Object Detection and Mapping on 3D Plane

Junwon Sung
Junghwan Yim

University at Buffalo
School of Engineering and Applied Sciences

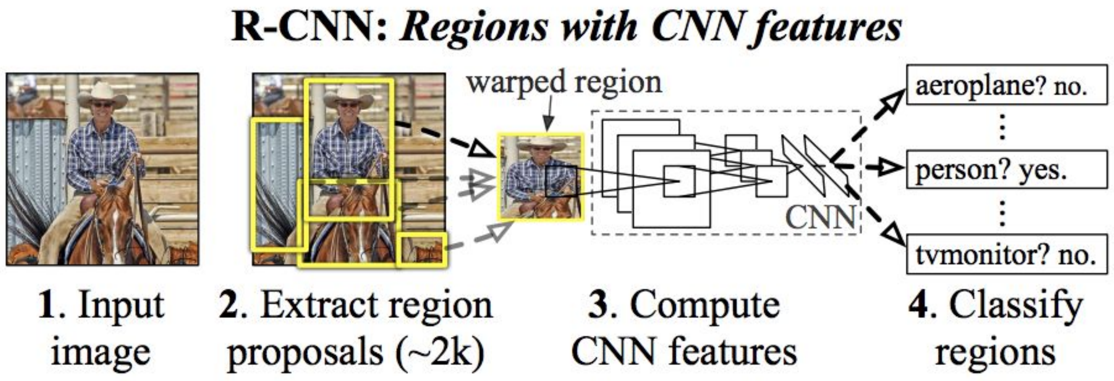
Contents

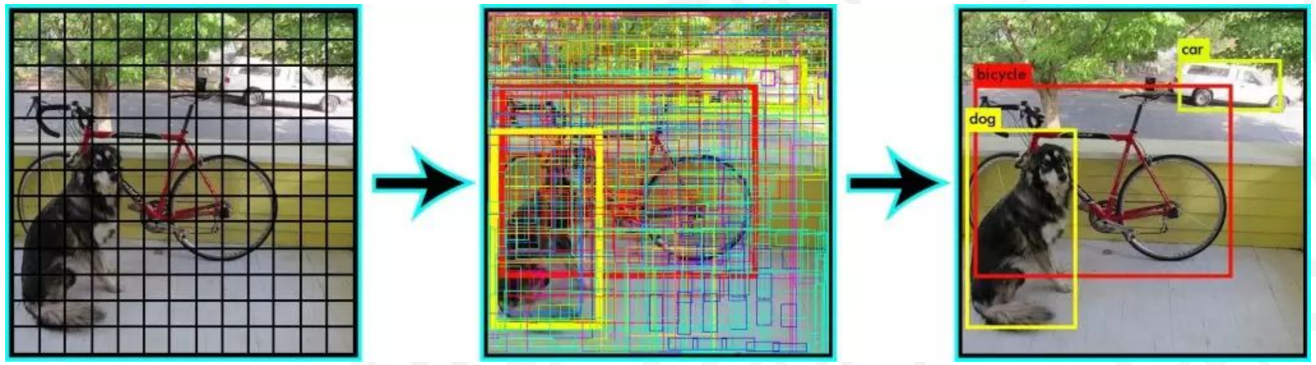
- Introduction
- Detection
- Prediction
- Mapping
- Future work

Introduction

How can we map the 2D detected objects into 3D plane (like HD-map)
with lower computing cost
and higher accuracy?

