多源融合导航及应用学术沙龙

MULTI-SOURCE FUSION NAVIGATION AND APPLICATION ACADEMIC SALON

武汉大学测绘学院
SCHOOL OF GEODESY AND GEOMATICS

报告主题

时间	报告主题	报告嘉宾
上午	主持人：李星星（武汉大学）、王博（北京理工大学）	
08:05–08:50	大规模点云地图的自动化构建	高翔（北京智行者科技有限公司）
08:50–09:20	专题讨论（一）	
09:20–10:05	多模多频实时GNSS软件接收机	刘刚（清华大学）
10:05–10:35	专题讨论（二）	
10:35–11:20	城市拒止环境下的视觉SLAM	邹丹平（上海交通大学）
11:20–11:50	专题讨论（三）	
	中场休息	
下午	主持人：牛小骥（武汉大学）、朱锋（武汉大学）	
14:00–14:45	On the Way to Autonomous Driving: AI Brings a New Dimension to Machine Vision	范睿（加州大学圣地亚哥分校）
14:45–15:15	专题讨论（四）	
15:15–16:00	3D LiDAR Aided GNSS Positioning in Urban Canyons	文伟松（香港理工大学）
16:00–16:30	专题讨论（五）	
16:30–17:15	基于多源传感器数据融合的动态场景SLAM研究	罗斌（武汉大学）
17:15–17:45	专题讨论（六）	

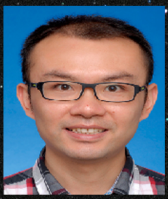
报告嘉宾

高翔

慕尼黑工业大学博士后，清华大学自动化系博士，长期从事SLAM研究，主要包括机器人中的视觉SLAM技术、机器学习与SLAM的结合。主编畅销书《视觉SLAM十四讲：从理论到实践》

刘刚

清华大学仪器科学与技术专业博士，先后在海军航空大学、阿里巴巴达摩院、清华大学任职。研究兴趣为惯性技术、卫星导航技术、室内外复杂环境多源融合定位技术



邹丹平

就职于上海交通大学感知与导航研究所，研究兴趣为视觉SLAM与微型无人机自主导航。代表工作有动态环境下群体协同CoSLAM，面向城市拒止环境下的StructSLAM以及StructVIO等

罗斌

就职于武汉大学测绘遥感信息工程国家重点实验室，长期从事图像检索、移动机器人自主导航、机器视觉等方向研究

范睿

加州大学圣地亚哥分校（UCSD）博士后研究员，研究方向为计算机视觉、无人驾驶、高性能计算。曾任港科大智能驾驶中心执行主任，负责港科大无人车的研发。在计算机视觉与机器人领域共收录、发表会议、期刊论文近50篇

文伟松

香港理工大学博士，主要研究方向为移动机器人、智能载体的鲁棒定位和导航。在IEEE Transactions on Intelligent Transportation System、IEEE Transactions on Vehicular Technology、ICRA、IEEE ITSC等知名期刊、会议上发表论文数篇

时间：1月5日 8:00-18:00

地点: 武汉大学测绘学院202报告厅

扫码进入直播间
GO LIVE SHOW HERE

On the Way to Autonomous Driving: AI Brings a New Dimension to Machine Vision

Rui Ranger Fan

Jan. 05, 2021

UC San Diego

Rui Fan [范睿]

- 09/2011-07/2015: BEng – Harbin Institute of Technology.
- 09/2015-06/2018: Ph.D. – VI-Lab, the University of Bristol.
- 07/2018-02/2020: Postdoc – RAM-Lab, RI, the Hong Kong University of Science & Technology.
- 02/2020-present: Postdoc – UC San Diego.

- I will join the Dept. of Control Science & Engineering at Tongji University in Fall 2021 as a Research Full Professor. I will have several Master/Ph.D./Postdoc openings.

Keynote Talk Outline

1. Introduction
2. 3D Geometry Model Reconstruction
3. Semantic Driving Scene Understanding
4. Object Detection/Recognition
5. Conclusion

Keynote Talk Outline

1. **Introduction**
2. 3D Geometry Model Reconstruction
3. Semantic Driving Scene Understanding
4. Object Detection/Recognition
5. Conclusion

What is **Artificial Intelligence (AI)**?

- AI is wide-ranging branch of computer science concerned with building smart machines capable of performing tasks that typically require human intelligence.
- AI is an interdisciplinary science with multiple approaches, but advancements in **machine learning (ML)** and **deep learning (DL)** are creating a paradigm shift in virtually every sector of the tech industry.

AI v.s. ML v.s. DL

AI: Engineering of making intelligent machines and programs

AI

ML

DL: Set of algorithm to model high-level of abstraction

DL

ML: Ability to learn without being explicitly programmed

What is Machine/Computer Vision (M/CV)?

- **Machine Vision or Computer Vision** is an interdisciplinary scientific field that deals with how computers can gain high-level understanding from digital images or videos.

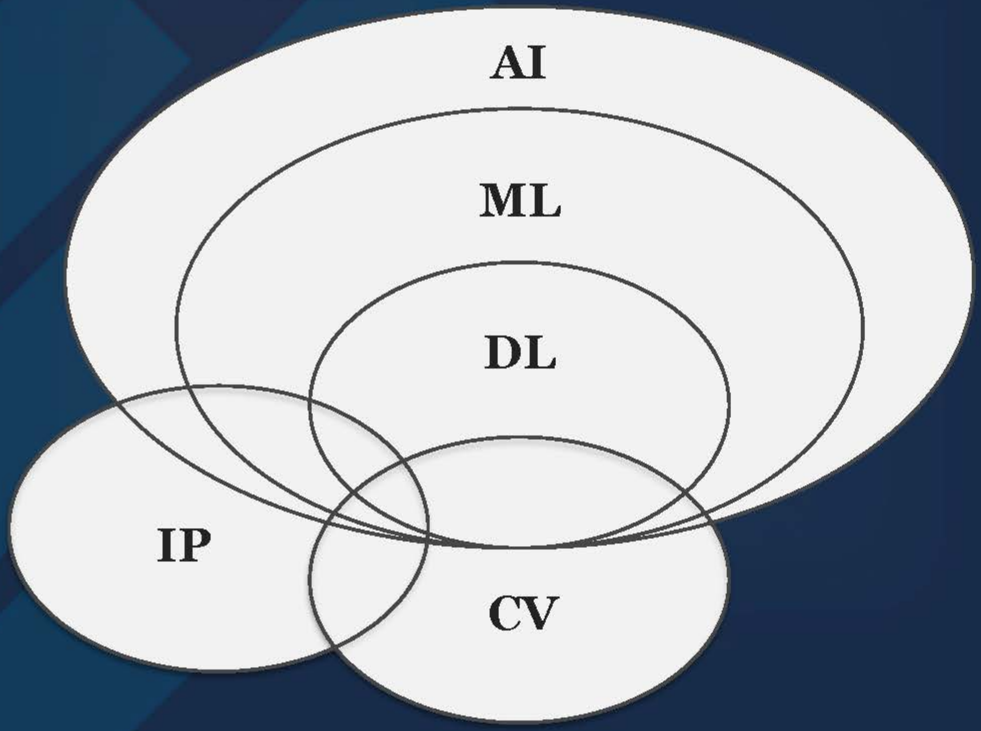
Image $\xrightarrow{\text{CV}}$ Understanding

- Difference between CV and Image Processing (IP)?

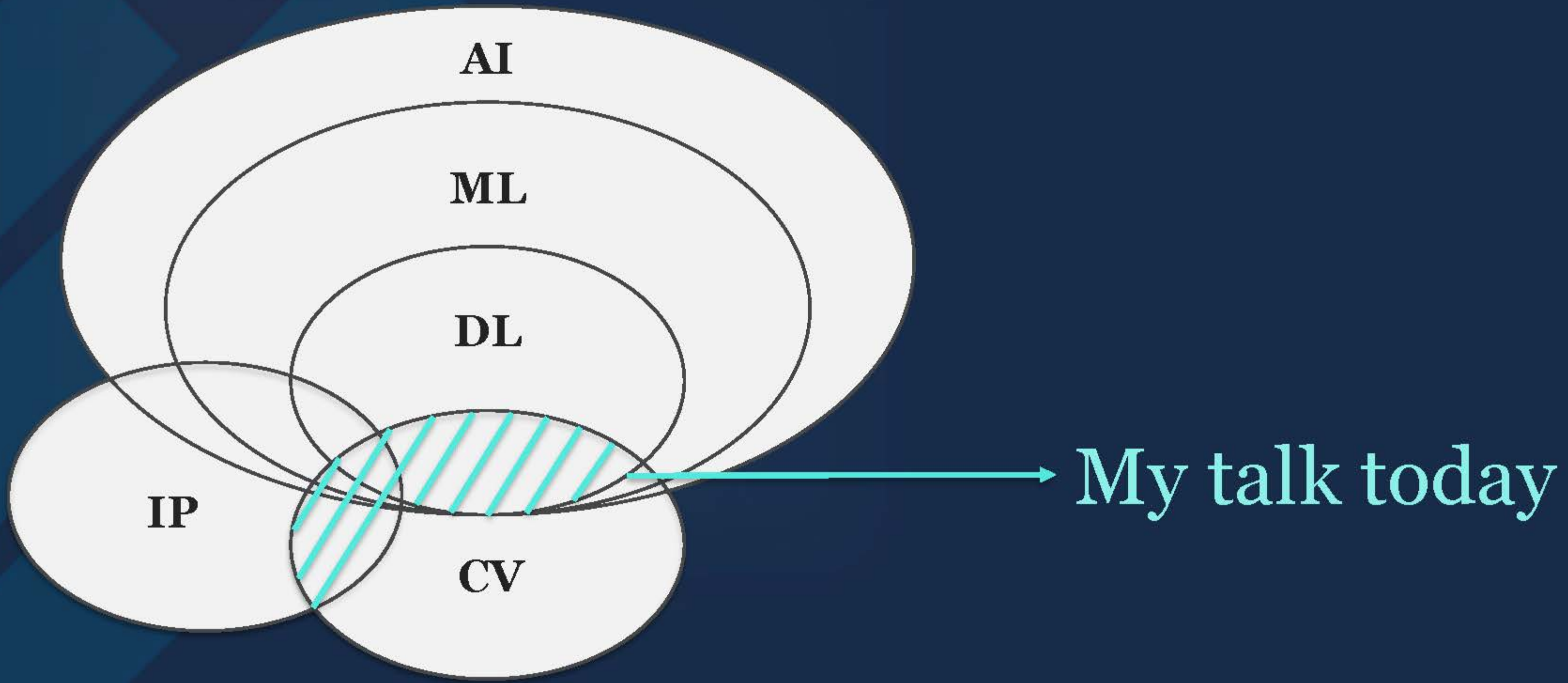
In IP, an image is "processed", namely, transformations are applied to an input image and an output image is obtained.

Image $\xrightarrow{\text{IP}}$ Image

AI v.s. ML v.s. DL v.s. CV v.s. IP



AI v.s. ML v.s. DL v.s. CV v.s. IP



Autonomous Vehicle

- An autonomous vehicle, also known as a robotic vehicle, self-driving vehicle, or driver-less vehicle, is a vehicle that is capable of sensing its environment and moving with little or no human input.

