



Lecture 1

Introduction to Computer Vision

Lin ZHANG, PhD
School of Software Engineering,
Tongji University
Fall 2023



Course Info

Contact Information

Room 418L, Jishi Building

Email: cslinzhang@tongji.edu.cn

TA: Linfei LI, wechat: lifambitious

Course information can be found at

<http://cslinzhang.gitee.io/home/>



Materials

- Major materials
 - My slides
- References
 - 张林等, 计算机视觉 : 原理算法与实践 (草稿)
 - Richard Hartley and Andrew Zisserman, Multiple View Geometry in Computer Vision (2nd Edition), Cambridge University Press, 2004
 - D.A. Forsyth and J. Ponce, Computer Vision: A Modern Approach (2nd Edition), Pearson Education, Inc., 2013
 - Milan Sonka, Vaclav Hlavac, and Roger Boyle, Image Processing, Analysis, and Machine Vision, Thomson, 2008



Examination

- Homework 30%: 3 times, and each time 10%.
- Project 20%: 2 or 3 people for one group
- Final paper exam: 40%
- Attendance: 10%; Being absent $\geq 1/3$ lectures, you will fail this course
- Bonus 5%: being active in class and answering my questions correctly



Today

- What is computer vision?
- Why do we need to study CV?
- Course overview



What is vision?

“The plain man’s answer (and Aristotle’s too) would be, to know what is where by looking. In other words, vision is the process of discovering from images what is present in the world, and where it is ” *David Marr*, Vision: A Computational Investigation into the Human Representation and Processing of Visual Information, 1982

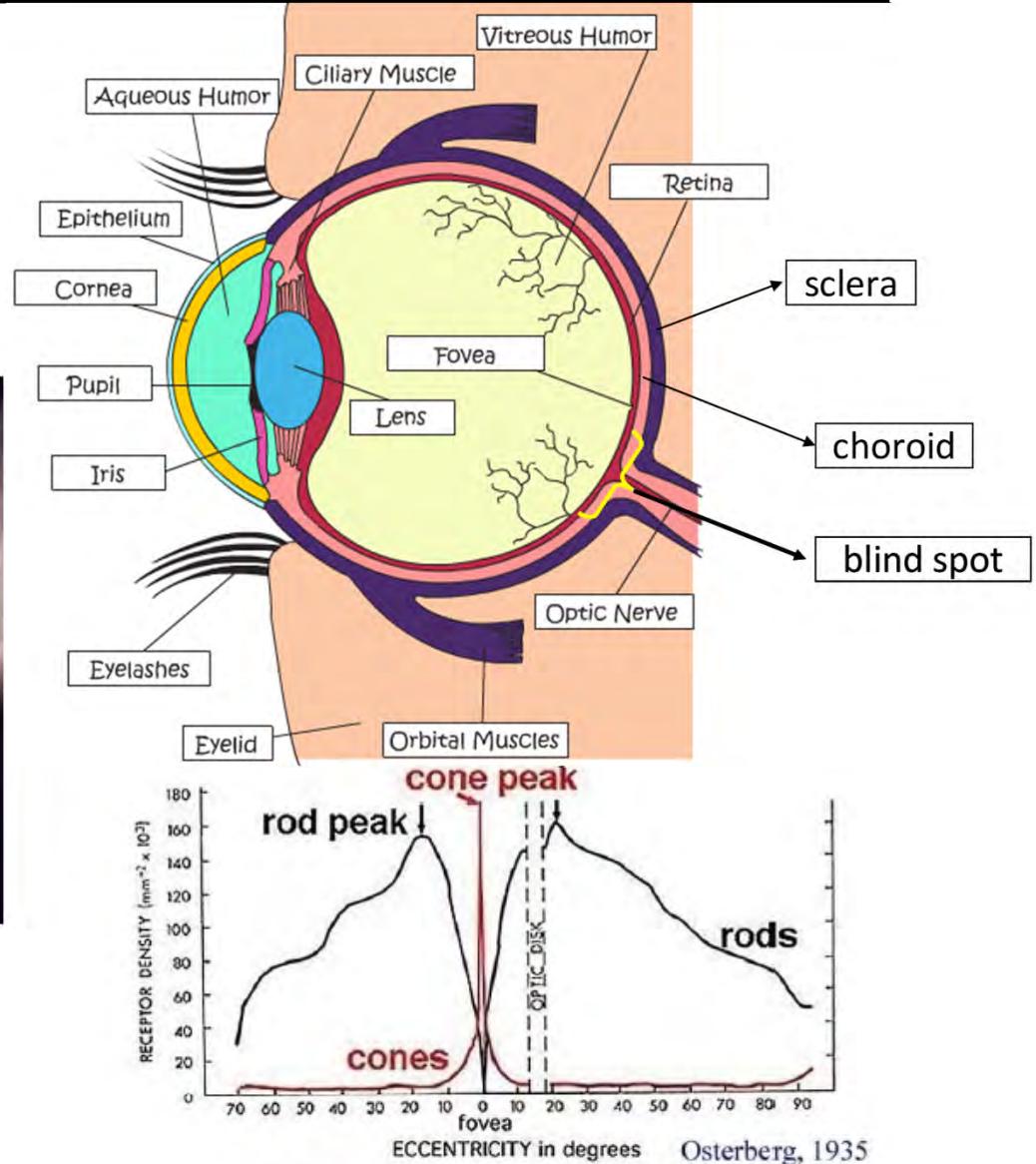
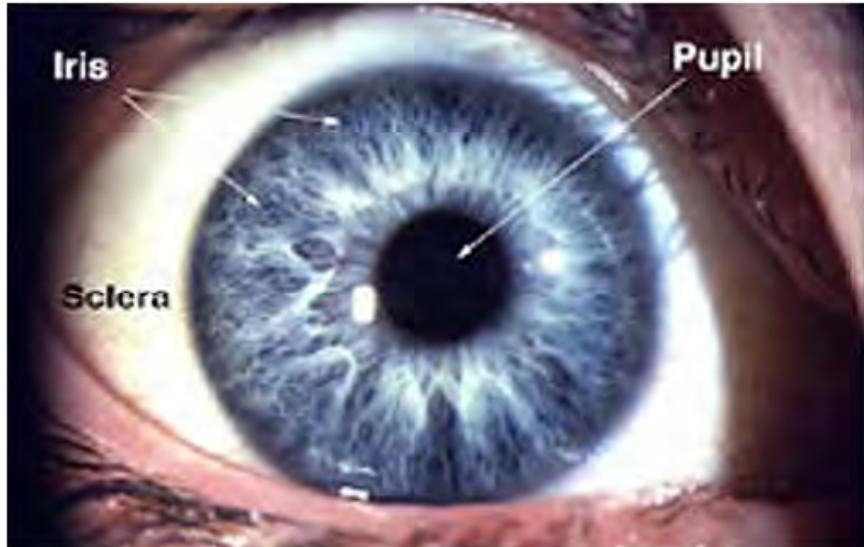


David Marr (1945.1.19 – 1980.11.17), was a British neuroscientist and psychologist. The Marr Prize, one of the most prestigious awards in computer vision, is named in his honor.



Human vision

Our Eyes

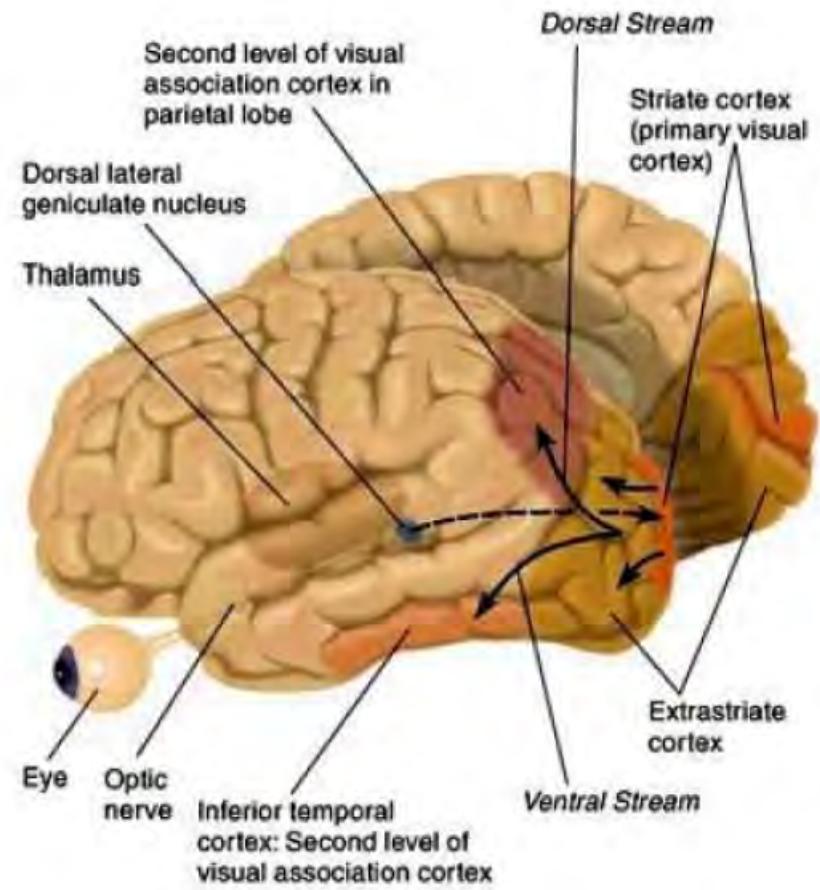




Human vision

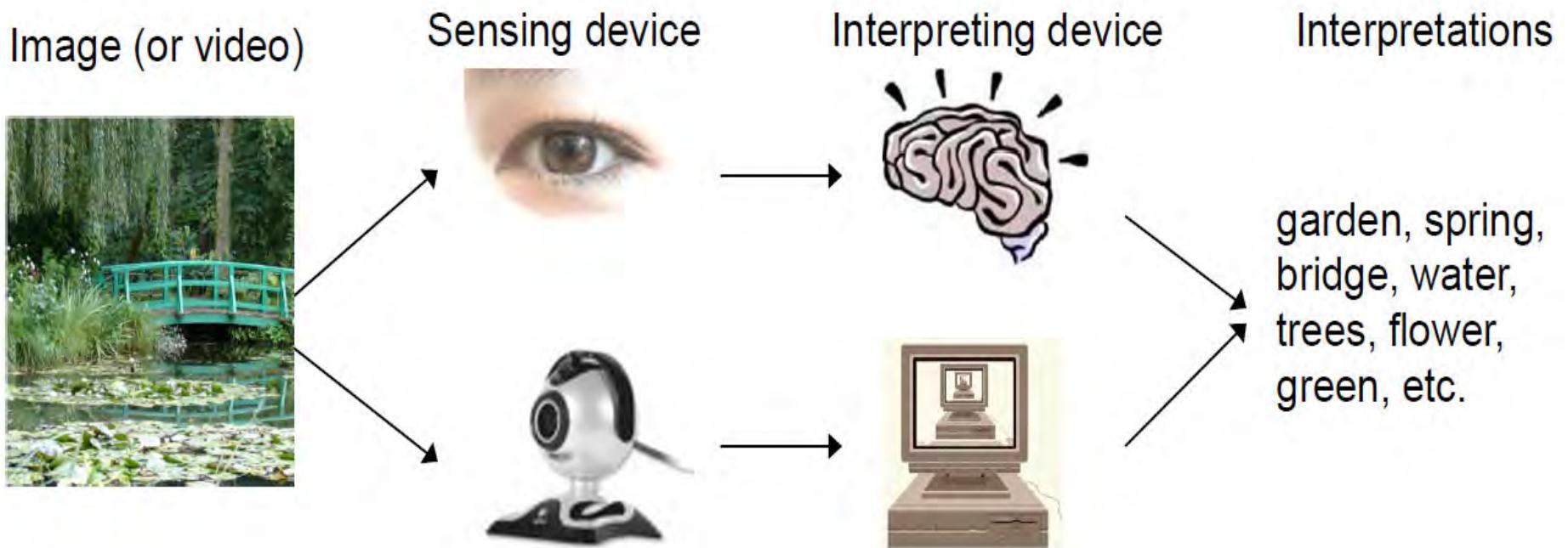
Human Visual System (HVS)

- Optical Receptors
 - Image formation
- Visual Pathway
 - Encoding
 - Representation
- Primary Visual Cortex
 - Interpretation





What is computer vision?