Detection - R-CNN, YOLO



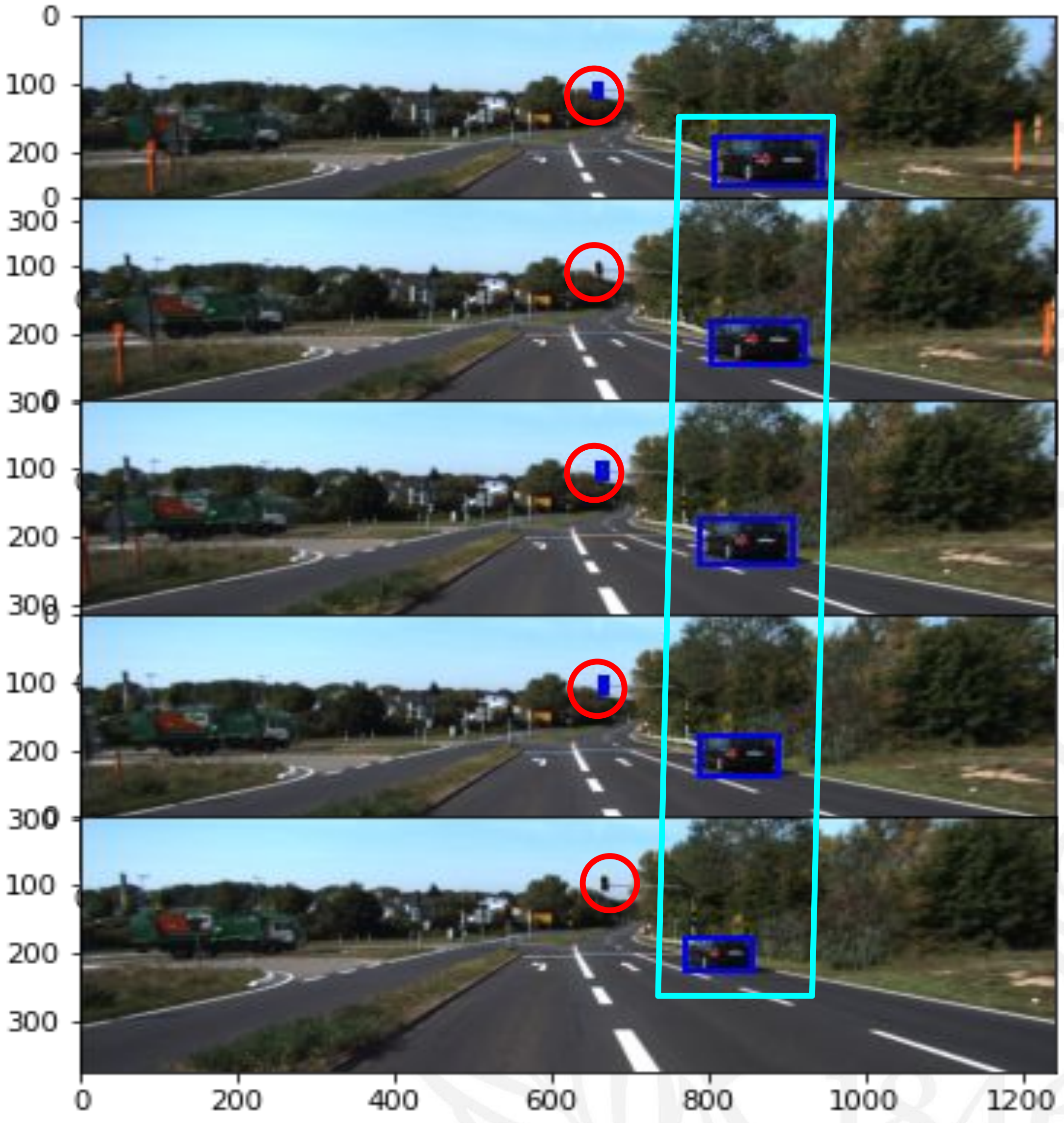


	Method	mAP	FPS
R-CNN	**2 stage** (classification after localization)	**82.5**	1.326
Yolov3	**1 stage** (classification, localization at the same time)	80.3	**2.809**

