ENGN 8501

Advanced (selected) topics in computer vision

2021 S2

Overview of ENGN8535

- Course overview
- Logistics, and administrative matters;
- Wattle configuration
- Lectures
- Tutorials

- Misc.

ENGN8535: Course Overview

- This is an 8000 series upper-level Mater's and entry level PhD course in Computer Vision.

- Targeted majors include Machine Learning and Computer Vision, Advanced Computing, Data Analysis and Science, Mechatronics, AI and Intelligent Systems.

- This course contains **supervised paper reading**, paper reviewing report, and group discussion, as well as a group research project.

Last year's offering

ENGN8501: Advanced Topics in Computer Vision

Week 1

Dylan Campbell

dylan.campbell@anu.edu.au

Australian National University

Semester 2, 2020

- **Course Staff:**

  Dylan (convener)

An excerpt form last year's lecture note

- **Warning:** This is an advanced course. You are expected to know your own abilities (and whether or not you have the right background for the course). You are expected to work diligently throughout the semester. We will endeavour to answer questions in lectures, tutorials and online forums, but research can be open-ended and ambiguous. This is part of the challenge, and intended to get you questioning and thinking like a researcher at the cutting-edge of computer vision and machine learning.

Aim of the course

- **Review current research topics** in computer vision
- **Develop your research skills** as a precursor to further study or work in computer vision, machine learning or AI
  - Critically reading research papers
  - Finding good research problems
  - Designing effective and novel solutions
  - Implementing using modern techniques

Expected learning outcomes

Upon successfully completing the course, you will have the knowledge and skills to:

- describe and analyse the **main research challenges** in the field of computer vision;
- **summarise research literature** and state-of-the-art techniques for solving the challenging research problems in those areas;
- **model and formulate problems**, **propose effective solutions** to the problem and **implement algorithms** using suitable programming languages;
- **design network structure and loss functions** in cases where problems need to be solved using deep learning techniques;
- **analyse the results** and effectively **evaluate the results** on benchmark datasets.

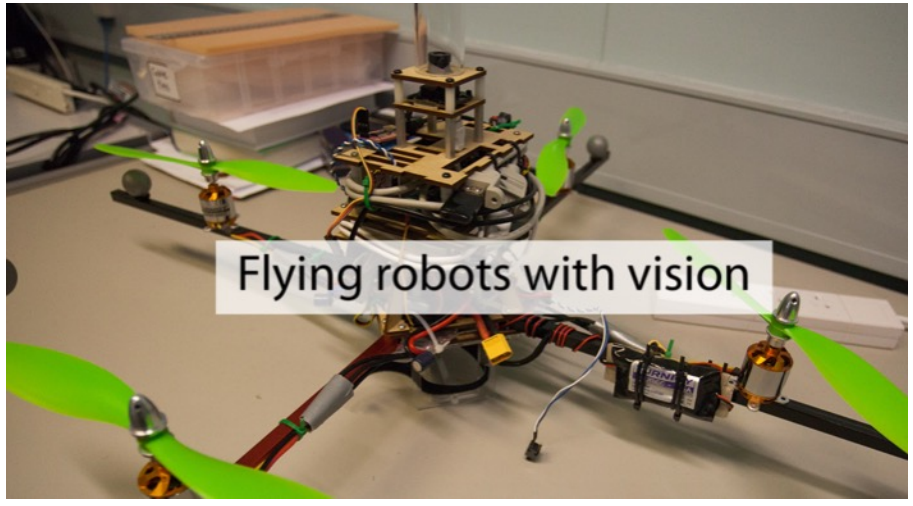
Know your Lecturer

- Professor Hongdong Li;
- Founding CI (Chief Investigator) for Australia ARC Centre of Excellence for Robotic Vision

- Office Hours: Monday 14:00-15:00, via Zoom or in the office.
- Monday Lecture time: 15:00—17:00, via Zoom @ lecture theatre.

ARC Centre of Excellence for ROBOTIC VISION

QUT Queensland University of Technology
MONASH University
ANU THE AUSTRALIAN NATIONAL UNIVERSITY
THE UNIVERSITY of ADELAIDE

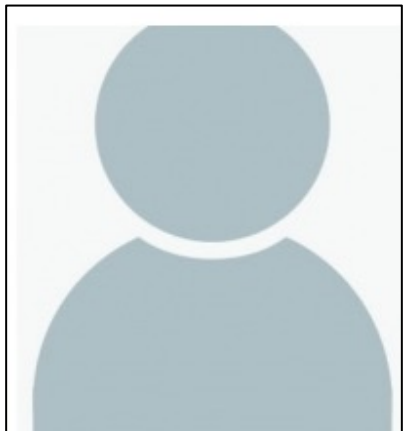


Know Your Tutors

- **Sahir Shrestha**
- Sahir.Shrestha@anu.edu.au

- Huiyu Gao <Huiyu.Gao@anu.edu.au>

- Tianyu Wang <Tianyu.Wang2@anu.edu.au>



Lecture and Tutorial timetable

- Zoom + Physical.
- **Lecture**: Monday, 15:00-17:00 (Zoom + Physical) , STB-S1 (307).
- **Tutorial**: Thursday, 13:00—15:00; CSIT N111 (Physical + Zoom)