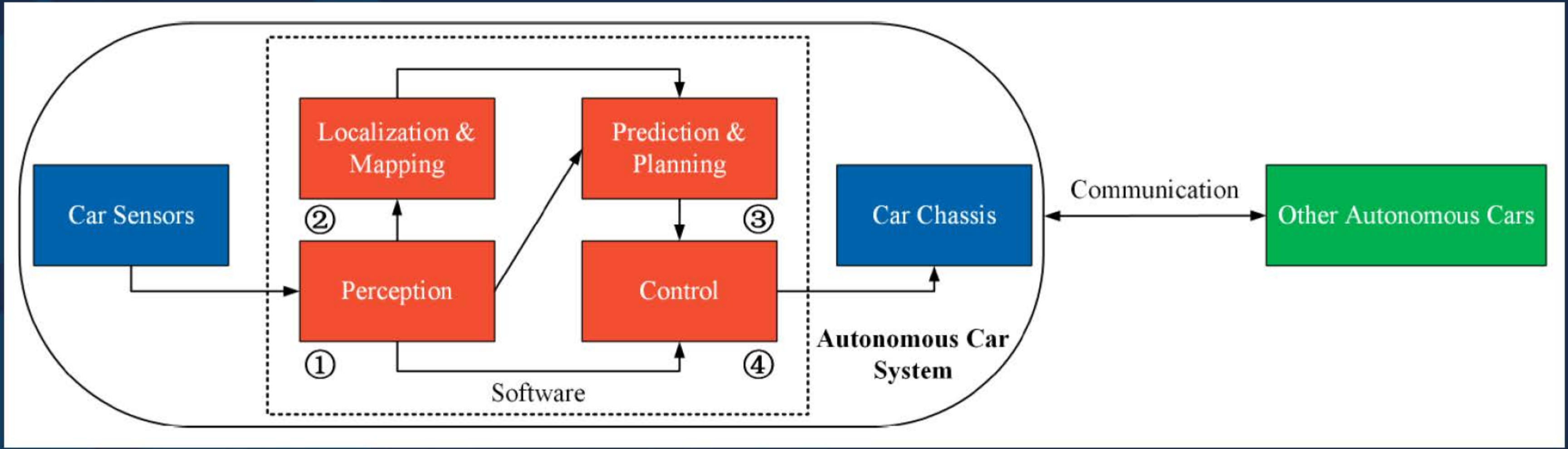
Which company is developing Autonomous Vehicles?

SAE Levels of Driving Automation

Full Automation

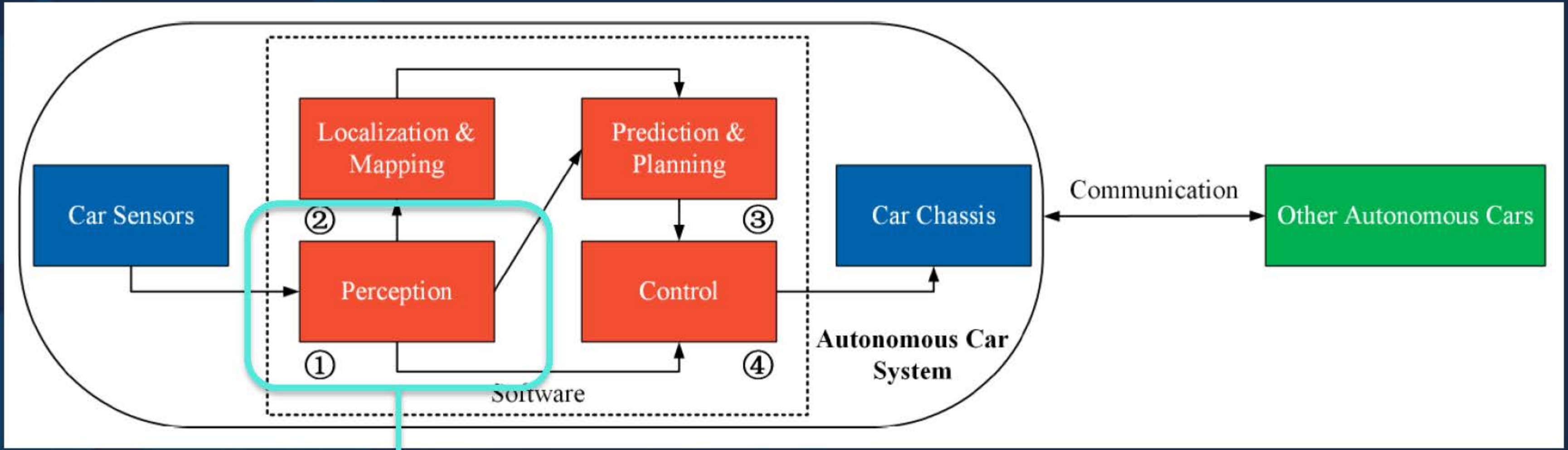
0	1	2	3	4	5
No Automation	**Driver Assistance**	**Partial Automation**	**Conditional Automation**	**High Automation**	**Full Automation**
Zero autonomy; the driver performs all driving tasks.	Vehicle is controlled by the driver, but some driving assist features may be included in the vehicle design.	Vehicle has combined automated functions, like acceleration and steering, but the driver must remain engaged with the driving task and monitor the environment at all times.	Driver is a necessity, but is not required to monitor the environment. The driver must be ready to take control of the vehicle at all times with notice.	The vehicle is capable of performing all driving functions under certain conditions. The driver may have the option to control the vehicle.	The vehicle is capable of performing all driving functions under all conditions. The driver may have the option to control the vehicle.

Autonomous Driving System



- The autonomous driving system consists of two main components: 1) **hardware (HW)** and 2) **software (SW)**.
- HW: Car Sensors and Car Chassis.
- SW: 1) Perception, 2) Localization & Mapping, 3) Prediction and Path Planning ; and 4) Control.

Autonomous Driving System



My talk today

Autonomous Vehicle Applications

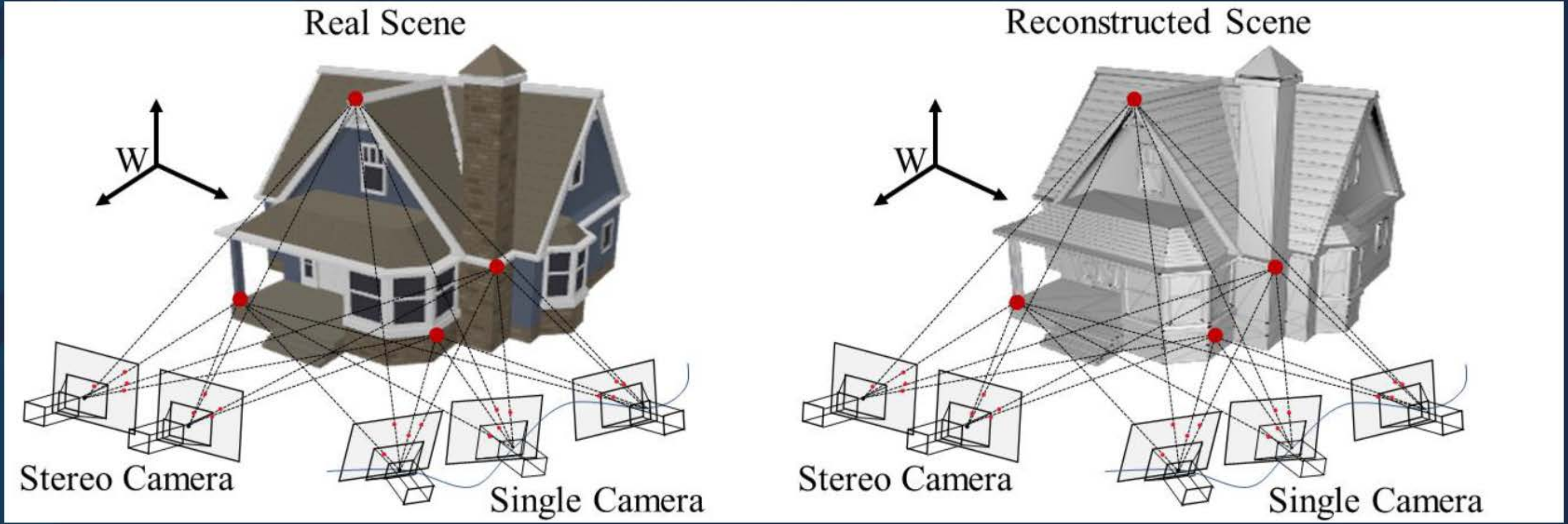
https://www.youtube.com/watch?v=0rc4RqYLtEU

Autonomous Driving Perception Tasks

1. 3D Geometry Model Reconstruction
2. Semantic Driving Scene Understanding
3. Object Detection/Recognition/Tracking

Autonomous Driving Perception Tasks

1. **3D Geometry Model Reconstruction**
2. Semantic Driving Scene Understanding
3. Object Detection/Recognition/Tracking



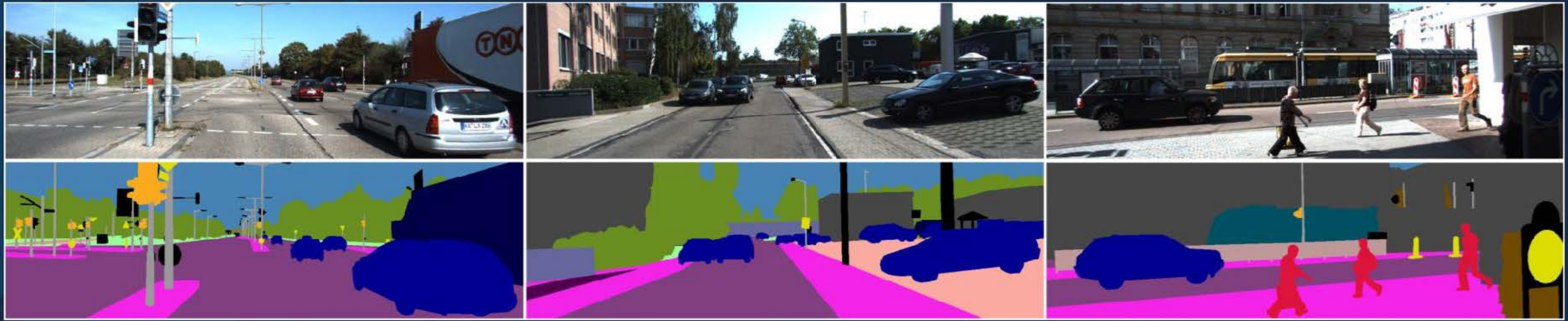
Autonomous Driving Perception Tasks

1. **3D Geometry Model Reconstruction**
2. Semantic Driving Scene Understanding
3. Object Detection/Recognition/Tracking

1.1. Traditional Dense Stereo
1.2. Self-Supervised Dense Stereo
1.3. Unsupervised Optical Flow Estimation

Autonomous Driving Perception Tasks

1. 3D Geometry Model Reconstruction
2. **Semantic Driving Scene Understanding**
3. Object Detection/Recognition/Tracking



