Chapter 01
Introduction to Computer Vision

Lin ZHANG
School of Computer Science and Technology
Tongji University

Course Information

Contact Information

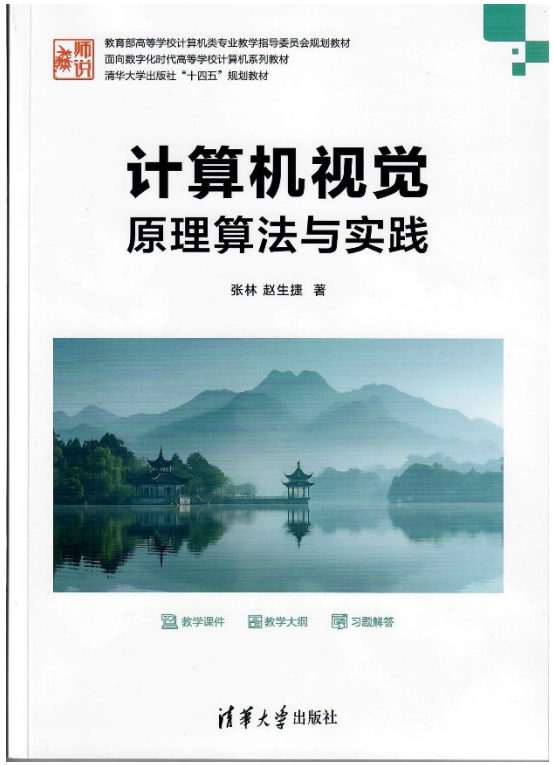
Room:
Email:

TA:

Course information can be found at
https://cslinzhang.github.io/home/CV/CV.htm

Slides and code can be found at
https://github.com/csLinZhang/CVBook

Materials



张林，赵生捷，计算机视觉：原理算法与实践，清华大学出版社，2025年5月

- References
  - 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

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 >=1/3 lectures, you will fail this course
- Bonus 5%: being active in class and answering my questions correctly

Today

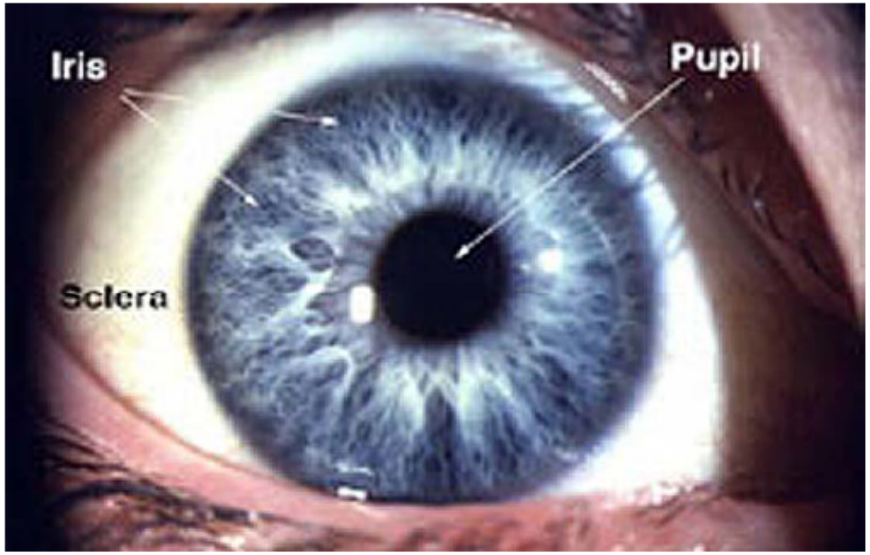
- What is computer vision?
- Why do we need to study CV?
- Course overview

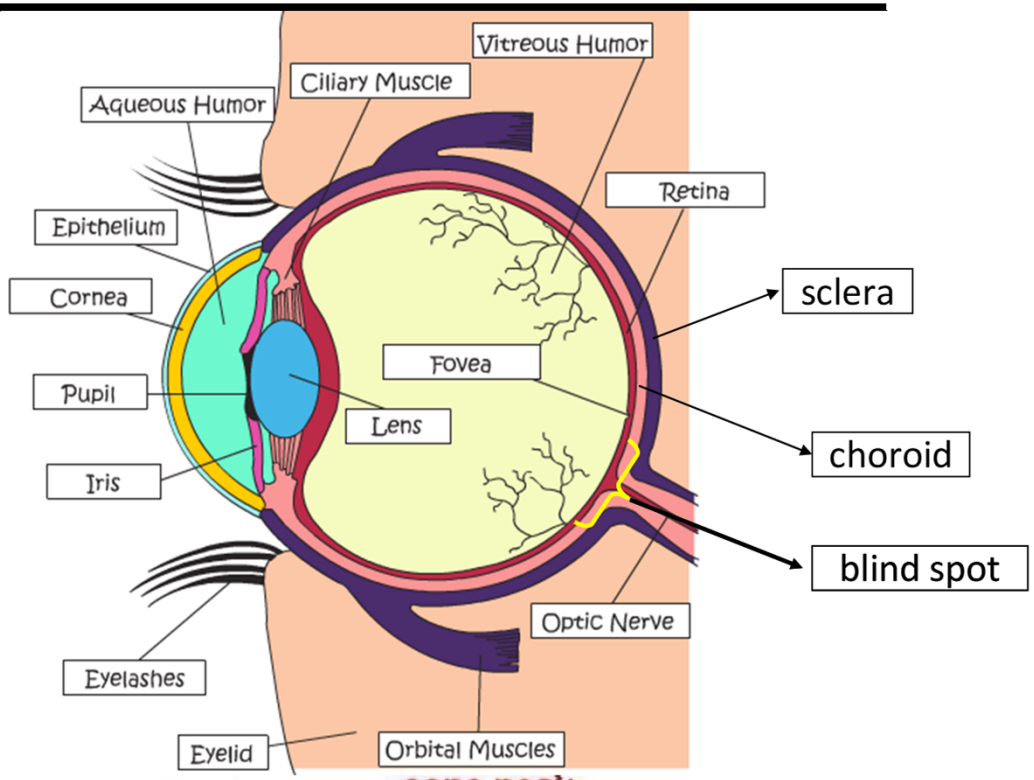
What is vision?

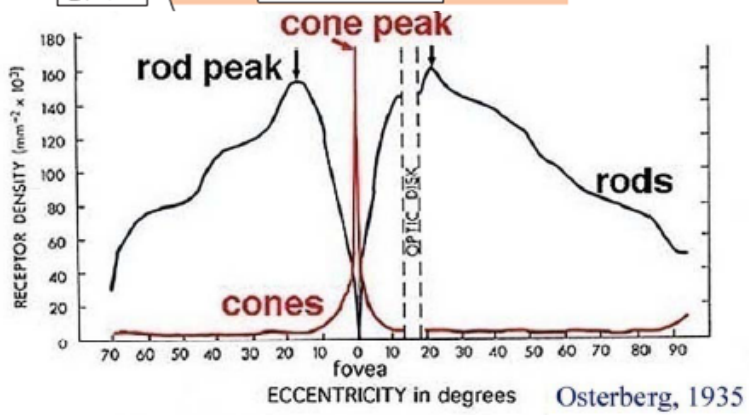
"The plain man's answer (and Aristotle's too) would be, to know what is where by looking. In other words, vision is the process of discovering from images what is present in the world, and where it is " *David Marr*, Vision: A Computational Investigation into the Human Representation and Processing of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

Our Eyes

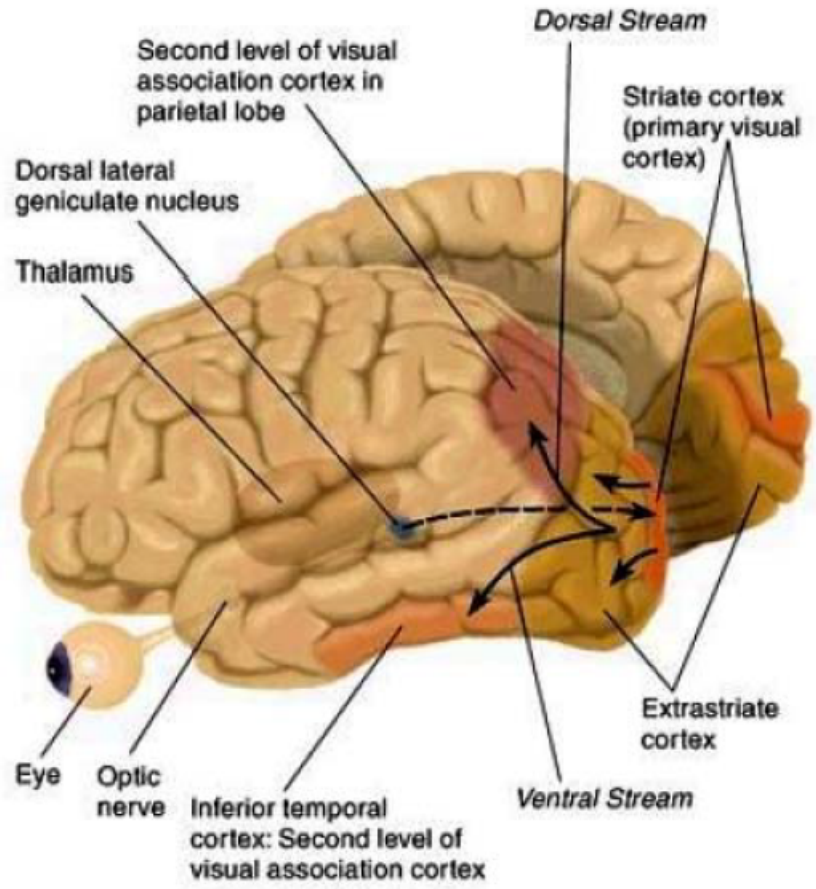

Iris
Pupil
Sclera


Vitreous Humor
Ciliary Muscle
Aqueous Humor
Retina
Epithelium
Cornea
sclera
Fovea
Pupil
Lens
choroid
Iris
blind spot
Optic Nerve
Eyelashes
Eyelid
Orbital Muscles