Wattle configuration and useful URL links

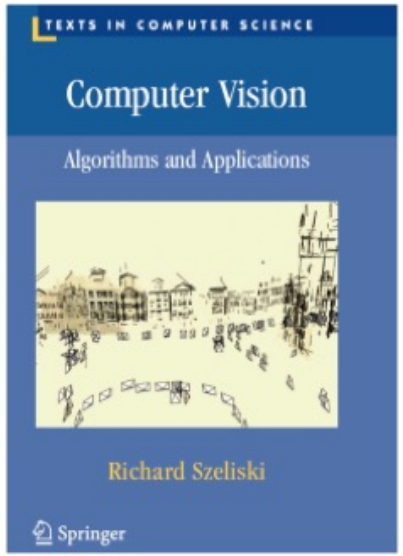
- Wattle page: https://wattlecourses.anu.edu.au/course
- Weekly Study plan: (living GoogleDoc) https://docs.google.com/document/d/1nb7jKzylIlL4Kz3Z1OGTUViulfCEZBoxGZd0trJPy2I/edit?usp=sharing
- Report Latex template: https://www.overleaf.com/read/tbvvpwgcfyyb
- Zoom link: https://anu.zoom.us/j/85451808066?pwd=NWRwN2svdkpqYzhDdzJWRTdJRTRKZz09
- **Meeting ID: 854 5180 8066**
- **Password: 306195**

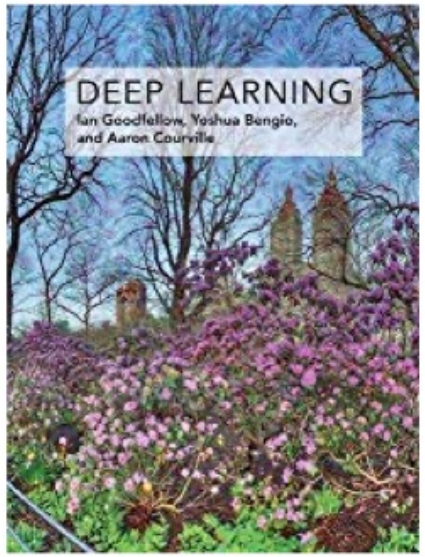
Prerequisites

- Calculus
- Linear algebra and matrix computation
- Probability theory
- Numerical optimization

- Have studied ENGN4528 or ENGN6528 before.

Recommended background readings





Reference textbooks:

- Szeliski, *Computer Vision: Algorithms and Applications*, 2011
  `szeliski.org/Book`
- Goodfellow et al., *Deep Learning*, 2016
  `deeplearningbook.org`

Assumed Computer Skills

- Python , numPy, SciPy (must)
- Github or GitLab (must)
- Latex (must)

- PyTorch/TensorFlow (optional)

Format of each lecture – supervised paper reading exercise

1. pre-lecture self-reading of 1~2 “in-class papers”. Bring the paper to classroom (printed, or if applicable, on a separate device (iPad/tablet)).

2. General technical background introduction;
3. Guided paper reading (1~2 papers) for about 0.5 hour to 1 hours.
4. Break-out room group discussion (10 minutes); remember the break-out room ID (by writing down on a piece of paper), the name of students in the same room, and for each breakout room please select/nominate one “speaker” (also indicate the name of the speaker on the piece of paper).
5. Questions to be discussed at the breakout room: will be explained later…
6. Selected groups report back, via zoom /shared screen (3 minutes per group, randomly choose 5~6 groups, totally about 20 minutes).

Format of tutorial sessions

- **Tutorial**: (physical + zoom).
- Thursday, 13:00—15:00; hybrid (physical CSIT N111 +Zoom).

- Read one extra paper (on the same topic as the current/preceding week), group discussion.

- **And/Or:** Free Q&A session.

- **Or:** research project mid-term presentation and checkpoints.

Assessment Components

- **Paper Reading Reports**
  - 10% x 5 = 50% marks.
  - No more than 2 pages long, in CVPR'21 Latex format.
  - → upload a single PDF file each report to Wattle. (PDF is the only accepted file format.)

- **Group Research Project:**
  - **Based on any one paper we have in the pool → Reproduce, tests, and improve.**
  - Group project (up to 3 students per group; I.e., 1 to 3 students per group are acceptable).

  - Project Proposal (1 page PDF, in CVPR format): 5% marks. (due in week-5).
  - Project report (up to 6 pages PDF (**excluding** references), in CVPR latex format) : 20%.
  - Source Code (in a single ZIP file with readme file included: 5%, based on originality and quality).
  - Project presentation (12 slides in 10 ~ 12 minutes live or recorded MP4 video. PPT + face video): 20%.
  - Coding environment: Python on CPU PC/Laptop (mostly), or GPU (not recommended) . No C/C++, no OpenCV non-I/O functions, no Matlab.

No Final Exam.