What is computer vision?

- Computer Vision has a dual goal^[1]
 - From the biological science point of view, computer vision aims to come up with computational models of the human visual system
 - **From the engineering point of view, computer vision aims to build autonomous systems which could perform some of the tasks which the human visual system can perform (and even surpass it in many cases) (our focus in this course!)**
 - Of course, the two goals are intimately related. The properties and characteristics of the human visual system often give inspiration to engineers who are designing computer vision systems. Conversely, computer vision algorithms can offer insights into how the human visual system works

[1] T.S. Huang, “Computer vision: Evolution and promise,” *Proc. 19th CERN School of Computing*, Geneva: CERN, pp. 21-25, 1996.



What is computer vision?

To bridge the gap between pixels and “meaning”



What we see

0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

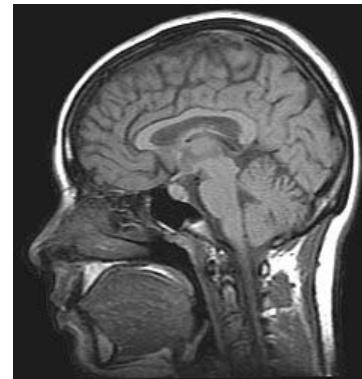
What a computer sees

Lin ZHANG, SSE, Tongji Univ.



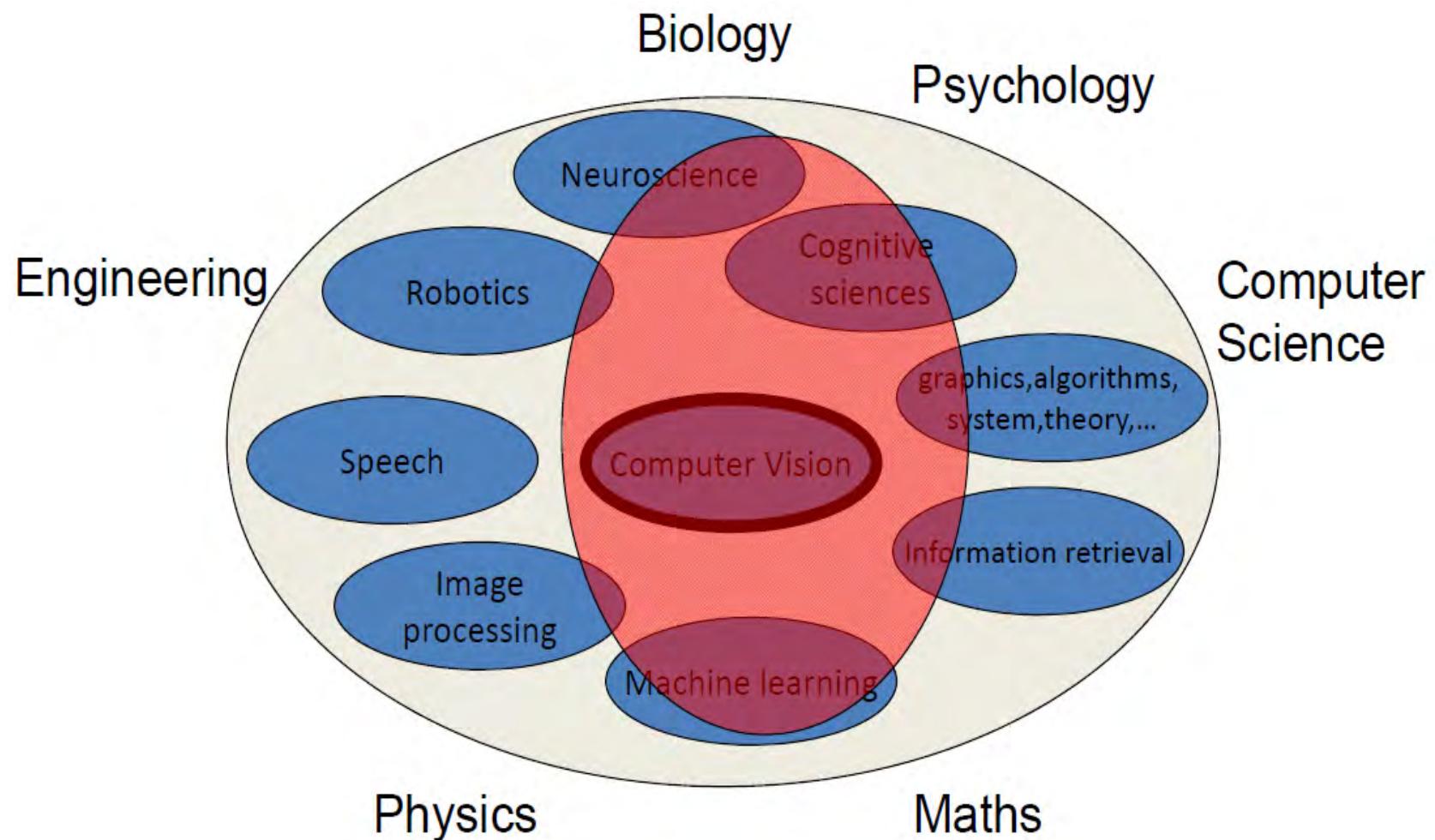
What is computer vision?

- Computer vision is the science and technology of machines that can see
- Concerned with the theory for building artificial systems that obtain information from images
- The image data can take many forms, such as a video sequence, depth images, views from multiple cameras, or multi-dimensional data from a medical scanner





What is it related to?





Vision as a measurement device

Real-time stereo

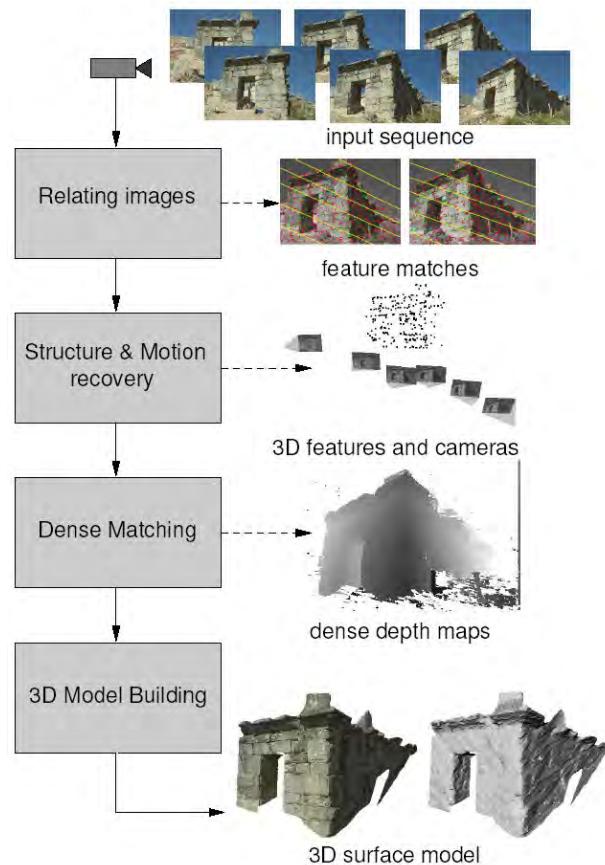


NASA Mars Rover

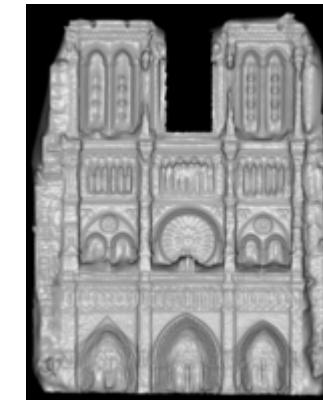
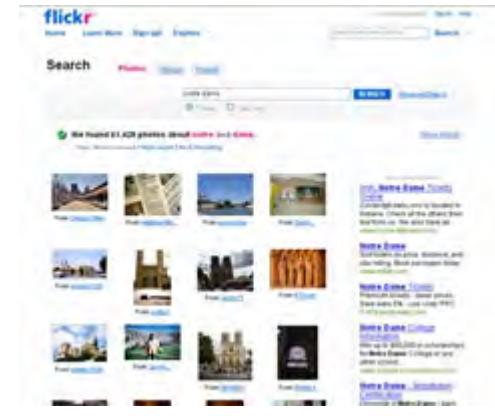


Pollefeys et al.