Detection Problem & Solution

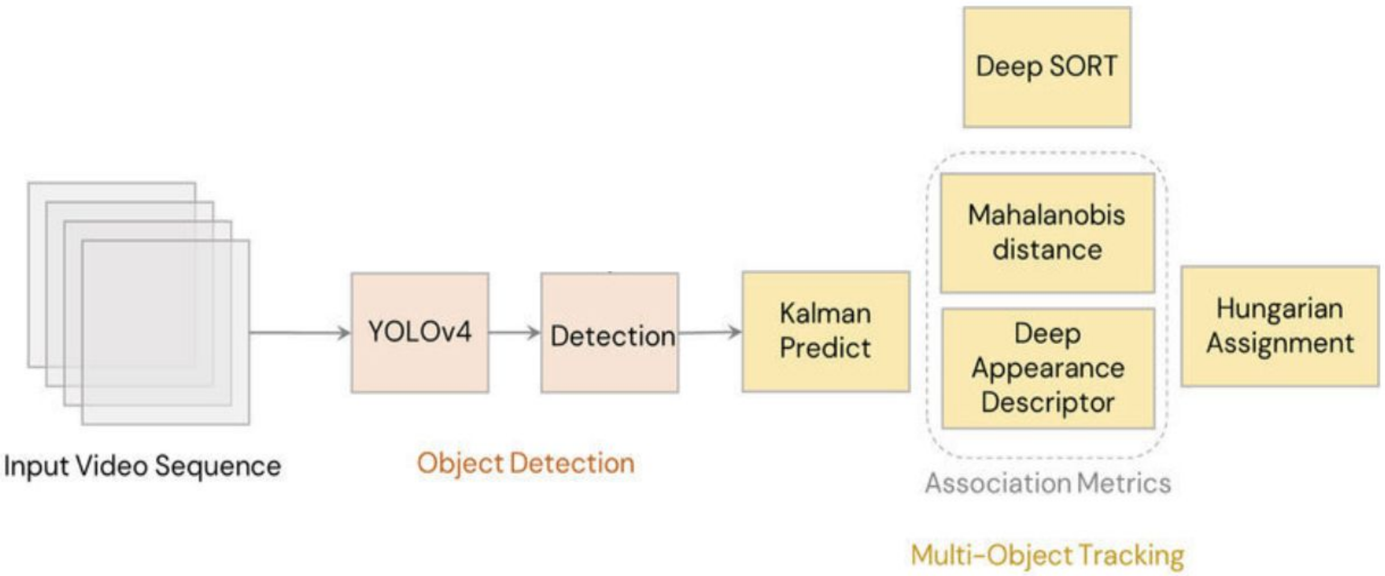
“Detection is NOT Consistent”

Detection Bounding Box can be predicted with the current car velocity and position from IMU sensor?



Prediction - SORT, Deep-SORT

SORT (SIMPLE ONLINE REAL-TIME TRACKING)