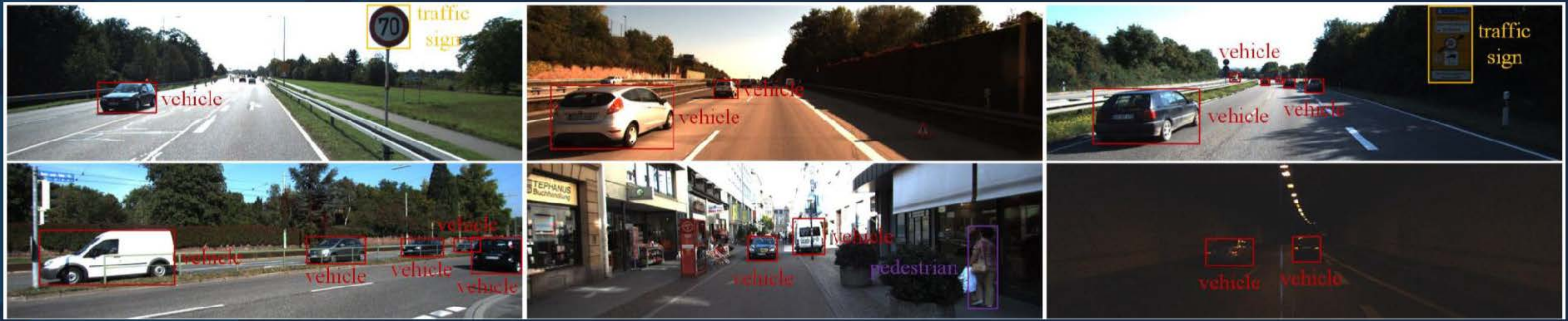
Autonomous Driving Perception Tasks

1. 3D Geometry Model Reconstruction
2. **Semantic Driving Scene Understanding**
3. Object Detection/Recognition/Tracking

2.1. Freespace Detection
2.2. Road Defect/Anomaly Detection

Autonomous Driving Perception Tasks

1. 3D Geometry Model Reconstruction
2. Semantic Driving Scene Understanding
3. **Object Detection/Recognition/Tracking**



Autonomous Driving Perception Tasks

1. 3D Geometry Model Reconstruction
2. Semantic Driving Scene Understanding
3. **Object Detection/Recognition/Tracking**

3.1. Intelligent Collaboration Among Air-Ground Robots for Parking Violation Detection

3.2. Lane Marking Detection

Keynote Talk Outline

1. Introduction
2. **3D Geometry Model Reconstruction**
3. Semantic Driving Scene Understanding
4. Object Detection/Recognition
5. Conclusion

3D Geometry Model Reconstruction

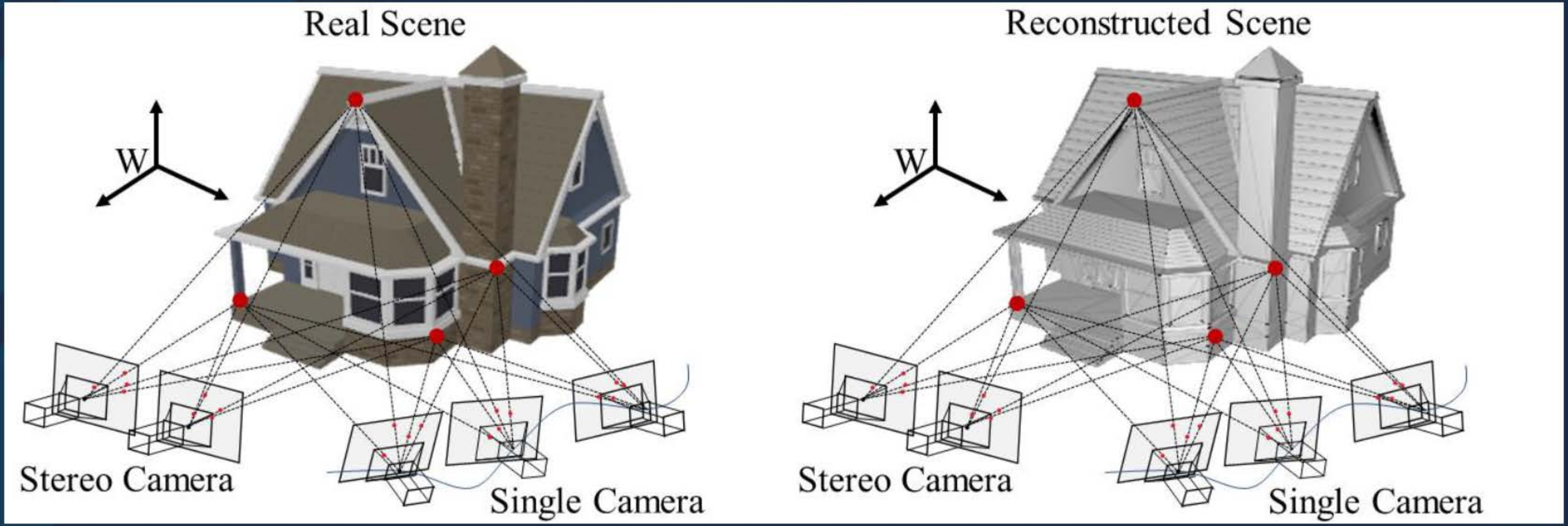
2D images

3D Geometry Model Reconstruction

2D images captured from different views

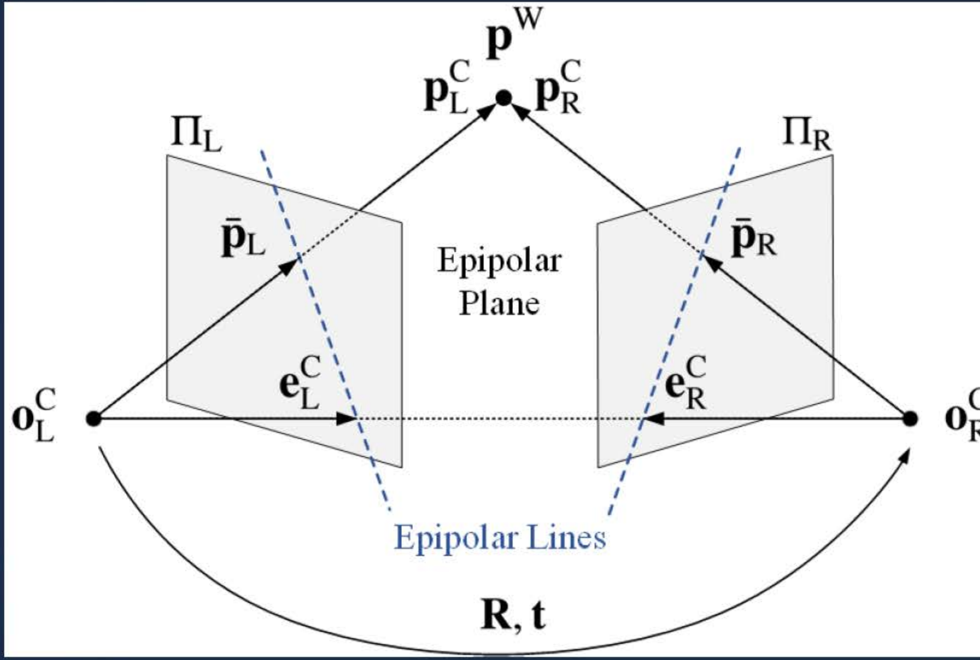
3D Geometry Model Reconstruction

- The geometry of a given 3D model can be reconstructed using either **an array of synchronized cameras (stereo vision)** or **a single movable camera (structure from motion, or optical flow)**.



Stereo Vision

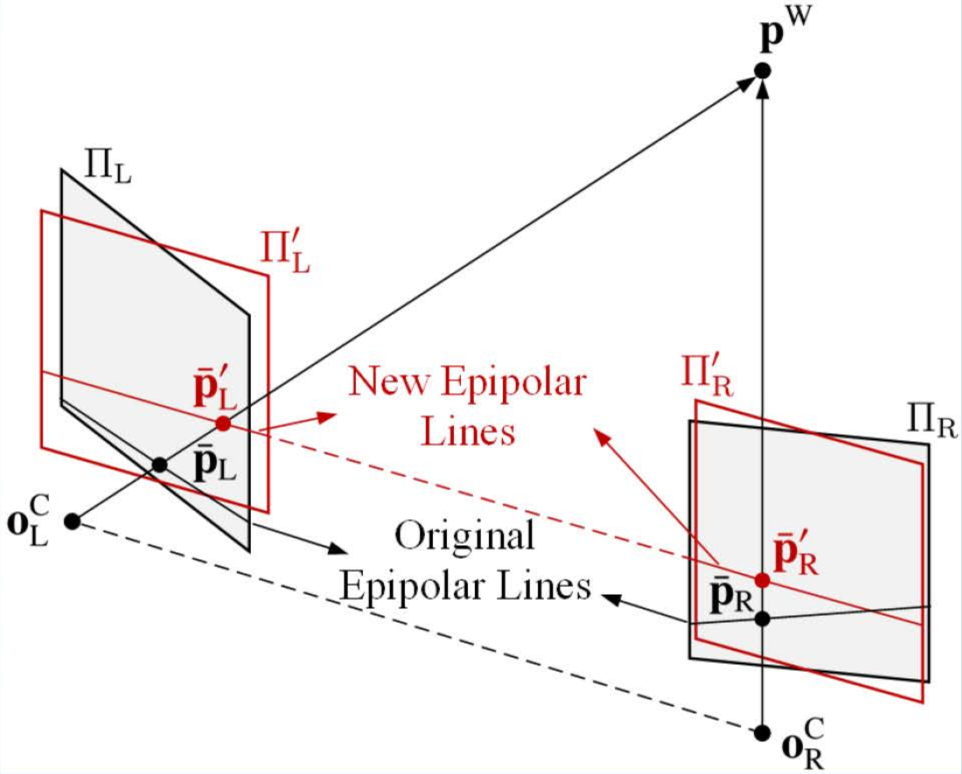
- 3D scene geometry reconstruction with a pair of synchronized cameras is based on determining pairs of correspondence pixels between the left and right images.
- For an uncalibrated stereo rig, finding the correspondence pairs is a 2D search process (***e.g.*, using optical flow estimation**), which is extremely computationally intensive.



Epipolar geometry

Stereo Vision

- If the stereo rig is calibrated, **1D search** should be performed along the epipolar lines.
- An image transformation process, referred to as **stereo rectification**, is always performed beforehand to reduce the dimension of the correspondence pair search.



