{z})\Big] = \sum_{i=1}^{t} \omega_i g(\mathbf{z}_i).$$

Proof. Since the randomness of the R_t is the randomness of the its elements, we have

$$\mathbb{E}_{R_t}\Big[\frac{1}{|R_t|}\sum_{\mathbf{z}\in R_t} g(\mathbf{z})\Big] = \frac{1}{|R_t|}\sum_{\mathbf{z}\in R_t} \mathbb{E}_{\mathbf{z}}[g(\mathbf{z})] = \mathbb{E}_{\mathbf{z}}[g(\mathbf{z})].$$

We know from [1] that $\forall i \leq t, \Pr(\mathbf{z}_i \in R_t) = \omega_i$, thus

$$\mathbb{E}_{\mathbf{z}}[g(\mathbf{z})] = \sum_{i=1}^{t} \omega_i g(\mathbf{z}_i).$$

Proposition 1 (Expected Reservoir Risk). *Minimising*

$$C\sum_{i=1}^{m} \lambda_i \ell(\mathbf{w}, \mathbf{z}_i) + \Omega(\mathbf{w}), \tag{1}$$

is equivalent to minimising

$$\frac{CZ_m}{|R_m|}\,\mathbb{E}_{R_m}\big[\sum_{\mathbf{z}\in R_m} \ell(\mathbf{w}, \mathbf{z})\big] + \Omega(\mathbf{w}), \tag{2}$$

[3] For clarity, we only show the result of five trackers in demonstration video, where these methods have better performance than others. The videos can be found on youtube: `http://www.youtube.com/woltracker12`.

where R_m *is a weighted reservoir with weights* $(\lambda_1, \lambda_2, \cdots, \lambda_m)$ *as in [1] and* $Z_m = \sum_{i=1}^{m} \lambda_i$.

Proof. Let $P_\lambda(\mathbf{Z} = \mathbf{z}_i) = \frac{\lambda_i}{\sum_{j=1}^{m} \lambda_j}$ and apply Lemma 1 , we have

$$\begin{aligned} &\frac{1}{Z_m} \sum_{i=1}^{m} \lambda_i \ell(\mathbf{w}, \mathbf{z}_i) \\ &= \sum_{i=1}^{m} P_\lambda(\mathbf{Z} = \mathbf{z}_i) \ell(\mathbf{w}, \mathbf{z}_i) \\ &= \frac{1}{|R_m|} \mathbb{E}_{R_m} [\sum_{\mathbf{z} \in R_m} \ell(\mathbf{w}, \mathbf{z})]. \end{aligned}$$

Multiply CZ_m on both side and add the $\Omega(\mathbf{w})$ yields the proposition.

Theorem 1 (Weighted Regret Bound). *At each iteration* i*, performing* OLaRank *on each* $\mathbf{b}_{ij} \in R_i, 1 \leq j \leq N$. *Assume that for each update, the dual increase after seeing the example* $\mathbf{b}_{ij}$ *is at least* $C\mu(\ell(\mathbf{w}_{ij}, \mathbf{b}_{ij}))$*, with*

$$\mu(x) = \frac{1}{C} \min(x, C)(x - \frac{1}{2} \min(x, C)),$$

then, we have for any $\mathbf{w}$*,*

$$\frac{1}{m} \sum_{i=1}^{m} \frac{1}{N} \mathbb{E}_{R_i} \Big[\sum_{j=1}^{N} \ell(\mathbf{w}_{ij}, \mathbf{b}_{ij}) \Big] \leq \frac{1}{m} \sum_{i=1}^{m} \Big(\frac{1}{Z_i} \sum_{j=1}^{i} \lambda_j \ell(\mathbf{w}, \mathbf{z}_j) \Big) + \frac{\Omega(\mathbf{w})}{CmN} + \frac{C}{2}, \tag{3}$$

where $Z_i = \sum_{j=1}^{i} \lambda_j$.