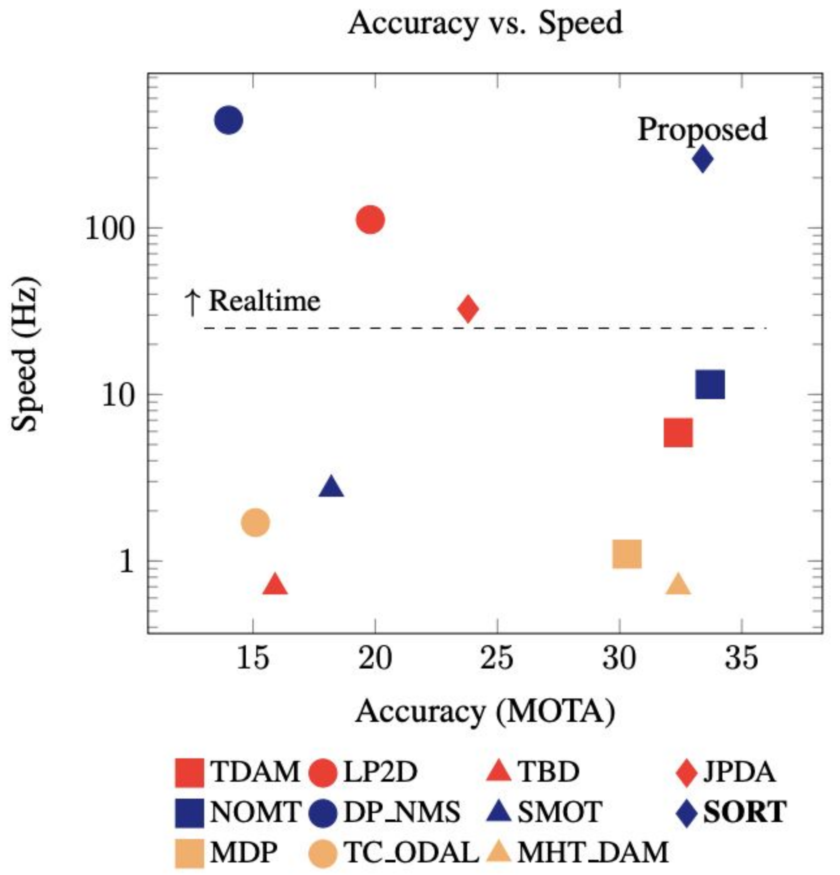


Fig. 1. Benchmark performance of the proposed method (**SORT**) in relation to several baseline trackers [6]. Each marker indicates a trackers accuracy and speed measured in frames per second (FPS) [Hz], i.e. higher and more right is better.

Prediction - Limit of SORT, Deep-SORT

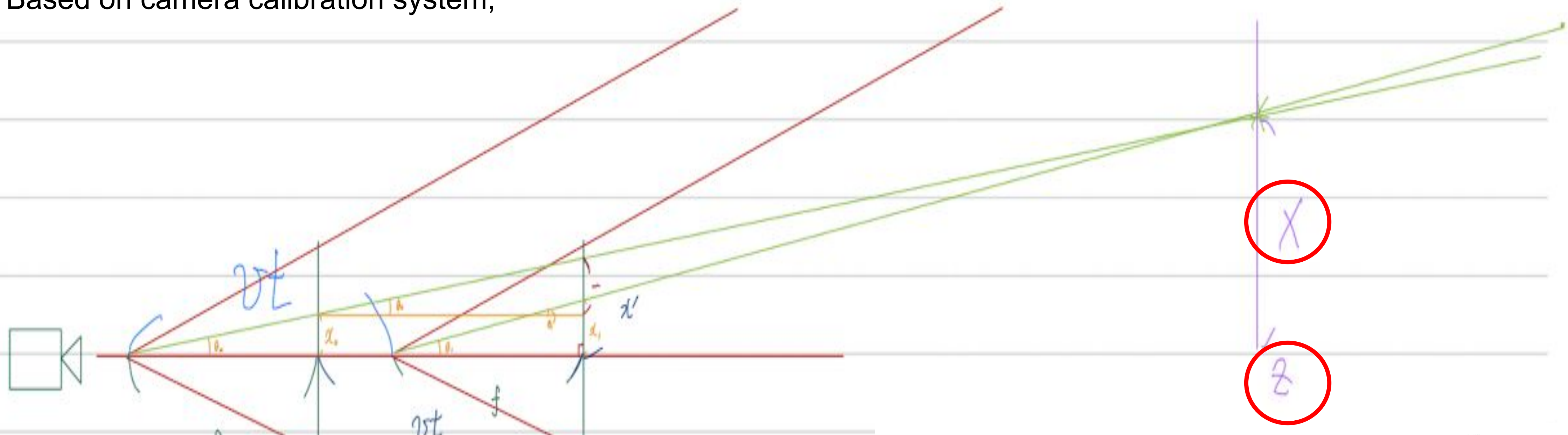
1. Kalman Filter designed with the constant velocity