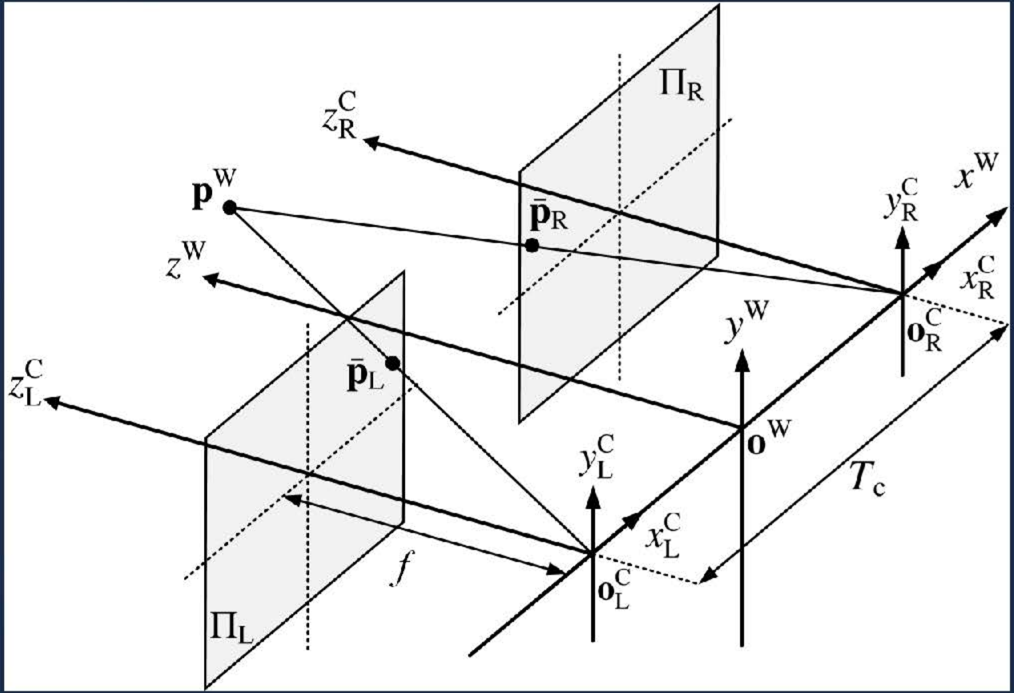
Stereo rectification

Stereo Vision

- Disparity is in inverse proportion to depth:

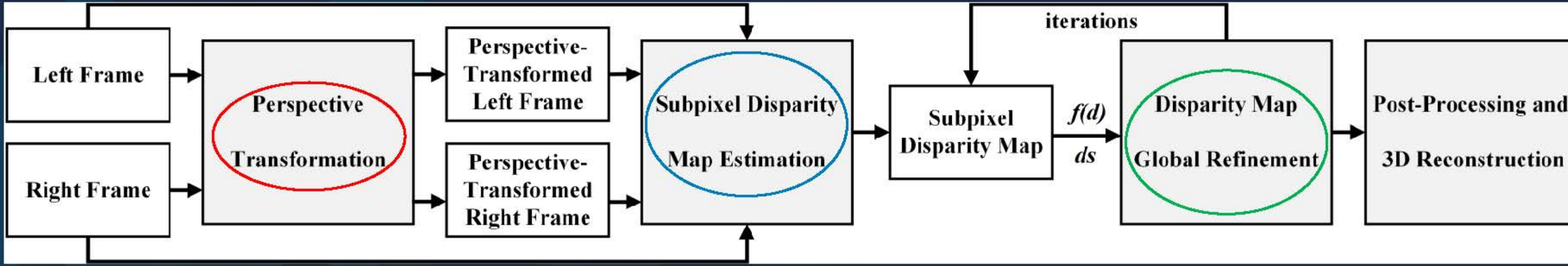
$$d = u_L - u_R = \frac{fT_c}{z^w}$$

- Using the camera intrinsic matrix, we can obtain the 2D pixel's 3D location in the World Coordinate System (WCS).



Basic stereo vision system

Traditional Dense Stereo



- In 2018, we proposed a novel dense subpixel disparity estimation algorithm [1] for dense road surface 3D reconstruction.
- The main contributions of this work include: 1) perspective transformation, 2) subpixel disparity map estimation, and 3) disparity map global refinement.

[1] **Fan, R.**, Ai, X. and Dahnoun, N., 2018. Road surface 3D reconstruction based on dense subpixel disparity map estimation. ***IEEE Trans on Image Processing***, 27(6), pp.3025-3035.

Traditional Dense Stereo

- The perspective transformation algorithm transforms the target image into the reference view, which enables better stereo matching in terms of both speed and accuracy.

Original left image

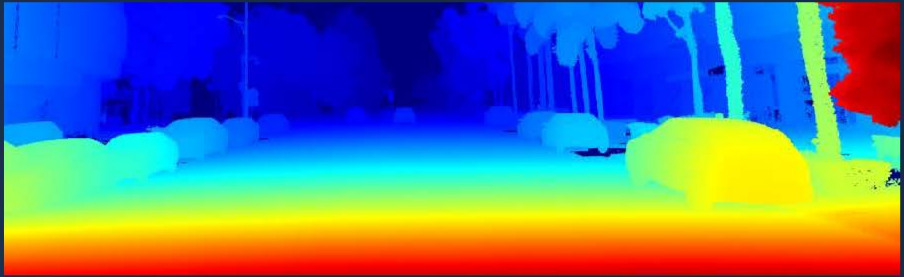
Original right image

Transformed left image

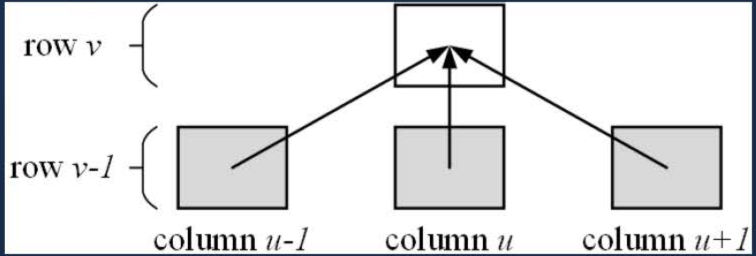
Transformed right image

Traditional Dense Stereo

- The road disparities decrease gradually from the bottom to the top, while the disparities of obstacles remain the same.
- Therefore, the disparities are then estimated iteratively, where the search range is propagated from three estimated neighboring disparities.



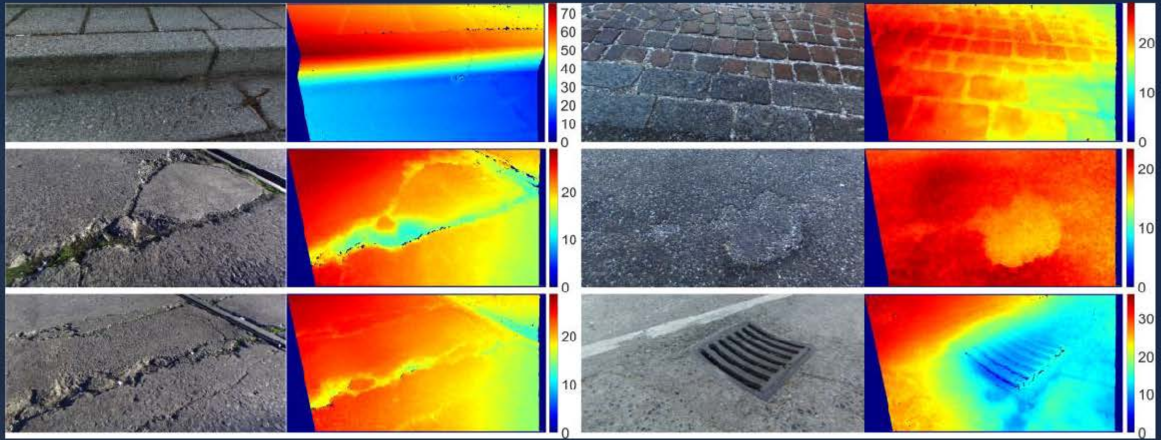
Road disparity image example



Iterative disparity estimation strategy

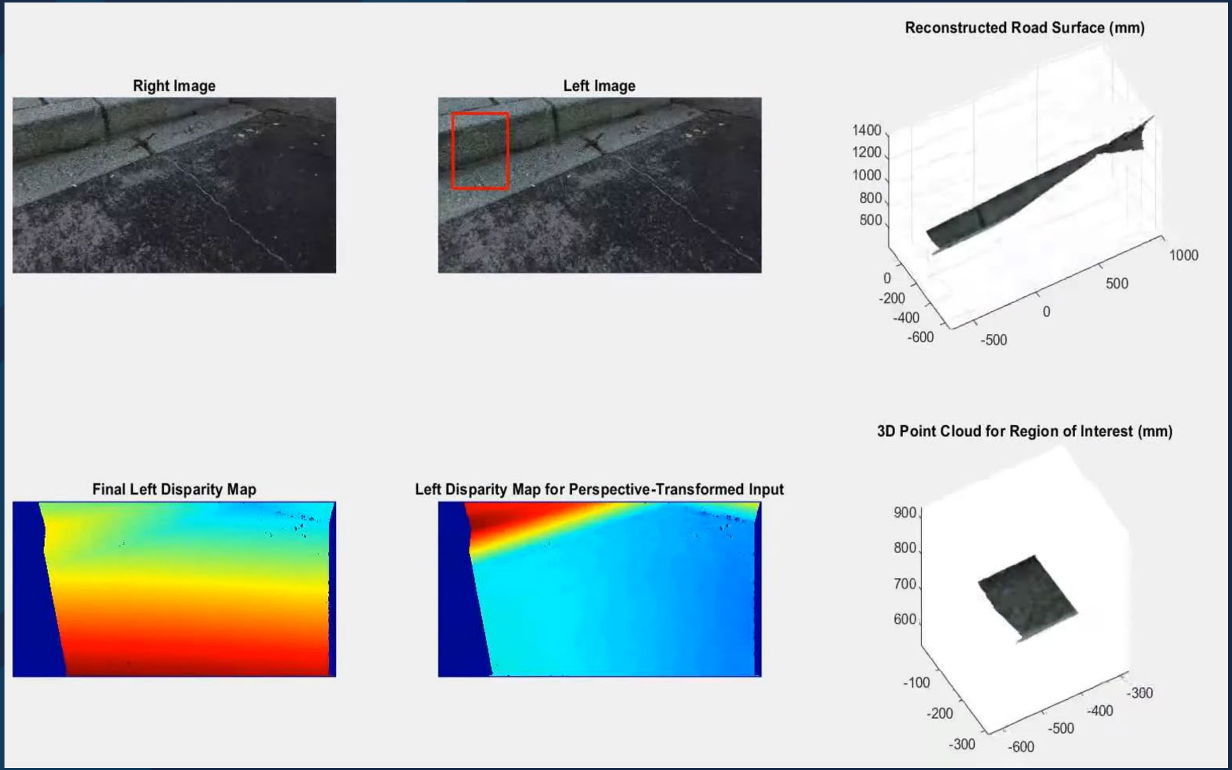
Traditional Dense Stereo

- A novel disparity global refinement approach developed from the Markov random fields (MRF) and fast bilateral stereo is introduced to further improve the accuracy of the estimated disparity map.