Structure from motion



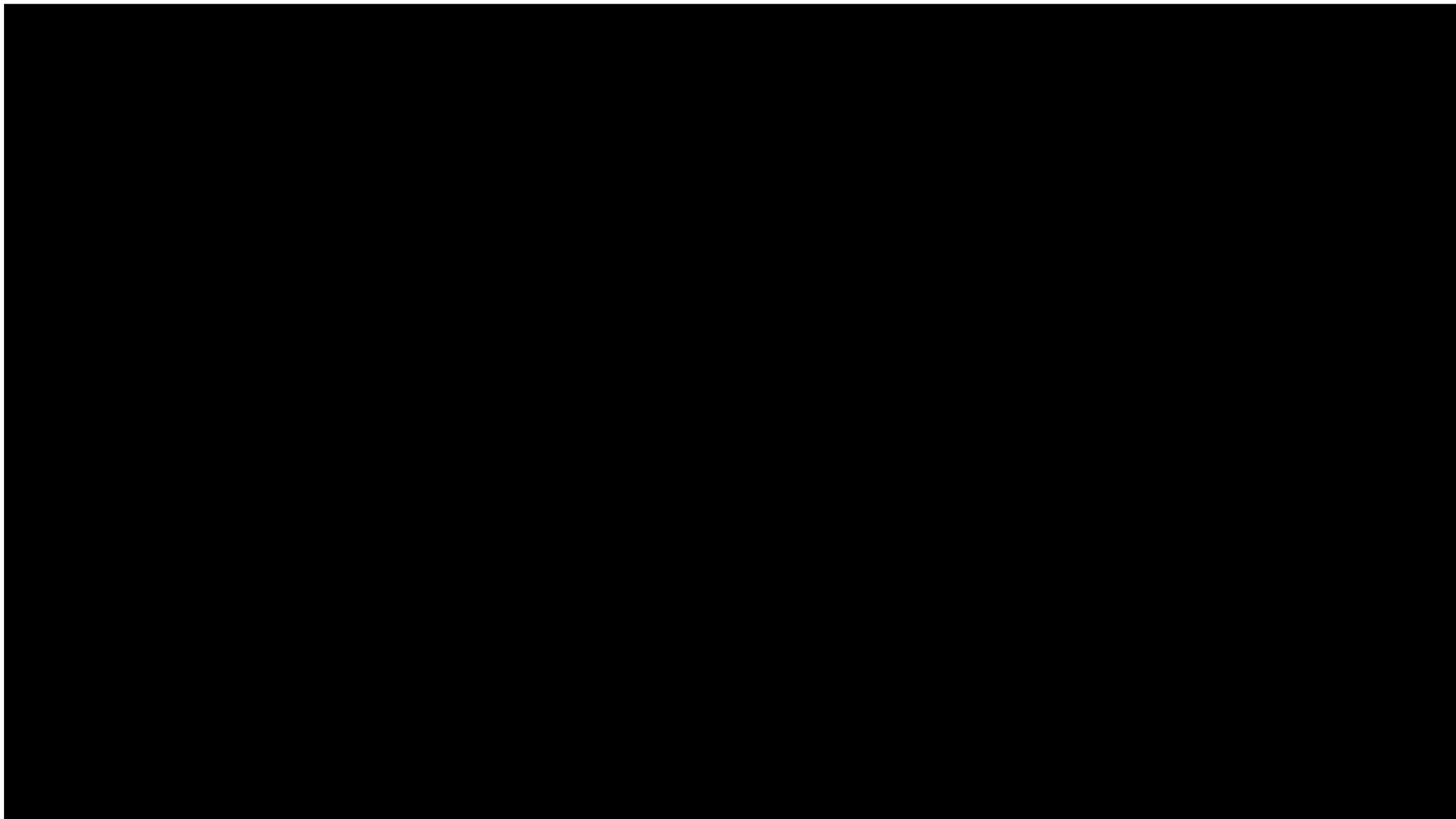
Reconstruction from Internet photo collections



Goesele et al.



Vision as a measurement device



3D Reconstruction using RGB-D input (C. Guo, L. Zhang et al, ICASSP 2022)

Lin ZHANG, SSE, Tongji Univ.



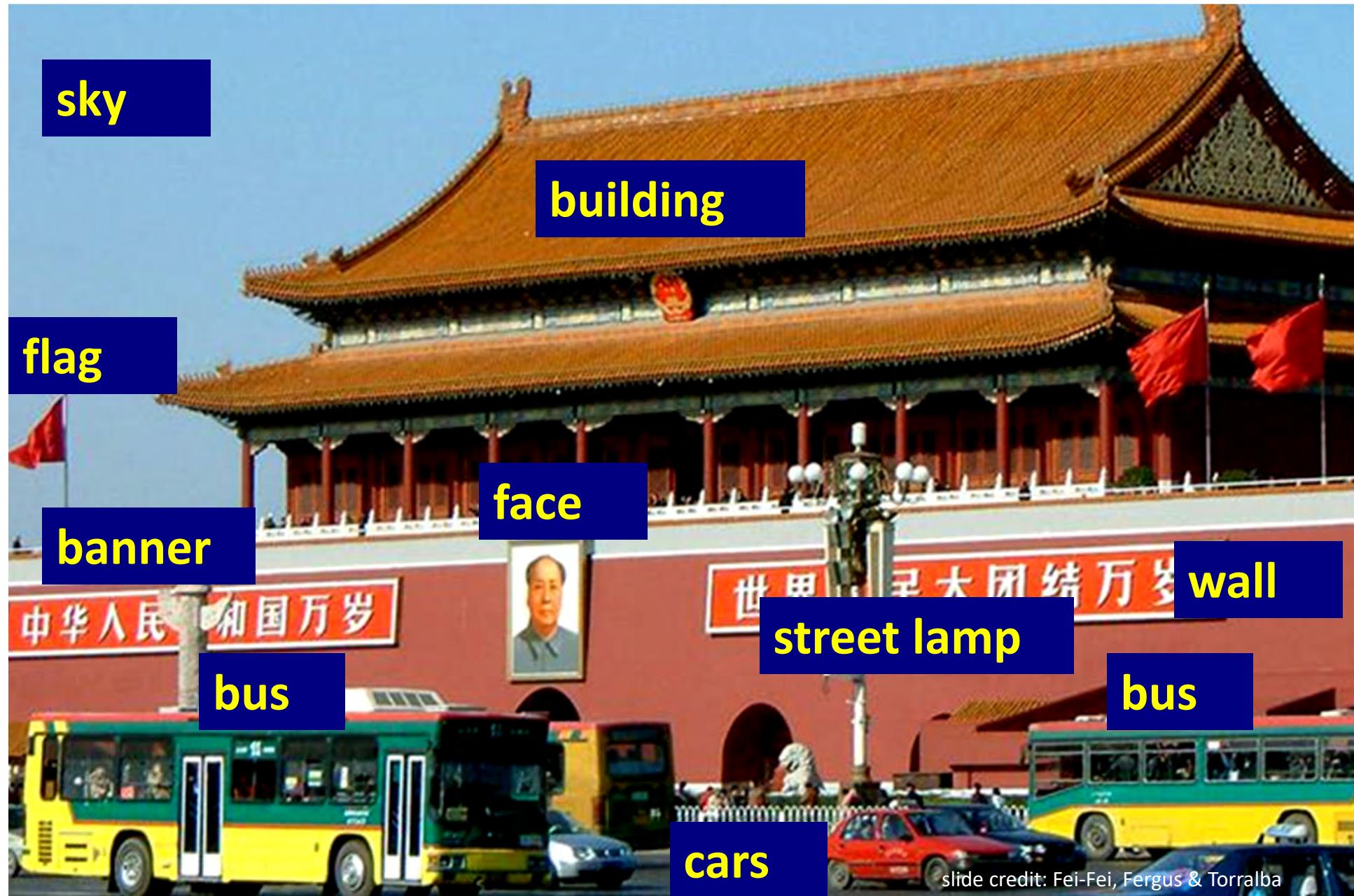
Vision as a source of semantic information



slide credit: Fei-Fei, Fergus & Torralba



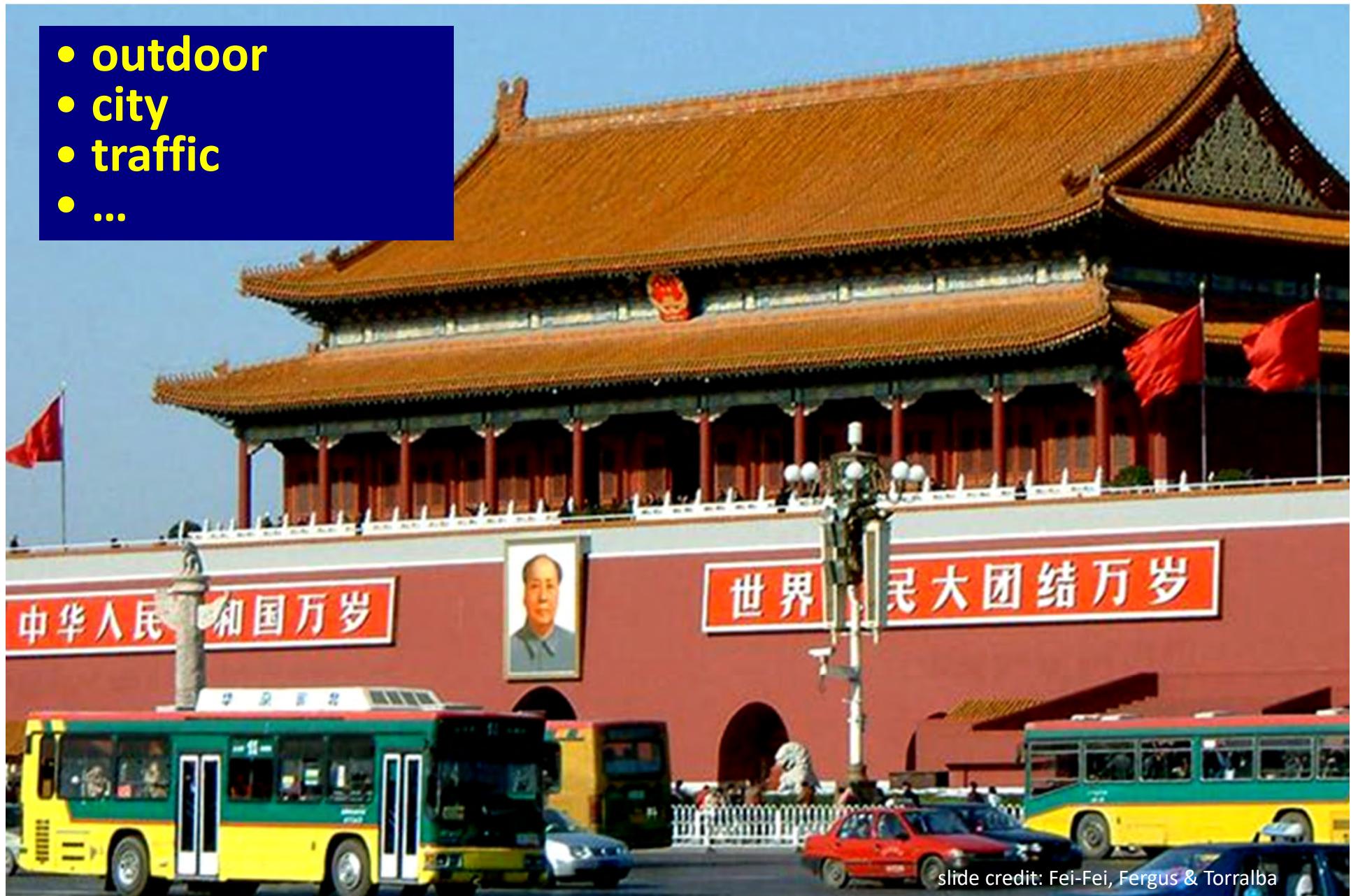
Object categorization





Scene and context categorization

- **outdoor**
- **city**
- **traffic**
- ...