Proof. Via Theorem 1 in [2] and Theorem 3 in [3], we have

$$\frac{1}{m} \sum_{i=1}^{m} \frac{1}{N} \Big[\sum_{j=1}^{N} \ell(\mathbf{w}_{ij}, \mathbf{b}_{ij}) \Big] \leq \frac{1}{m} \sum_{i=1}^{m} \frac{1}{N} \Big[\sum_{j=1}^{N} \ell(\mathbf{w}, \mathbf{b}_{ij}) \Big] + \frac{\Omega(\mathbf{w})}{CmN} + \frac{C}{2}. \tag{4}$$

Take expectation on both side, we have

$$\frac{1}{m} \sum_{i=1}^{m} \frac{1}{N} \mathbb{E}_{\mathbf{b}_{ij}} \Big[\sum_{j=1}^{N} \ell(\mathbf{w}_{ij}, \mathbf{b}_{ij}) \Big] \leq \frac{1}{m} \sum_{i=1}^{m} \frac{1}{N} \mathbb{E}_{\mathbf{b}_{ij}} \Big[\sum_{j=1}^{N} \ell(\mathbf{w}, \mathbf{b}_{ij}) \Big] + \frac{\Omega(\mathbf{w})}{CmN} + \frac{C}{2}. \tag{5}$$

By Lemma 1, we have

$$\frac{1}{m} \sum_{i=1}^{m} \frac{1}{N} \mathbb{E}_{\mathbf{b}_{ij}} \Big[\sum_{j=1}^{N} \ell(\mathbf{w}, \mathbf{b}_{ij}) \Big] = \frac{1}{m} \sum_{i=1}^{m} \Big(\frac{1}{Z_i} \sum_{j=1}^{i} \lambda_j \ell(\mathbf{w}, \mathbf{z}_j) \Big), \tag{6}$$

where $Z_i = \sum_{j=1}^{i} \lambda_j$. Plugging (6) into (5) yields the theorem.

Corollary 1 (Tracking Bound). *Assume that for each update in Algorithm 1, the dual increase after seeing the example* $\mathbf{b}_{ij} \in R_i$ *is at least* $C\mu(\ell(\mathbf{w}_{ij}, \mathbf{b}_{ij}))$*, then we have for any* $\mathbf{w}$*,*

$$\frac{1}{tN}\sum_{i=1}^{t}\mathbb{E}_{R_i}\Big[\sum_{j=1}^{N}\ell(\mathbf{w}_{ij},\mathbf{b}_{ij})\Big] \le \frac{1}{t}\sum_{i=1}^{t}\Big(\frac{1}{Z_i}\sum_{j=1}^{i}\lambda_j\sum_{l=1}^{n_j}\ell(\mathbf{w},\mathbf{x}_j,\mathbf{y}_j,\mathbf{y}_{jl})\Big) + \sqrt{\frac{2\Omega(\mathbf{w})}{tN}}, \tag{7}$$

where $Z_i = \sum_{j=1}^{i}\lambda_j n_j$*.*

Proof. Apply Theorem 1 to the data stream $((\mathbf{x}_i, \mathbf{y}_i, \mathbf{y}_{ij}))_{1\le i\le t, 1\le j\le n_i}$ with weight λ_i for all $((\mathbf{x}_i, \mathbf{y}_i, \mathbf{y}_{ij}))_{1\le j\le n_i}$, we have

$$\frac{1}{t}\sum_{i=1}^{t}\frac{1}{N}\mathbb{E}_{R_i}\Big[\sum_{j=1}^{N}\ell(\mathbf{w}_{ij},\mathbf{b}_{ij})\Big] \le \frac{1}{t}\sum_{i=1}^{t}\Big(\frac{1}{Z_i}\sum_{j=1}^{i}\lambda_j\sum_{l=1}^{n_j}\ell(\mathbf{w},(\mathbf{x}_j,\mathbf{y}_j,\mathbf{y}_{jl})\Big) + \frac{\Omega(\mathbf{w})}{CtN} + \frac{C}{2}. \tag{8}$$

Letting $C = \sqrt{2\Omega(\mathbf{w})/tN}$ yields the result.