Left: RGB images
Right: disparity images

Traditional Dense Stereo



Reconstruction accuracy: 3mm

Our paper is downloadable!

https://www.youtube.com/watch?v=pypPI7fsctg

Traditional Dense Stereo

Real-Time Dense Stereo Embedded in A UAV for Road Inspection

Rui Fan, Jianhao Jiao, Jie Pan,

Huaiyang Huang, Shaojie Shen, Ming Liu

CVPR LONG BEACH CALIFORNIA June 16-20, 2019

香港科技大學 THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

香港科大機器人研究所 ROBOTICS INSTITUTE HKUST

This system was improved and embedded on an NVIDIA Jetson TX2 GPU.

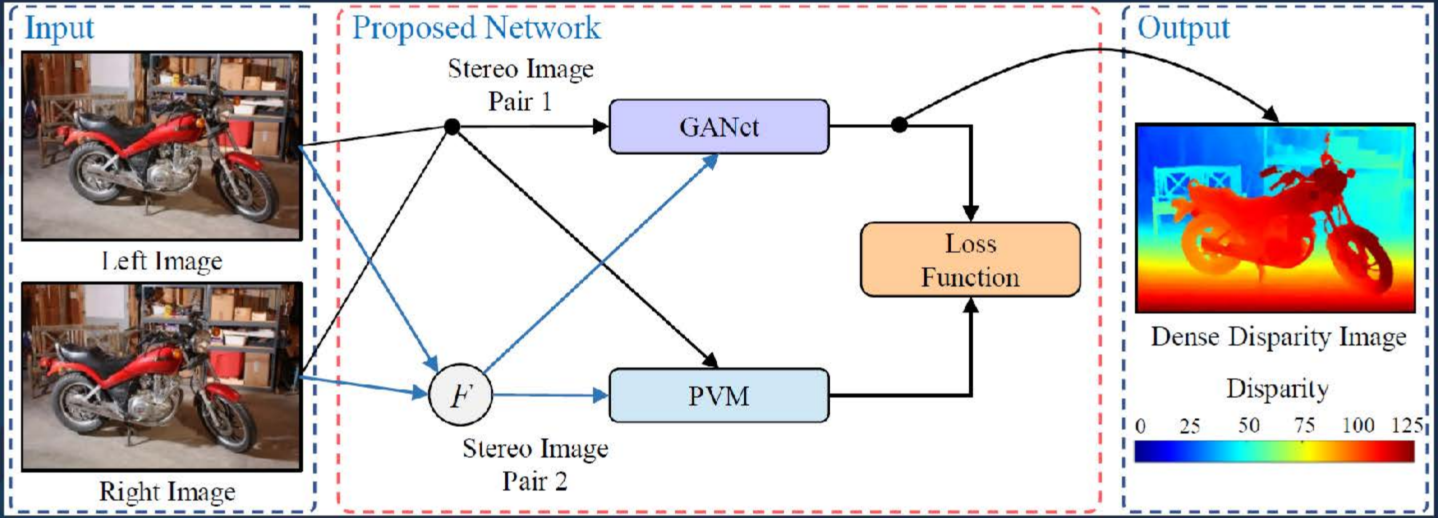
Our paper is downloadable!

https://www.youtube.com/watch?v=_-YmlxojVMI&t=61s

Self-Supervised Dense Stereo

- The existing stereo matching approaches are classified as either traditional or data-driven ones.
- The former generally formulate stereo matching as block matching (local methods) or energy minimization (global methods) problems.
- The latter typically employ data-driven classification and/or regression models, *e.g.*, convolutional neural networks (CNNs), to learn a feasible solution for stereo matching.
- With recent advances in deep learning, many researchers have resorted to deep CNNs (DCNNs) for stereo matching.
- However, these approaches generally require a large amount of hand-labeled training data to learn the best DCNN parameters.

Self-Supervised Dense Stereo



- We propose a novel approach for self-supervised stereo matching. Specifically, we develop a module named Pyramid Voting Module (PVM), which can be deployed in any existing supervised stereo matching DCNN, converting it into a self-supervised approach.

Self-Supervised Dense Stereo

The Demo of the KITTI Raw Data [1]

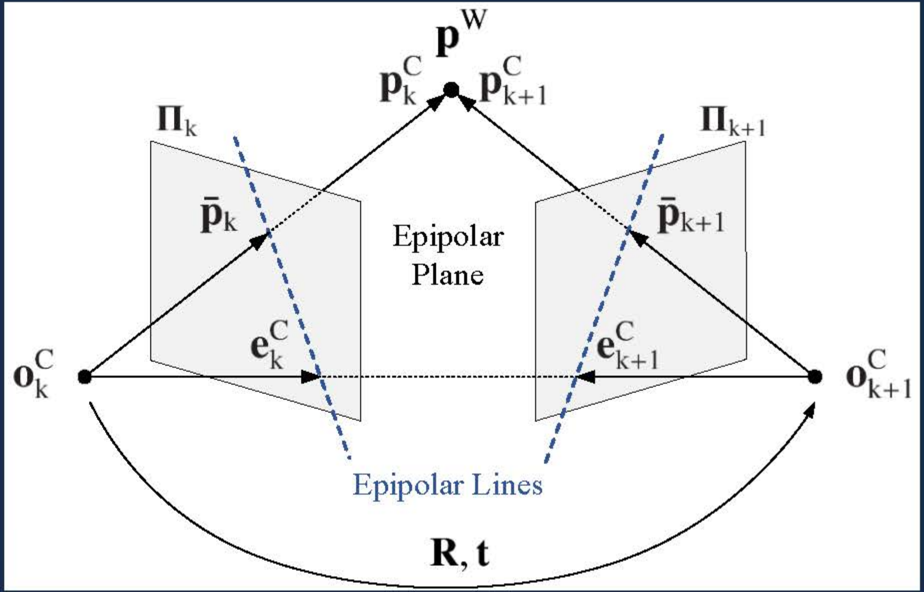
[1] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In 2012 IEEE Conference on Computer Vision and Pattern Recognition, pages 3354–3361. IEEE, 2012.

- The state-of-the-art self-supervised stereo algorithm.
- Ranked the 3rd on the KITTI benchmark.

Unsupervised Optical Flow Estimation

- Optical flow describes the motion of pixels between consecutive frames of a video sequence.
- Optical flow is a 2-channel visual information:
- 1st channel (F_u): horizontal positional difference;
- 2nd channel (F_v): vertical positional difference.

$$\begin{cases} F_u = u^k - u^{k-1} \\ F_v = v^k - v^{k-1} \end{cases}$$



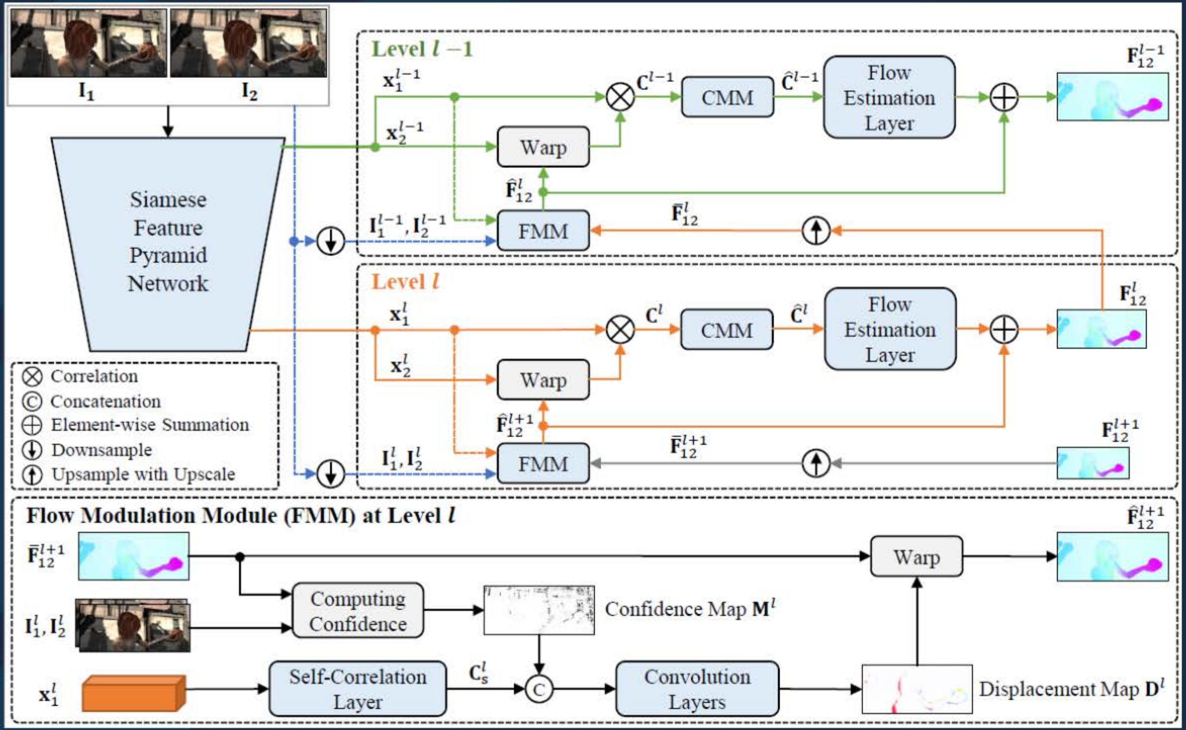
Epipolar geometry

Unsupervised Optical Flow Estimation

- In 2020, we proposed CoT-AMFlow [2], an unsupervised optical flow estimation approach.
- In terms of the network architecture, we develop an adaptive modulation network to remove outliers in challenging regions.
- As for the training paradigm, we adopt a co-teaching strategy, where two networks simultaneously teach each other about challenging regions to further improve accuracy.