A brief history about CV (in 1960s)

- In 1963, Lawrence Roberts, a PhD student at the MIT, completed his thesis "Machine concept of three-dimensional solids"^[2], which is considered as the **first professional thesis in the field of computer vision**; this thesis describes the process of deriving three-dimensional information from two-dimensional images in the ideal block world
- In view of his groundbreaking contribution, Roberts is widely considered as **the father of computer vision**



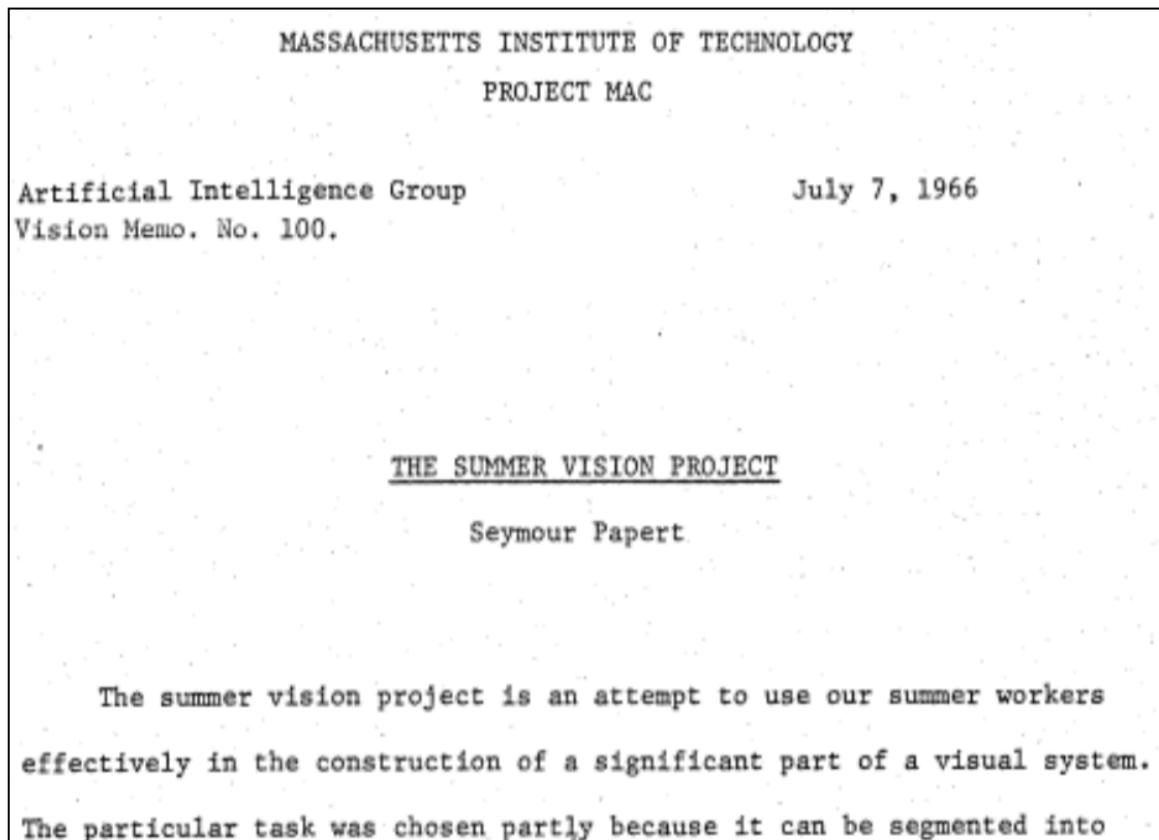
Lawrence Gilman Roberts (December 21, 1937 to December 26, 2018), an American engineer, has completed the first doctoral thesis in the field of computer vision, and is generally considered as the father of computer vision; At the same time, he is also a pioneer in Internet technology. He and his team designed and managed the world's first packet switching network ARPANET.

[2] L. G. Roberts, Machine perception of three-dimensional solids, PhD thesis, Dept. of Electrical Engineering, Massachusetts Institute of Technology, 1963



A brief history about CV (in 1960s)

- In 1966, Seymour Papert at MIT asked his undergraduate student Gerald Jay Sussman to “spend the summer linking a camera to a computer and getting the computer to describe what it saw”



Seymour Aubrey Papert (Feb. 29, 1928~ Jul. 31, 2016) is an American mathematician, computer scientist and educator born in South Africa



A brief history about CV (in 1970s)

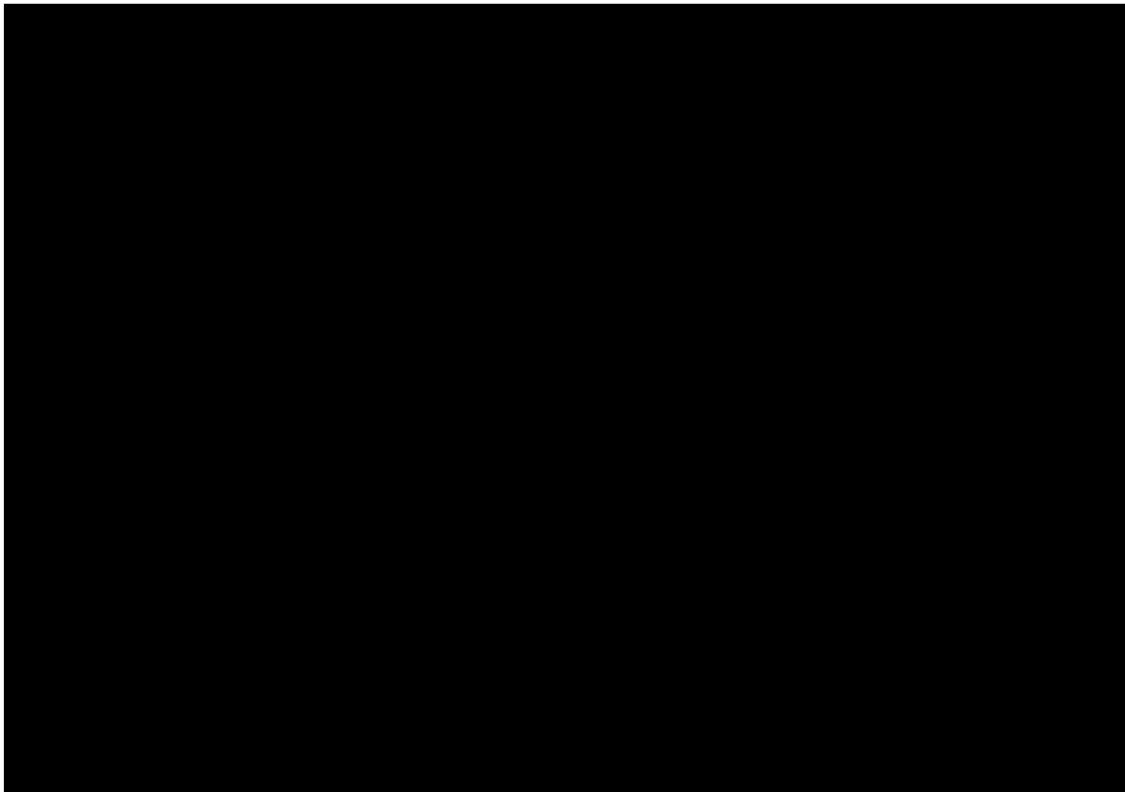
- In the 1970s, there was a landmark work, the computational vision theory proposed by Professor David Marr of MIT
- David Marr gave the development direction and some basic algorithms of computer vision from a rigorous and long-term perspective, providing a clear system for the research of this discipline
- Marr described vision as proceeding from a two-dimensional visual array (on the retina) to a three-dimensional description of the world as output. His stages of vision include^[3]:
 - a primal sketch of the scene, based on feature extraction of fundamental components of the scene, including edges, regions, etc. Note the similarity in concept to a pencil sketch drawn quickly by an artist as an impression.
 - a 2.5D sketch of the scene, where textures are acknowledged, etc. Note the similarity in concept to the stage in drawing where an artist highlights or shades areas of a scene, to provide depth
 - a 3D model, where the scene is visualized in a continuous, 3-dimensional map.

[3] David Marr, Vision: A Computational Investigation into the Human Representation and Processing of Visual Information, W. H. Freeman and Company, 1982



A brief history about CV (in 1980s)

- In 1989, Yann LeCun invented the convolutional neural network and applied it to hand-written digits recognition
- 29 years later, He was awarded ACM Turing Award in 2018



A demo from 1993 of 33-year-old Yann LeCun showing off the world's first convolutional network for text recognition



Yann LeCun, July 8, 1960~, a French computer scientist, is currently a professor at New York University

Lin ZHANG, SSE, Tongji Univ.



A brief history about CV (in 1980s)

- CVPR (IEEE International Conference on Computer Vision and Pattern Recognition) was founded in 1983



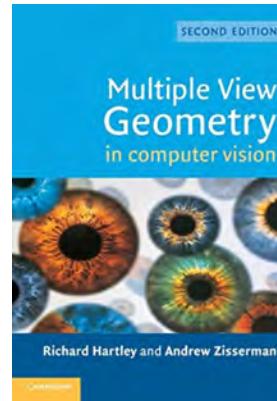
- ICCV (IEEE International Conference on Computer Vision) was founded in 1987





A brief history about CV (from 1990 to and 2010)

- In 1999, David Lowe invented SIFT (scale invariant feature transform)
- In 1999, Zhengyou Zhang, a Chinese researcher in Microsoft, invented the camera calibration algorithm using planar calibration patterns
- In 2001, Paul Viola and Michael Johns invented Adaboost+Haar features based real-time face detection algorithm
- In 2000, Richard Hartley (University of Queensland, Australia) and Andrew Zisserman (Oxford University, Britain) published their famous textbook “Multiple View Geometry in Computer Vision”



Richard, Andrew, and their famous book



A brief history about CV (from 1990 to and 2010)

- In 2009, Li Fei-fei (Stanford University, USA) released the ImageNet dataset which is a large-scale dataset for benchmarking image classification and object detection algorithms



Li Fei-Fei, a professor of Stanford University. Li was born in Beijing, China in 1976 and grew up in Chengdu. When she was 12, her father moved to the US; when she was 15, she and her mother joined him in Parsippany-Troy Hills, New Jersey. Dr. Li is an elected Member of the National Academy of Engineering (NAE), USA

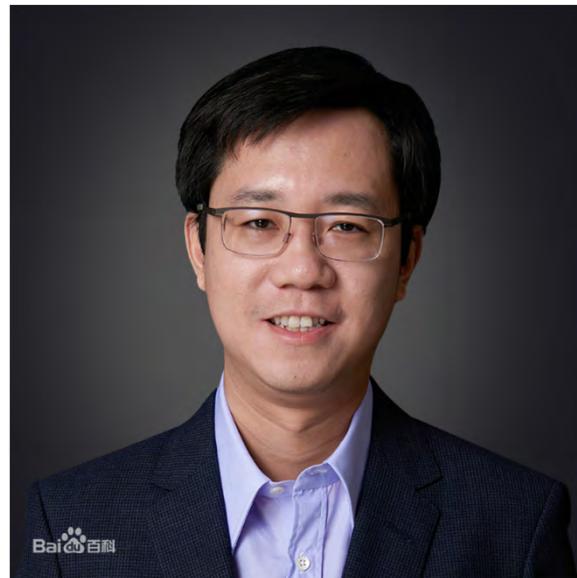


A brief history about CV (Since 2010s)

- In 2012, Alex Krizhevsky (Toronto University, Canada) et al. built the first deep CNN and won the ILSVRC challenge of that year. The year 2012 has been deemed as the starting year of the deep learning era
- In 2016, Kaiming He, Jian Sun et al. invented the ResNet, which is now the basic module of almost all the modern network architectures



Xiaou Tang and Kaiming He, in CUHK, 2009



Kaiming will join EECS, MIT as a faculty member in 2024



A brief history about CV (Since 2010s)

- In 2014, Ian Goodfellow (Montreal University, Canada) et al. proposed Generative Adversarial Networks, which is now widely used for self-supervised learning
- In 2020, Ben Mildenhall (University of California, Berkeley, USA) et al. proposed NeRF (Neural radiance field) based rendering, which is now a new paradigm for image based reconstruction and rendering



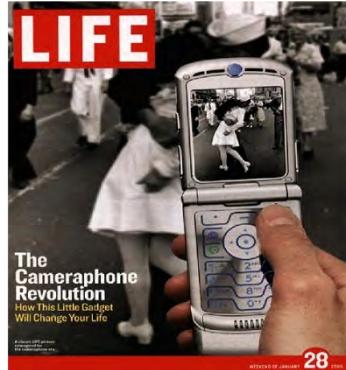


Today

- What is computer vision?
- Why do we need to study CV?
- Course overview



Why study computer vision?