Expected study load (10~12 hours per week)

- Total 10 hours =
  - Pre-reading in-class paper: 0.5 hours per week.
  - + Attend lecture/tutorial: 3 hours per week.
  - + Home-reading papers: 2 hours per week.
  - +Writing paper reading report: 2 hours.
  - +Discussing and working on your group Research Project: 2.5 hours per week.

“Group” Research Project

- Form your group. (1~3 students per group).
- Pick any one paper from all the papers listed on “Weekly Study Plan” (note: the list may add new papers as the semester goes on).
- Implement the main method, re-do necessary experiments on new date, think about whether you can further improve it, and test your improvement and extensions.
- Write a project report (6 pages in CVPR format, excluding references), and prepare a 10 ~ 12 minutes seminar presentation.
- What if you have found some reference codes on the Internet ?
  - *If you want to use existing codes, it is okay. You need to make sure the others’ codes are properly referenced, understood, modified and rewritten, and prove that your modifications are major and substantial.*
  - *In this case, new and substantial extensions/improvements are essential. Tutors and I will specifically look for new improvements and extensions. You must clearly explain them in your project report.*
  - *For more detailed instruction, please read the “Project Guideline”– to be released in week-3.*

Marks distribution for some ENGN4xxx courses

- HD: >= 80% ; ~ 10%
- D: >= 70%; ~30%
- CR: >=60%, 40%
- P: >=50% , 15%
- F: <50%, 5%

- average mark about 65%-69% after possible mark moderation.

Questions from you ?

Topics to be covered:

- Low Level Vision
- Computational Photography

- Physics based Vision

- Human Pose
- Graphical Models

- Scene Understanding & 3D vision.

Topics not covered:

- Medical Imaging
- Face recognition
- Vision and Language (NLP, GPT-3, CLIP)
- Remote sensing (satellite image)
- Light-field Camera
- Image classification, object detection, instance and semantic segmentation
- …

Selected Topics in (advanced) Computer Vision

- 1. Low level computer vision-A: **image deblur/super-resolution**
- 2. Low level computer vision-B: **camera shake removal**
- 3. Computational Photography-A: **coded capture**
- 4. Computational Photography-B: **natural image matting**
- 5. Physics based vision-A: **photometric stereo**
- 6. Physics based vision-B: **Photograph relighting**
- ===
- 7. Human Pose: 2D pose, 3D pose, **body shape**
- 8. Graphical models-A : **MRF and LBP**
- 9. Graphical models-B: **max flow and graph cut**

- 10. Indoor scene understanding-A: **3D room Layout**
- 11. Indoor scene understanding-B: **3D object parsing**
- 12. Outdoor scene understanding: **Autonomous Driving**

Papers to read

- **In-class reading papers: (12~24 papers)**
  - About 1~2 papers per lectures per week → Supervised reading and analysis.
  - Before attending the lecture, you must pre-read the paper
  - Bring the papers to classroom (either in printed version, or on iPad/tablet if applicable).
- **Home-reading papers: (many ..)**
  - You must pick one paper to read very carefully every two weeks, and write a 2-page (in CVPR format) reading report based on it, and submit the report to Wattle before due.
  - So totally, each student will submit in total 5 reading reports, totally 10 pages, worth 50% full marks.
- **Format of a Reading Report** (See Latex template. Will explain later…)

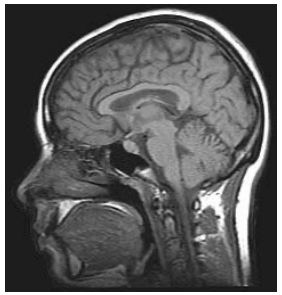
Questions ?

What is Computer Vision ?

What is Computer Vision?

- **Computer Vision is concerned with the theory of building artificial vision systems that obtain useful information from images.**

