Supplementary

A Improved IOU Tracker Algorithm

A detailed description of our improved IOU tracker method is shown in Algorithm 1, where D_f denotes the detections at frame f, d_j the j^{th} detection at the frame, Δ_f the displacements of corresponding detections in frame f and $f+1$, δ_j the displacements of j^{th} detection in frame f, T_α active tracks, T_f finished tracks, ttl the maximum number of virtual detections and F the number of frames in the sequence.

Fed with a set of detections and displacements, the algorithm first filters detections with a confidence threshold σ_{low}. Then for each track t_i in T_α, if IOU of t_i and a detection in current frame is larger than threshold σ_{iou}, it will be treated as a good association and the track will be updated. Otherwise, a virtual detection will be assigned to t_i. If the number of virtual detections in t_i is greater than ttl, then t_i will be considered as a finished track and added to T_f after filtered by track confidence and length. Virtual detections can fill the gaps caused by missing detection and thus reduce fragmentation in multi-object tracking. Compared with original algorithm, a 3D IOU is used in our approach, and each detection d_j is adjusted with its displacement δ_j from previous frame. The improvement is highlight in red.

Algorithm 1: Improved IOU Tracker

Input: $D = \{D_0, D_1, ..., D_{F-1}\} = \{\{d_0, d_1, ..., d_{N_0}\}, ..., \{d_0, d_1, ..., d_{N_{F-1}}\}\}$,
$\Delta = \{\Delta_0, \Delta_1, ..., \Delta_{F-1}\} = \{\{\delta_0, \delta_1, ..., \delta_{N_0}\}, ..., \{\delta_0, \delta_1, ..., \delta_{N_{F-1}}\}\}$
Output: T_f
Initialize: $T_a = \phi, T_f = \phi, D = \{\{D_i \mid d_i \in D_j, d_j \leq \sigma_{low}\} \mid D_j \in D\}$
for $f = 0$ *to* F **do**
  for $t_i \in T_a$ **do**
    $d_{best} = d_j$ where $max(IOU_{3d}(d_j + \delta_j, t_i)), d_j \in D_f, \delta_j \in \Delta_f$
    if $IOU_{3d}(d_{best}, t_i) \geq \sigma_{iou}$ **then**
      replace virtual detections in t_i through interpolation
      add d_{best} to t_i
      remove d_{best} from D_f
    else if *number of virtual detections in* $t_i \leq ttl$ **then**
      add new virtual detection to the end of t_i
    else if $highest\_score(t_i) \geq \sigma_{high}$ *and* $len(t_i) \geq t_{min}$ **then**
      remove virtual detections in t_i
      add t_i to T_f
      remove t_i from T_a
  for $d_j \in D_f$ **do**
    start a new track t with d_j and insert into T_a
for $t_j \in T_a$ **do**
  remove visual detections in t_j
  if $highest\_score(t_j) \geq \sigma_{high}$ *and* $len(t_j) \geq t_{min}$ **then**
    add t_i to T_f
return T_f