2 Quantitative and Qualitative Results

In this section, we show more results of our tracking algorithm. Fig. 1 shows the frame-by-frame centre location errors (pixel) of tracking result obtained by the nine trackers on the twelve video sequences. As we can see from Fig. 1, our tracker indicates a robust performance, the CLE of our result is always lower than others for the most part. Fig. 2 – Fig. 13 show the visual result over some representative frames. For clarity, we only show the result of five trackers. These sequences contain various scenes and a variety of object motion events. The representative frame shown in Fig. 2 – Fig. 13 contains challenging lighting(sequence *shaking,singer1*, *etc.*), large variation in pose and scale(sequence *basketball,bird*, *etc.*), frequent half or full occlusions(sequence *coke,tiger1,tiger2*, *etc.*), fast motion(sequence *walk,tiger1*, *etc.*), shape deformation and distortion(sequence *basketball,iceball*, *etc.*), and similar interference(sequence *basketall,iceball*, *etc.*). Our tracker can keep track of object in these situations, and achieves robust results.

References

1. Efraimidis, P.S., Spirakis, P.G.: Weighted random sampling with a reservoir. Inf. Process. Lett. **97** (2006) 181–185
2. Bordes, A., Usunier, N., Bottou, L.: Sequence labelling svms trained in one pass. In: Proc. ECML/PKDD. (2008) 146–161
3. Shalev-Shwartz, S., Singer, Y.: A primal-dual perspective of online learning algorithms. Machine Learning **69** (2007) 115–142

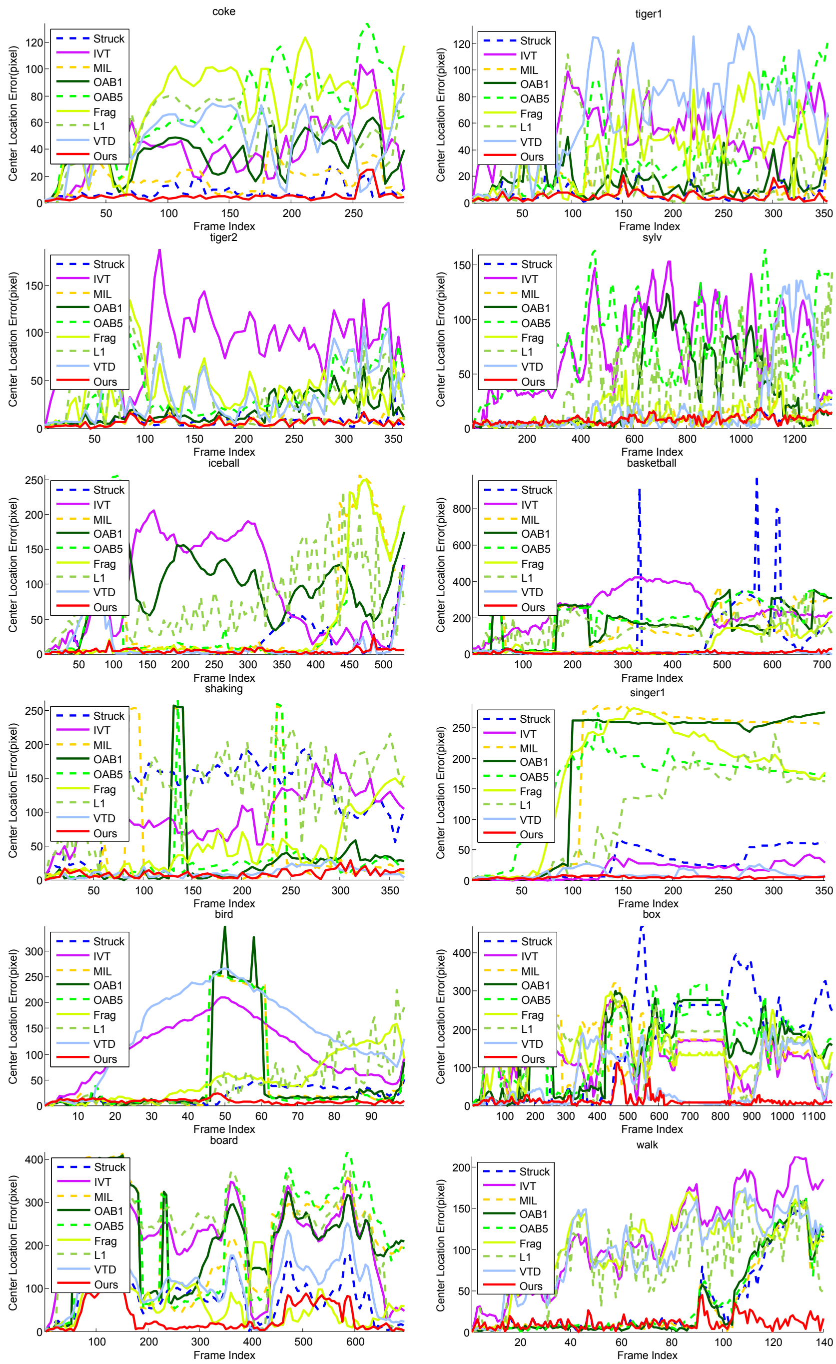

Fig. 1. Quantitative evaluation of different trackers in centre location error plots on twelve sequences.

