Progress Towards Improved Multi-Object Tracking (MOT) Performance Based on Imperfect Detection System

UH Hilo Undergraduate: Taishi Nammoto
Supervisor: Dr. Travis Mandel

Introduction

1. Motivation

KPD tracker: Kalman tracking based on Probabilistic Detections

What are the goals of our tracking system?
There are two major goals of KPD tracker. One is to build a tracking system that works with **limited data** in terms of the labeled videos and pictures. This causes the low recall and precision of the detector. Another goal is to make a **long term** tracking system. A system that doesn't lose track of fish will help divers.

We want a general tracking system that works not only for fish but many other objects. We choose cars for another tracking object because there are a lot of existing tracking models specifically for cars.

Bobbie Suarez

2. Previous work

The tracking system was evaluated by Intersection over Union (**IoU**) approach.

evaluated value = (overlapping area of a true box and a predicted box) / (Combined area)

Overlapping area Combined area

Fig. 1: IoU approach

Although the video evaluated by IoU approach seemed great, the performance in terms of the recall and precision were poor compared to other systems.

The system was just developed to track fish. The tracking is based on the motion, the detection, and the color.

Improvements

3. New Evaluation Approach

IoU approach was replaced with a new evaluation method, "Box distance approach."

$$Evl = \frac{\sqrt{(C_{true}-C_{predicted})\cdot(C_{true}-C_{predicted})}}{\sqrt{(C_{true}-O_{true})\cdot(C_{true}-O_{true})}}$$

Evl: evaluated value
C_{true}: the position at the center of a true box
$C_{predicted}$: the position at the center of a predicted box
O_{true}: the position at the corner of a true box

Fig. 2: Box distance approach

This method is better to match intuition because things are close rather than requiring near-perfect overlap.

KPD has a minimum confidence to start tracking, but KPD used all of them including low confidence one.

4. New Tracking Object

We trained deep learning system with just 200 images of car. It resulted in a mediocre detection performance. This allows us to determine how the tracker works in another case of imperfect detection. Initially, the tracker's result was poor on cars even though it worked well on fish.

5. Performance

Divers want us to track off-screen fishes. However, the system did not allow to annotate off-screen fishes. This caused the poor evaluation. By removing the box of off-screen objects, a huge improvement of the score was acquired.

Measurements of accuracy

Tracking systems

Car	F-score	Recall	Precision	MOT Accuracy	MOT Precision
KPD	**84.8**	**83.9**	85.8	**69.8**	74.3
GOG	71.5	71.3	71.8	37.4	**85.0**
KIOU	56.6	49.3	66.3	23.3	66.2
GMMCP	69.2	54.0	**96.5**	52.0	6.0

Fish	F-score	Recall	Precision	MOT Accuracy	MOT Precision
KPD	**87.0**	**82.8**	91.6	**74.7**	53.0
GOG	76.3	67.3	87.6	55.9	55.9
KIOU	71.3	57.2	94.6	53.6	**58.6**
GMMCP	48.7	32.4	**100.0**	31.9	15.0

6. Video

The detection system sometimes detected fish in a single high confidence box with no boxes before and after the frame. Our tracker would see a sequence of detector confidences like [-2, 0.74, -2]. -2 means no object detection.

By default, the tracker starts at the highest confidence box. However, starting at these isolated detections did not work well as the tracker could not determine motion. Therefore, we require the tracker to identify a detection on at least one side before committing to a initial box, rather than just picking the one with highest confidence. For instance, the good sequence of the detector confidences would be like [0.67, 0.34, 0.58].

Questions

What is the Kalman? (KPD tracker: Kalman tracking based on Probabilistic Detections)
Kalman comes from "**Kalman filtering**," an algorithm that produces estimates of unknown variables based on multiple measurements

What is precision and recall and F-score?
TP: number of evaluated values more than a threshold (prediction is 1) and the label is 1
FP: number of evaluated values less than a threshold (prediction is 0) and the label is 1
FN: number of evaluated values more than a threshold (prediction is 1) and the label is 0
Precision: Fraction of predicted values that are actually correct, $precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$
Recall: Fraction of actual values that are predicted correctly, $recall = \frac{True\ Positive}{True\ Positive true + False\ Negative}$
F-score: a measurement of accuracy based on precision and recall, $F - score = \frac{2 \cdot precision \cdot recall}{precision + recall}$

What is the detection? (The tracking is based on the motion, the detection, and the color.)
The detection refers to an independent neural network applied to all image frames to obtain object detections.