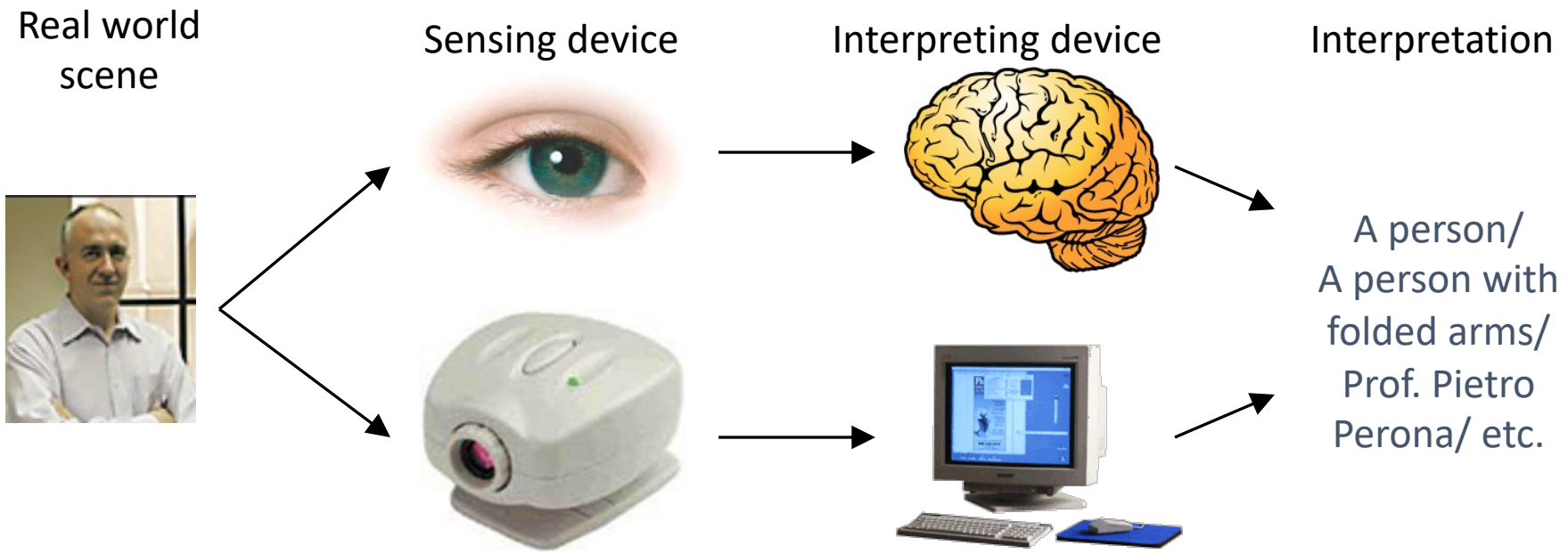
- The image data can take many different forms, such as a video sequence, depth images, views from multiple cameras, or multi-dimensional data from a medical scanner



The computer vision problem

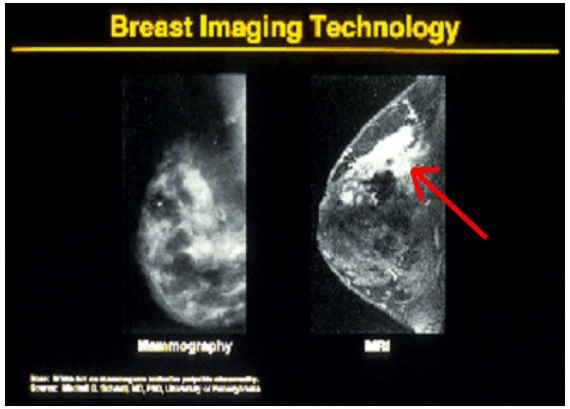
- Make a computer to see and to understand images

- We know it is physically possible – we do it every day and effortlessly!