[2] Wang, H., **Fan, R.** and Liu, M., 2020. CoT-AMFlow: Adaptive modulation network with co-teaching strategy for unsupervised optical flow estimation. *CoRL 2020*. (Acceptance rate: 34%)

Unsupervised Optical Flow Estimation



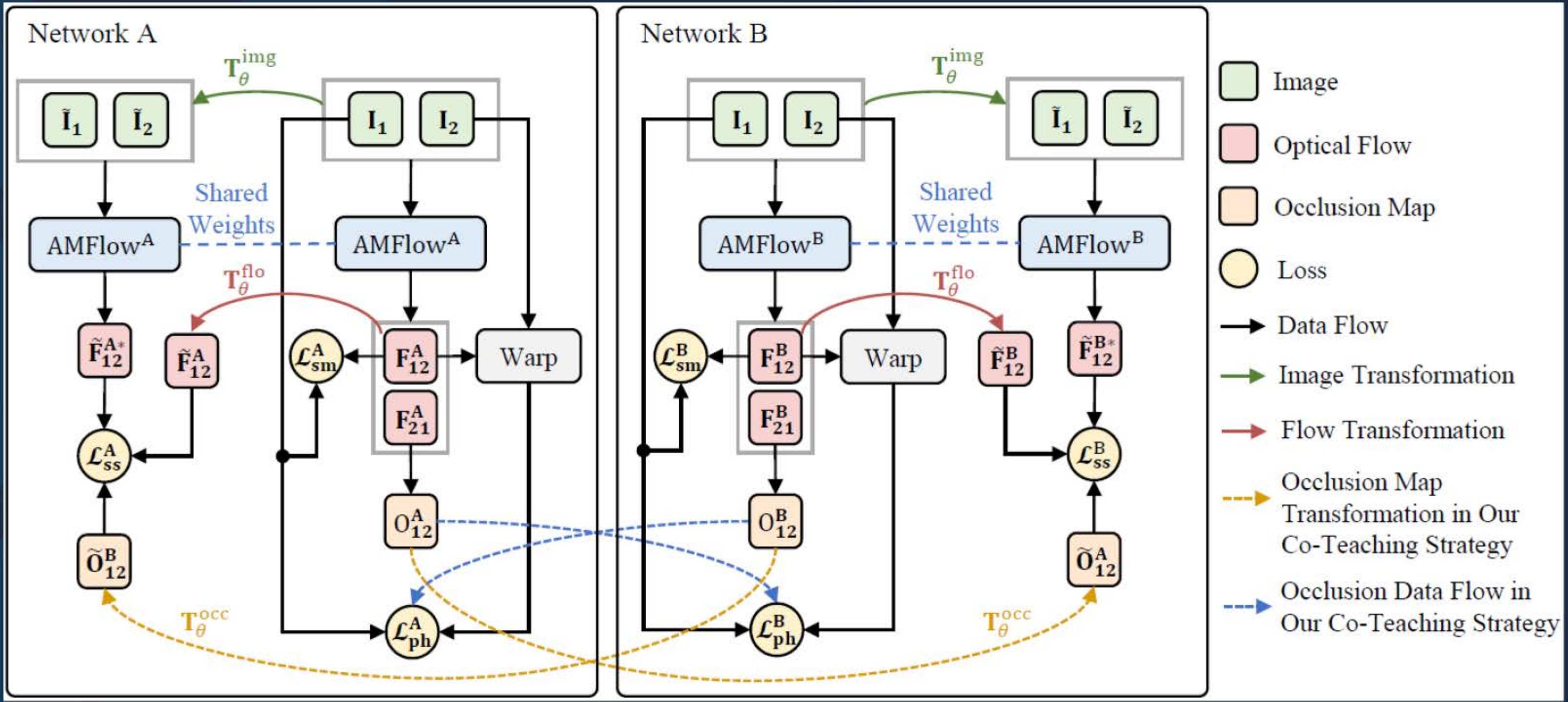
AMFlow for self-supervised optical flow estimation

- FMMs refine the flow initialization from the preceding pyramid level using local flow consistency.
- CMMs explicitly reduce outliers in the cost volume using a flexible and efficient sparse point-based scheme.

Rui Ranger Fan

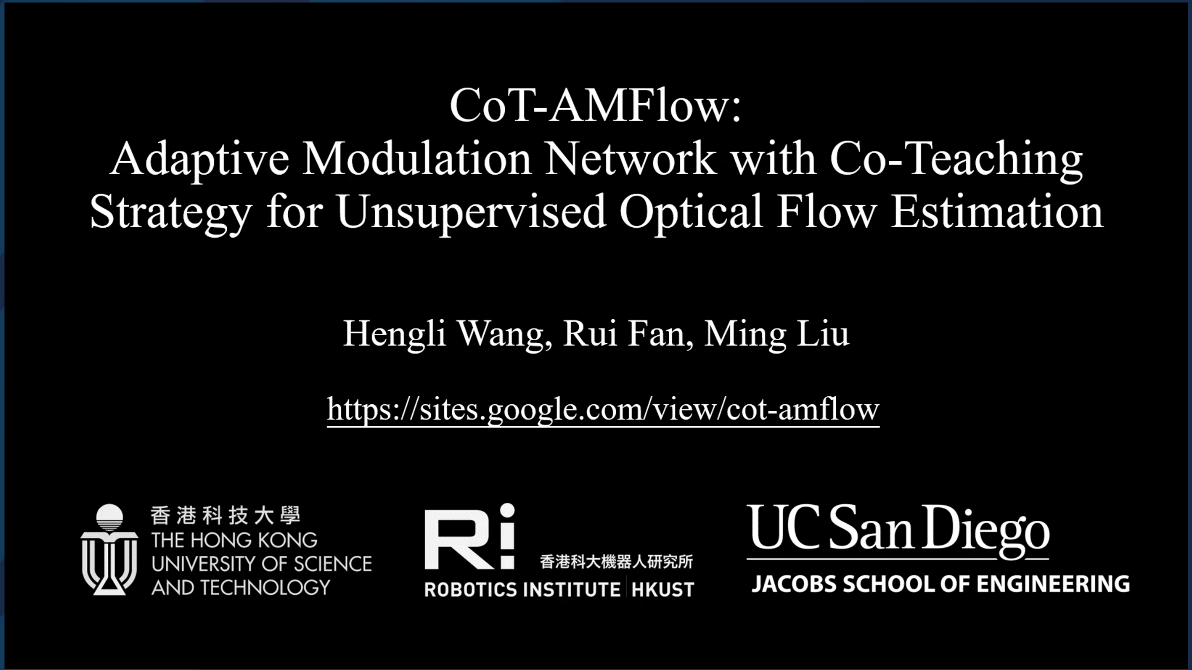
UC San Diego

Unsupervised Optical Flow Estimation



Co-Teaching Training Strategy

Unsupervised Optical Flow Estimation



https://www.youtube.com/watch?v=LzL7QZhwFjE&feature=youtu.be

CoT-AMFlow outperforms all other unsupervised methods on the MPI Sintel, KITTI Flow 2012/2015 and Middlebury Flow datasets.

Our paper is downloadable!

Keynote Talk Outline

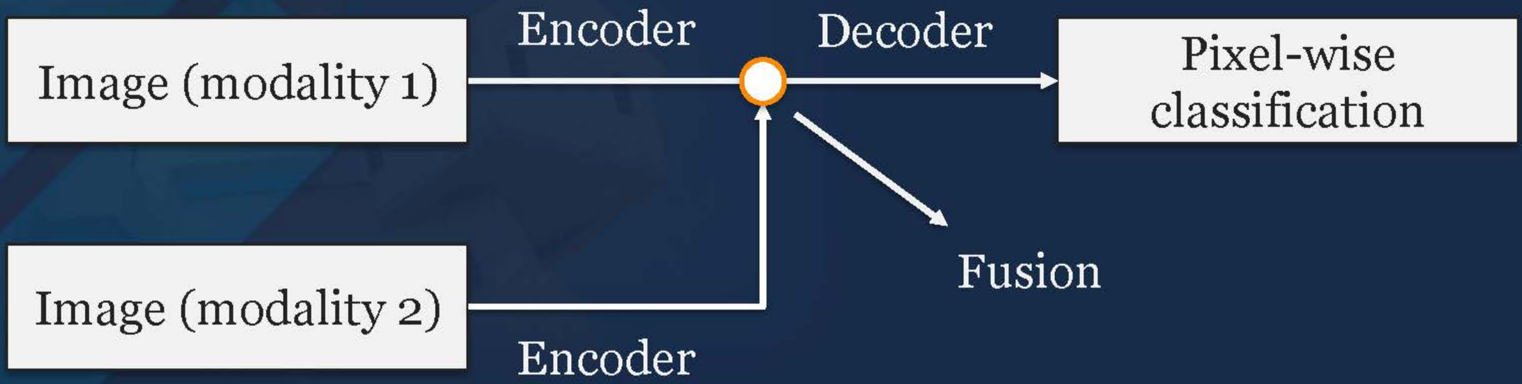
1. Introduction
2. 3D Geometry Model Reconstruction
3. **Semantic Driving Scene Understanding**
4. Object Detection/Recognition
5. Conclusion

Semantic Driving Scene Understanding

- Semantic Segmentation CNNs can be categorized as two groups: 1) single-modal and 2) data-fusion.
- The former typically segments RGB images with an encoder-decoder CNN architecture. In recent years, many popular single-model semantic image segmentation algorithms, such as Fully Convolutional Network (FCN), U-Net, SegNet, DeepLabv3+, DenseASPP, DUpsampling, *etc.*, have been proposed.

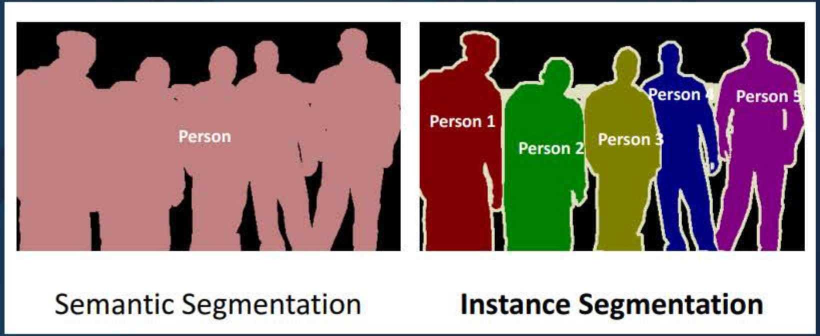
Semantic Driving Scene Understanding

- Data-fusion semantic image segmentation approaches generally learn features from two different types of vision data, such as RGB and depth images in FuseNet, RGB and surface normal image in SNE-RoadSeg, *etc.*



Semantic Driving Scene Understanding

- **Semantic segmentation:** Objects shown in an image are grouped based on defined categories.
- **Instance segmentation:** Consider instance segmentation a refined version of semantic segmentation — instance segmentation detects the instances of each category.



Semantic segmentation
v.s.
instance segmentation

Freespace Detection

- Freespace detection is an essential component of visual perception for self-driving cars.
- Freespace detection can be formatted as a binary semantic driving scene segmentation problem.
- Freespace detection approaches generally classify each pixel in an RGB or depth/disparity image as drivable or undrivable.
- Such pixel-level classification results are then utilized by other modules in the autonomous system, such as trajectory prediction and path planning, to ensure that the autonomous car can navigate safely in complex environments.

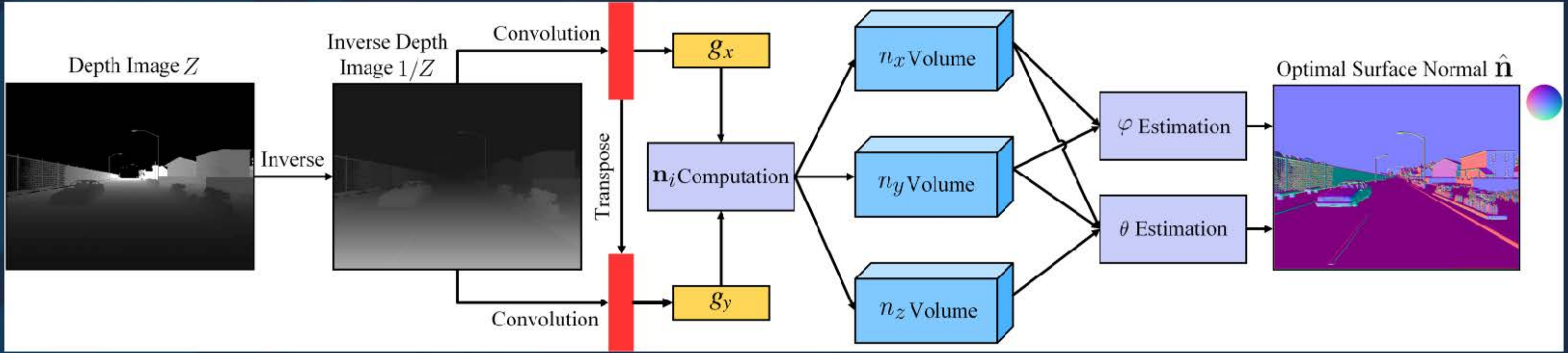
Freespace Detection

- In 2020, we proposed a novel freespace detection approach, referred to as **SNE-RoadSeg** [3].
- It consist of 1) a novel module, named surface normal estimator (SNE), which can infer surface normal information from dense depth/disparity images with high accuracy and efficiency; and 2) a data-fusion CNN architecture, referred to as RoadSeg, which can extract and fuse features from both RGB images and the inferred surface normal information for accurate freespace detection.