Fig. 2. Sequence *basketball*: The tracking results of different tracking approaches over representative frames.

Fig. 3. Sequence *bird*: The tracking results of different tracking approaches over representative frames.

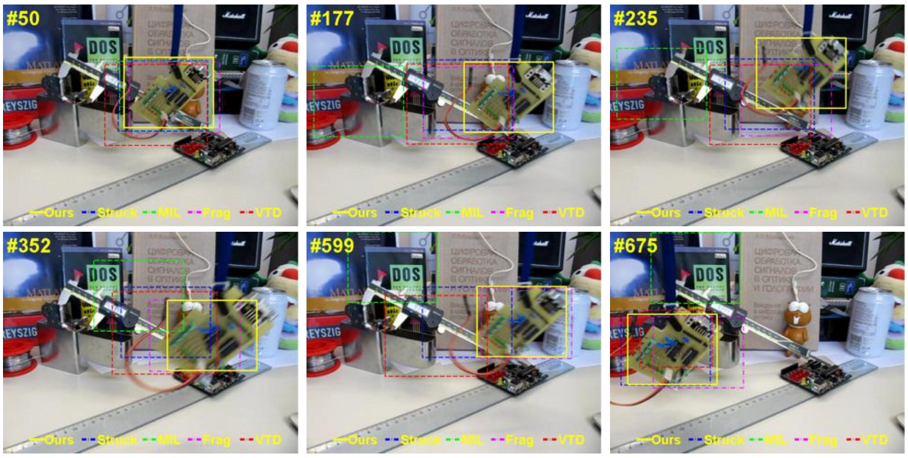

Fig. 4. Sequence *board*: The tracking results of different tracking approaches over representative frames.

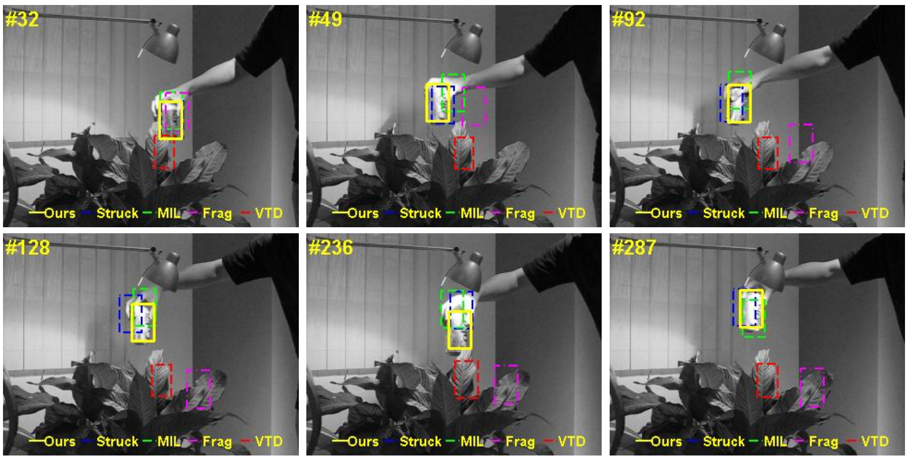

Fig. 5. Sequence *coke*: The tracking results of different tracking approaches over representative frames.

