Fusion with Diffusion for Robust Visual Tracking (Supplementary Material)

Yu Zhou[1], **Xiang Bai**[1], **Wenyu Liu**[1], **Longin Jan Latecki**[2]
[1] Dept. of Electronics and Information Engineering, Huazhong Univ. of Science and Technology, P. R. China
[2] Dept. of Computer and Information Sciences, Temple Univ., Philadelphia, USA
`{zhouyu.hust,xiang.bai}@gmail.com,liuwy@hust.edu.cn,latecki@temple.edu`

1 Frame Results of Our Tracking Algorithm

In Fig.1, frame results for *Cliff Bar* are shown, this video include 328 frames, the main challenge in this video include cluttered background, scale change and motion blur.

At the beginning of this video, i.e.,frame No.029, many trackers draft greatly, like MS and Farg(Chi). In frame No.085 and No.211, motion blur appeared, most of the other trackers lose the target completely and our tracker could track the target correctly. For the background is very similar to the foreground target, at the end of this video, most of the other trackers lose the target, and our tracker could track the target all the times.

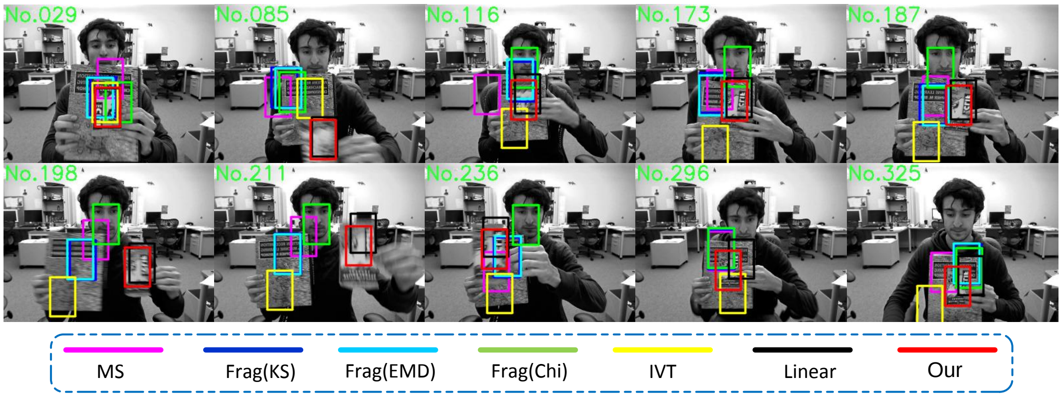

Figure 1: Tracking results on *Cliff Bar*.

In Fig.2, frame results for *Coke Can* are shown, this video include 292 frames in total. The main challenge in this video include appearance change greatly, lighting variance, partly or fully occlusion.

In frame No.015, many trackers draft greatly, and our tracker also draft a bit. In frame No.046, all the other tracker lose the target completely expect our tracker, but our tracker is still draft. In frame No.055, No.081, No.111, although the lighting and the appearance changed greatly, our tracker could correct track the target. At the end of this video, we still best track the target compare with other methods. Although our tracker draft in this video, we should claim that we never lose the target. Linear could achieve comparable results in this video.

In Fig.3, frame results for *Coupon Book* are shown, this video include 327 frames in total. The main challenge in this video is the cluttered background, a background target exists in this video which is very similar with the true foreground target.