[3] **Fan, R.**, Wang, H., Cai, P. and Liu, M., 2020, August. SNE-RoadSeg: Incorporating surface normal information into semantic segmentation for accurate freespace detection. In ECCV (pp. 340-356). Springer, Cham.

Freespace Detection

- Our SNE module can efficiently produce a surface normal map from a depth/disparity image in an end-to-end way. This module is developed based on our previously proposed surface normal estimator, named three-filters-to-normal (3F2N) [4].

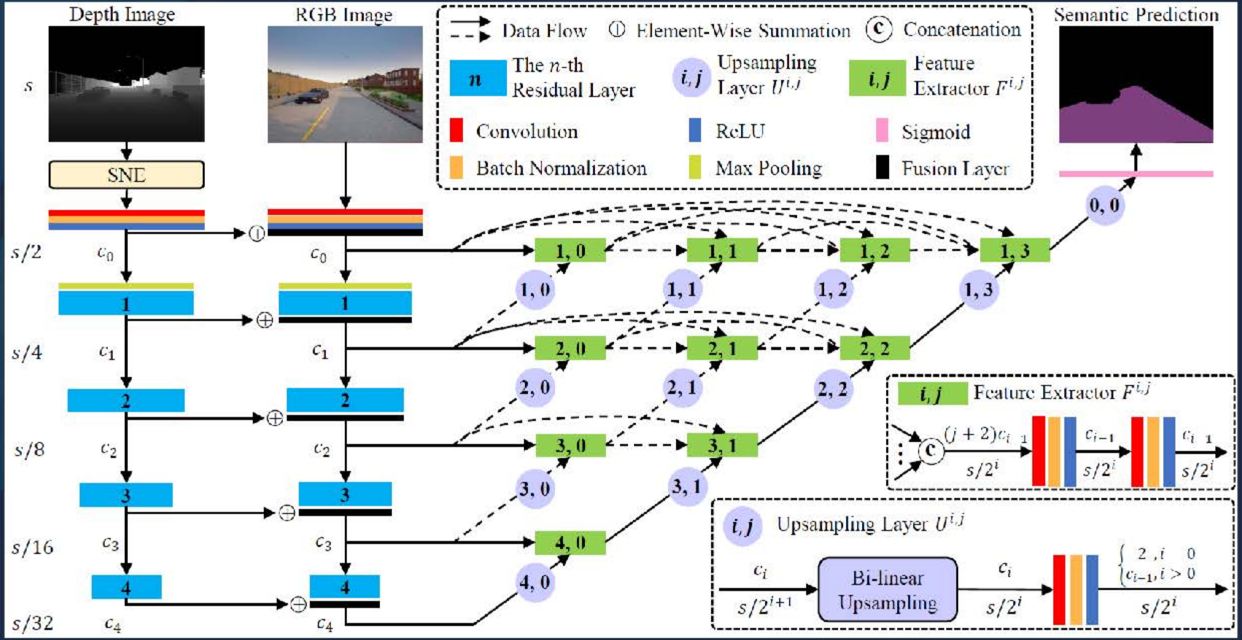


SNE

[4] Fan, R., Wang, H., Xue, B., Huang, H., Wang, Y., Liu, M. and Pitas, I., 2020. Three-Filters-to-Normal: An Accurate and Ultrafast Surface Normal Estimator. *arXiv preprint arXiv:2005.08165.*

Freespace Detection

- RoadSeg incorporates both RGB and surface normal information into semantic segmentation for accurate freespace detection.



RoadSeg Architecture

Freespace Detection

SNE-RoadSeg: Incorporating Surface Normal Information into Semantic Segmentation for Accurate Freespace Detection

Rui Fan*, Hengli Wang*, Peide Cai, Ming Liu

sites.google.com/view/sne-roadseg

ECCV'20

UC San Diego

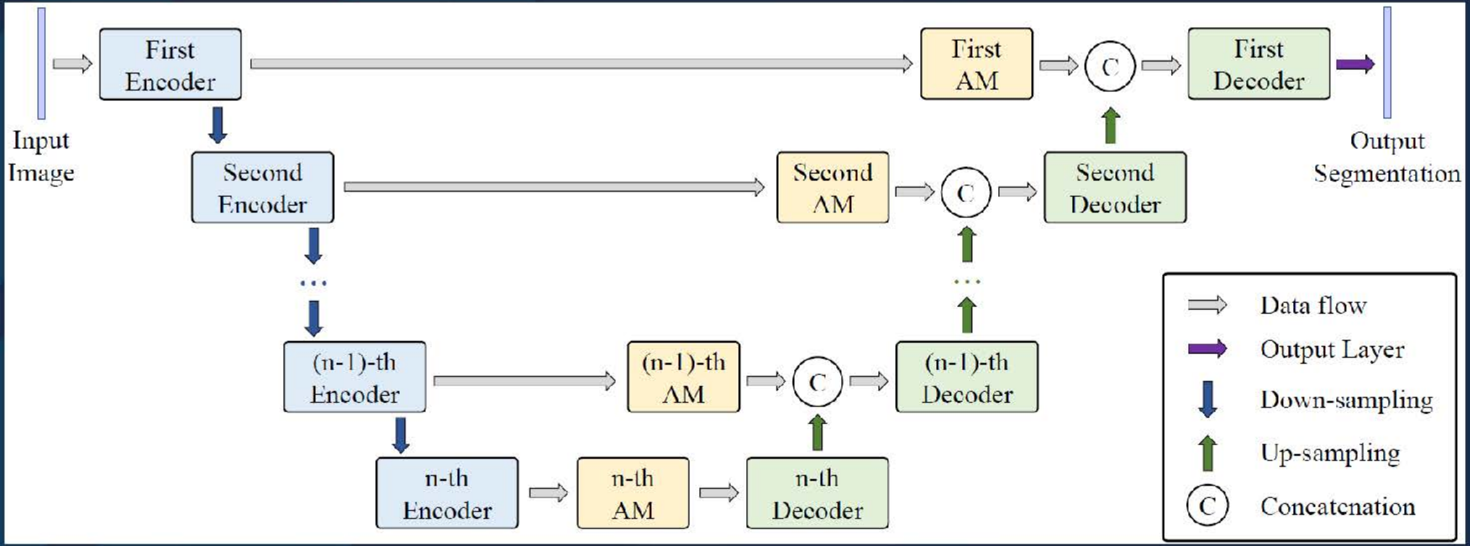
香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

SNE-RoadSeg was ranked the 2nd on the KITTI road benchmark.

Our paper is downloadable!

https://www.youtube.com/watch?v=wWrZhDuh6xc

Freespace Detection

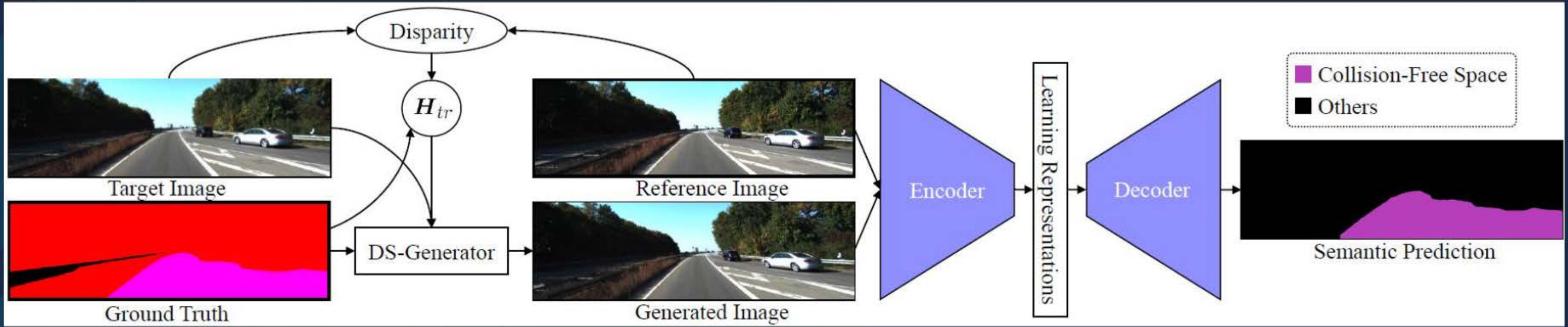


Our paper is downloadable!

Incorporating attention modules can further improve the semantic segmentation accuracy [5].

[5] Fan, R., Wang, H., Bocus, M.J. and Liu, M., 2020. We learn better road pothole detection: from attention aggregation to adversarial domain adaptation. ECCV Workshop 2020.

Freespace Detection

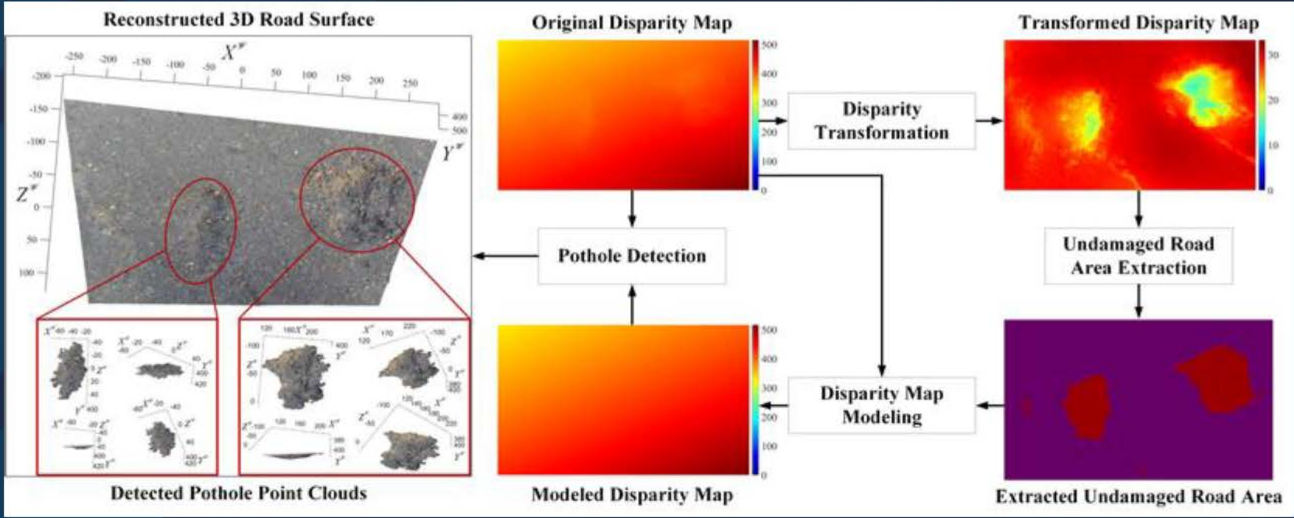


Using images captured from multiple views can also improve the accuracy of freespace detection [6].