What is MOTA and MOTP?
Multiple Object Tracking Precision (MOTP): the total position error for matched object hypothesis pairs over all frames, averaged by the total number of matches made.
t frame
d_t distance between object for frame t
c_t number of object-hypothesis correspondences made for frame t

$$MOTP = \frac{\sum_{i,t} d_t^i}{\sum_t c_t}$$

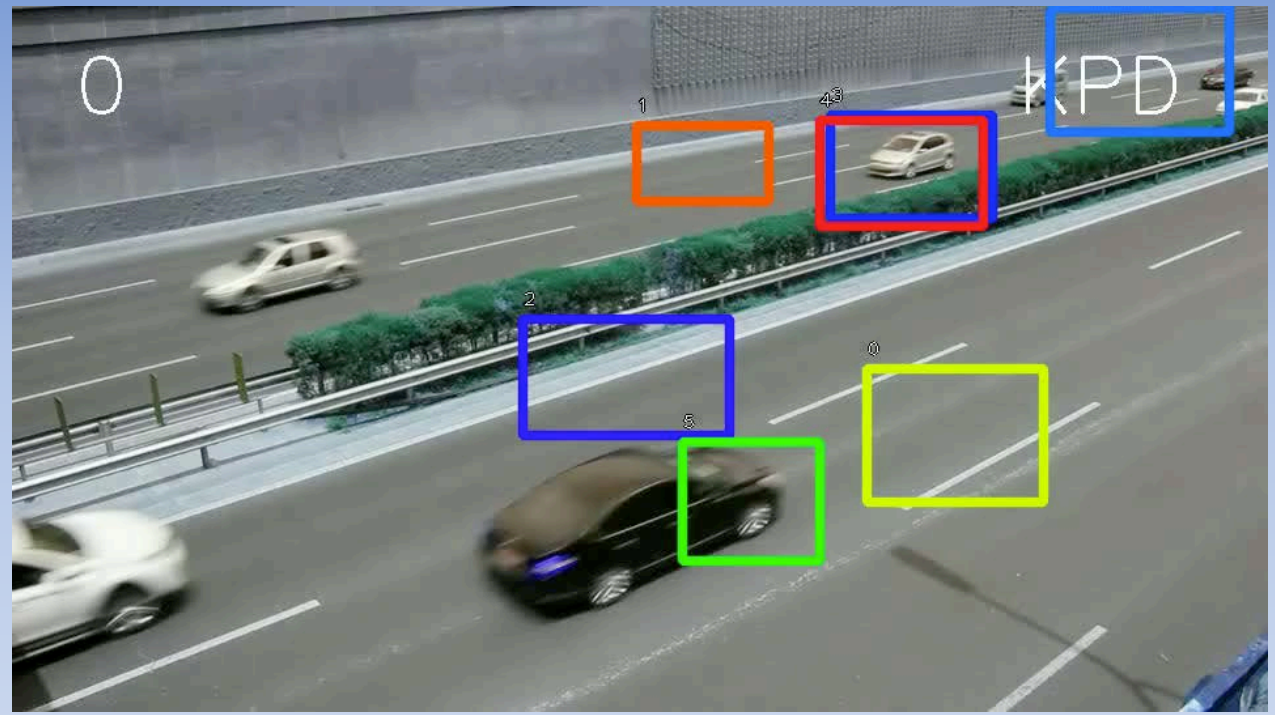
Multiple Object Tracking Accuracy (MOTA): a very intuitive measure of the tracker's performance at keeping accurate trajectories, independent of its precision in estimating object position.
m_t misses for frame t
fp_t number of false positive for frame t
mme_t track ID mismatch errors made for frame t
g_t number of objects for frame t