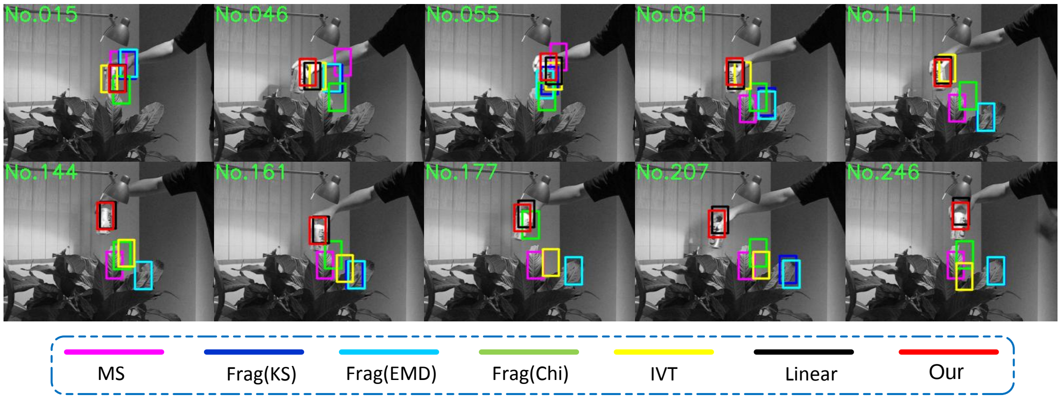

Figure 2: Tracking results on *Coke Can*.

In frame No.061, the appearance changed compare with frame No.047, then MS and IVT draft greatly. In frame No.135, No.175, and No.191, when the true target is moving, MS, IVT, Frag(KS) and Frag(EMD) lose the target, and our tracker could always track the true target correctly. At the end of this video, i.e., in frame No.310, Frag(EMD), Frag(Chi),Frag(KS) are all wrongly track the background target which is very similar with the true target. MS, Linear and our tracker could always track the target.

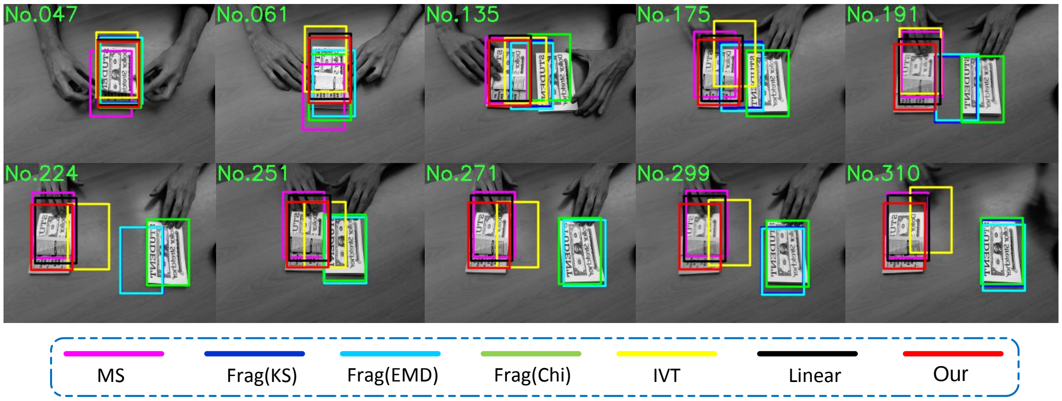

Figure 3: Tracking results on *Coupon Book*.

In Fig.4, frame results for *PETS01D1* are shown, this video include 412 frames in total. The main challenge is that the appearance and scale change greatly.

In frame No.014, Linear and IVT drafted, and after frame No.145, IVT and Linear lose the target completely, Frag achieves comparable results in this video, but we should claim that fragment-based image representation is used in this method, which is more discriminative than the template based image representation used in our method, and we could see that only combine different template representation using our method, we could achieve more stable results compare with Frag.

In Fig.5, frame result for *Surfer* are shown, this video include 376 frames in total, and the main challenges include cluttered background and greatly appearance variance.

In this video, we want to track the head of the surfer. we could see that the background is always changed and the pose of the surfer also variance greatly. In frame No.027, most of the other methods draft greatly, and we also could see that in frame No.167, No.189, our tracker could track the target accurate, and at the end of the video, i.e., frame No.357, only our tracker still cover the target accurate.

