Welcome!

COMP 558 Fundamentals of Computer Vision
Fall 2021

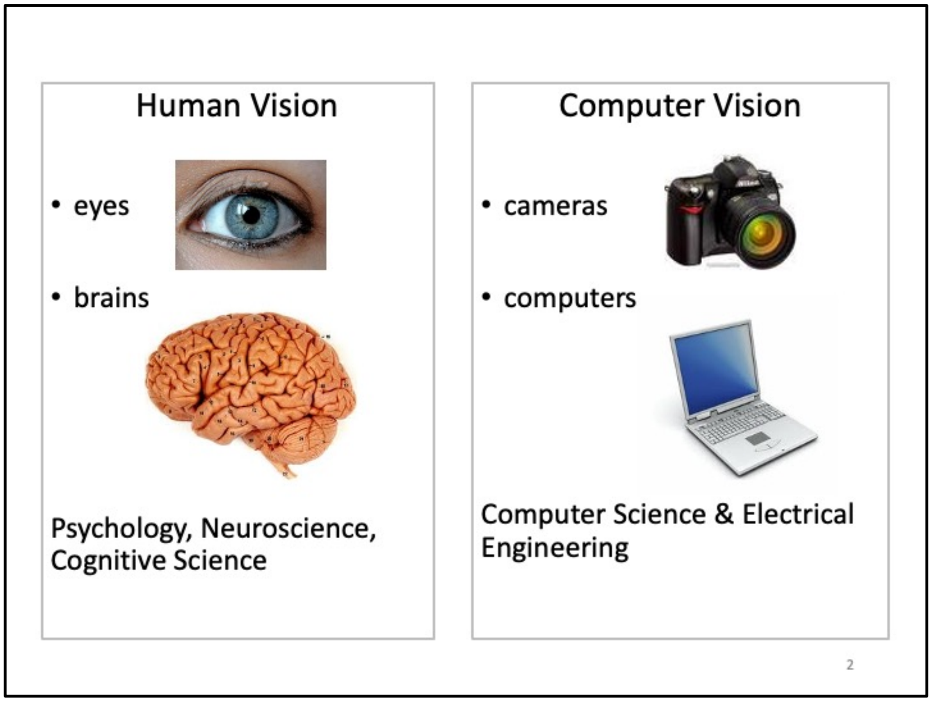
(Adapted from past Langer-Siddiqi content.
Detailed notes from Mike Langer.)

Instructor:

Prof. Kaleem Siddiqi

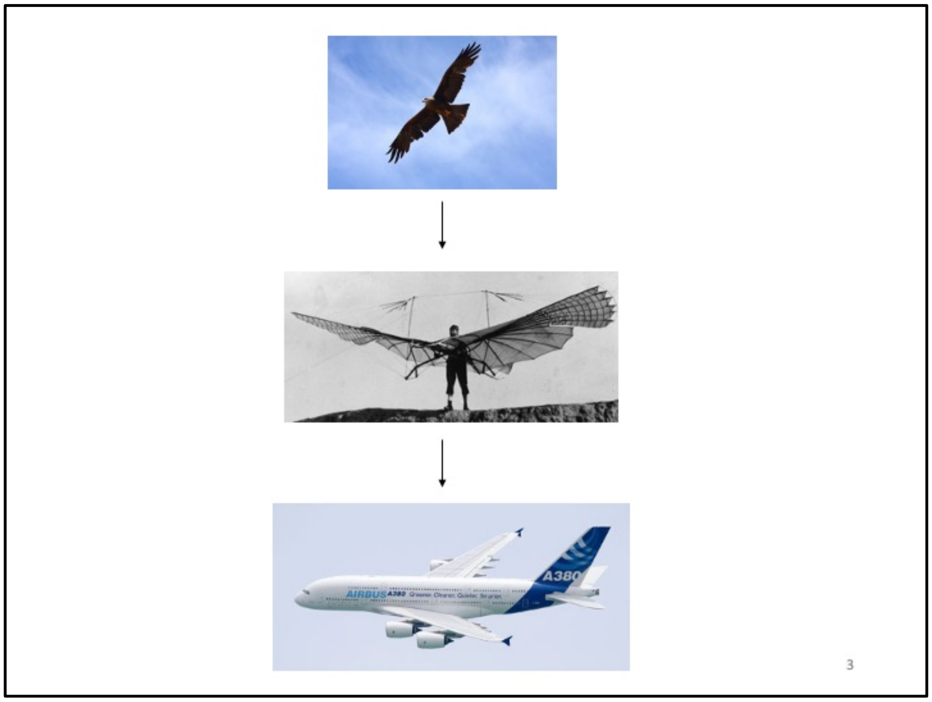
1

Hi everybody. Welcome to the course Fundamentals of Computer Vision. Fall 2021. I am your instructor Kaleem Siddiqi. Today I'm just going to give a brief overview of the course.