Lin ZHANG, SSE, Tongji Univ.



Why study computer vision?

Vision is useful: Images and video are everywhere!



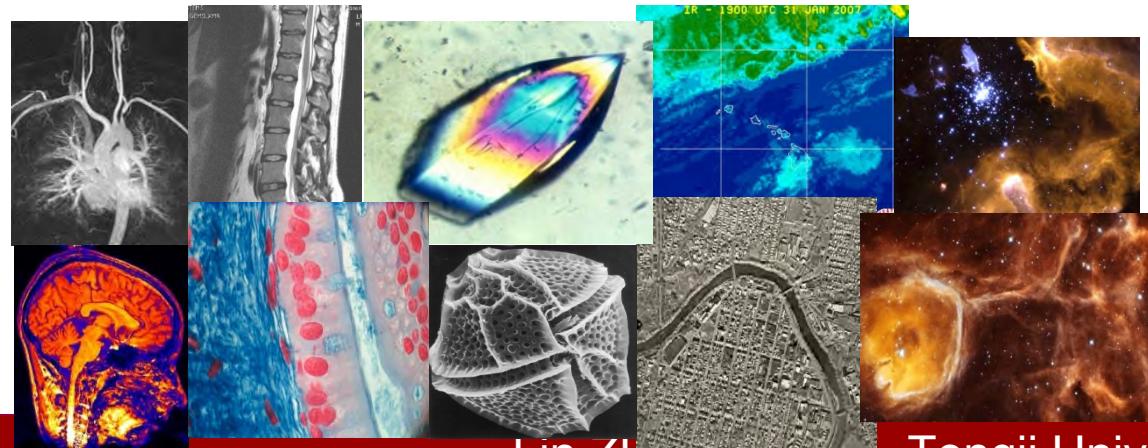
Personal photo albums



Movies, news, sports



Surveillance and security



Lin ZHANG, SSE, Tongji Univ.



Why study computer vision?

Vision is useful: Images and video are everywhere!





Structure from motion

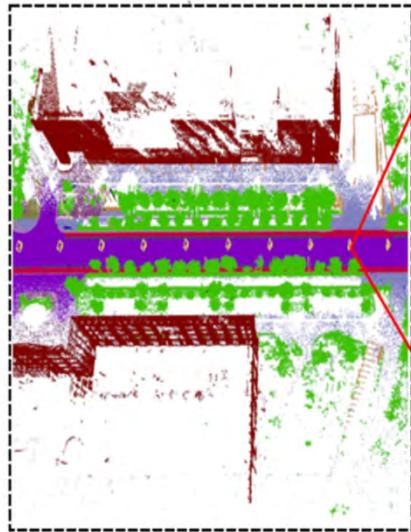


Bundler: Structure from Motion (SfM) for Unordered Image Collections (<https://www.cs.cornell.edu/~snavely/bundler/#S3>)

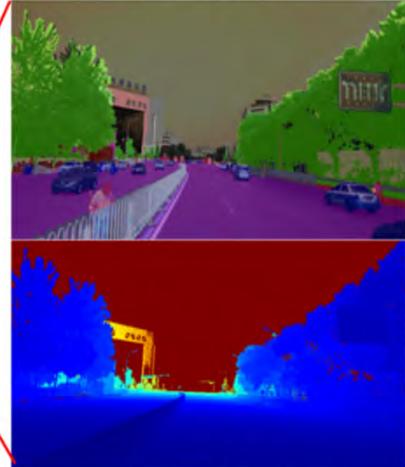


Automotive safety

3D semantic map & camera location

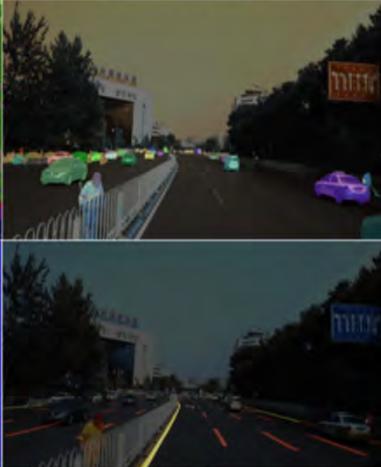


Semantic mask



Projected depth

Instance mask



Lanemark mask



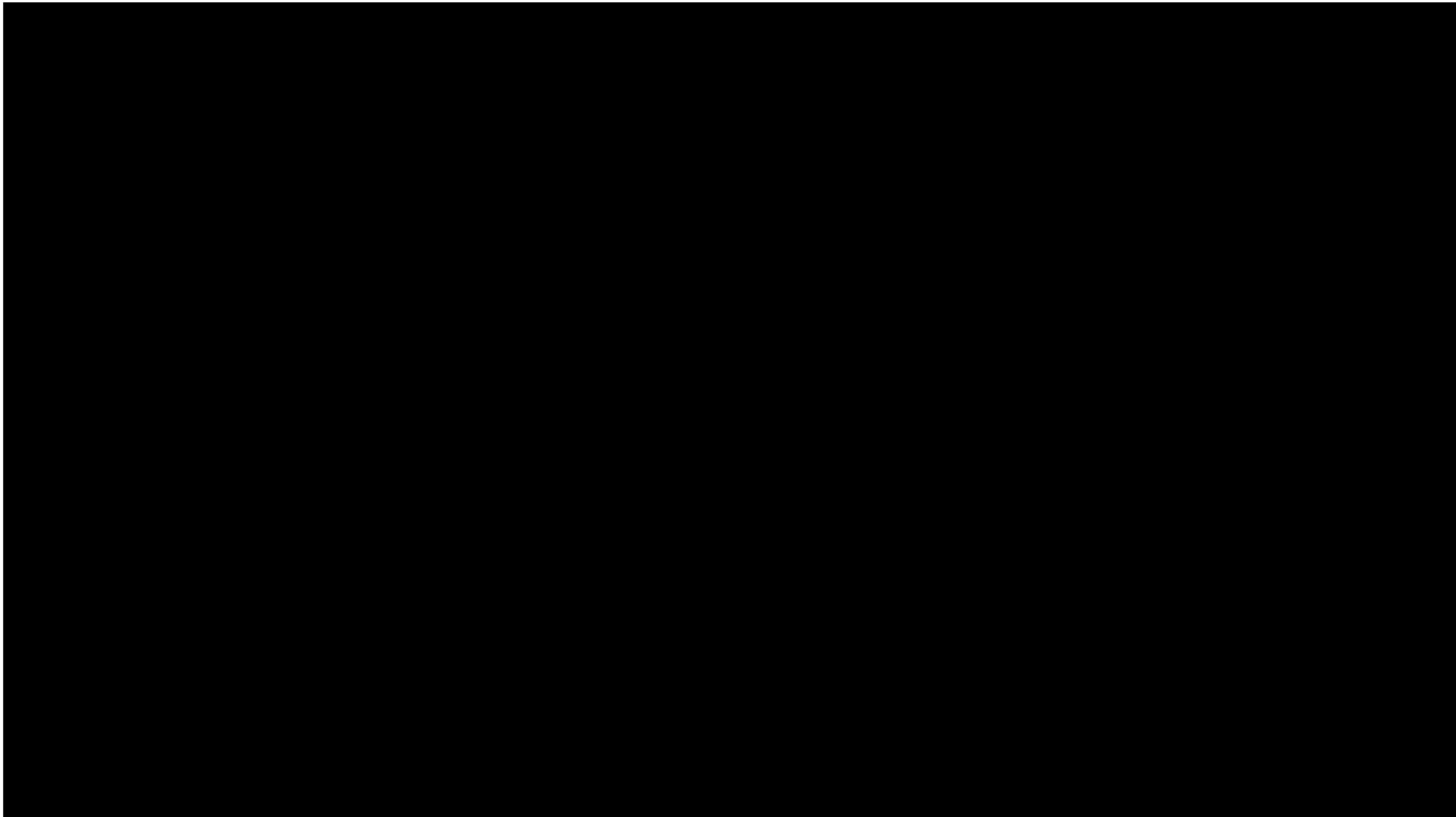
ApolloScape Open Dataset for Autonomous Driving^[1]

- ✓ holistic semantic dense point cloud for each site
- ✓ Stereo
- ✓ per-pixel semantic labelling
- ✓ lanemark labelling
- ✓ instance segmentation
- ✓ 3D car instance
- ✓ high accurate location for every frame in various driving videos

[4] X. Huang et al., The ApolloScape Open Dataset for Autonomous Driving and Its Application, IEEE T-PAMI, vol. 42, pp. 2702-2719, Oct. 2020



AVP (Automated Valet Parking)



2021年4月，威马W6车型已量产百度AVP系统

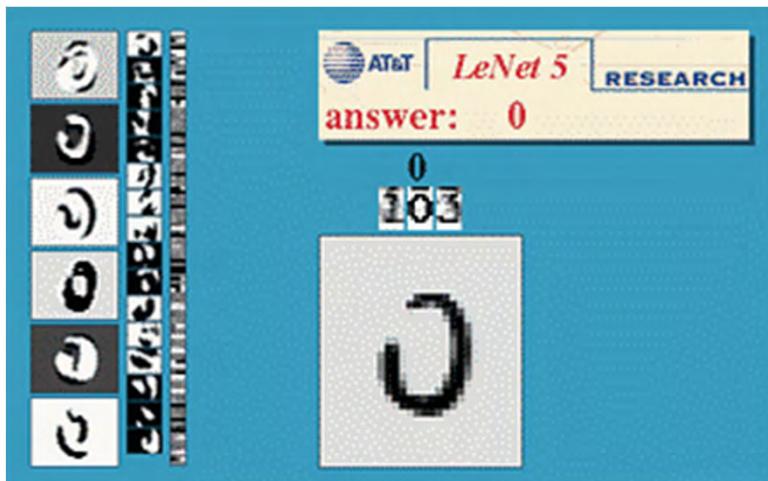
Lin ZHANG, SSE, Tongji Univ.