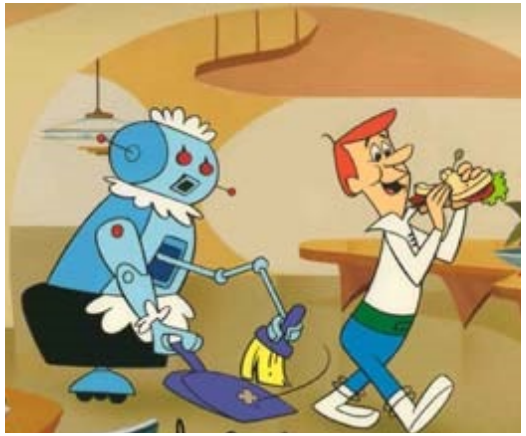
Why computer vision matters

Safety



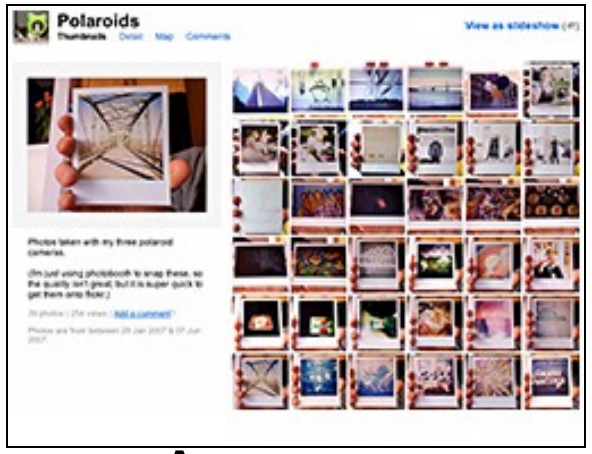
Health

Security



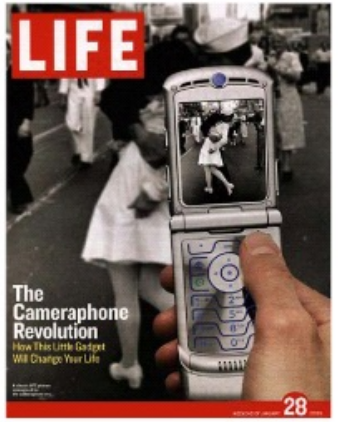
Comfort

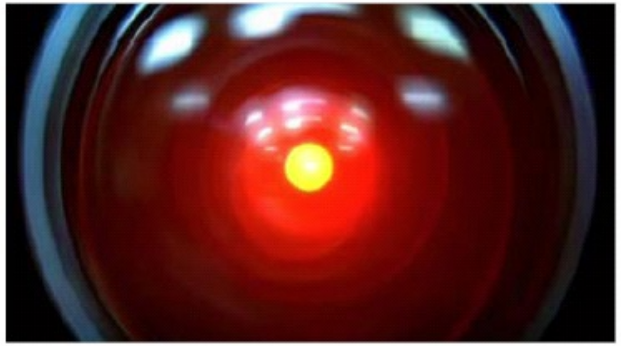
Entertainment



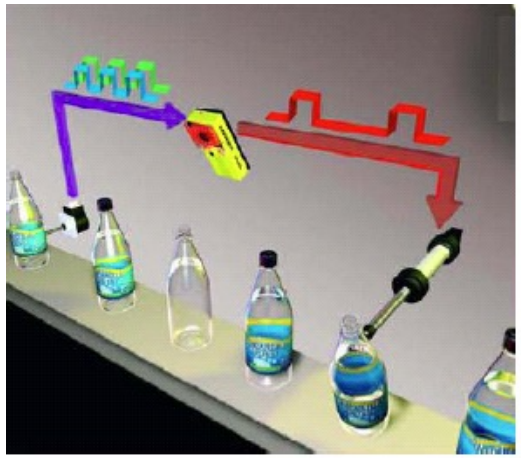
Access

Cameras Everywhere



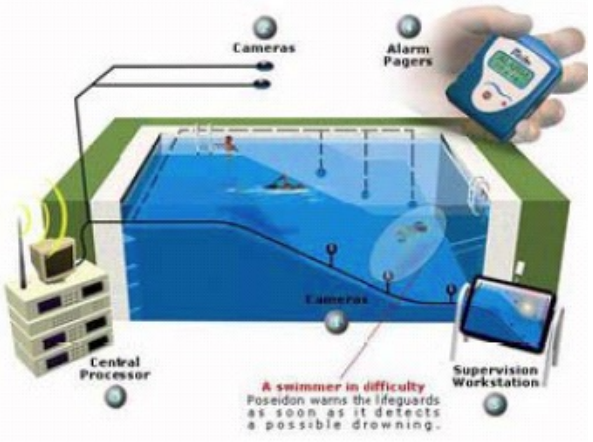


Sample Applications



Factory inspection

Reading license plates, checks, ZIP codes



Monitoring for safety (Poseidon)

Surveillance

Autonomous driving, robot navigation

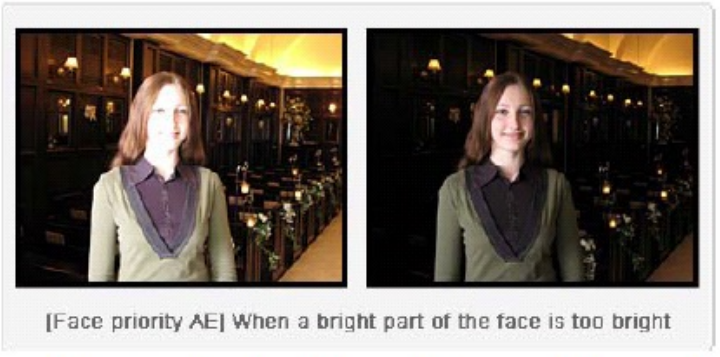
Driver assistance (collision warning, lane departure warning, rear object detection)

Sample Applications

Assistive technologies

Entertainment
(Sony EyeToy)

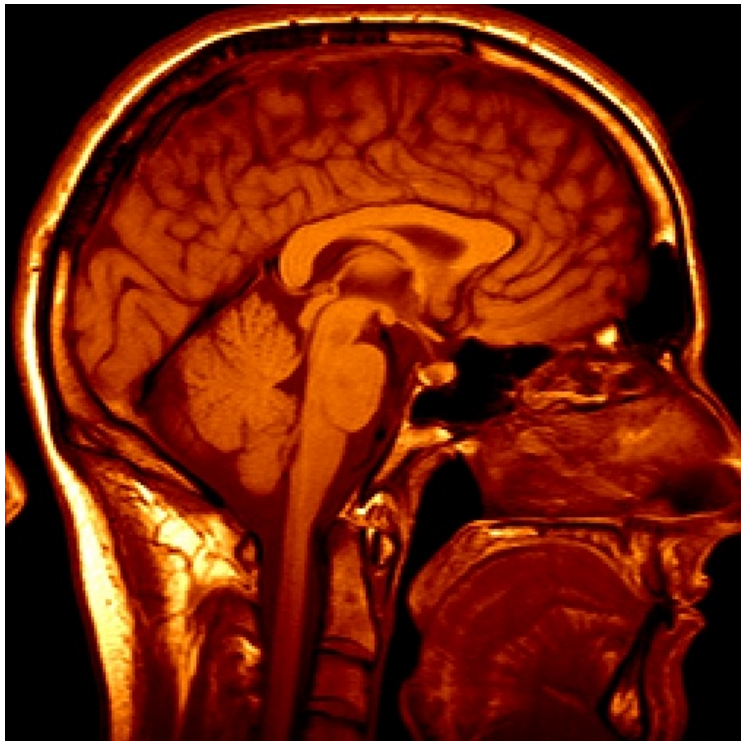
Movie special effects



Digital cameras (face detection for setting focus, exposure)