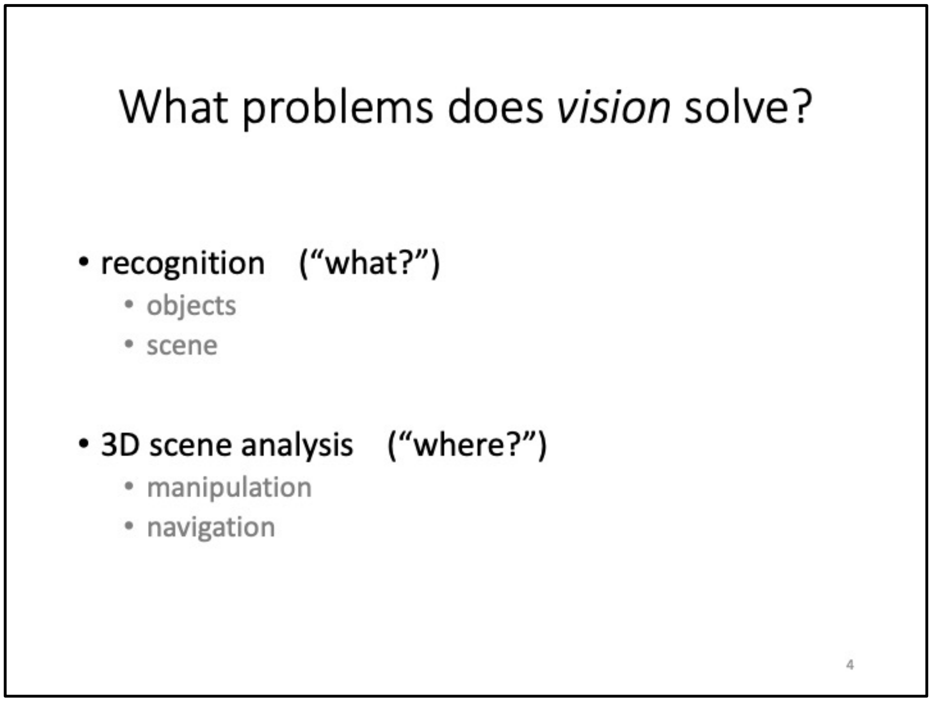


Before we talk about what computer vision is, let's make a distinction between computer vision and human vision.   People have been trying to understand how human vision works for centuries. People have studied how the eyes work and the brain works, not just in humans but in many other types of animals as well.   There are now several huge fields of vision science devoted to this –psychology, neuroscience, and cognitive science.

So what is computer vision?   Computer vision is a field within computer science and engineering.   Computer vision uses cameras rather than eyes. Modern cameras capture images as digital files. Then the computer processes the images using techniques that you'll learn about in this course.   Computer Vision has become a huge field as well. The computer vision conferences every year have several thousand attendees, and there are many applications of computer vision in industry as well.