Optical character recognition (OCR)

Technology to convert scanned docs to text

- If you have a scanner, it probably came with OCR software



Digit recognition, AT&T labs
<http://www.research.att.com/~yann/>



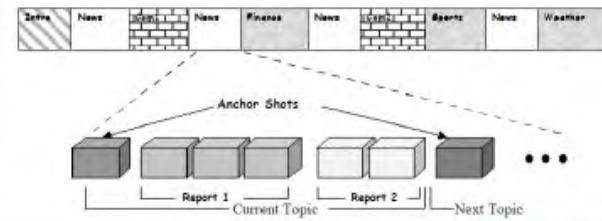
License plate readers
http://en.wikipedia.org/wiki/Automatic_number_plate_recognition



Videos based applications

视频拆分系统

- 检测片头片尾
- 检测新闻故事边界
- 检测广告和看点



新闻视频

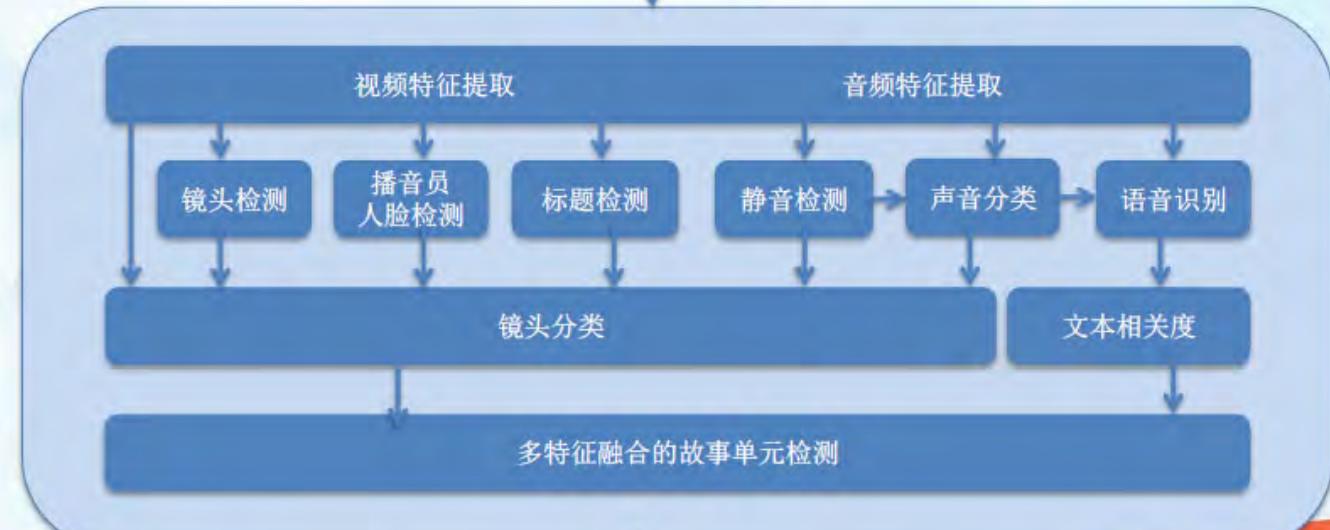


图 视频拆分标注系统





Face and mask detection

YOLOv4目标检测教程系列-最佳速度和精度中英文字幕

5497 10 2020-08-31 08:00:17



1人正在看，已装填 0 条弹幕



A 发个弹幕见证当下

弹幕礼仪 >

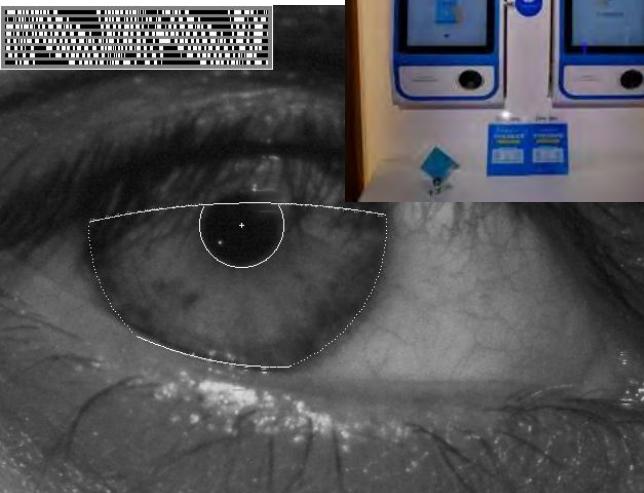
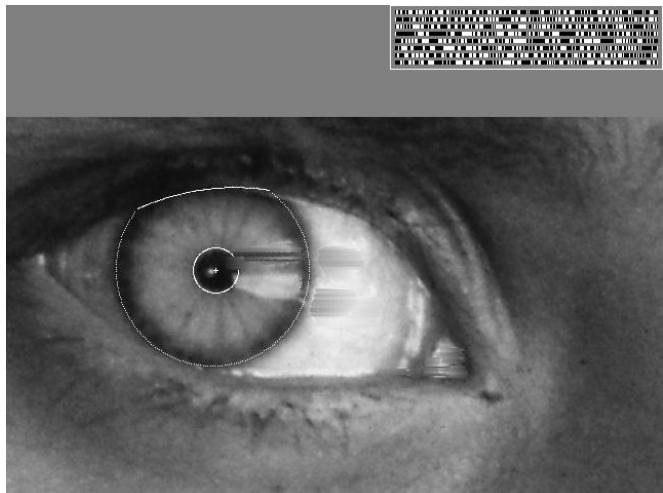
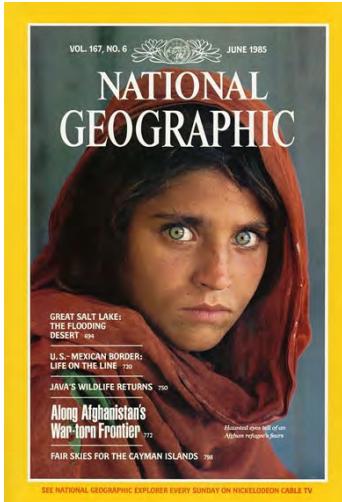
发送

<https://www.bilibili.com/video/av926885838?p=15>

Lin ZHANG, SSE, Tongji Univ.



Vision-based biometrics



Lin ZHANG, SSE, Tongji Univ.



Robotics



Robotics in MAP (Machine Autonomous Perception) lab

Lin ZHANG, SSE, Tongji Univ.



Work by our students: Photosynth

Project products of students from 2009 Media&Arts



Lin ZHANG, SSE, Tongji Univ.



Recent work of our group



Lin Zhang, Lida Li, Anqi Yang et al., "Towards contactless palmprint Recognition: A novel device, a new benchmark, and a collaborative representation based identification approach", Pattern Recognition, vol. 69, pp. 199-212, 2017

Lin ZHANG, SSE, Tongji Univ.



Recent work of our group



Yingyi Zhang, Lin Zhang* et al., Pay by showing your palm: A study of palmpoint verification on mobile platforms, in: Proc. ICME, pp. 862-867, 2019.

Lin ZHANG, SSE, Tongji Univ.



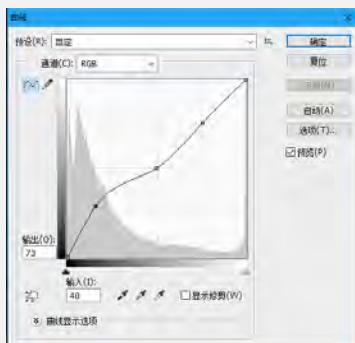
Recent work of our group

图像曝光度的自动校正



研究目标

- 图像曝光度失真的校正
- 利用无监督学习解决马尔可夫随机场问题，设计端到端的自动化算法
- 部署至移动端
- 实时？视频曝光度校正？



图像曝光度校正原理——调整图像 α 曲线

校正前



校正后



图像曝光度校正典型示例



我们的贡献

- ✓ 首次提出了MRF as CNN算法，利用神经网络解决图像曝光度校正问题
- ✓ 部署至移动端，开发具有实用价值的移动应用



Recent work of our group

图像曝光度的自动校正

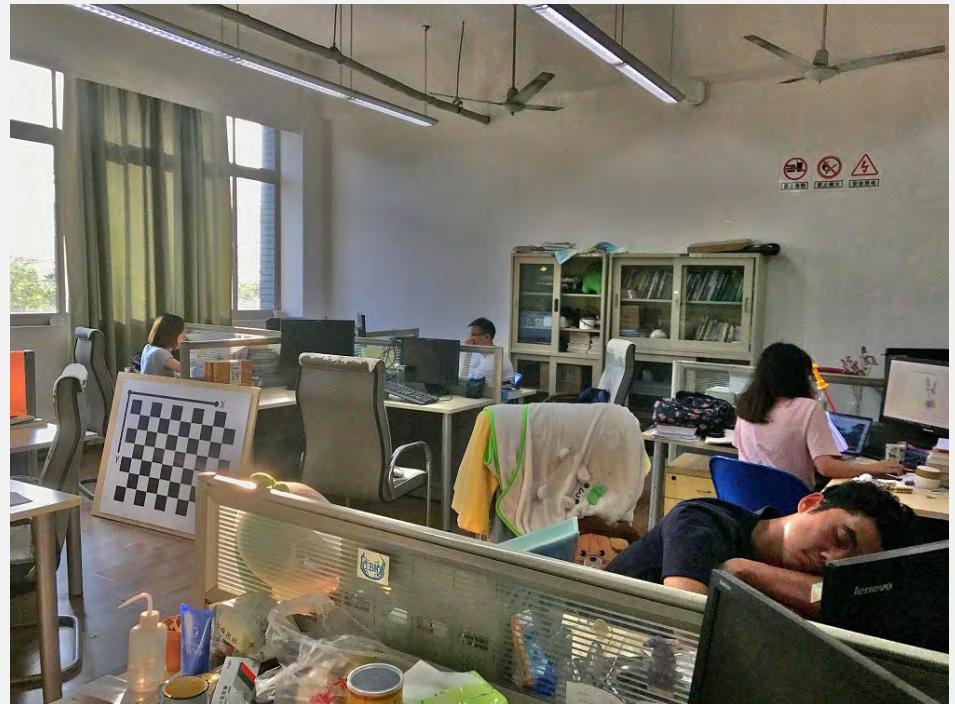


Lin ZHANG, SSE, Tongji Univ.



Recent work of our group

图像曝光度的自动校正



Lin ZHANG, SSE, Tongji Univ.