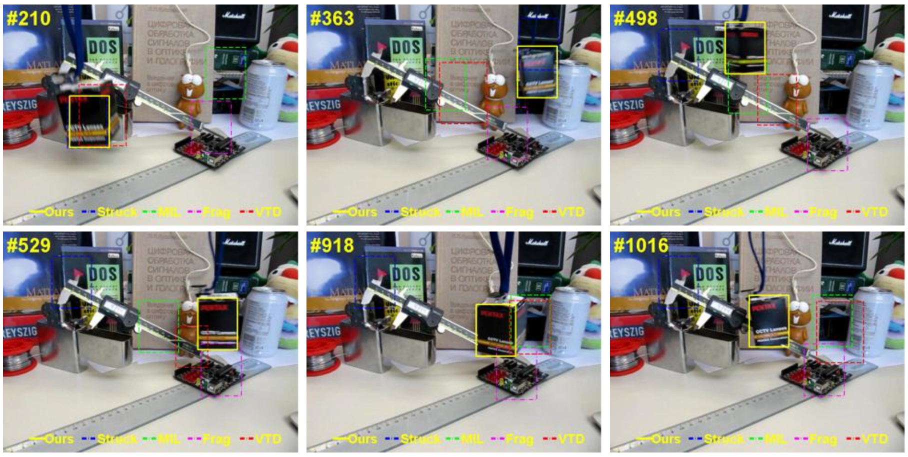

Fig. 6. Sequence *box*: The tracking results of different tracking approaches over representative frames.

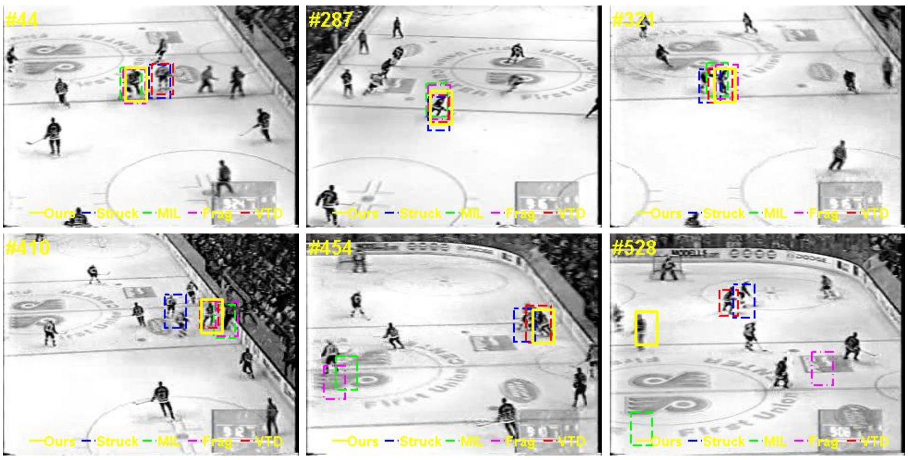

Fig. 7. Sequence *iceball*: The tracking results of different tracking approaches over representative frames.

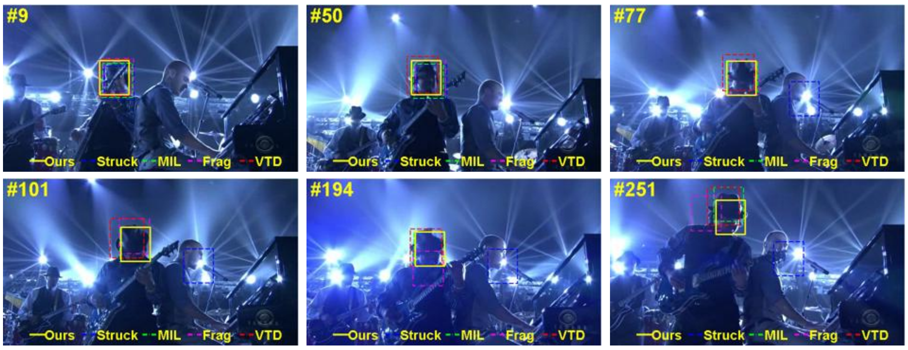

Fig. 8. Sequence *shaking*: The tracking results of different tracking approaches over representative frames.