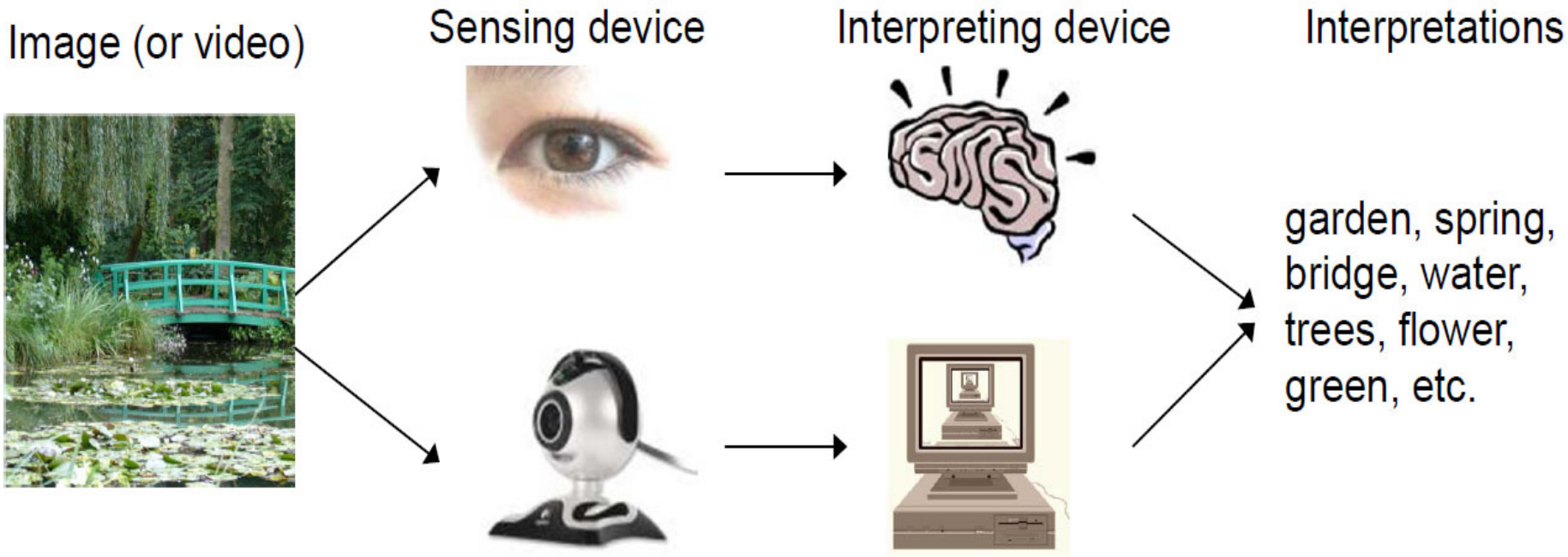

cone peak
rod peak
rods
cones
OPTIC DISK
RECEPTOR DENSITY (mm⁻² x 10³)
fovea
ECCENTRICITY in degrees
Osterberg, 1935

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”

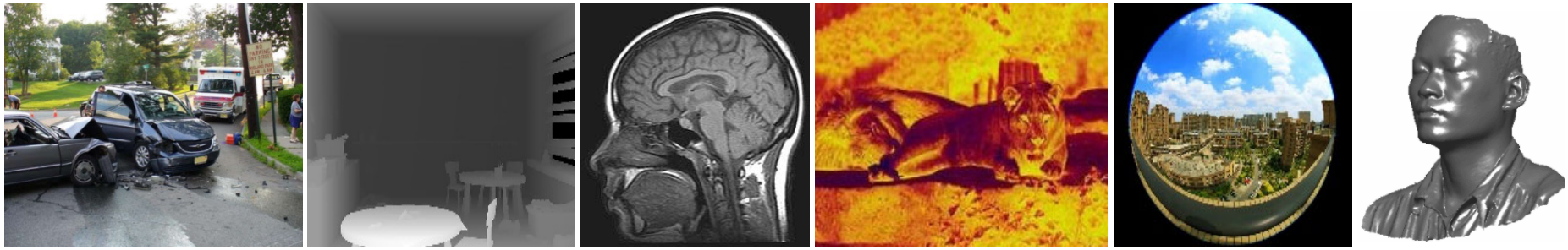
What we see

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

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 computer vision?

camera on mobile phone

surveillance camera

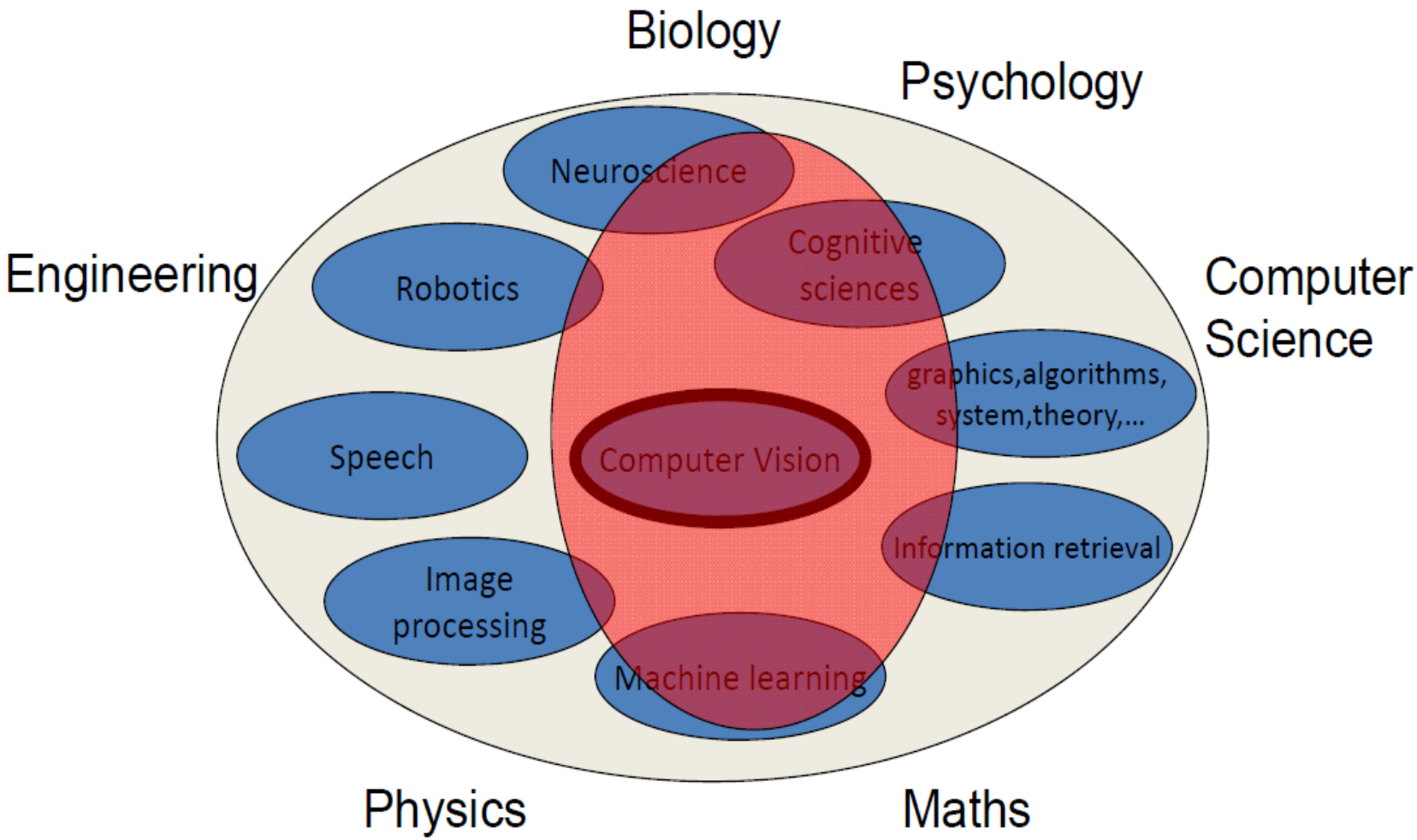
fish-eye camera

RGB-D camera

3D scanner

Commonly used sensors in CV

What is it related to?



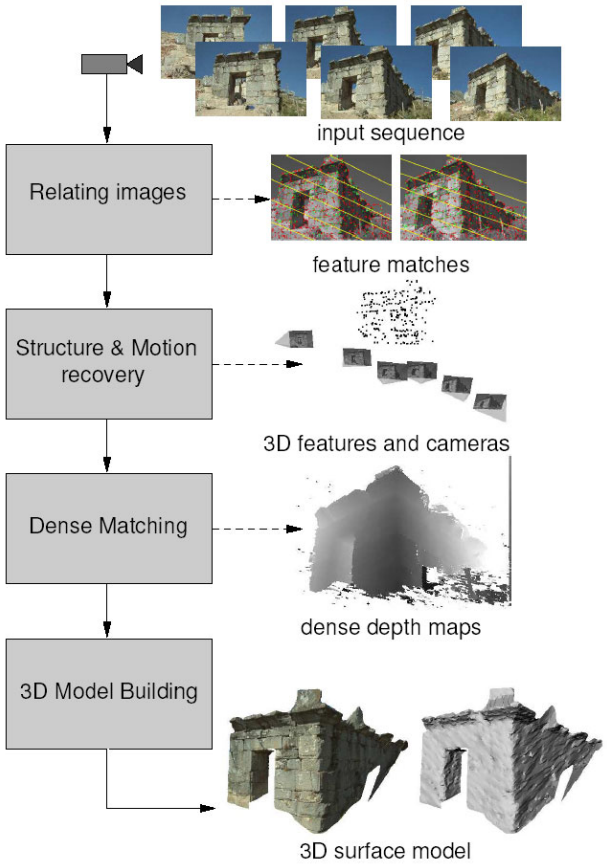
Vision as a measurement device

Real-time stereo

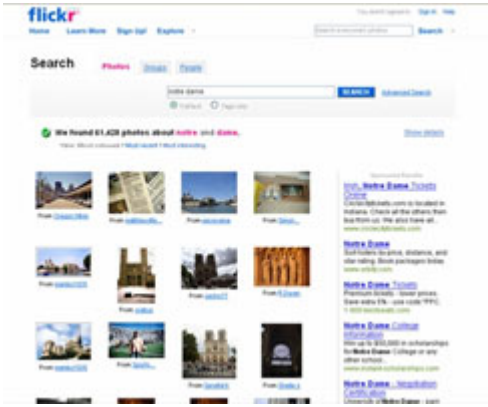
NASA Mars Rover

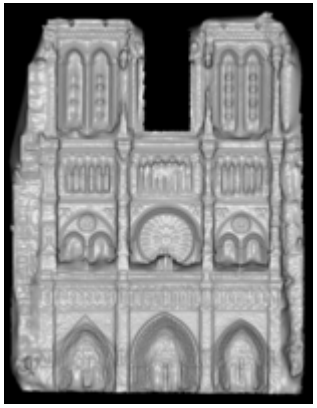
Pollefeys et al.