Recent work of our group

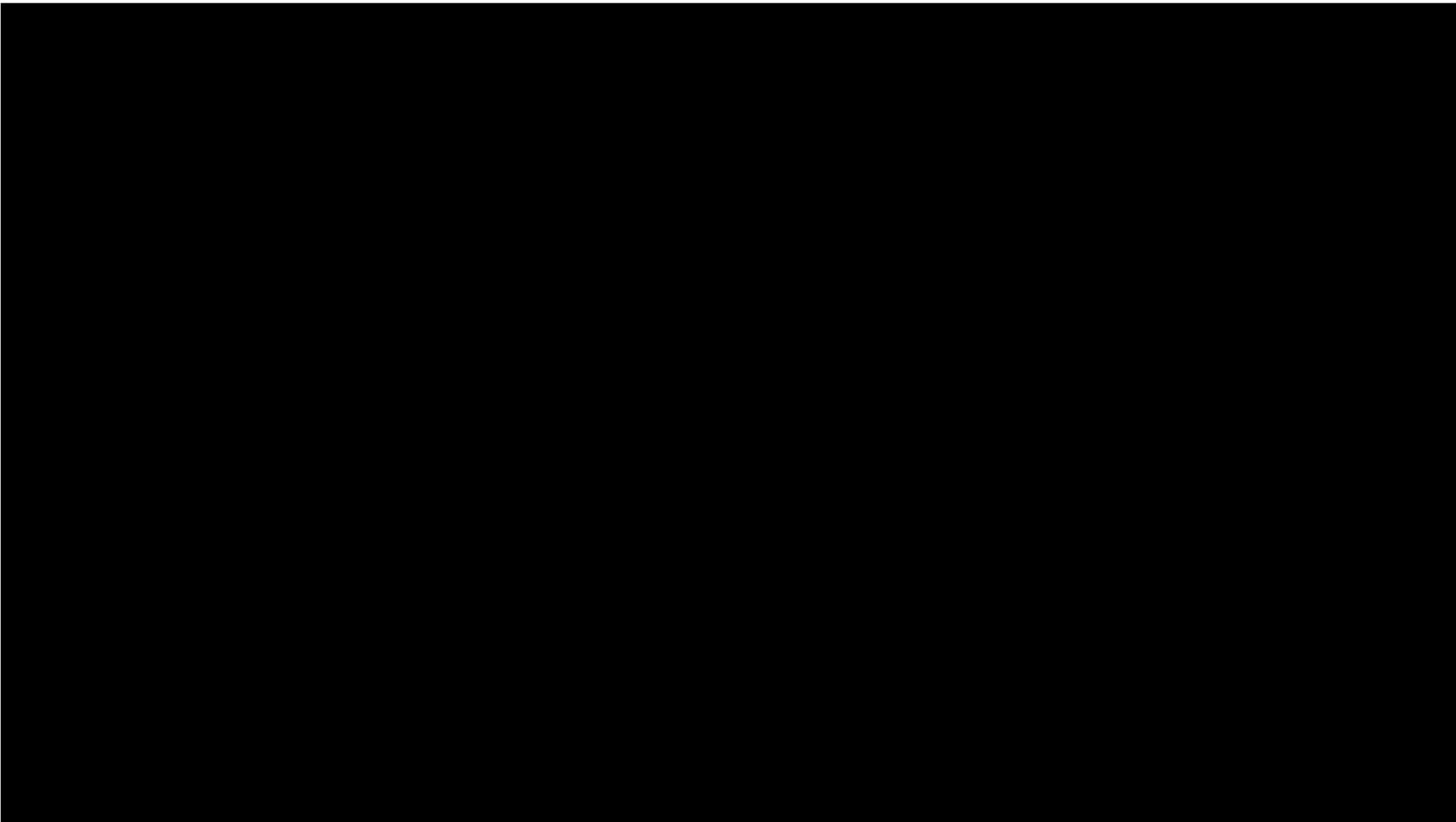
**Zero-Shot Restoration
of Back-lit Images
Using Deep Internal Learning**

ACM MM 2019 - Paper ID 1014

Lin Zhang, Lijun Zhang et al., "Zero-Shot Restoration of Back-lit Images Using Deep Internal Learning",
ACM Int'l Conf. Multimedia, 2019



Recent work of our group



Lin Zhang, Anqi Zhu et al., "Simulation of atmospheric visibility impairment", IEEE Trans. Image Processing, vol. 30, pp. 8713-8726, 2021.

Lin ZHANG, SSE, Tongji Univ.



Recent work of our group

Short-range Self-parking



Lin Zhang, Junhao Huang et al., "Vision-based parking-slot detection: A DCNN-based approach and a large-scale benchmark dataset", IEEE Trans. Image Processing, vol. 27, no. 11, pp. 5350-5364, 2018.

Lin ZHANG, SSE, Tongji Univ.



Recent work of our group

2018年9月，人民网采访报道



在学术界和工业界都产生了较大影响

■ 在学术界的影响

- ✓ CSDN、知乎等多家网络技术媒体对申报人的泊车位检测技术进行了大篇幅介绍和正面评价
- ✓ 公开的数据集和工具已经被美国南德州大学、韩国汉阳大学、北交大、华南理工、湖南大学等多所研究机构的人员下载使用

■ 在工业界的影响

- ✓ 华为、科大讯飞、纵目、天瞳威视等企业下载使用了我们的数据或复现了我们的算法

知乎帖子截图 (https://www.zhihu.com/question/370205708)

知乎用户 [Liaoxiaodong \(baconboy\)](#) 在 2020-12-08 13:39 发布了一个回答，标题为：《自主泊车系统中停车位检测与避让》。该回答展示了关于停车位检测的研究，并附有相关的代码和论文链接。

该回答的内容包括：

- 发送停车位信息给决策模块。
- 决策模块根据停车位信息进行决策。
- 基于深度学习的停车位检测算法流程。
- 引用的论文：“Vision-Based Parking-Slot Detection: A DCNN-Based Approach and a Large-Scale Benchmark Dataset.” IEEE Transactions on Image Processing 27.11 (2018): 5350-5364.
- 时间：2018年 IEEE TIP
- 单位：同济
- github: DeepPS

评论区有一条评论：“应该是第一篇公开基于深度学习的停车位检测算法，同时公开了数据集和开源，工作很赞。”

发件人：[Liaoxiaodong \(baconboy\)](#)
发送时间：2020-12-08 13:39
收件人：[calimhao@tongji.edu.cn](#)
妙语：[Fenqbalan](#)
主题：Tongji Parking-slot Dataset 2.0数据集使用申请
张林老师，您好，
我是华为诺亚方舟实验室的一名研究员，在做停车位检测方面相关的工作，最近拜读了您的文章Vision-Based Parking-Slot Detection:
A DCNN-Based Approach and a Large-Scale Benchmark Dataset., 对这篇论文很感兴趣，深受启发，
希望您能允许我们使用您的公开的数据集，继续做一些工作，
以便我们发表一些相关论文。
祝好！
李晓东
2020.12.08
华为诺亚方舟实验室

Lin ZHANG, SSE, Tongji Univ.



Recent work of our group

Pedestrian and speed-bump detection and distance measurement



Lin ZHANG, SSE, Tongji Univ.

Recent work of our group

Online optimization of camera poses in a surround-view system

ROECS:
A Robust Semi-direct Pipeline Towards Online
Extrinsics Correction of the Surround-view System

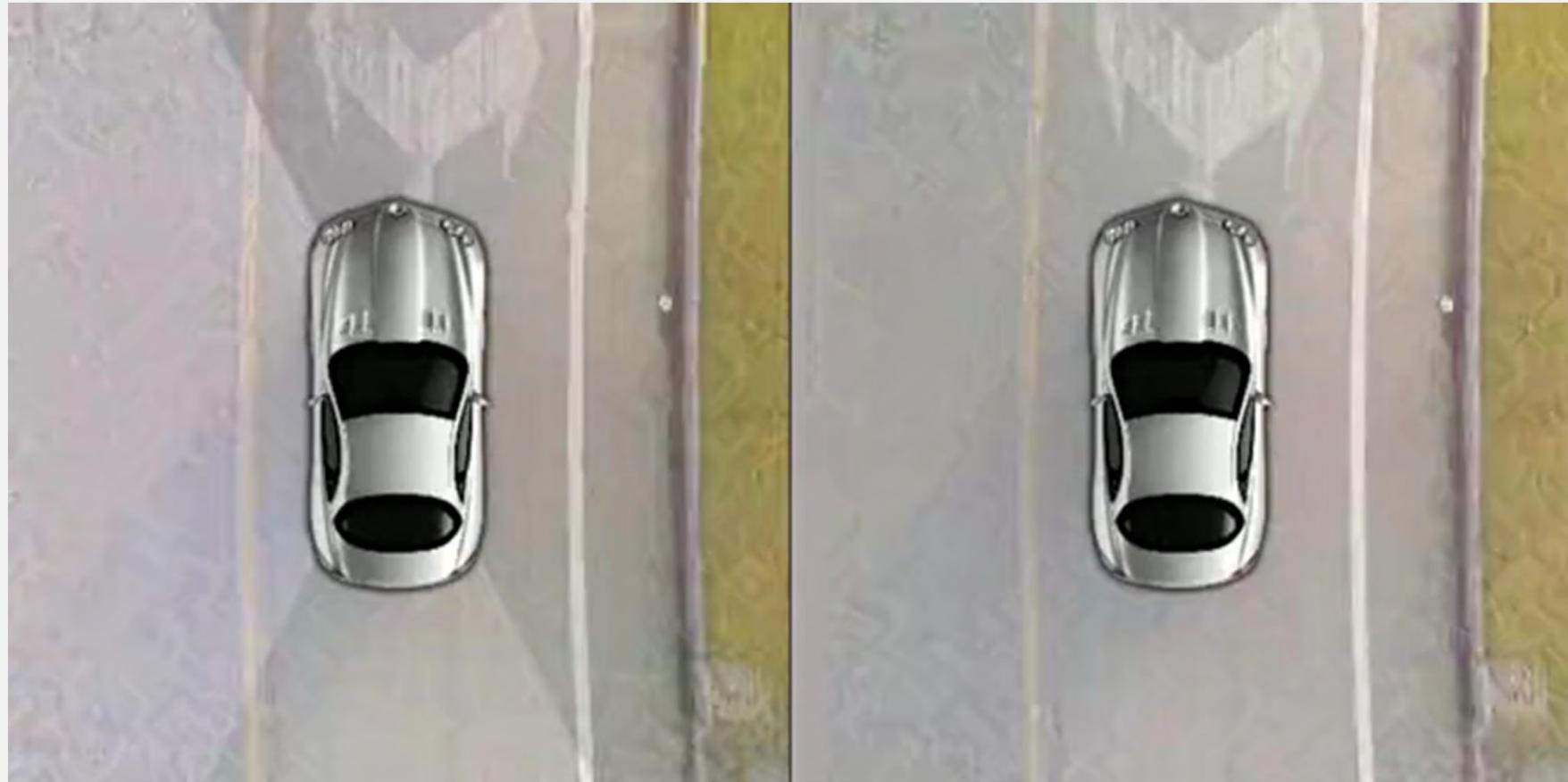
ACM MM 2021 Paper ID: 1640

Tianjun Zhang, Lin Zhang* et al., "ROECS: A Robust Semi-direct Pipeline Towards Online Extrinsics Correction of the Surround-view System", in Proc. ACM Int'l Conf. Multimedia, 2021.