Visual search
http://www.kooaba.com/

Medical Imaging



3D imaging
MRI, CT

Image guided surgery
Grimson et al., MIT

Automotive Safety and Self-driving Cars

- Mobileye: Vision systems in high-end BMW, GM, Volvo models
  - "In mid 2010 Mobileye will launch a world's first application of full emergency braking for collision mitigation for pedestrians where vision is the key technology for detecting pedestrians."

Source: A. Shashua, S. Seitz

Autonomous Vehicles/ Self-driving cars

Vision in supermarkets

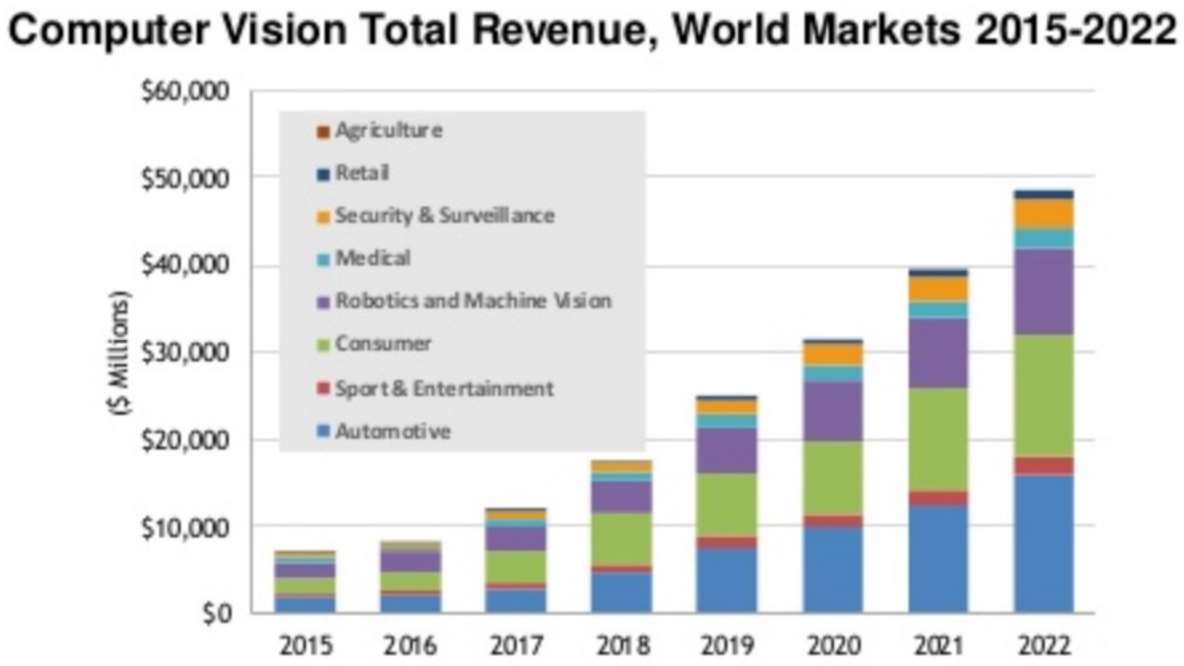
[LaneHawk by EvolutionRobotics]

"A smart camera is flush-mounted in the checkout lane, continuously watching for items. When an item is detected and recognized, the cashier verifies the quantity of items that were found under the basket, and continues to close the transaction. The item can remain under the basket, and with LaneHawk,you are assured to get paid for it… "

Source: S. Seitz

The best time studying Computer Vision (and Machine Learning)

- Over 50,000+ new jobs tagged “computer vision” listed on LinkedIn at the moment.
- Investment grows by 30 folds since 2012



Global Tech Firms sponsors for major computer vision conferences

How To Read a Paper ?

(or How to Write a Paper Reading Report)

Why read papers?

- Primary form in which research results are disseminated in sciences.
- Conference papers (shorter, but fast) : generally considered solid and of high quality in Major Conferences of computer science.
- Journal papers (longer, and slow)
  - Often the complete version of a conference paper
  - May come out several years after the conference paper

Major Conferences in computer vision

- CVPR – Computer Vision and Pattern Recognition, since 1983
  - Annual, held in US
- ICCV – International Conference on Computer Vision, since 1987
  - Every other year, alternate in 3 continents
- ECCV – European Conference on Computer Vision, since 1990
  - Every other year, held in Europe

Major Conferences in ML&CV (Cont.)

- ACCV – Asian Conference on Computer Vision
- SIGGRAPH-ACM SIGGRAPH, SIGGRAPH-Asia.
- Neurlps (NIPS) – Neural Information Processing Systems
- ICML, ICLR (deep learning).