Structure from motion



Reconstruction from Internet photo collections





Goesele et al.

Vision as a measurement device

Paper ID: 1706
Submission to ACM Multimedia 2024

3D Reconstruction (L. Li, L. Zhang et al, ACMMM 2024)

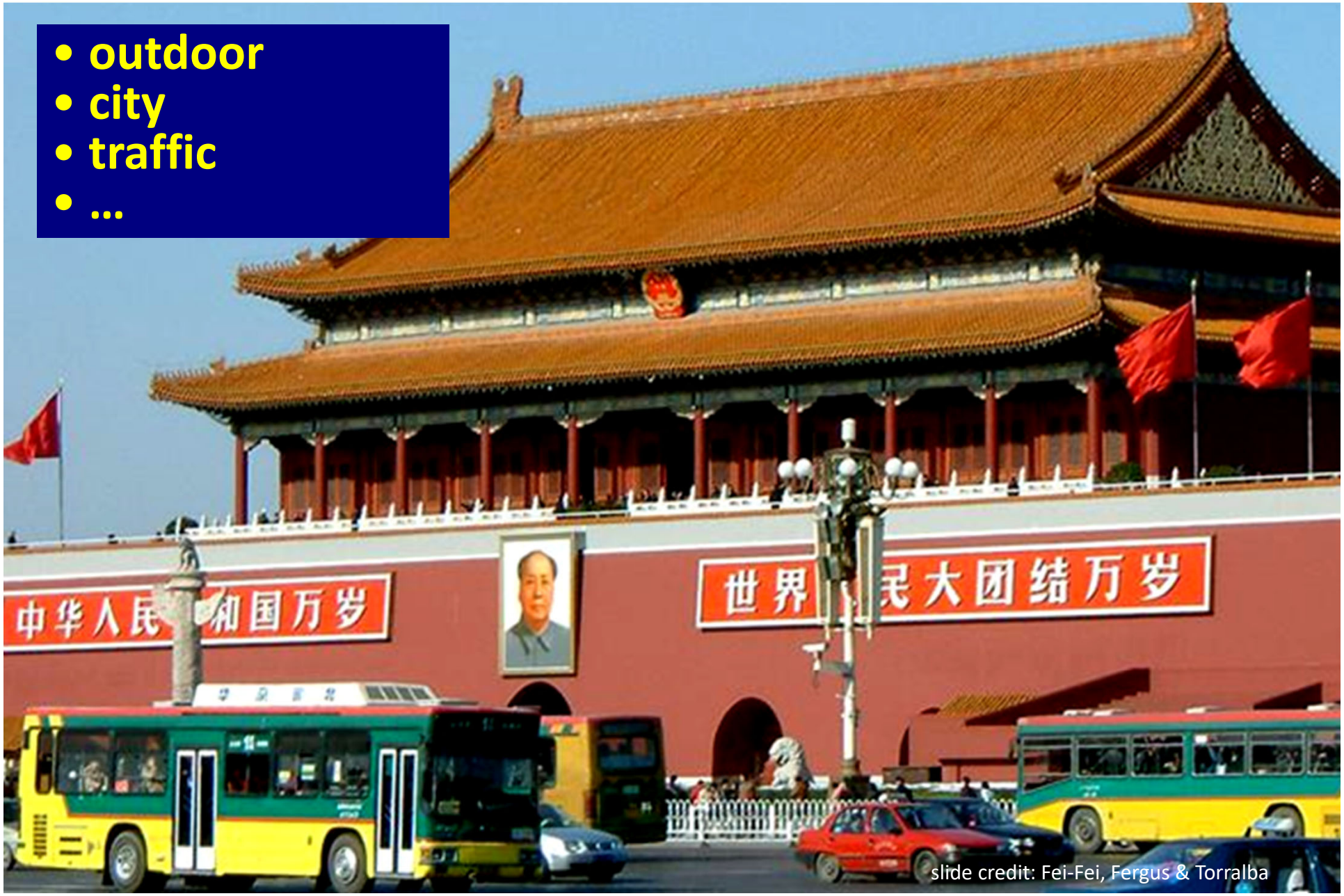
Vision as a source of semantic information

slide credit: Fei-Fei, Fergus & Torralba

Object categorization

slide credit: Fei-Fei, Fergus & Torralba

Scene and context categorization



- outdoor
- city
- traffic
- …

slide credit: Fei-Fei, Fergus & Torralba

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”

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

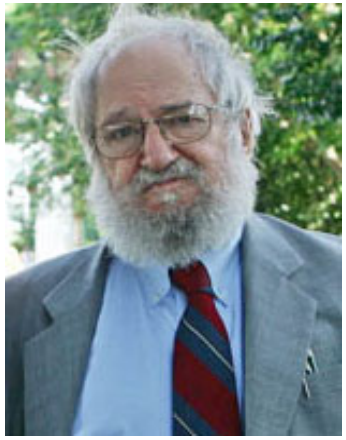
Artificial Intelligence Group July 7, 1966

Vision Memo. No. 100.

THE SUMMER VISION PROJECT

Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into



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

A brief history about CV (in 1980s)

- CVPR (IEEE International Conference on Computer Vision and Patter 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)

Prof. David Lowe. From 2015-2018, David Lowe was a Senior Research Scientist with Google in the Machine Intelligence Group. From 2009-2015 he was co-founder and chairman of Cloudburst Research, a computer vision startup in Vancouver that was acquired by Google in May 2015. From 1987-2015, he was a professor of computer science at the University of British Columbia

SIFT features matched between the two images and the transform computed from the matches using RANSAC

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

- In 1999, Zhengyou Zhang, a Chinese researcher in Microsoft, invented the camara calibration algorithm using planar calibration patterns

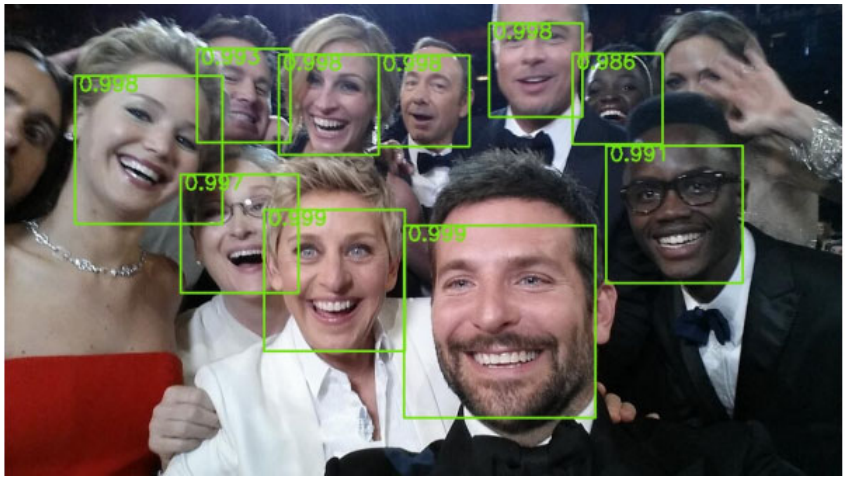
Zhengyou Zhang, IEEE and ACM Fellow, a recipient of the 2013 Helmholtz Test of Time Award which was awarded to him by the ICCV. Zhang earned a degree in electronic engineering from the Zhejiang University in 1985 and two years later earned a master's degree in computer science from the University of Nancy where his advisor was Jean-Paul Haton. In 1990 he defended his PhD at the University of Paris-Sud where his adviser was Olivier Faugeras and then received the doctor of science degree at the same place in 1994. After 20 years of work at Microsoft Research in Redmond, Dr. Zhang joined Tencent in 2018 and is currently leading both the Robotic X Lab and AI Lab at Tencent

Converting a fisheye image into a panoramic, spherical or perspective projection

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

- In 2001, Paul Viola and Michael Johns invented Adaboost+Haar features based real-time face detection algorithm

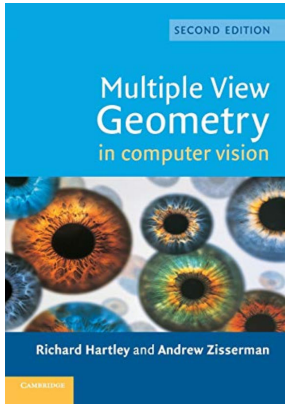
Paul Viola is a computer vision researcher, and Distinguished Engineer at Zoox. He is a former MIT professor, a former vice president of science for Amazon Prime Air and a former Distinguished Engineer at Microsoft. He is best known for his seminal work in facial recognition and machine learning. He is the co-inventor of the Viola–Jones object detection framework along with Michael Jones. He won the Marr Prize in 2003 and the Helmholtz Prize from the ICCV