2. Performance strongly depend on the detection model

So, **make the Kalman Filter based on IMU sensor**

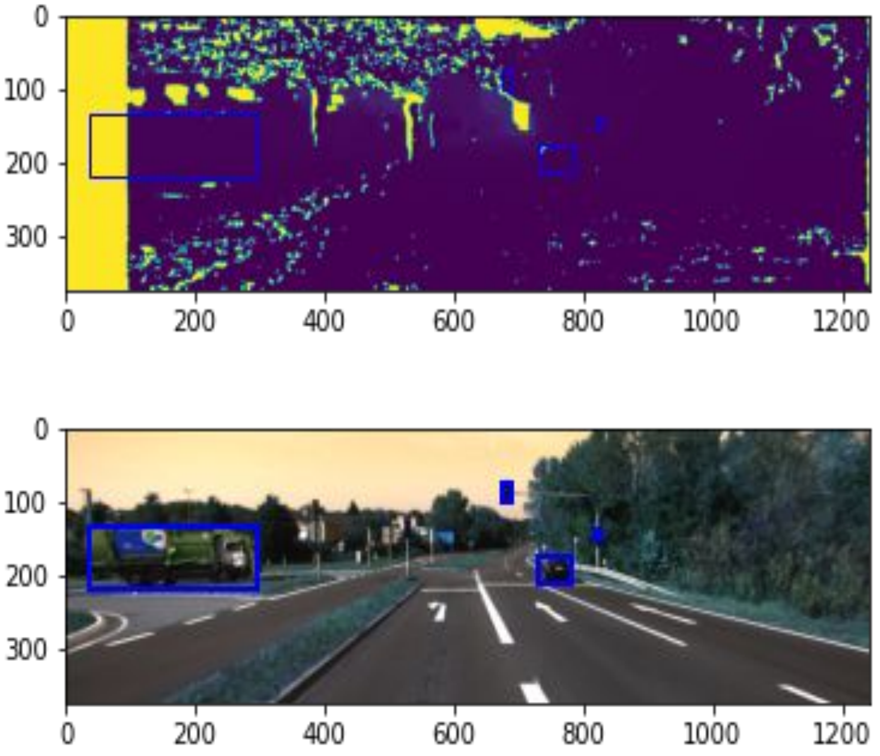
Prediction - Master equation of problem

Based on camera calibration system,



$$\frac{X}{Z} = \frac{x_0}{f} \qquad \frac{X}{Z-vt} = \frac{x_1}{f}$$

Prediction - Measuring Distance between The Object




```
125
125 : car        minimum distance :    (824.0, 179.0) 141.0    74.0    13.516368873914562
125 : tra        minimum distance :    (655.0, 102.0) 7.0      18.0    48.04646748149317
126
126 : car        minimum distance :    (807.0, 179.0) 139.0    71.0    14.436497271434567
126 : tra        minimum distance :    (657.0, 99.0) 8.0       19.0    46.24020930549718
127
127 : car        minimum distance :    (804.0, 181.0) 119.0    64.0    14.677679803415574
128
128 : car        minimum distance :    (789.0, 175.0) 119.0    66.0    15.413403101832394
128 : tra        minimum distance :    (661.0, 94.0) 8.0       20.0    27.702467737077143
129
129 : car        minimum distance :    (788.0, 178.0) 100.0    60.0    16.184073256924012
129 : tra        minimum distance :    (662.0, 92.0) 8.0       22.0    4.075512152174371
130
130 : car        minimum distance :    (778.0, 181.0) 97.0     50.0    14.748076349235312
130 : tra        minimum distance :    (663.0, 91.0) 8.0       23.0    4.072813137504056
131
131 : car        minimum distance :    (771.0, 180.0) 87.0     47.0    19.218586992597267
132
132 : car        minimum distance :    (765.0, 175.0) 83.0     50.0    17.47144272054297
132 : tra        minimum distance :    (665.0, 88.0) 10.0      22.0    4.479204543067098
133
133 : car        minimum distance :    (760.0, 175.0) 76.0     50.0    20.364065687520284
```

Problem : **Too Large Uncertainty**

Prediction - Estimating Distance between The Object

Input : (vf, vl, vu, af, al, au, x, y, w, h)