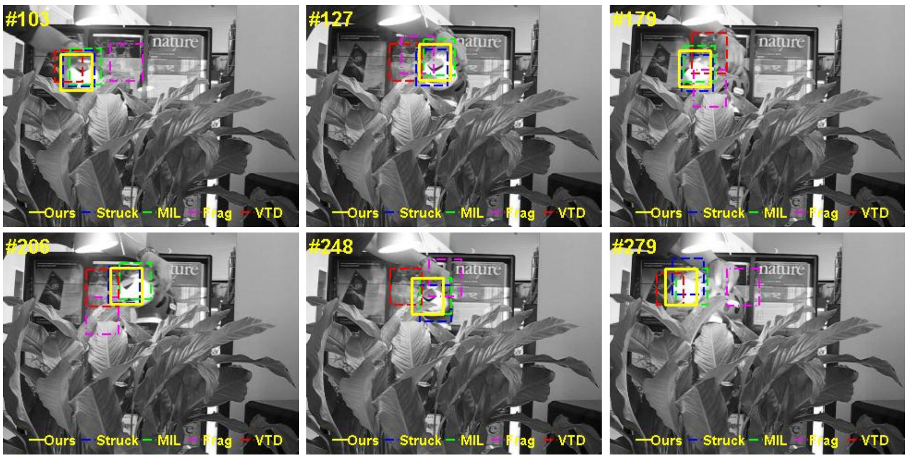

Fig. 9. Sequence *tiger2*: The tracking results of different tracking approaches over representative frames.

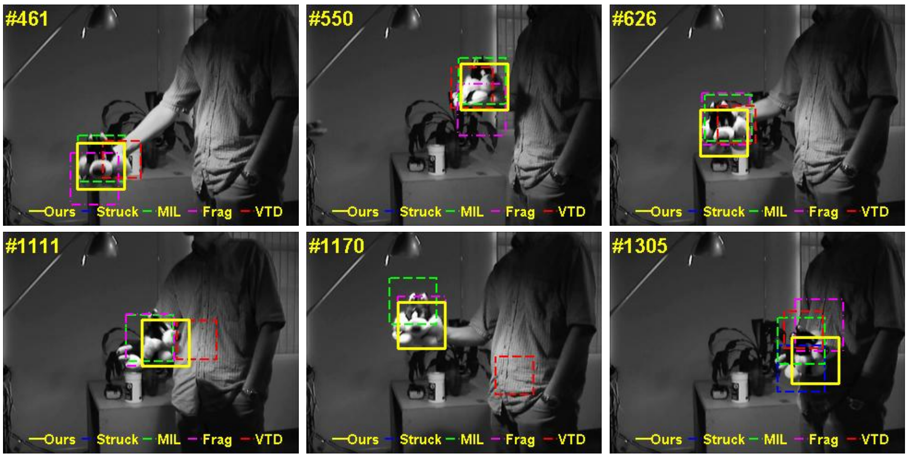

Fig. 10. Sequence *sylv*: The tracking results of different tracking approaches over representative frames.

Fig. 11. Sequence *singer1*: The tracking results of different tracking approaches over representative frames.

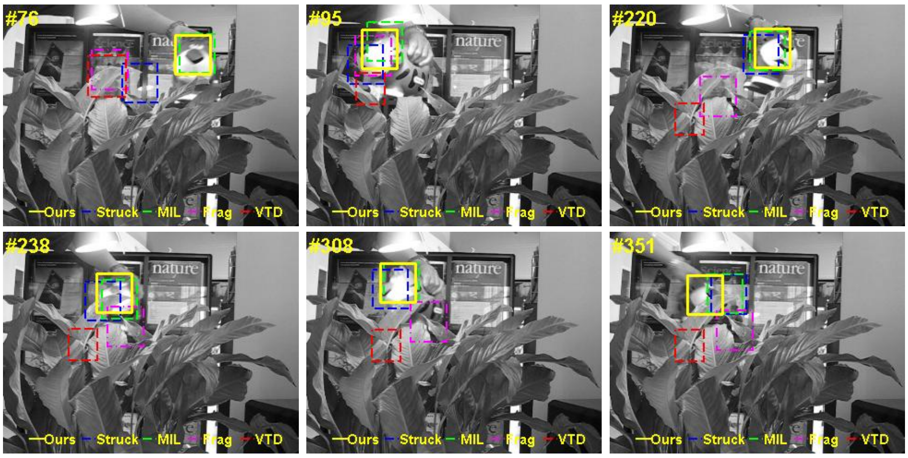

Fig. 12. Sequence *tiger1*: The tracking results of different tracking approaches over representative frames.

Fig. 13. Sequence *walk*: The tracking results of different tracking approaches over representative frames.