Real-time face detection

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

- 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

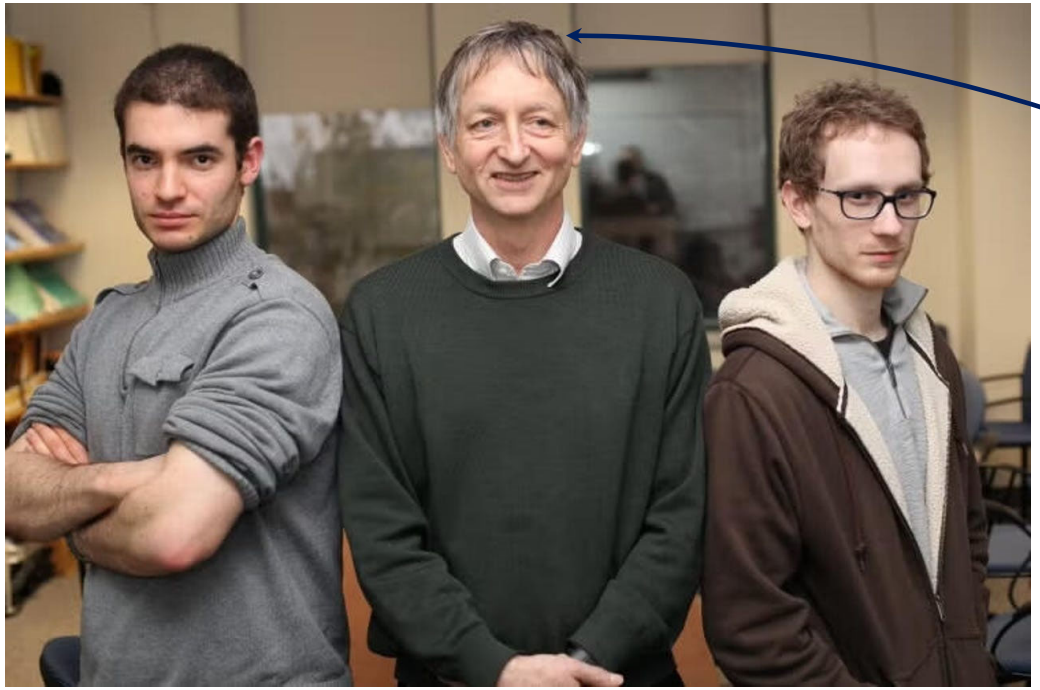
A brief history about CV (from 1990 to and 2010)

- 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, Ilya Sutskever, and Geoffrey Hinton (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



Won the ACM Turing Award in 2018

University of Toronto professor Geoffrey Hinton, center, and graduate students Ilya Sutskever, left, and Alex Krizhevsky, right, in 2013

A brief history about CV (Since 2010s)

- In 2016, Kaiming He, Jian Sun, and Xiaoou Tang invented the ResNet, which is now the basic module of almost all the modern network architectures

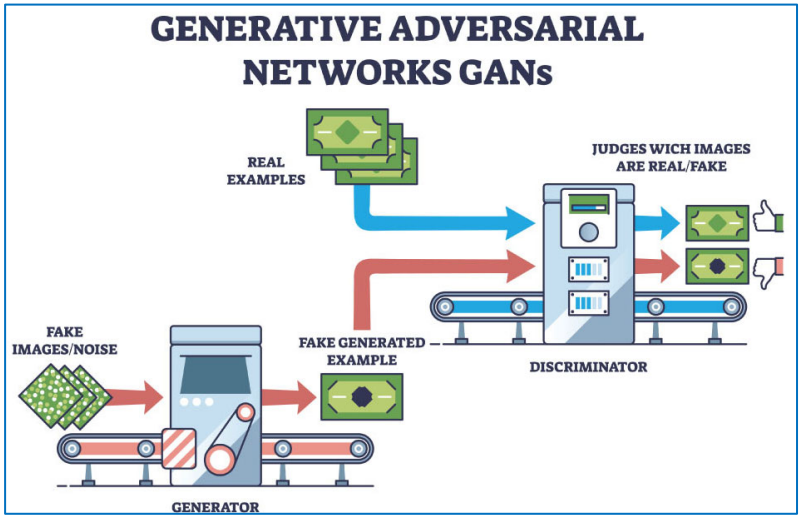
Xiaoou Tang and Kaiming He, in CUHK, 2009

Kaiming joined EECS, MIT as a faculty member in 2024

A brief history about CV (Since 2010s)

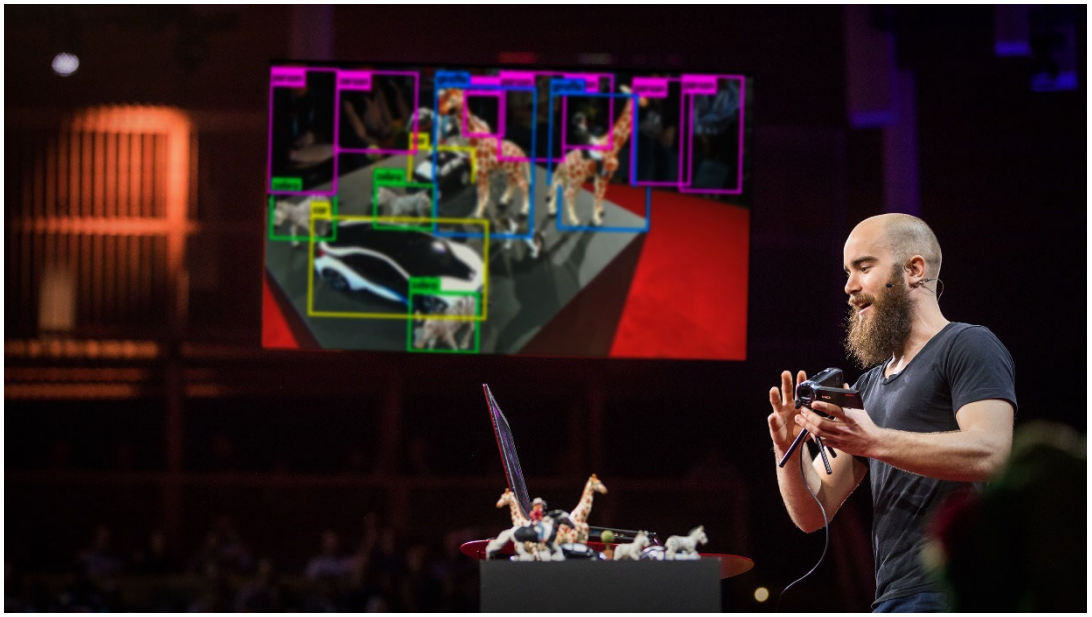
- In 2014, Ian Goodfellow (Montreal University, Canada) et al. proposed Generative Adversarial Networks, which is now widely used for self-supervised learning

Ian J. Goodfellow (born 1987) is an American computer scientist, engineer, and executive, most noted for his work on artificial neural networks and deep learning. He is a research scientist at Google DeepMind, was previously employed as a research scientist at Google Brain and director of machine learning at Apple as well as one of the first employees at OpenAI, and has made several important contributions to the field of deep learning, including the invention of the generative adversarial network (GAN)



A brief history about CV (Since 2010s)

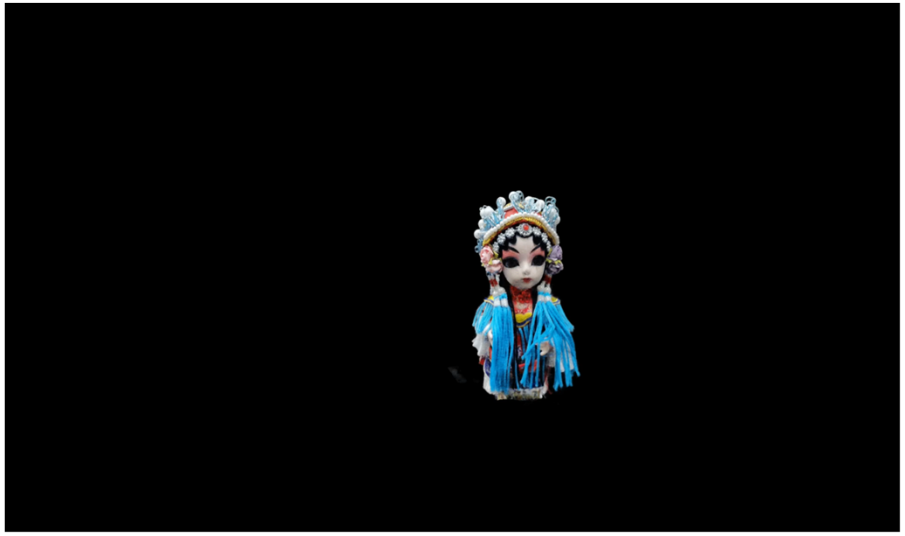
- In 2016, Joseph Redmon invented the YOLO algorithm, which is now a widely used (the most successful in my opinion) algorithm for real-time object detection



Joseph Redmon is an American computer scientist best known for developing the YOLO (You Only Look Once) object detection algorithm, which revolutionized real-time computer vision. His work has had a profound impact on fields such as autonomous driving, surveillance, robotics, and more. Redmon earned his Bachelor's degree in Computer Science from the University of Washington. He pursued his Ph.D. at UW under the supervision of Prof. Ali Farhadi (a prominent AI researcher now at Apple).

A brief history about CV (Since 2010s)

- 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



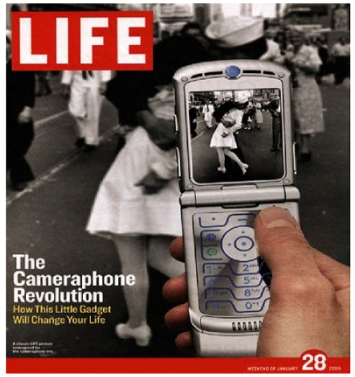
A brief history about CV (Since 2010s)

- In 2023, Bernhard Kerbl (Inria, France) et al. proposed to use 3D Gaussians to represent the scene

Today

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

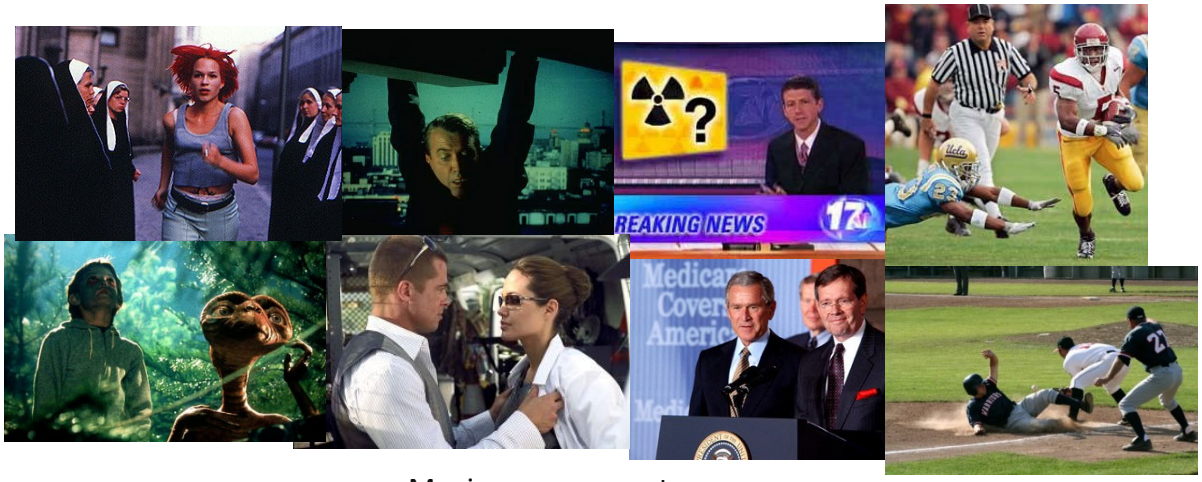
Why study computer vision?