Recent work of our group

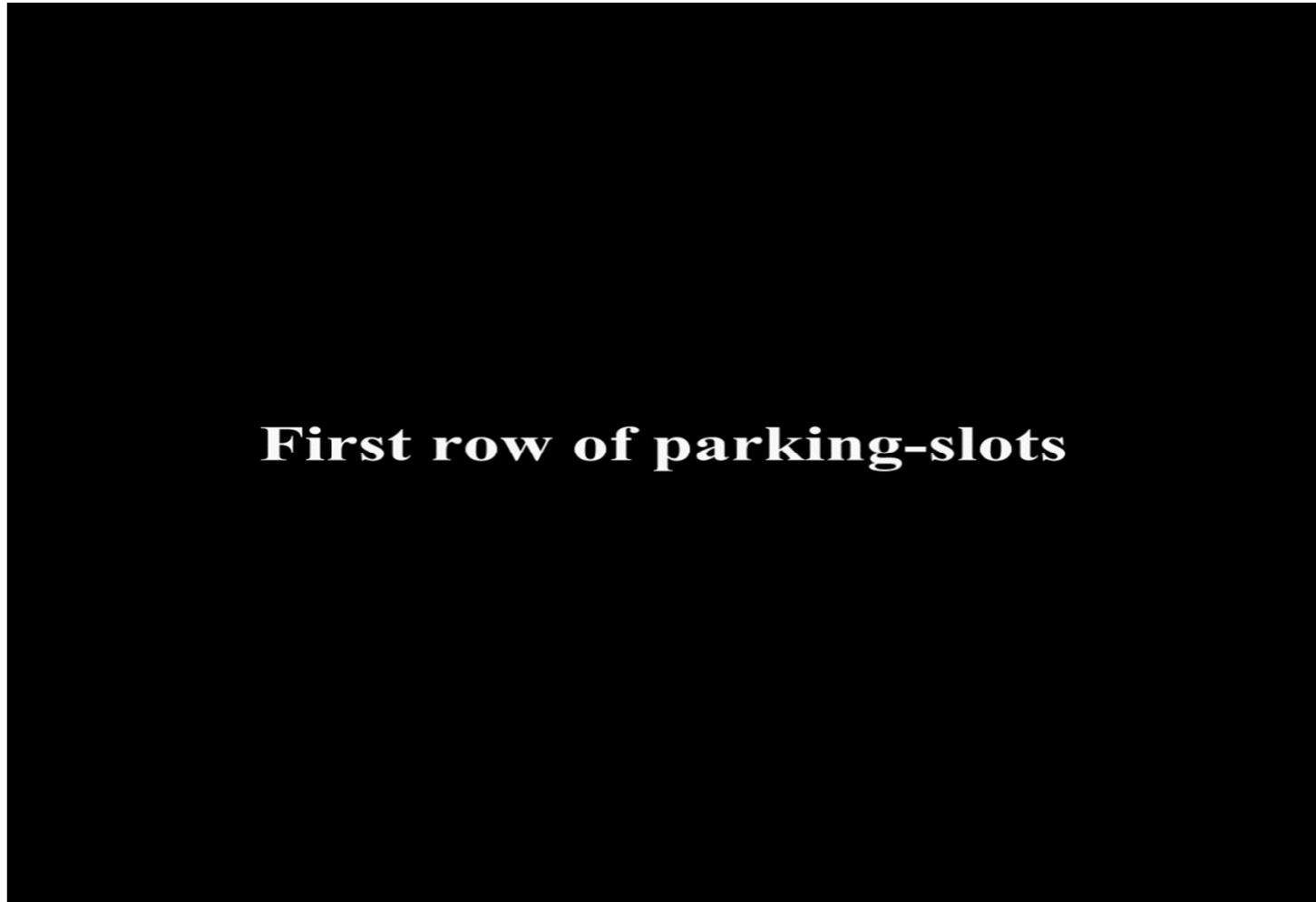
Photometric adjustment in the surround-view



Lin ZHANG, SSE, Tongji Univ.



Recent work of our group



First row of parking-slots

Xuan Shao, Lin Zhang* et al., "MOFISSLAM: A multi-object semantic SLAM system with front-view, inertial and surround-view sensors for indoor parking", IEEE Trans. CSVT, 2022

Lin ZHANG, SSE, Tongji Univ.



Recent work of our group



Anqi Zhu, Lin Zhang* et al, Pedestrian-Aware Panoramic Video Stitching Based on a Structured Camera Array, ACM Transactions on Multimedia Computing, Communications, and Applications, 17 (4): 136, 2021.

Lin ZHANG, SSE, Tongji Univ.



Recent work of our group

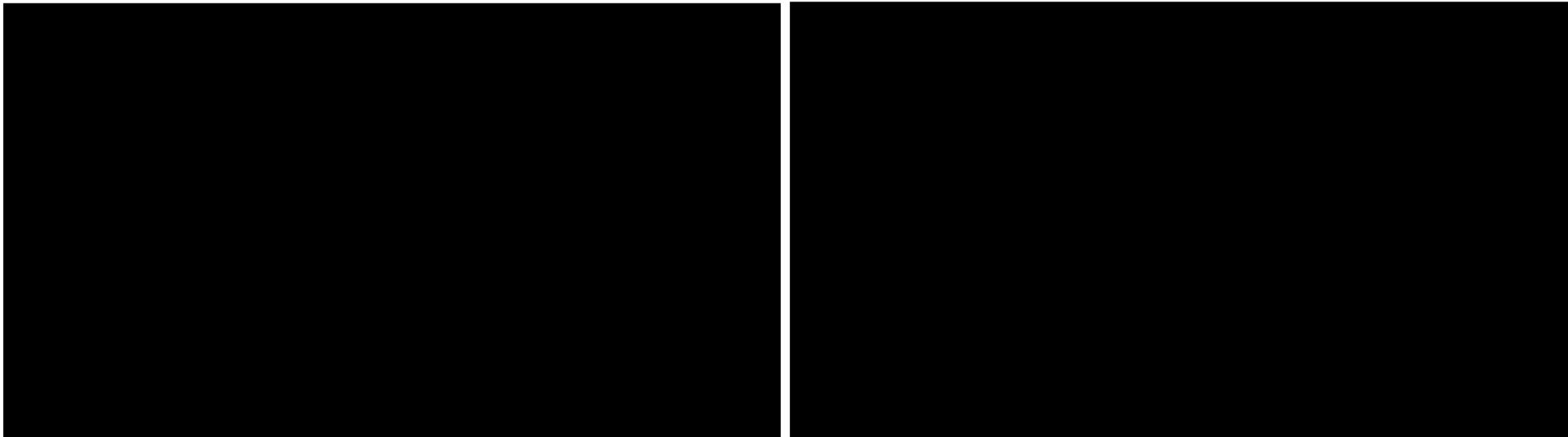
Single-agent Indoor Mapping

Tianjun Zhang, Lin Zhang* et al., "CVIDS: A collaborative localization and dense mapping framework for multi-agent based visual-inertial SLAM," IEEE Transactions on Image Processing, vol. 31, pp. 6562-6576, 2022

Lin ZHANG, SSE, Tongji Univ.



Recent work of our group



Zhong Wang, Lin Zhang* et al., "D-LIOM: Tightly-coupled direct LiDAR-inertial odometry and mapping," IEEE Transactions on Multimedia, 2023

Lin ZHANG, SSE, Tongji Univ.



Recent work of our group

An Audio-Visual Speaker Localization System
for Cocktail Party Problem

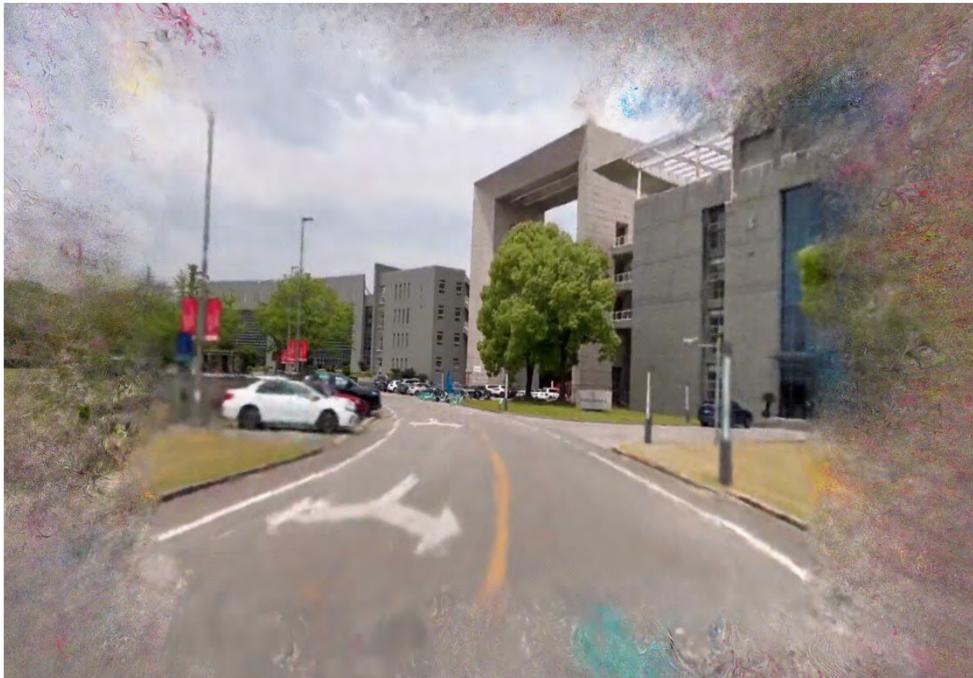
*School of Software Engineering, Tongji University
Machine Autonomous Perception Group*

Zhanbo Shi, Lin Zhang* et al., Audio–Visual Sound Source Localization and Tracking Based on Mobile Robot for The Cocktail Party Problem, Appl. Sci., 13(10), 2023

Lin ZHANG, SSE, Tongji Univ.



Recent work of our group



Lin ZHANG, SSE, Tongji Univ.



Recent work of our group



Lin ZHANG, SSE, Tongji Univ.



You can find a good job!

- Many first-class companies now are developing CV related applications, to name a few
 - Google
 - Ali
 - Facebook
 - SenseTime
 - megvii
 - Tencent
 - Baidu
 - DJI
 - Huawei
 - ...



Today

- What is computer vision?
- Why do we need to study CV?
- Course overview



Course content

- Introduction
- Local interest point detectors
- Local feature descriptors and matching
- Projective geometry
- Nonlinear least squares
- Measurement using a single camera
- Basics for machine learning and its applications
- Applications of DCNNs
- Introduction to numerical geometry

The lectures are grouped into four themes



Some tips

- Prerequisites
 - Linear algebra
 - Calculus
 - Matlab Programming
 - C++ Programming
- Knowledge sources
 - IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)
 - IEEE Transactions on Image Processing (TIP)
 - International Journal of Computer Vision (IJCV)
 - IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)
 - IEEE International Conference on Computer Vision (ICCV)
 - European Conference on Computer Vision (ECCV)



Lin ZHANG, SSE, Tongji Univ.