Top Journals in Computer Vision

- PAMI – IEEE Transactions on Pattern Analysis and Machine Intelligence, since 1979 (impact factor: 19.96, #1 in all engineering and AI, top-ranked IEEE and CS journal)

- IJCV – International Journal on Computer Vision, since 1988 (impact factor: 11.36, #2 in all engineering and AI)

- CVIU – Computer Vision and Image Understanding, since 1972 (impact factor: 2.20) .

Why read research papers ?

- Learn to do research
- Learn to think critically about quality of research papers
  - Someone will be thinking critically about your own work!
  - In any discipline, there are fads and there are lasting ideas… learn to tell the difference!
- Gain perspective
- Key issue: what are the questions to ask?

How to read a paper ?

Objective

To understand the scientific contributions of the paper

Challenges: It can be difficult to work out how to start reading academic papers. They are often dense, poorly-written, contain many references to previous work, poorly-explained acronyms, terms, and concepts, and unfamiliar mathematics.

Reading a paper critically

- Understand the problem
- Understand the proposed solution
- Understand competing approaches / methods
- Evaluate the paper (any flaw, or limitations ?)

- Peer review is the cornerstone of the scientific publishing process

Be skeptical

- *“Be skeptical. But when you get [enough] proof, accept the proof.”*

–Michael Specter

- What constitutes enough proof?

- To form a truly educated opinion on a scientific subject, you need to become familiar with current research in that field. And to do that, you have to read the “primary research literature”.

Evaluating a paper

- What is the problem being solved?
  - Is it important? Relevant? Why?
  - What is the prior work in this area?
- Is the proposed solution clever?
  - Cleverness is orthogonal to importance!
- Are the assumptions and model sound/solid ?
- Impact
  - Easier to evaluate for older papers
  - Does other work build on it? Do other papers use similar techniques and solutions proposed in this paper?

Evaluation Process

- Read slowly, take notes as you read
  - Question assumptions, importance of the problem
  - Write questions to track what you don't understand
- Sometimes what is not in the paper is more important than what is in it
  - Is there something the authors have overlooked?
- Don't let ideas or design details pass until you understand them!
- Do not assume the paper is correct, even if published in a prestigious peer-reviewed venue

Approach

Approach:

- Read the title and abstract: this will summarise the paper and make its case
- Look at the pictures: they typically illustrate the problem, the solution (often with a flowchart), and the results
- Read the introduction: this will provide the context, motivate the problem and the solution, and list the contributions
- Read the method: skip any proofs, just try to understand the general approach at first
- Read the results: how well does the method work; can you tell from what the authors have presented?
- Read the rest: related work, discussion and conclusion
- Read critically: be suspicious; what have the authors not considered; what are their assumptions; is there a simpler approach?
- Make notes: scribble in the margins, underline, draw pictures

At the end, you should be answer the following questions:

1. What is the problem?
2. What is the proposed solution?
3. What are the contributions?
4. Is the approach technically correct?
5. What are the assumptions and limitations?
6. How well does the method perform?
7. Is it a good paper? Does it lead anywhere?

Resources:

- S. Keshav, "How to Read a Paper"
- M. Mitzenmacher, "How to Read a Research Paper"

Use LATEX to write reports

- What is LaTeX? A typesetting system where you write your document in a mark-up language which gets compiled into a PDF file (unlike WYSIWYG programs like Microsoft Word)
- Why LaTeX? A powerful tool for typesetting academic papers, with particular advantages in writing mathematics, managing citations, and creating vector graphics for diagrams and plots
- Templates: See Wattle for templates for the different assessment items
- Tutorial: Using Overleaf, an in-browser editor

Format of your reading reports

- Report Latex template: https://www.overleaf.com/read/tbvvpwgcfyyb

- Format:
  - → (See right hand side)

- Two pages (in PDF) maximum.

2. Problem Statement

What is the research problem the paper attempts to address ? Is the problem important or significant and why?

What is the current state of the art (i.e other competing methods/solution) ?

What is the motivation of this paper? What is the niche of this paper?

If necessary you may use a picture to better illustrate the research problem.

3. Summarise the paper's main contributions

What are the claimed contributions of the paper?

What is new in this paper? What is innovative about this paper?

Are the authors over-claiming their contribution, and if so how?

4. Method and Experiment

What method/idea/insight do they propose? Explain the crux (key idea/insight/cleverness) of their method.

What is the main point/argument?

How do the authors substantiate their claims ? What experiments are conducted? What are the main results or

Critical Analysis

5.1. Are the paper's contributions significant?

Are the contribution/improvement trivial, incremental?
Why previous efforts failed ?

5.2. Are the authors' main claims valid?

Have they convincingly validated their main idea?
Any hole in their arguments, derivation, experiments ?

5.3. Limitation and weaknesses

Any limitation/weakness of their method? What can be done to improve the work ?

What would you do to address/overcome their weaknesses?

5.4. Extension and future work

What *extra* experiments that you would suggest the authors to conduct in order to strengthen the result ?

Can you think of other possible applications of the method/ideas (assuming valid) presented in the paper?