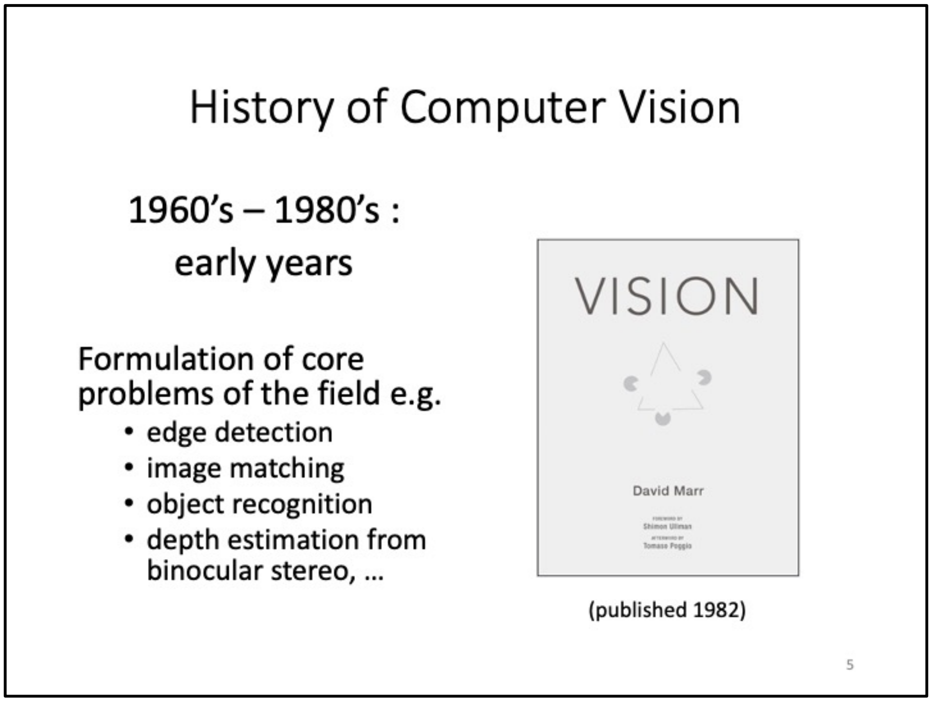
One of the goals of computer vision has always been to build systems that solve vision problems as well or better than humans. But as we learned in many other domains, we cannot just copy nature. In the case of flying, the original inventors and engineers tried to make flying machines that mimicked birds do. But that didn't work so well. Rather, it was necessary to understand the underlying principles, and to build systems based on those principles. Now we have airplanes and helicopters and drones, and the resemblance to birds is very weak. You don't *need* feathers to fly. Well, let's return to the problem of vision. What do you need to see and to make inferences about the visual world? That's what computer vision researchers have been trying to understand.



So what problems does vision solve? Most vision researchers would say there are two basic problems.

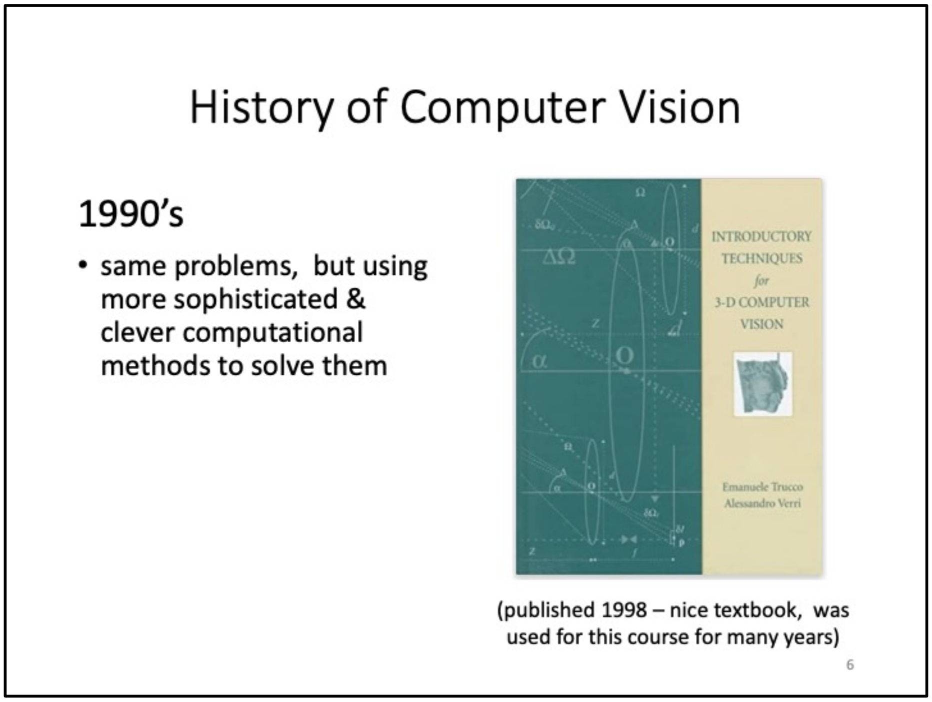
One is to answer questions about WHAT is out there? This is sometimes called the recognition problem. We need to decide what are the objects we are seeing – people, cars, trees, desks, lamps, etc. We also need to decide what sort of scene we are in. We can visually recognize manmade scenes – offices, cars, kitchens, bedrooms. And we can recognize many types of natural scenes too - forests, deserts, beaches, etc. It would be great to know how our vision systems do this. But in computer vision, we really want to build artificial vision systems to do this. Maybe we can borrow something from human vision, but maybe we can just figure out the underlying principles of what is need to do it. These are hard problems and we’re not yet there.