Dense Layer(12, 10), Lelu

Dense Layer(8, 12), Lelu

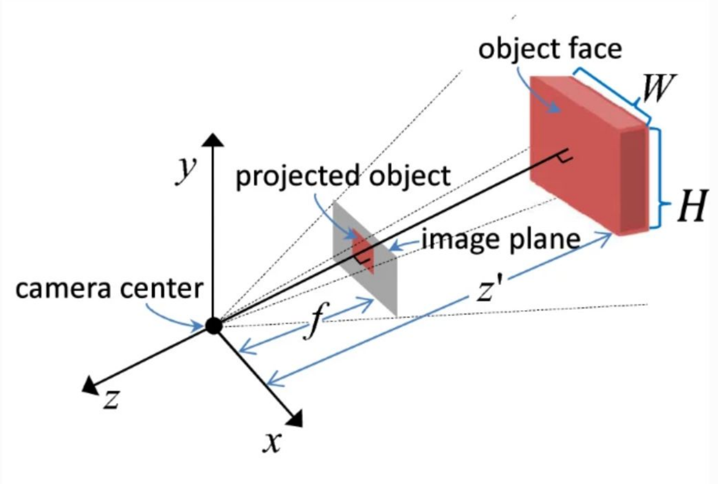
Dense Layer(4, 8), Lelu

Dense Layer(2, 4), Lelu

Dense Layer(1, 2), Lelu

loss : mean sqrt error
optimizer : adam

Mapping



As we getting the z-value from network and Kalman Filter, the height of object also can be determined.

If we getting the z-value from pixels placed in Bounding Box, we can inference the shape of the detected objects like scanned with LiDAR sensor

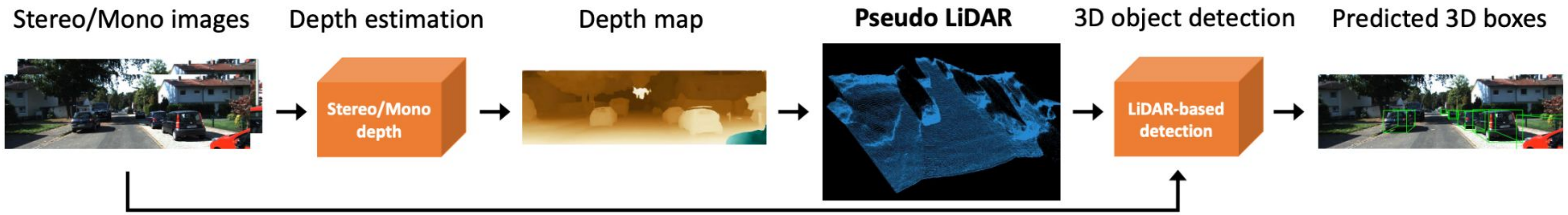
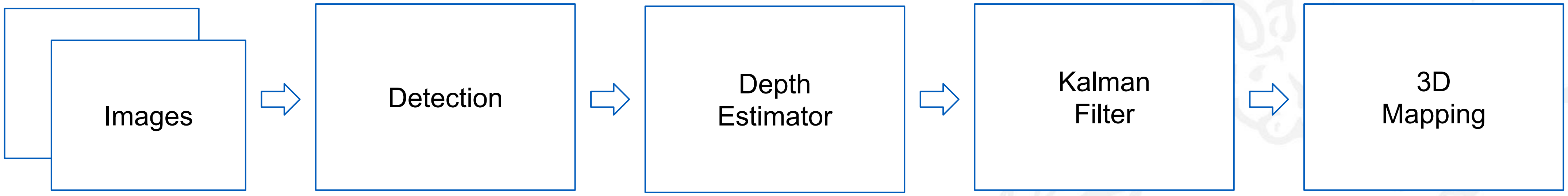


Figure 2: **The proposed pipeline for image-based 3D object detection.** Given stereo or monocular images, we first predict the depth map, followed by back-projecting it into a 3D point cloud in the LiDAR coordinate system. We refer to this representation as *pseudo-LiDAR*, and process it exactly like LiDAR — any LiDAR-based detection algorithms can be applied.

Future Work / Conclusion



- This study suggests the method of detecting objects first in 2D and converting to 3D plane.
- This approach is possibly more efficient than the pseudo-LiDAR by reducing the computing pixels to convert into 3D and using Kalman Filter instead of the Neural network approach.
- With this approach, computing power can be concentrated on the 2D object detection model to improve mapping and detection performance.
- This study could get cost-effectiveness and FPS performance at the same time.

Thank you