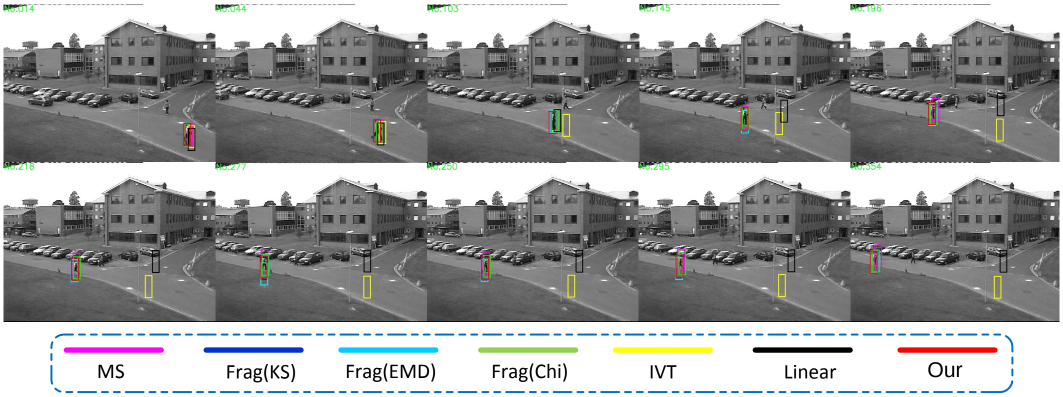

Figure 4: Tracking results on *PETS01D1*.

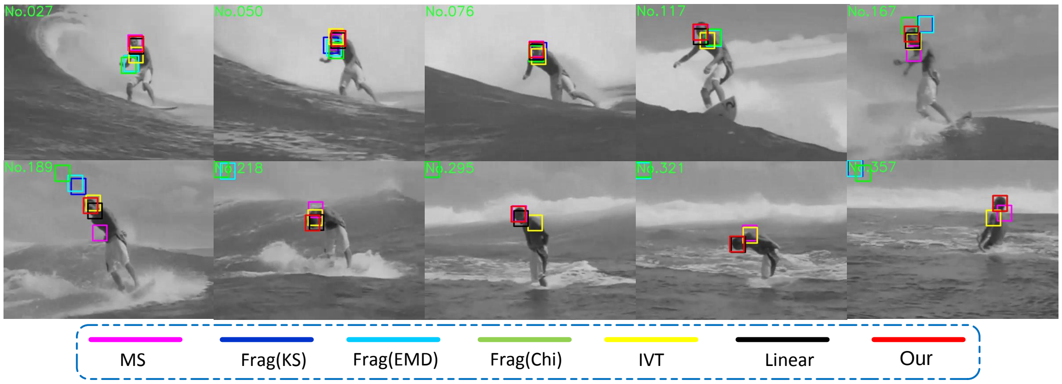

Figure 5: Tracking results on *Surfer*.

In Fig.6, frame results for *Sylvester* are shown, this video have 1344 frames in total, which make it very difficult for all the tracking algorithms, the main challenge in this video include appearance change greatly, lighting variance, scale variance.

At the begin, i.e., frame No.079, IVT begin to draft, and in frame No.382, Frag draft. In frame No.604, No.850, the appearance of the target changes greatly, our tracker could also track the target accurate. Even at the end of this video, i.e., frame No.1204 and No.1299, our tracker could also track the target the target correctly.

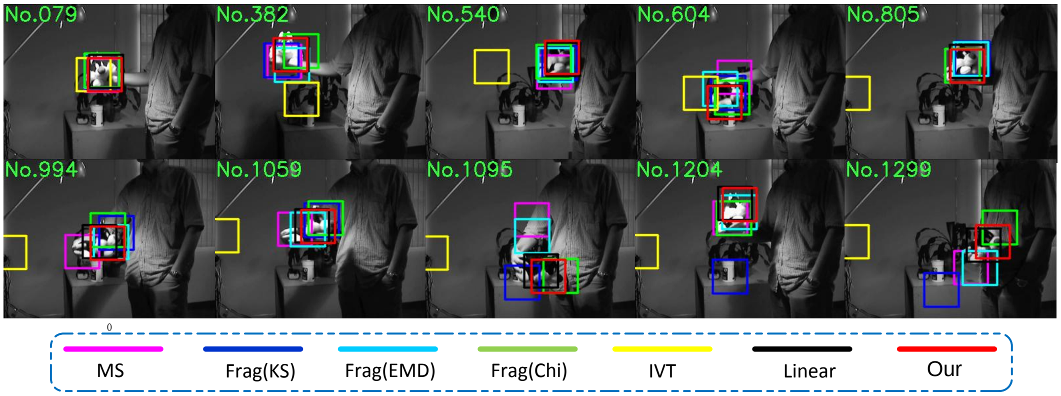

Figure 6: Tracking results on *Sylvester*.