Another important set of vision problems involves answering questions about WHERE things are. This usually involves a 3D scene analysis. We want to identify where things are in space and how they are layed out relative to each other. This helps us when we want to manipulate objects with our hands, or when we want to navigate through a scene. This is particularly relevant in the context of autonomous vehicles, and research in that direction.



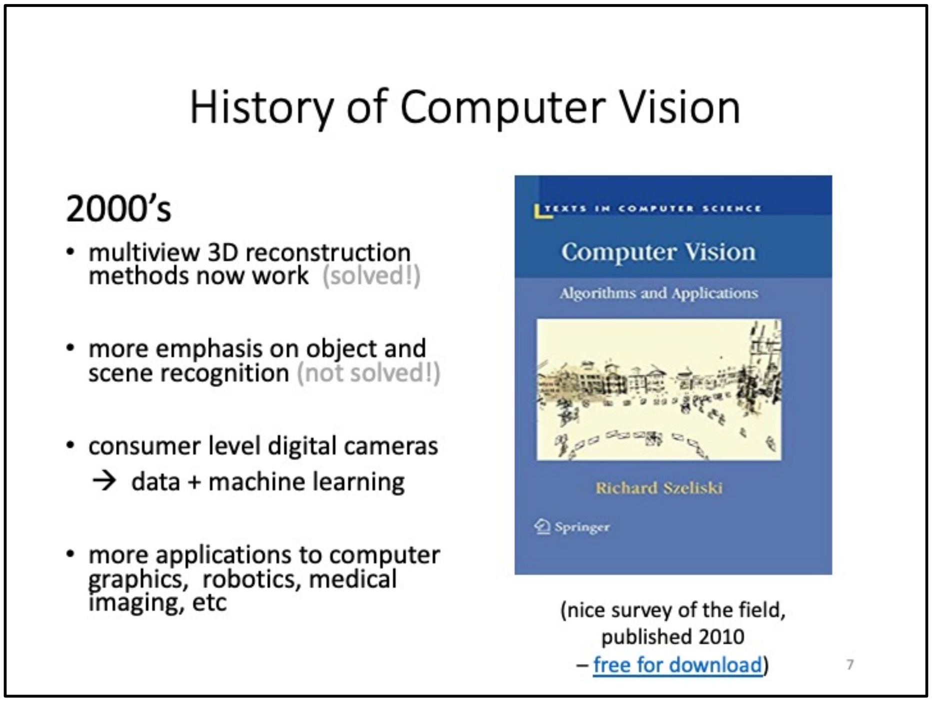
The field of computer vision began in the 1960s. The first few decades of computer vision were very important, in that they defined many of the main problems that are still studied today. Some of these problems are listed here including edge detection, image matching, object recognition, camera calibration, and depth estimation from various sources of information such as having multiple views of an object. In the case of having two cameras (which is like having two eyes), the problem is called binocular stereopsis.

In the 1970's in particular, there was a remarkable burst of creative activity in computer vision, especially at MIT. This work is summarized in a famous book by David Marr. Marr sadly died in 1980 at the young age of 35. But while he was alive, he and his colleagues published a wonderful set of papers that are still worth reading today. At the International Conference on Computer Vision, which takes place every two years, the best paper award is called the Marr Prize, in honour of David Marr. Although most of the methods he and his colleagues developed are no longer considered useful (at least not compared to what's been invented since), the field of computer vision owes a huge debt to these early researchers in getting the ball rolling.



There was much progress in computer vision in the 1990's. Mostly the same problems were studied, but they were studied in much greater depth. And many new computational methods were brought into the field. You know the saying, "if all you have is a hammer, then everything looks like a nail"? Well many people came to the field of computer vision from many areas of mathematics and computer science and engineering and they brought different tools with them. Optimization, graph theory, probabilistic models, signal processing, differential geometry, you name it. Sometimes these tools seems like overkill. But people used what they knew, and the field moved forward.