Why study computer vision?

Vision is useful: Images and video are everywhere!

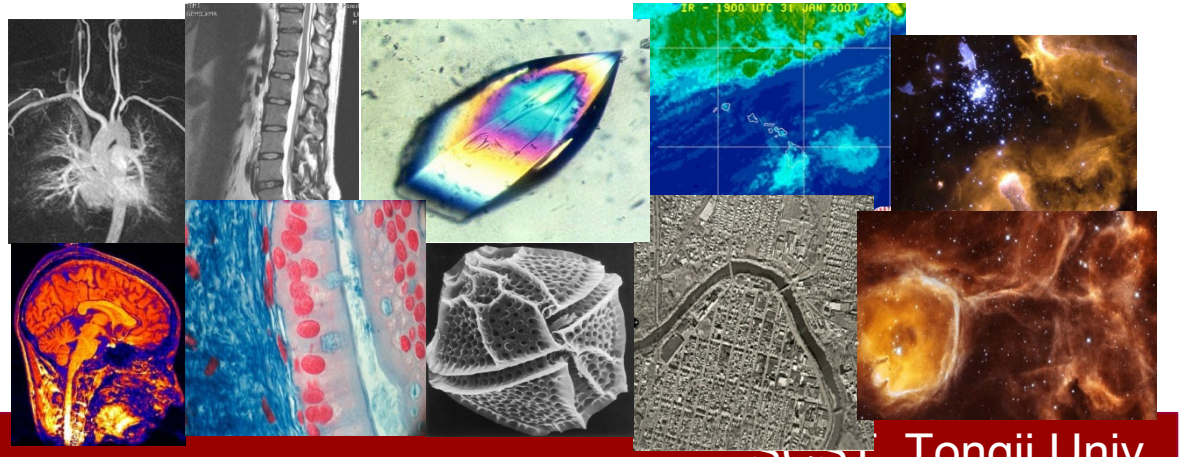
Personal photo albums



Movies, news, sports

Google Image Search · Picasa · flickr · webshots · picsearch · YouTube Broadcast Yourself

Surveillance and security



Medical and scientific images

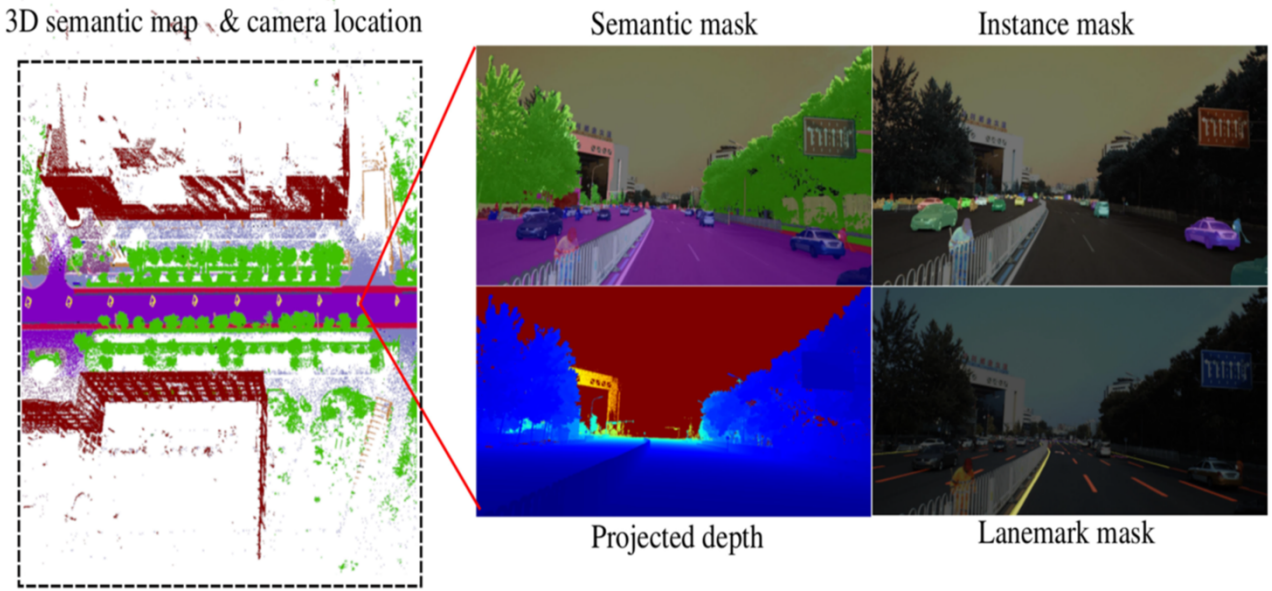
Why study computer vision?

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



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

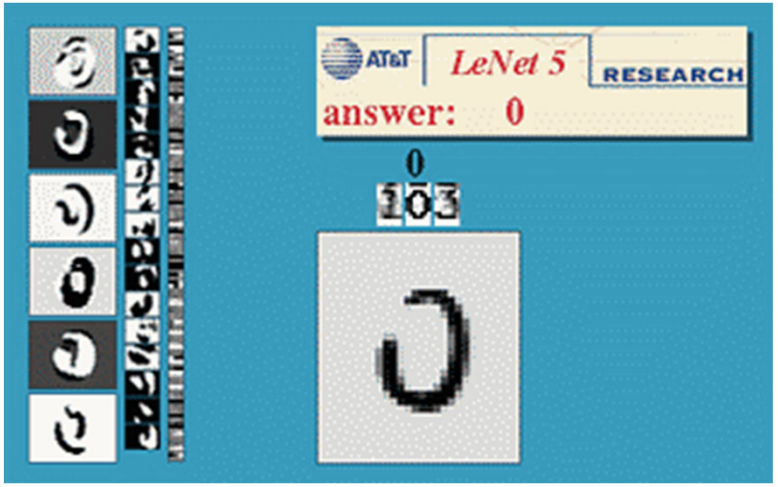
AVP (Automated Valet Parking)

百度AVP演示系统

Optical character recognition (OCR)

Technology to convert scanned docs to text

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



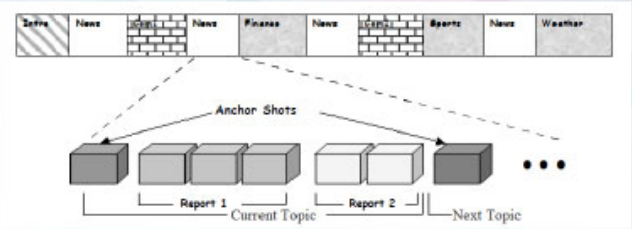
Digit recognition, AT&T labs
http://www.research.att.com/~yann/