B More Tracking Examples

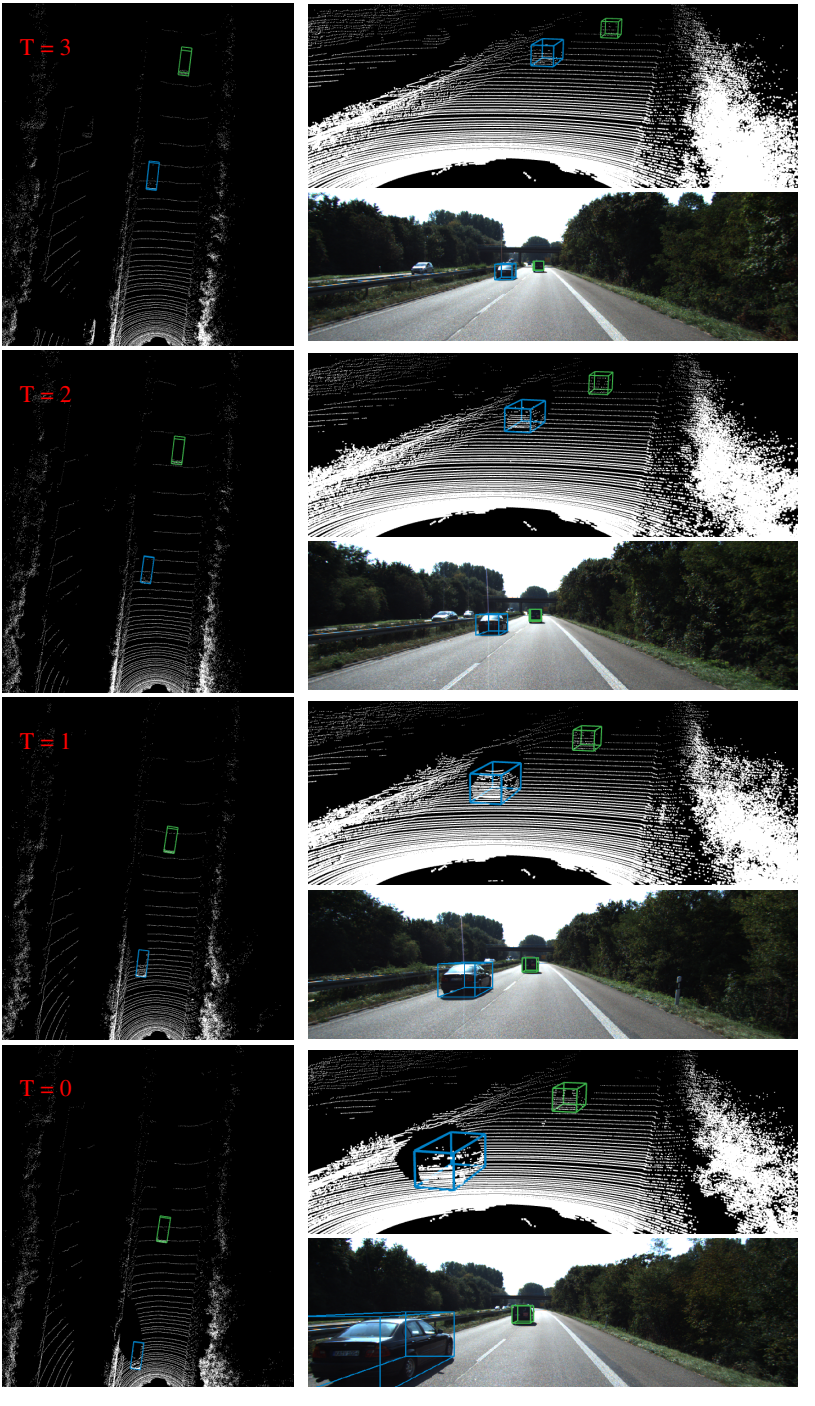

Figure 1: A set of trajectories of sequence 6.