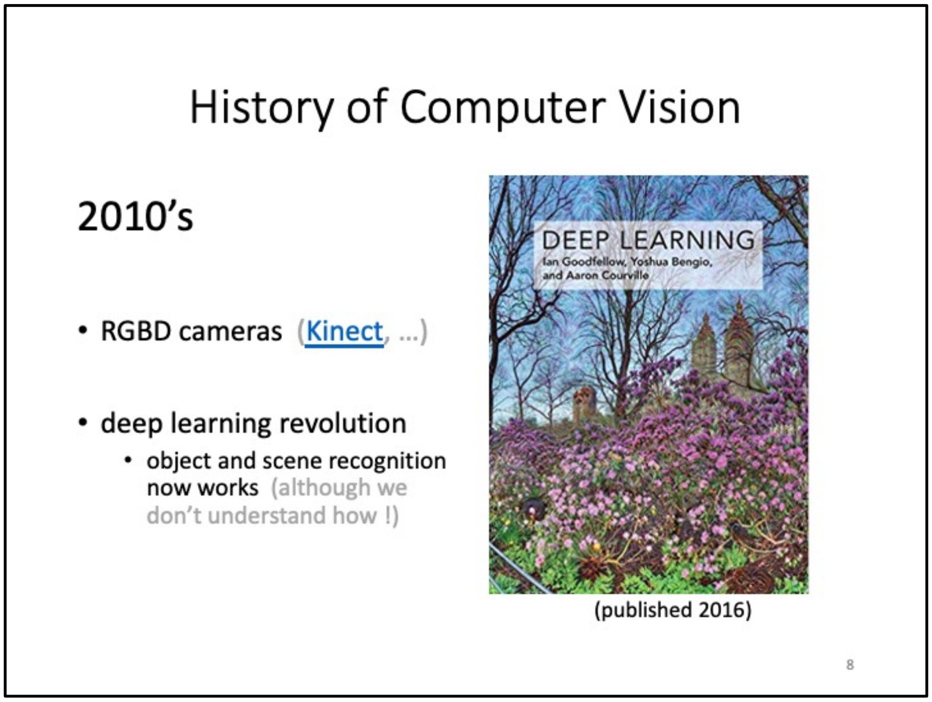
The image on the right is the cover of a very nice book that was published in 1998. It is by Trucco and Verri. It was the textbook for this course for many years. While the course has changed somewhat in the past few years, we still cover many of the classical methods, and they are very well described in this book.



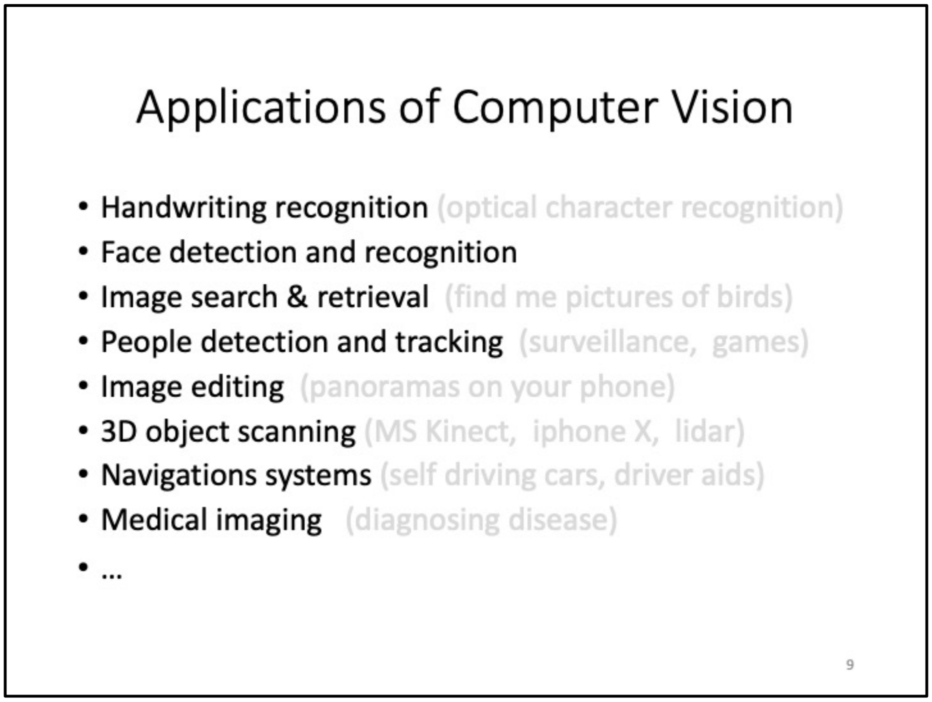
In the early 2000's, the field of computer vision started to change. Digital cameras became affordable to the general consumer, and so the number of digital photos that were available for research increased massively. The world wide web was expanding, and so people could more easily share images and data sets. How did the field change?