$$MOTA = 1 - \frac{\sum_t (m_t + fp_t + mme_t)}{\sum_t g_t}$$

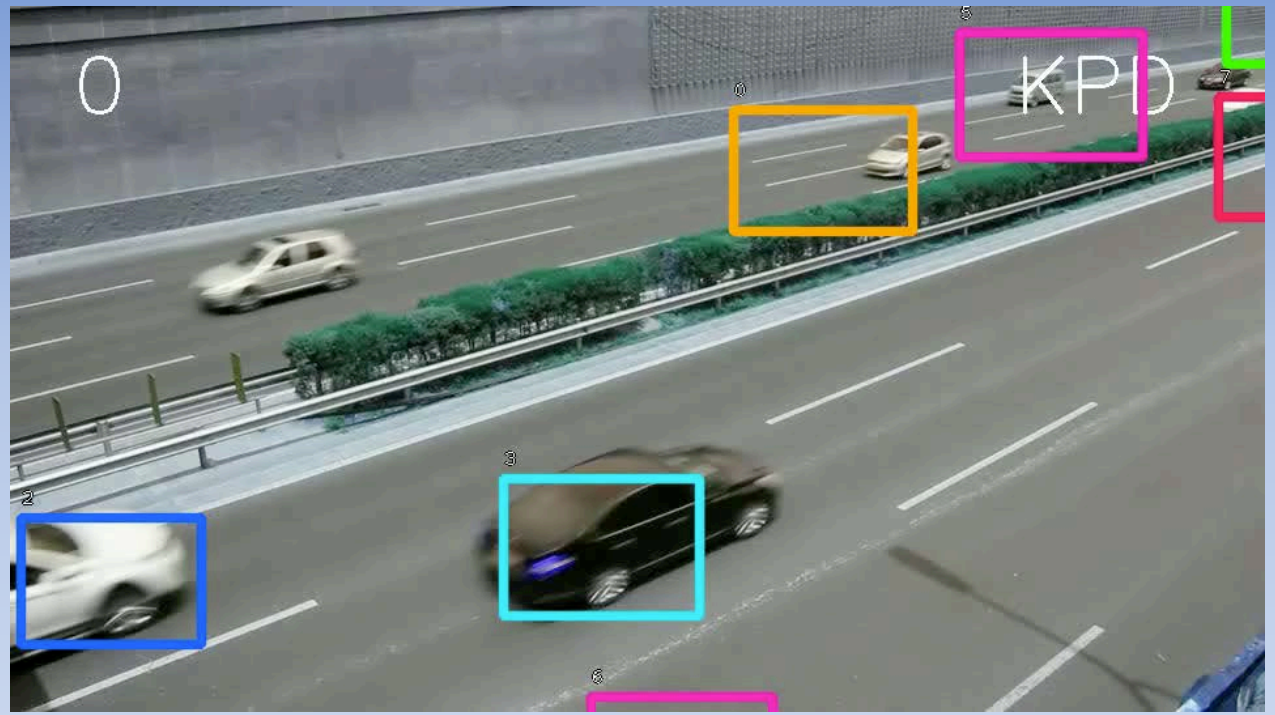
The fish video generated by the KPD

Fish	F-score	Recall	Precision	MOTA	MOTP	MOTAL
KPD (New eval)	**87.0**	**82.8**	**91.6**	**74.7**	53.0	**75.1**
KPD (IoU method)	67.3	66.2	68.5	35.7	**63.4**	35.7

The car video generated by the unimproved KPD



The car video generated by the improved KPD



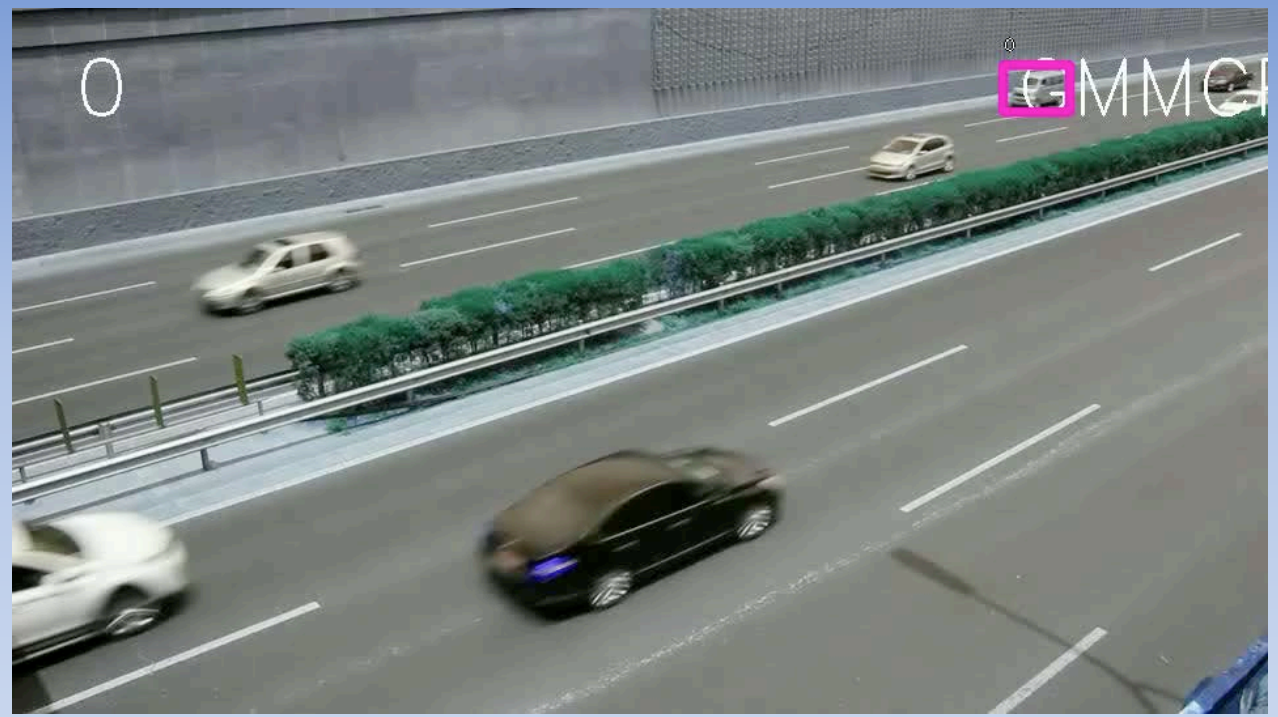
Car	F-score	Recall	Precision	MOTA	MOTP
Improved KPD	**84.8**	**83.9**	**85.8**	**69.8**	74.3
Unimproved KPD	73.3	74.8	71.9	45.0	**121.1**