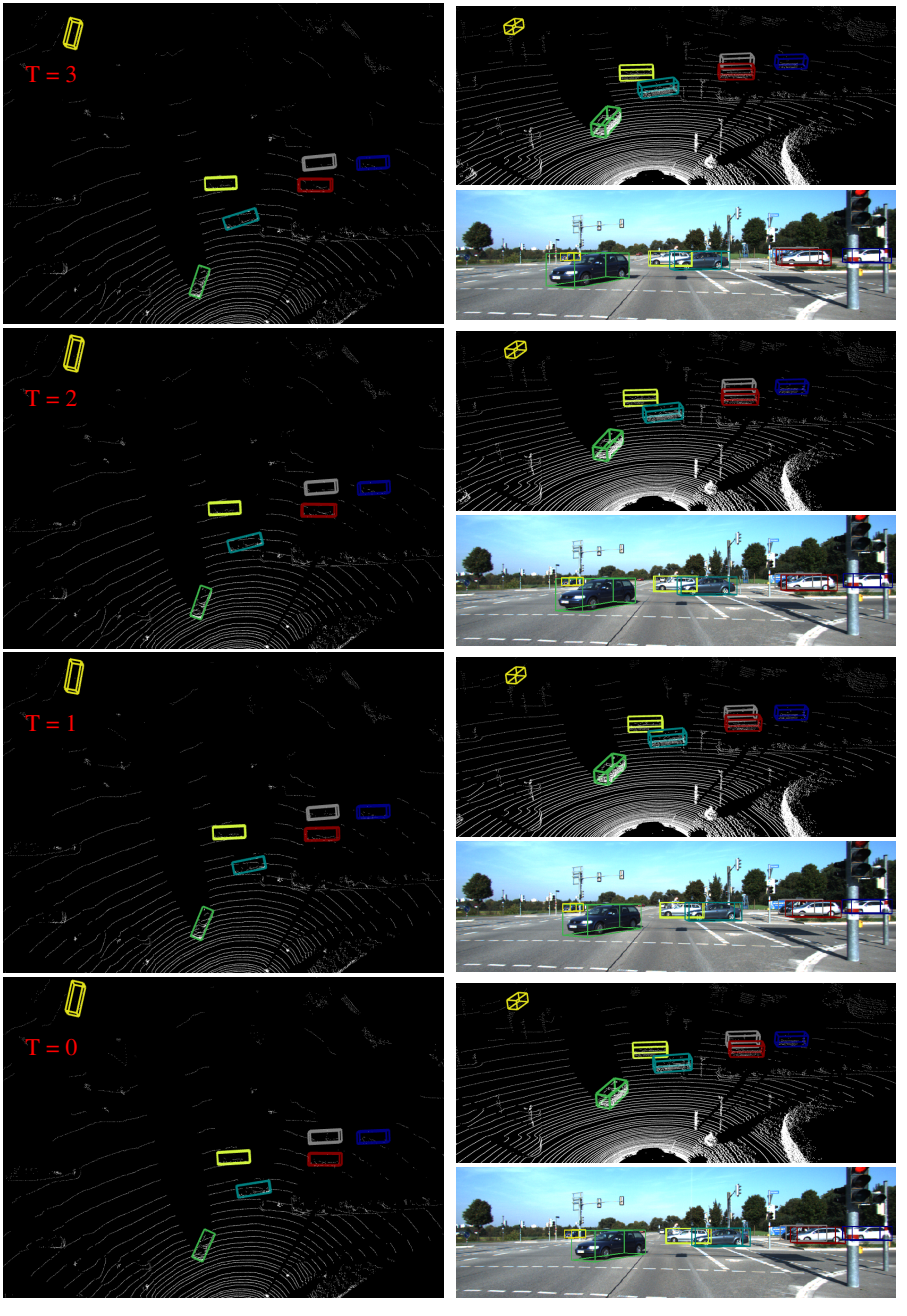

Figure 2: A set of trajectories of sequence 10.