First of all, some of the classical computer vision problems were now basically considered to be solved. For example, many of the WHERE problems were solved. Some top researchers who worked on these problems now turned their full time attention to problems of object and scene recognition. These problems were not solved !

What gradually happened over the first decade of the 2000's is that researchers completely changed their methods. Rather than trying to design computer vision models from scratch, they turned to the massive amounts of image data that were becoming available. At first they applied classical pattern recognition methods (such as PCA which you may know) but they soon started to incorporate more recent advances in machine learning. For a nice survey of the state of the field around 2010, see the book by Rick Szeliski which is free for download. Rick was a undergrad at McGill and has been a senior research at Microsoft Research and now Facebook Research for many years.



In the last decade, specifically starting around 2012, there have been more major changes in the field. First, new cameras have been developed which allow one to measure 3D directly. These are called RGBD cameras where RGB is the standard color image representation that we'll learn about next lecture, and D stands for depth. An example is the Kinect device https://en.wikipedia.org/wiki/Kinect which was used in the Microsoft Xbox.
Second, and now more significant is the revolution in Deep Learning which has occurred in the last decade. Deep Learning is an extension of classical neural networks which were developed in the 1980's. As you may know, a lot of this work was done here in Canada at the University of Toronto and at the University of Montreal. The right panel shows the cover of the well known book Deep Learning which was published in 2016 by Goodfellow, Bengio, and Courville who are at the University of Montreal. Goodfellow was a PhD student there, and he has now moved on. Bengio was a undergrad and graduate student at McGill. The use of deep learning methods has significantly changed how object and scene recognition is done. I will spend about a few lectures on these methods and how they are applied in object recognition and other application domains towards the end of the course.



There are many applications of computer vision. I've listed some of them here. Many applications use object recognition, for example, handwriting, face recognition, or image retrieval. There are also applications in surveillance, for example, tracking people's motion in a video. Computer graphics is another huge application: many advanced image and video editing tools are based on computer vision methods. Another emerging application is vision systems for autonomous vehicles. Medical imaging is a huge applicatoin. I think any of you who have had an MRI or ultrasound have wondered how computer vision could be used to improve these systems. This is an active research area in our department and in ECE and in Biomedical Engineering and Neurology and Neurosurgery. Rick Szeliski's book describes many of these applications in more detail. We'll talk about some of these as we go through the course.

COMP 558 Overview

Part 1 : 2D Vision

RGB
Image filtering
Edge detection
Least Squares Estimation
Robust Estimation: Hough transform & RANSAC
Features 1: corners
Image Registration: the Lucas-Kanade method
Scale spaces (Gaussian and Laplacian)
Histogram-based Tracking:
Features 2: SIFT, HOG
Features 3: CNN's
Object classification and detection
Segmentation