The fish video generated by the GMMCP

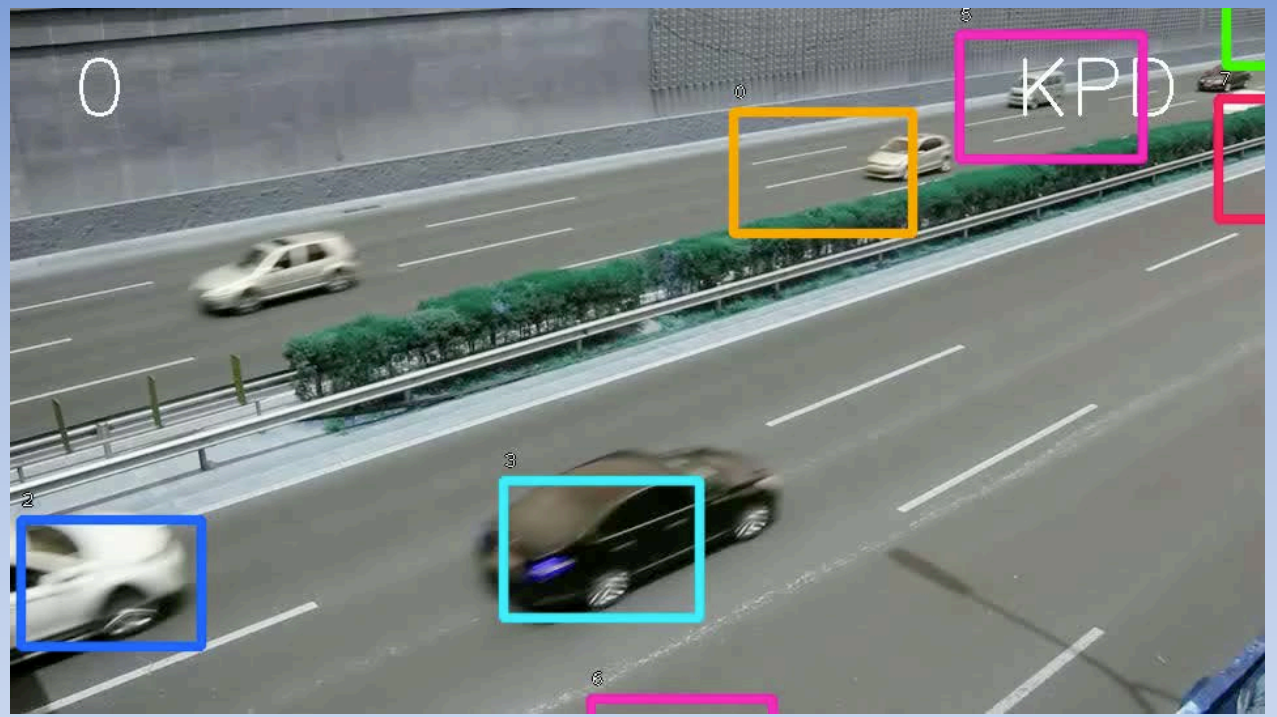
The fish video generated by the KPD (Long term track)

Fish	F-score	Recall	Precision	MOTA	MOTP
KPD	**87.0**	**82.8**	91.6	**74.7**	53.0
GMMCP	48.7	32.4	**100.0**	31.9	15.0

The car video generated by the GMMCP



The car video generated by the KPD



Car	F-score	Recall	Precision	MOTA	MOTP
KPD	**84.8**	**83.9**	85.8	**69.8**	74.3
GMMCP	69.2	54.0	**96.5**	52.0	6.0