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



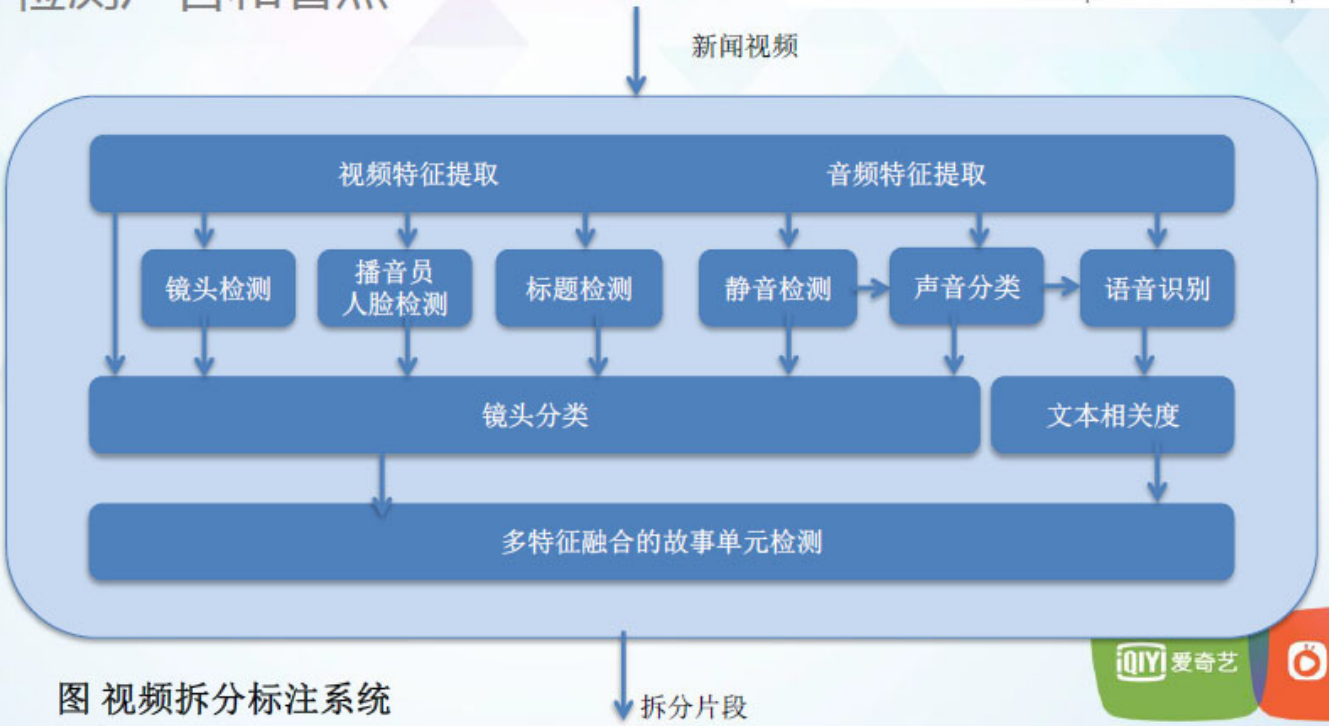
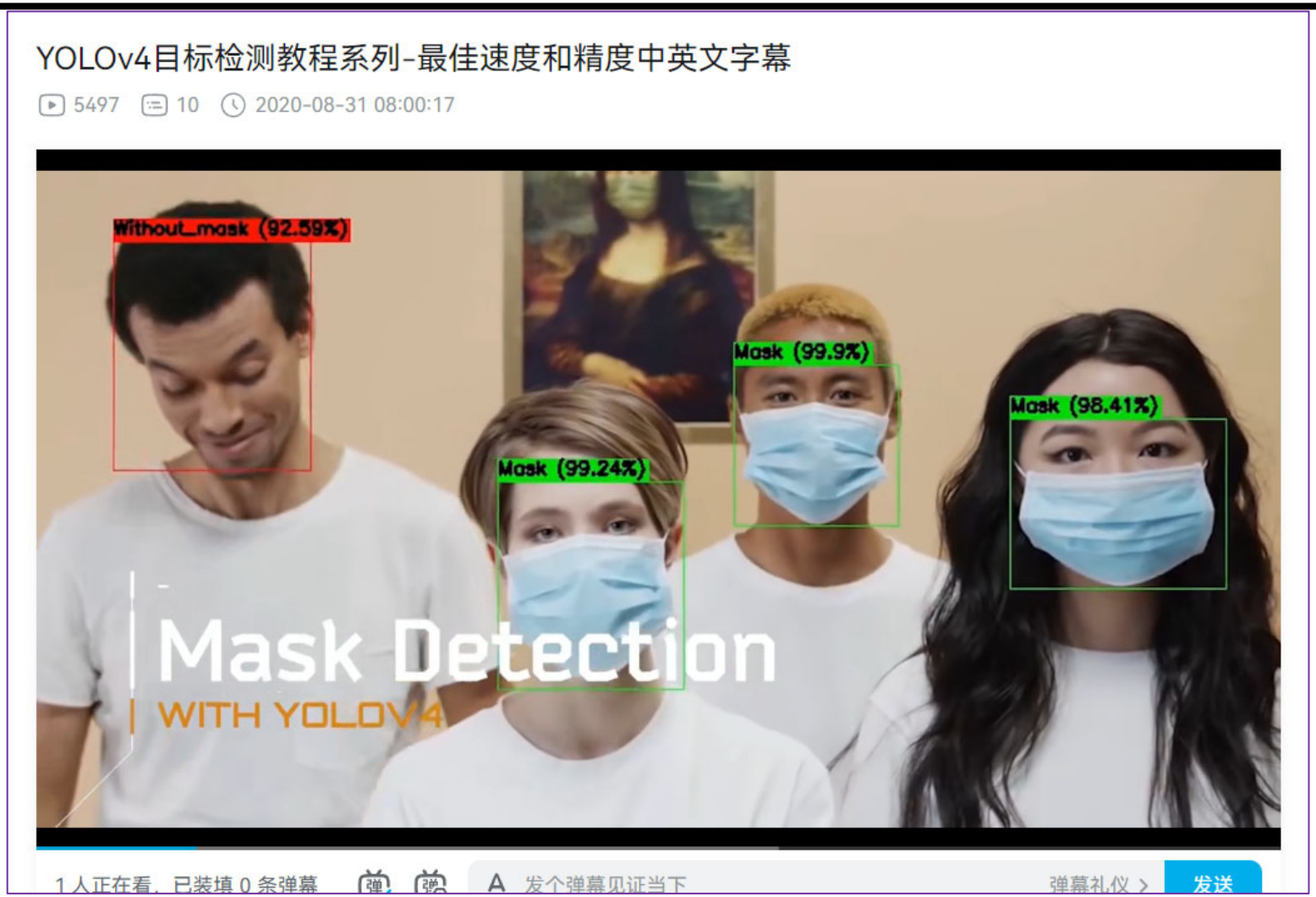


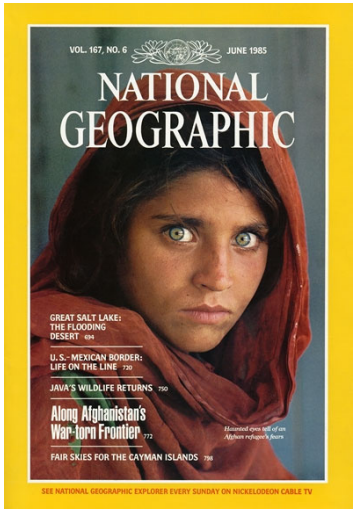
图 视频拆分标注系统

Face and mask detection

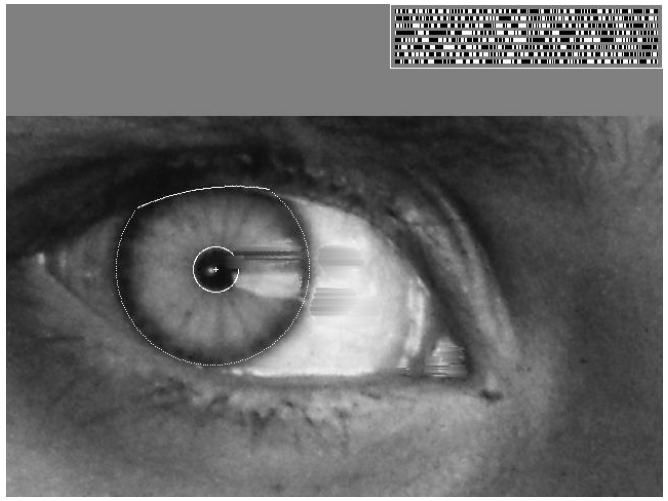


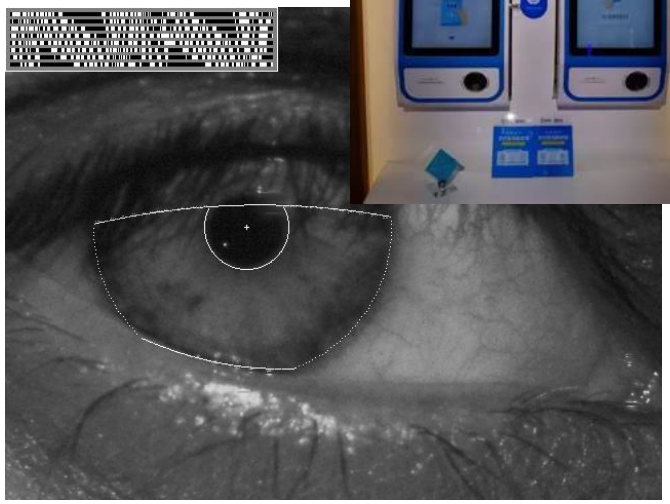
https://www.bilibili.com/video/av926885838?p=15

Vision-based biometrics

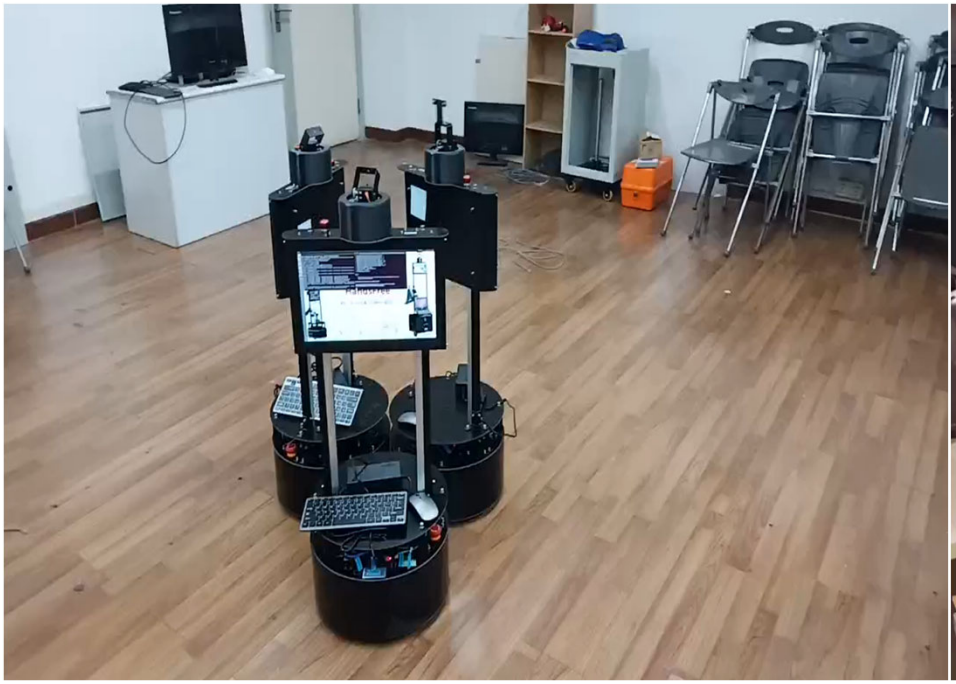


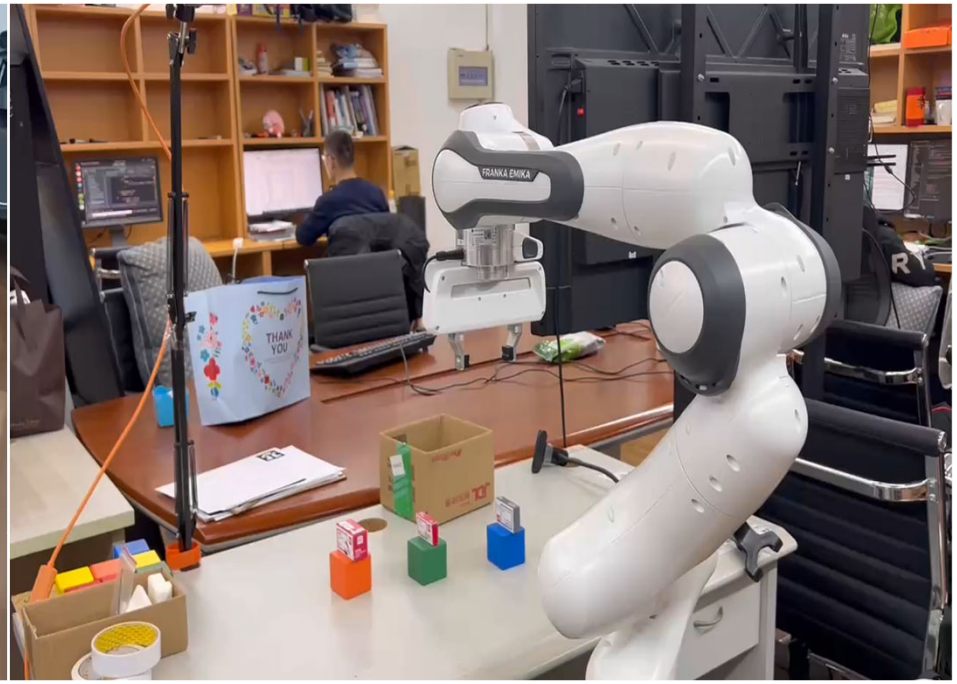
"How the Afghan Girl was Identified by Her Iris Pattern