[6] Fan, R., Wang, H., Cai, P., Wu, J., Bocus, M.J., Qiao, L. and Liu, M., 2020. Learning collision-free space detection from stereo images: Homography matrix brings better data augmentation. IEEE Trans on Mechtronics (to be published).

Road Defect/Anomaly Detection

- To ensure traffic safety, it is crucial and necessary to frequently inspect and repair road potholes
- In 2019, we proposed a novel road pothole detection system [7].

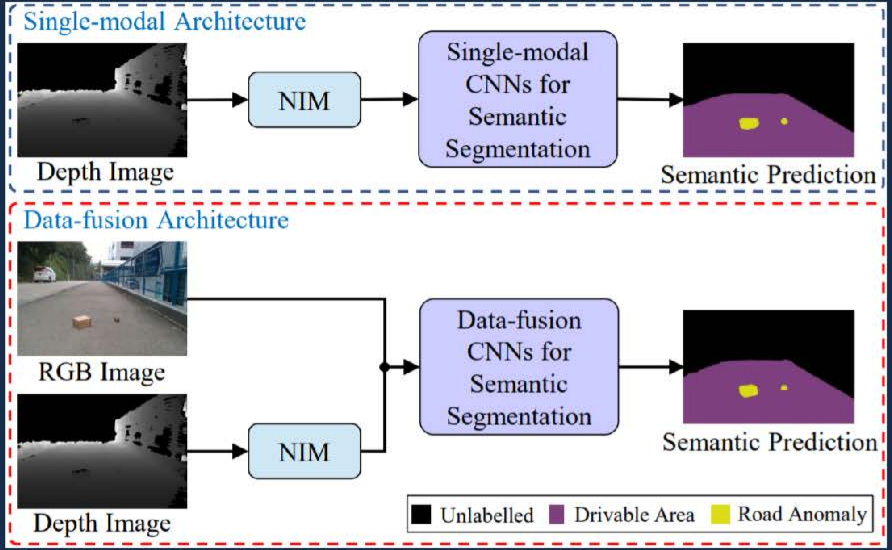


Our paper is downloadable!

[7] Fan, R., Ozgunalp, U., Hosking, B., Liu, M. and Pitas, I., 2019. Pothole detection based on disparity transformation and road surface modeling. IEEE Trans on Image Processing.

Road Defect/Anomaly Detection

- In IROS 2020, we also presented a robust road defect anomaly detection algorithm [8], designed for mobile robots, such as intelligent wheelchairs.
- Such a road anomaly detection algorithm can help navigation system plan a safe trajectory for the mobile robot.



Single-modal and data-fusion road anomaly detection

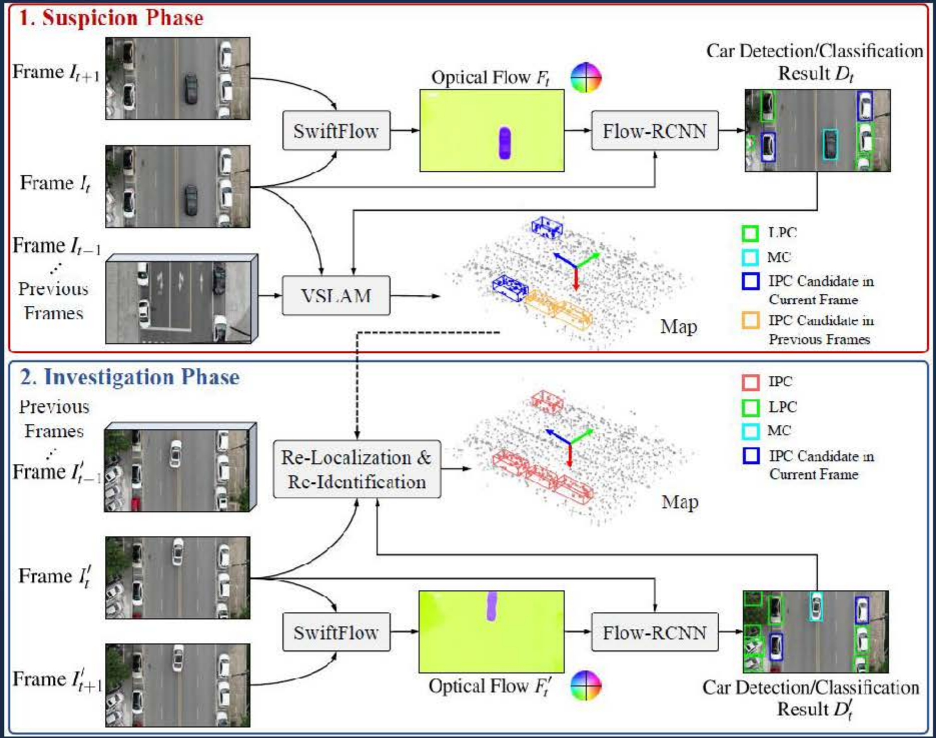
[8] Wang, H.*, Fan, R.*, Sun, Y. and Liu, M., 2020. Applying surface normal information in drivable area and road anomaly detection for ground mobile robots. *IROS 2020*.

Keynote Talk Outline

1. Introduction
2. 3D Geometry Model Reconstruction
3. Semantic Driving Scene Understanding
4. **Object Detection/Recognition**
5. Conclusion

Intelligent Collaboration Among Air-Ground Robots for Parking Violation Detection

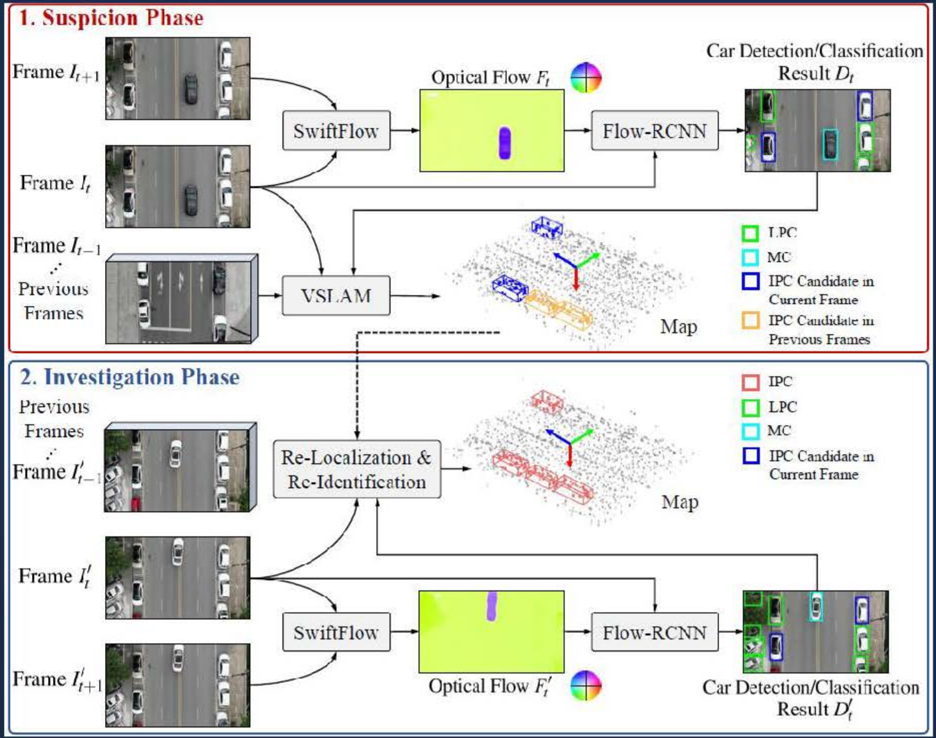
- In 2020, we introduce a novel suspect-and-investigate framework [10], which can be easily embedded in a drone for automated parking violation detection (PVD).



[10] Wang, H., Liu, Y., Huang, H., Pan, Y., Yu, W., Jiang, J., Lyu, D., Bocus, M.J., Liu, M., Pitas, I. and Fan, R., 2020. ATG-PVD: Ticketing parking violations on a drone. ECCV Workshop 2020.

Intelligent Collaboration Among Air-Ground Robots for Parking Violation Detection

- Our proposed framework consists of: 1) SwiftFlow, an efficient and accurate convolutional neural network for unsupervised optical flow estimation; 2) Flow-RCNN, a flow-guided CNN for car detection and classification; and 3) an illegally parked car (IPC) candidate investigation module developed based on visual SLAM.



Intelligent Collaboration Among Air-Ground Robots for Parking Violation Detection

ATG-PVD: Ticketing Parking Violations on A Drone

Hengli Wang[1*], Yuxuan Liu[1*], Huaiyang Huang[1*], Yuheng Pan[2*], Wenbin Yu[2], Jialin Jiang[2], Dianbin Lyu[2], Mohammud J. Bocus[3], Ming Liu[1], Ioannis Pitas[4], and Rui Fan[2,5]

[1] HKUST Robotics Institute
[2] ATG Robotics
[3] University of Bristol
[4] Aristotle University of Thessaloniki
[5] UC San Diego

sites.google.com/view/atg-pvd

Our paper is downloadable!

https://www.youtube.com/watch?v=byfa7-fT4j4

Lane Marking Detection

- In 2016, we developed a stereo vision-based multiple lane marking detection algorithm [9].
- This algorithm can estimate multiple dense vanishing points from disparity maps acquired using a stereo camera.
- By optimizing two accumulators with respect to the horizontal and vertical coordinates of the vanishing points using dynamic programming (DP), we can obtain a vanishing point model (VPM). The lane markings can then be detected using the information of VPM.

Our paper is downloadable!

[9] Ozgunalp, U., Fan, R., Ai, X. and Dahnoun, N., 2016. Multiple lane detection algorithm based on novel dense vanishing point estimation. *IEEE Trans on Intelligent Transportation Systems*, *18*(3), pp.621-632.

Keynote Talk Outline

1. Introduction
2. 3D Geometry Model Reconstruction
3. Semantic Driving Scene Understanding
4. Object Detection/Recognition
5. **Conclusion**

Conclusion

- Recent AI technologies, such as deep learning, have greatly enhanced machine vision algorithms for driving scene understanding.
- The combination of CNNs and traditional computer vision algorithms provides a feasible solution to un/self-supervised driving scene understanding, as labeled training data are no longer required.
- With better driving scene understanding outputs, obtained by data-driven autonomous car perception techniques, other subsystems, such as location & mapping, navigation, and path planning, can also be significantly improved.

Thank you very much for your attention!

Rui Ranger Fan

Q & A

UC San Diego