In Fig.7 and Fig.8, the frame results for *Tiger1* and *Tiger2* are shown, respectively. For *Tiger1*, we have 354 frames in total, for *Tiger2*, 356 frames we have. The main challenges in those two videos include appearance change, lighting variance, partly occlusion, and motion blur.

In Fig.7, we could see that in frame No.047, Frag, MS, IVT are all lose the target completely. Linear could achieve comparable results in frame No.157, No.175, No.194, but it lose the target from No.229. In frame No.243 and Frame No.325, only our tracker could track the target accurate.

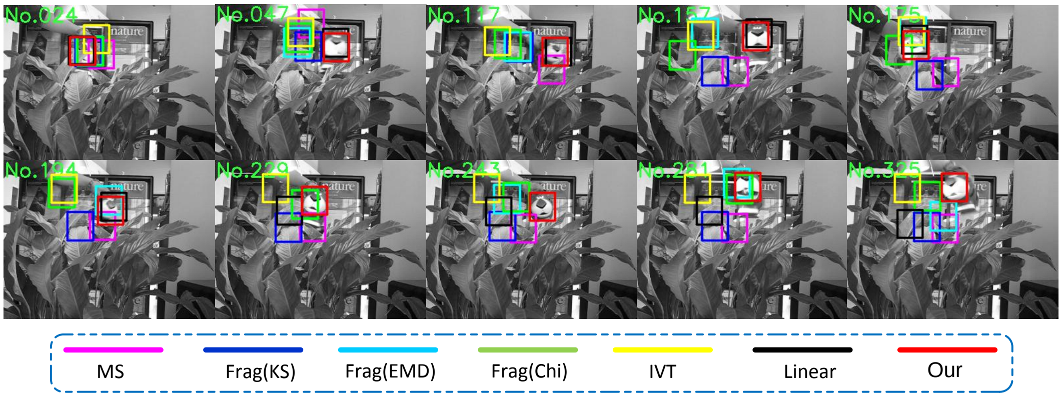

Figure 7: Tracking results on *Tiger1*.

In Fig.8, Frag, MS and IVT also very unstable, i.e., as shown in frame No.030, No.094, No.115. Linear could also achieve comparable results with our tracker in many frames, but at the end of this video, i.e., frame No.361, Linear lose the target, but we still cover the target accurate.

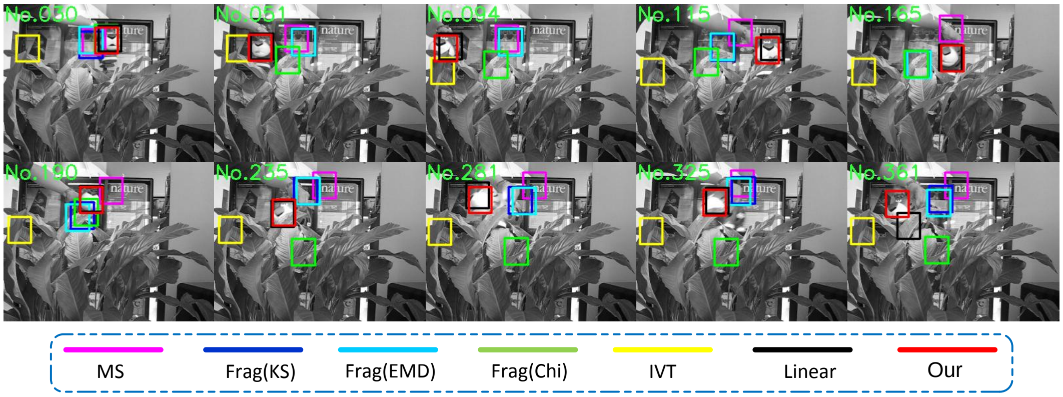

Figure 8: Tracking results on *Tiger2*.