Robotics





Robotics in MAP (Machine Autonomous Perception) lab

Work by our students: Photosynth

Project products of students from 2009 Media&Arts

Recent work of our group

Palmprint and palmvein recognition

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

Recent work of our group

Palmprint verification on mobilephones

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

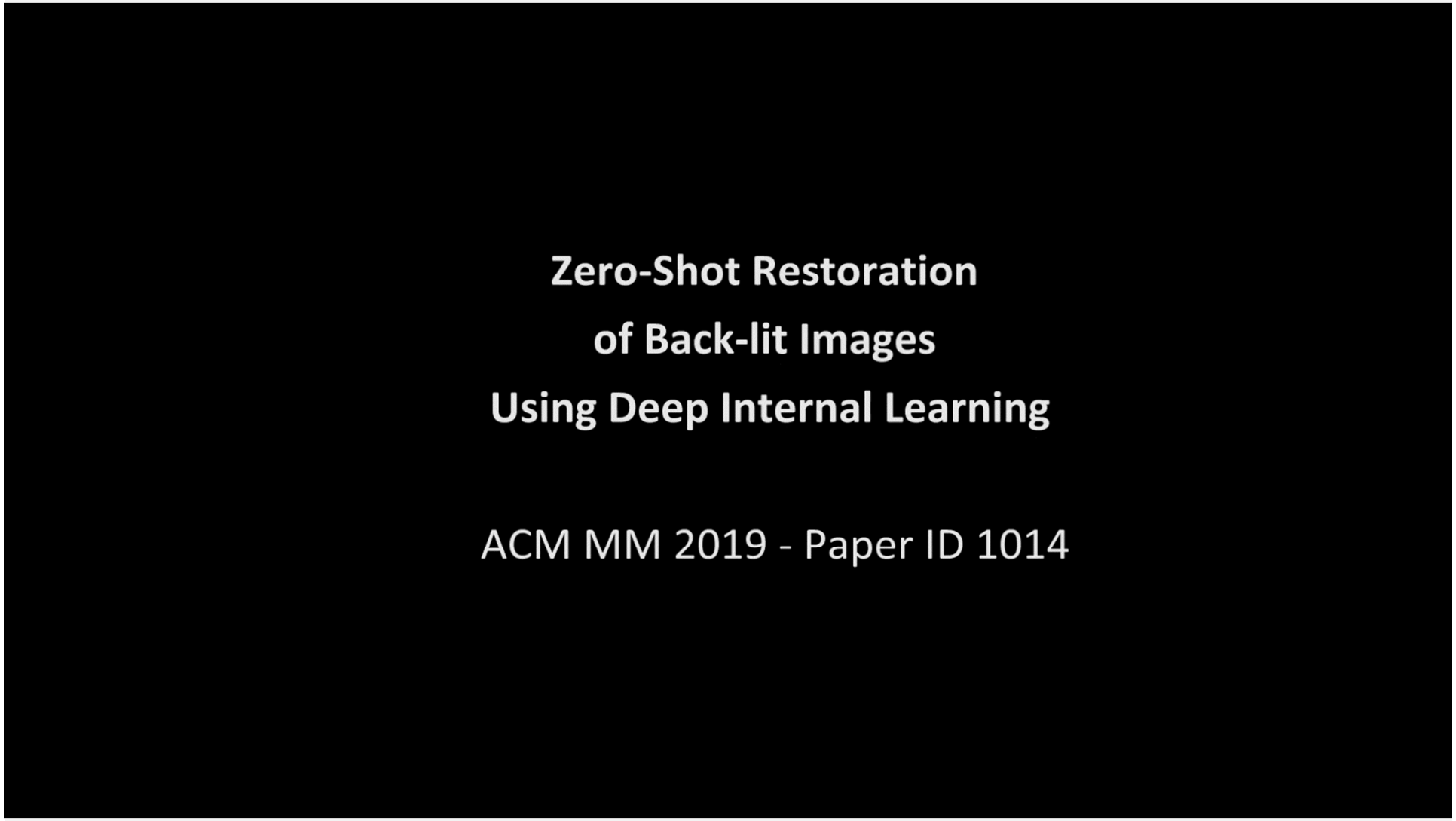
Recent work of our group

图像曝光度的自动校正

Recent work of our group

图像曝光度的自动校正

Recent work of our group



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

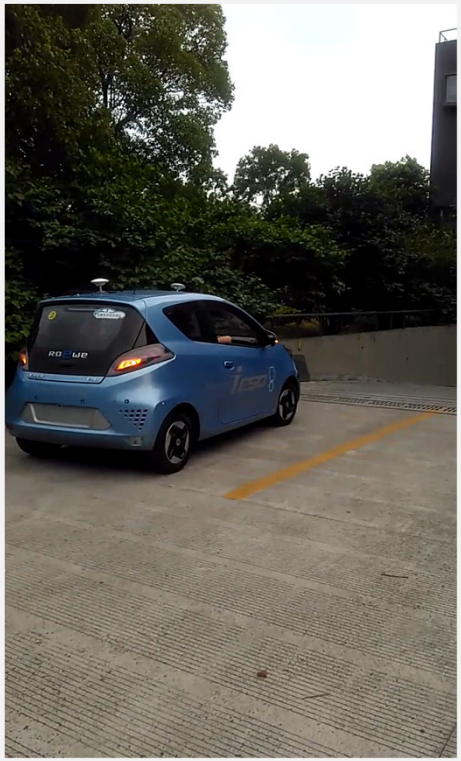
Recent work of our group

Simulation of atmospheric visibility impairment

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

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.