What are possible future works?

5.5. Is the paper stimulating or inspiring ?

Many papers (even those published ones) are dull, while some are exciting. What is your opinion about this paper and why?

5.6. Conclusion and personal reflection

First, draw a short conclusion about this paper.

Then, if you were tasked to solve the research problem, what would you do differently? an alternative solution ?

Finally, in one sentence, summarize what you have learn from reading this paper.

Q&A; then 5 minutes break, before we move on to Topic-1: low level vision-1.

Next:

Low level vision-1:
image defocus deblur/super-resolution

General format of each guided paper reading session

- Before lecture:
  - Pre-lecture self-reading of the in-class papers (for at least 30 minutes up to 1 hour).

- During Lecture/Tutorial:

  - **General technical background introduction** (20~30 minutes)
  - Guided paper reading session (30 minutes)
  - Breakout room group discussion (10 minutes)
  - Group report back (20 minutes) : randomly pick up 3~5 groups.

Papers to read today:

Image Pre-conditioning for Out-of-Focus Projector Blur

Michael S. Brown Peng Song Tat-Jen Cham
School of Computer Engineering
Nanyang Technological University, Singapore 639798
msbrown@ntu.edu.sg psong@pmail.ntu.edu.sg astjcham@ntu.edu.sg

Abstract

We present a technique to reduce image blur caused by out-of-focus regions in projected imagery. Unlike traditional restoration algorithms that operate on a blurred image to recover the original, the nature of our problem requires that the correction be applied to the original image before blurring. To accomplish this, a camera is used to estimate a series of spatially varying *point-spread-functions (PSF)* across the projector's image. These discrete *PSFs* are then used to guide a pre-processing algorithm based on Wiener filtering to condition the image before projection. Results

Figure 1. (Left) Original image with blurring due to regions of the projected image being out-of-focus. (Right) The same image with our deblurring pre-conditioning.

Papers to read today:

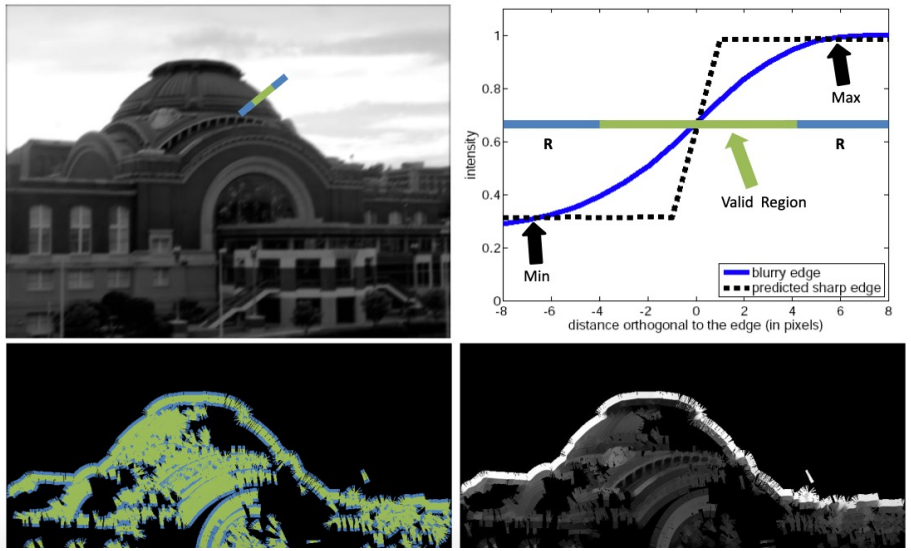
PSF Estimation using Sharp Edge Prediction

Neel Joshi† Richard Szeliski* David J. Kriegman†

†University of California, San Diego *Microsoft Research

Abstract

Image blur is caused by a number of factors such as motion, defocus, capturing light over the non-zero area of the aperture and pixel, the presence of anti-aliasing filters on a camera sensor, and limited sensor resolution. We present an algorithm that estimates non-parametric, spatially-varying blur functions (i.e., point-spread functions or PSFs) at sub-pixel resolution from a single image. Our method handles blur due to defocus, slight camera motion, and inherent aspects of the imaging system. Our algorithm can be used to measure blur due to limited sensor resolution by estimating



A 3rd paper to read (CVPR 2021):

Controllable Image Restoration for Under-Display Camera in Smartphones

Kinam Kwon* Eunhee Kang* Sangwon Lee Su-Jin Lee Hyong-Euk Lee
ByungIn Yoo Jae-Joon Han
Samsung Advanced Institute of Technology (SAIT), South Korea

Abstract

Under-display camera (UDC) technology is essential for full-screen display in smartphones and is achieved by removing the concept of drilling holes on display. However, this causes inevitable image degradation in the form of spatially variant blur and noise because of the opaque display in front of the camera. To address spatially variant blur and noise in UDC images, we propose a novel controllable image restoration algorithm utilizing pixel-wise UDC-specific kernel representation and a noise estimator. The kernel representation is derived from an elaborate optical model that reflects the effect of both normal and oblique light inci