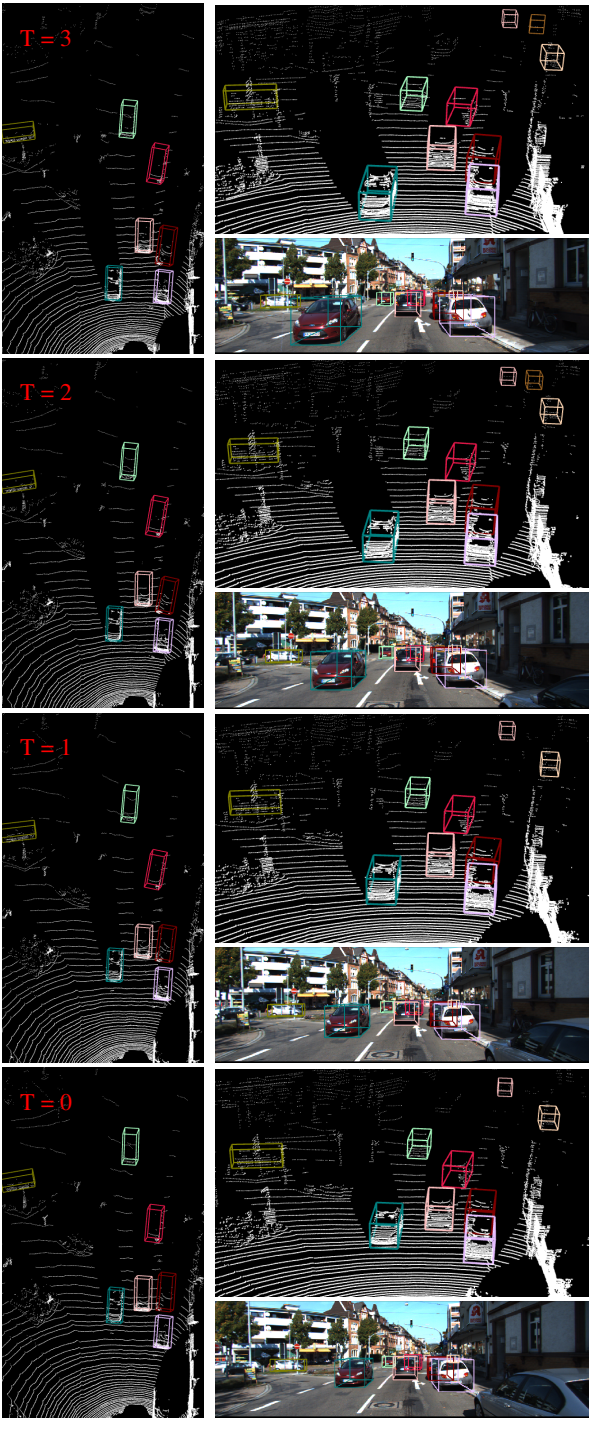

Figure 3: A set of trajectories of sequence 11.

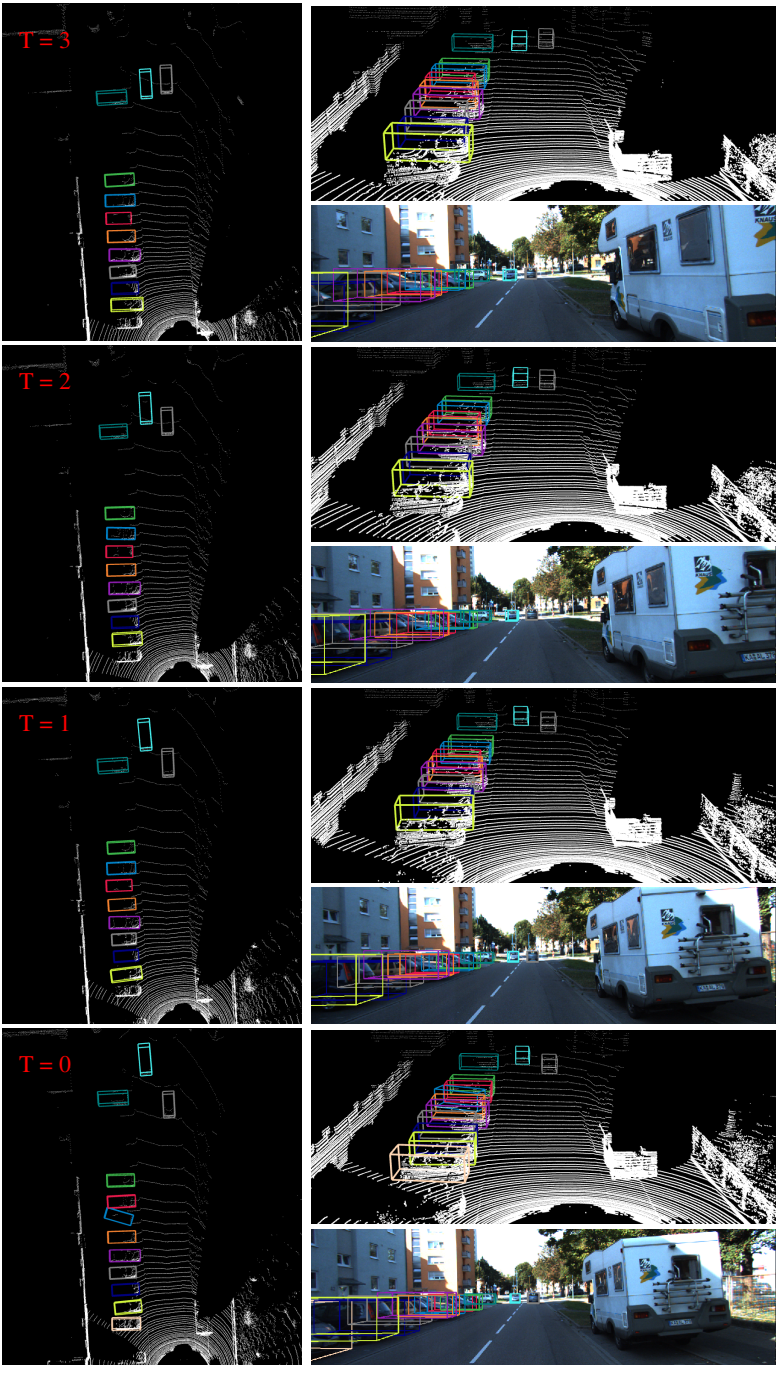

Figure 4: A set of trajectories of sequence 12.