Recent work of our group

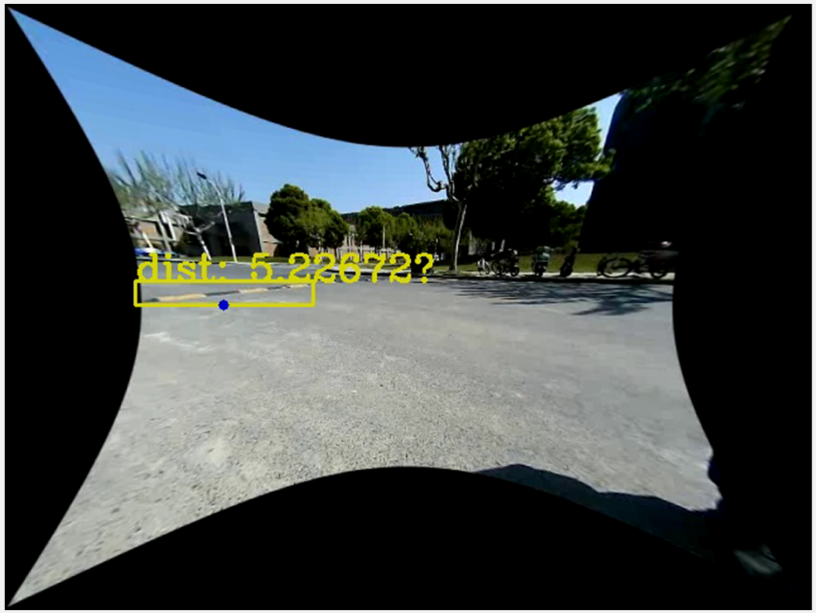
2018年9月，人民网采访报道

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

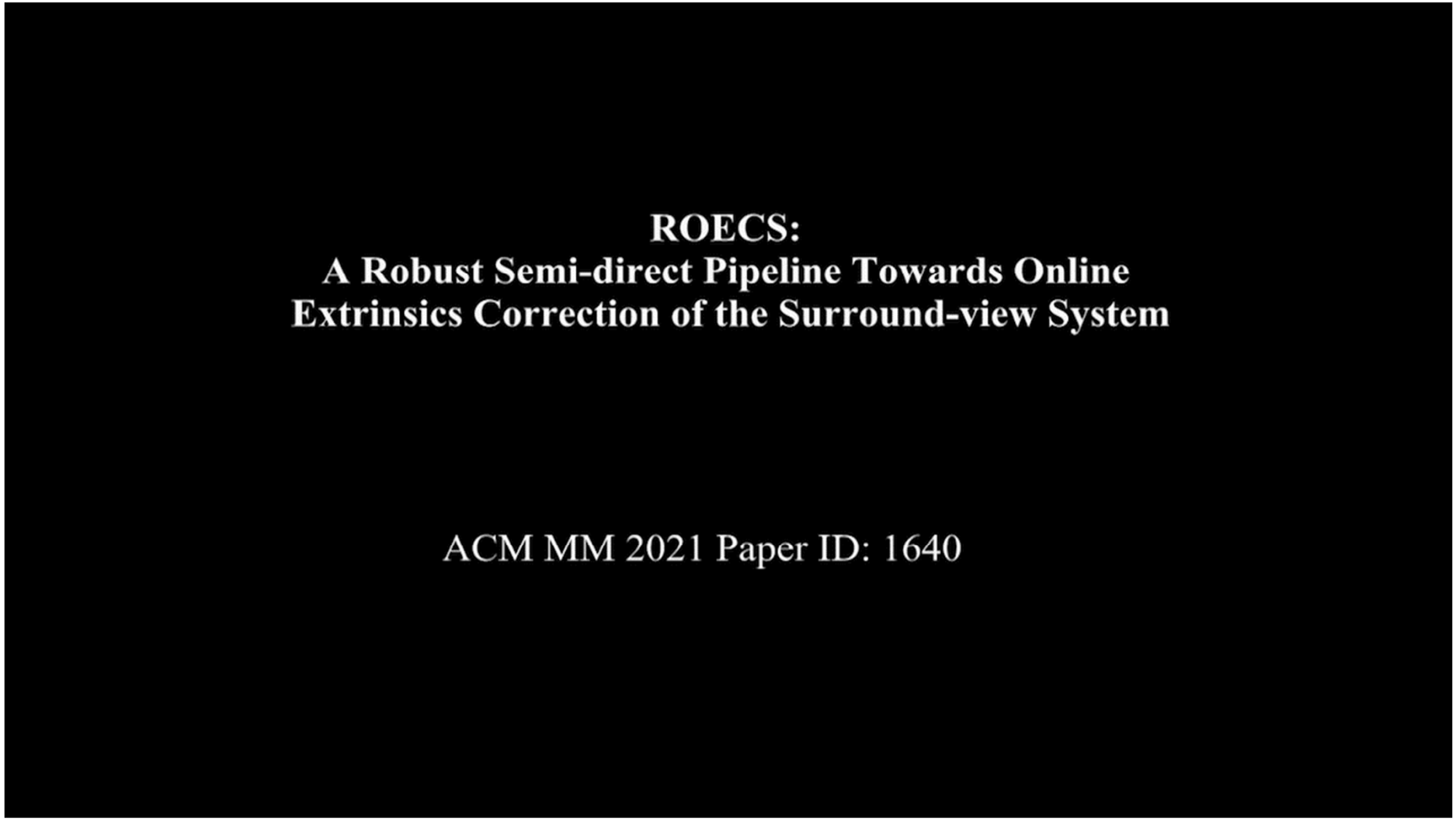
Recent work of our group

Pedestrian and speed-bump detection and distance measurement



Recent work of our group

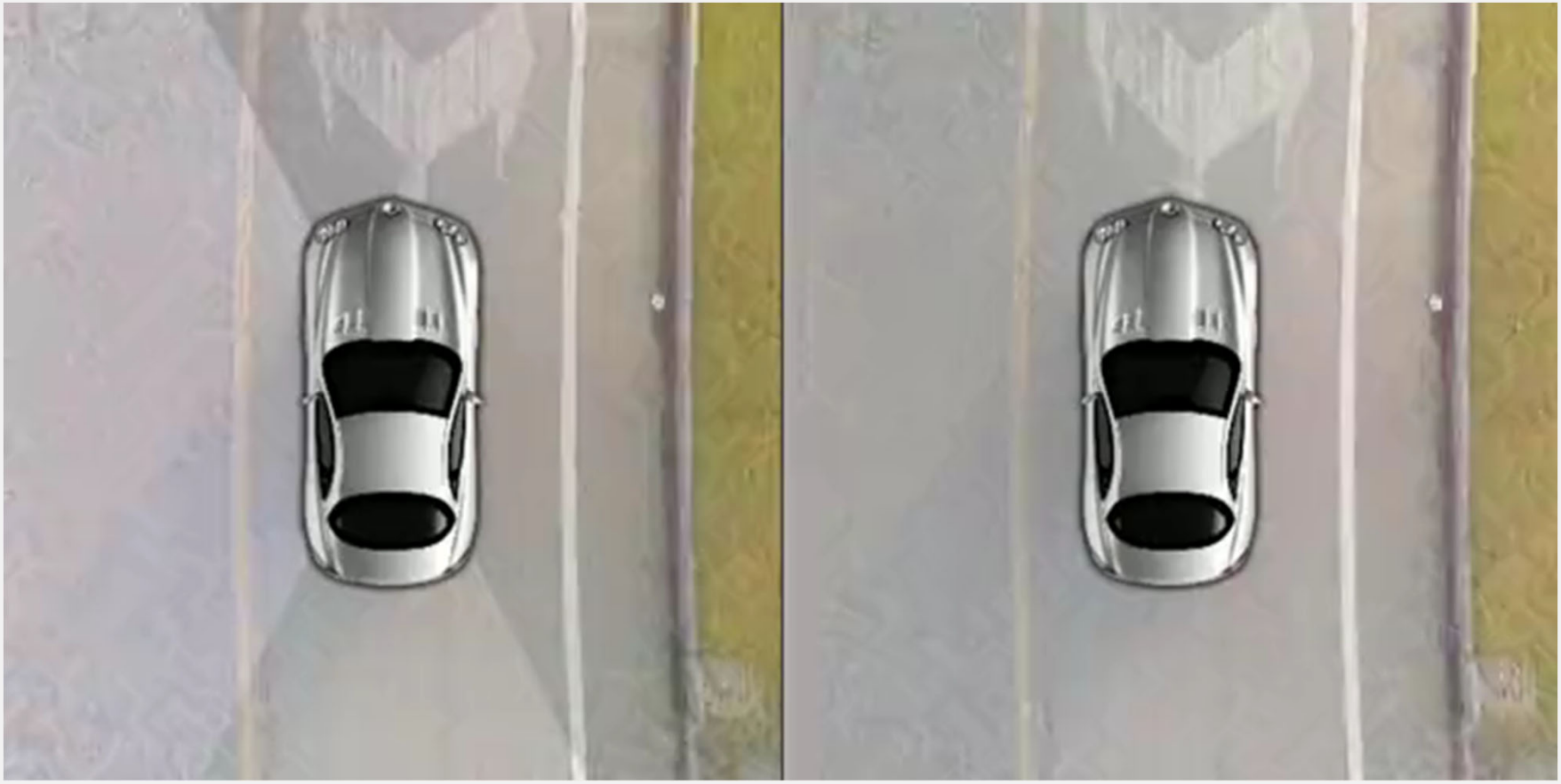
Online optimization of camera poses in a surround-view system



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



Recent work of our group

SLAM for indoor parking environments

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

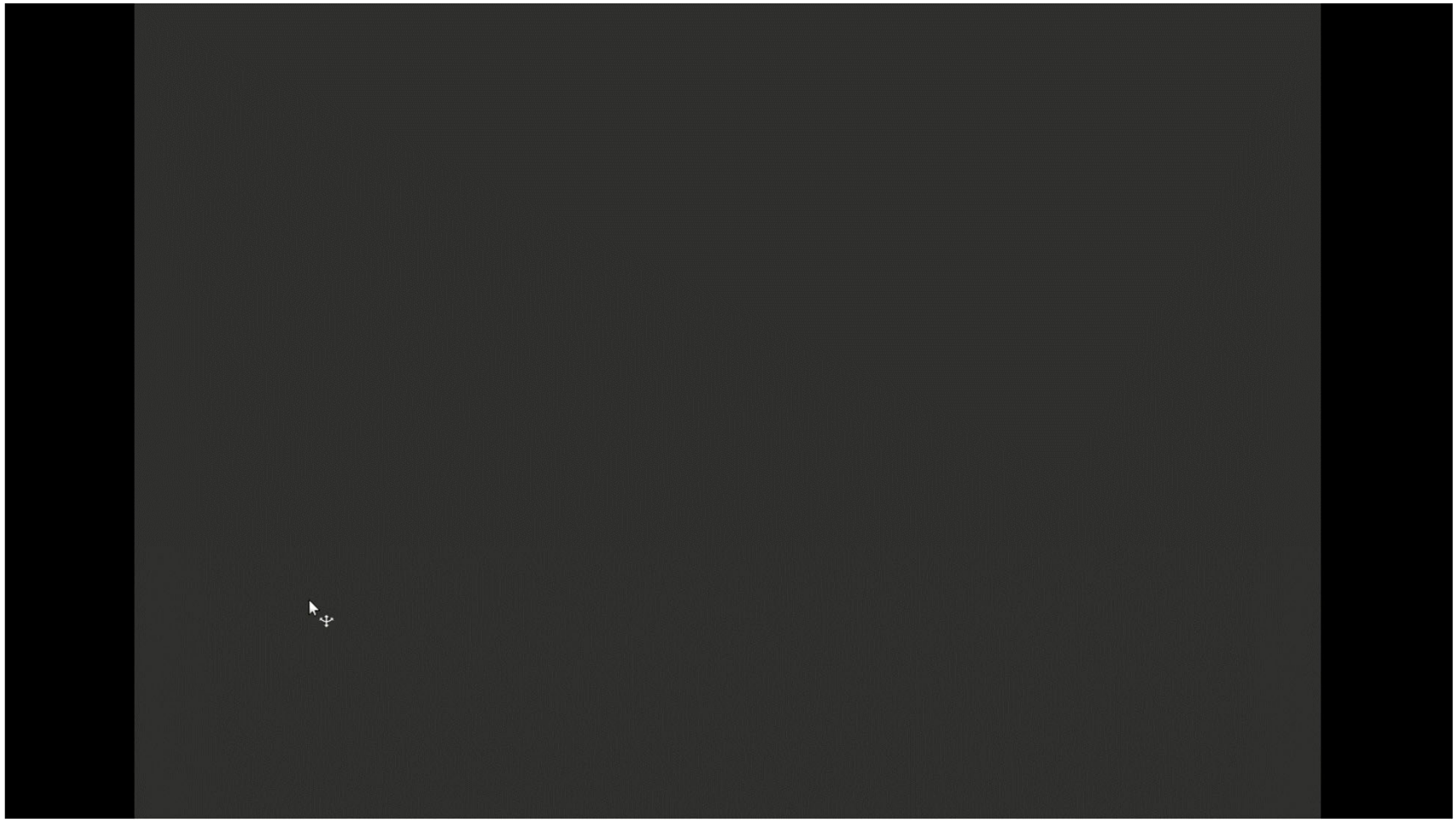
Recent work of our group

Panoramic video stitching

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.

Recent work of our group

Multi-agent 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, 2022

Recent work of our group

Tightly-coupled direct LiDAR-inertial odometry and mapping

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

Recent work of our group

Sound localization in noisy environments

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

Recent work of our group

Human pose tracking using sparse IMUs

Kaixin Chen, Lin Zhang* et al., "Skeleton-aware Graph-based Adversarial Networks for Human Pose Estimation from Sparse IMUs," ACM Transactions on Multimedia Computing, Communications, and Applications, 2024.

You can find a good job!

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

Today

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

Course content

- Introduction
- Linear geometric transformations
- Local interest point detectors
- Local feature descriptors and matching
- Homography estimation
- Projective geometry
- Nonlinear least squares
- Imaging model and camera calibration
- Surround-view
- Convex optimization and SVM
- DCNN-based object detection
- Binocular stereo vision
- Radiance field based rendering and NeRF

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)

Q+A