Part 2 : 3D Vision

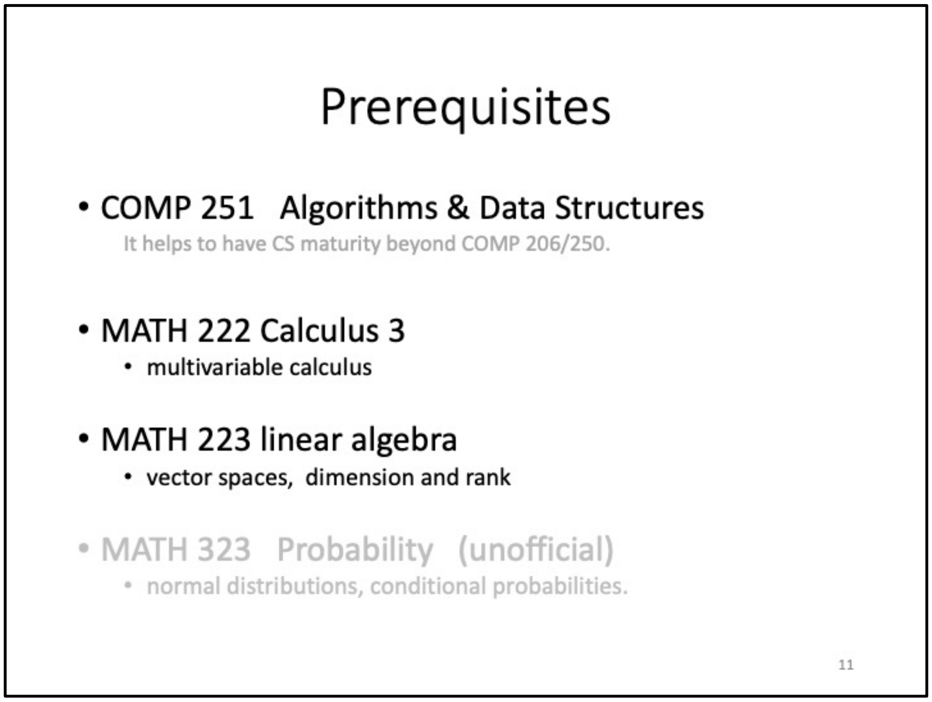
Linear perspective, camera translation
Vanishing points, camera rotation
Homogeneous coordinates, camera intrinsics
Least Squares methods (eigenspaces, SVD)
Camera Calibration
Homographies & rectification
Stereo and Epipolar Geometry
Stereo correspondence
RGBD Cameras

10

So, what is in this course? There are two parts to the course and I've listed the detailed topics here. Roughly speaking the first part concerns 2D image analysis. We will look at problems such as edge detection and feature detection. We'll learn about various least squares techniques and more robust techniques. I'll take you through some of the developments of the field from the 1970s to the modern day. The last few lectures of part 1 will be about how modern deep learning methods are used to solve problems related to object recognition.

In part 2, we'll look at 3D vision problems. We'll study the mathematics of linear perspective, and we'll study the relationship between 3D positions of scene points and 2D image points. We'll examine how images change as the camera changes position and orientation. Most of this material is from the 1990's – well before deep learning. These methods really are understood now, and they work quite well.

Finally, I'll finish the course by discussing more about related topics including RGBD cameras.



What do you need for this course? What are the prerequisties. We've listed COMP 251 as a prereq. But frankly this is overkill. There is a lot of great stuff in 251 and a lot of it is used in computer vision, such as dynamic programming and graph matching and network flow algorithms. However, we will not have time to discuss such computer vision methods in this course. So, in fact, COMP 250 is enough. What important from a CS prereq is that you know how to program and debug code. We'll be using Matlab for the assignments, so if you don't know Matlab, then you'll have to teach yourself. We'll provide some help there, but this is a 500 level course so we're assuming you can learn a lot on your own.

The Math prereqs are more important. You need to have a solid background in Calculus and Linear Algebra. For the latter, you need to have good understanding of vector spaces. You need to know what the rank of a matrix is, for example. You need to know what eigenvectors are. These linear algebra concepts will be heavily used in the second part of the course. MATH 223 is plenty. If you only have 100 level linear algebra, then you'll need to fill in some gaps so make sure you have time to do so.

Finally, it would help to have some background in probability. Math 323 is overkill. But you should know some basic probability, what a probability density function is, and you should familiar with the idea of conditional probabilities.

Remote Teaching and Learning

I am *not* planning to do live zoom lectures.

Instead we'll use zoom for:

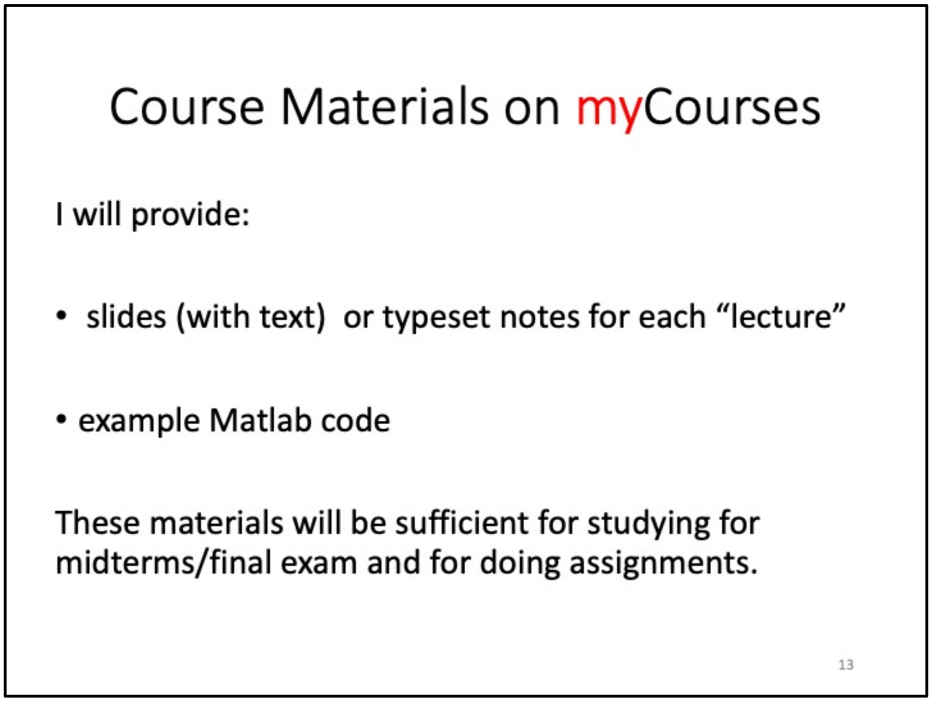
- Tutorials (size and number depending on enrollment)
- office hours (individual with me and TAs)

I *encourage* but do *not require* you to participate.

12

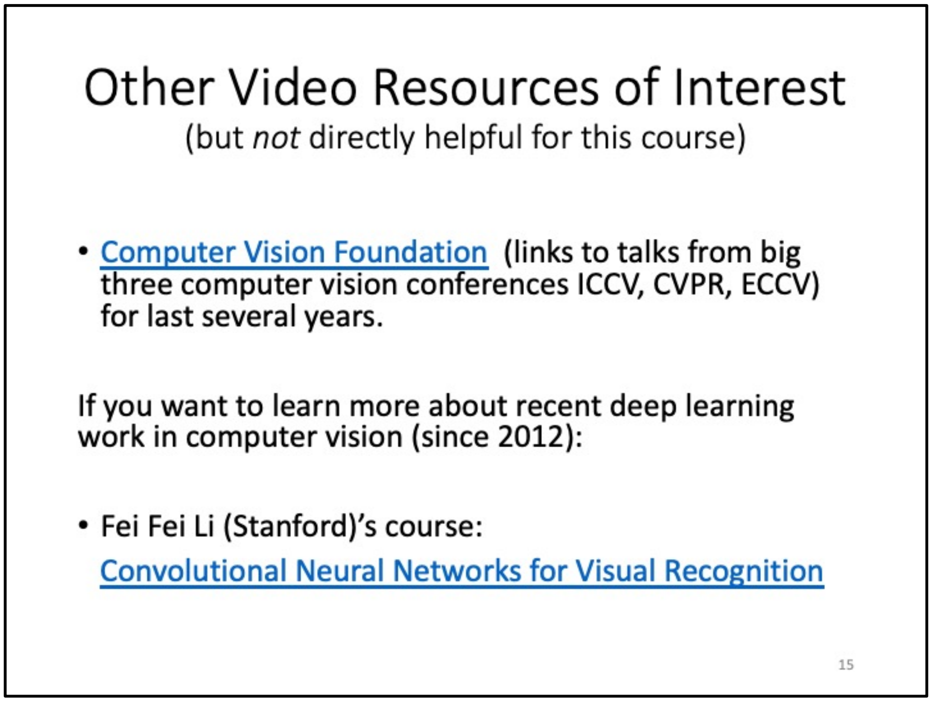
I lectured on zoom for the last few weeks of Winter 2020, and during fall 2020. and frankly I am not a fan. Most students turned off their video, and the main interaction was the occasional question in the chat bar. I don't want to rely on Zoom technology for lecturing and you probably don't either. However, zoom is the best we have, so we will use it. I will definitely hold office hours on zoom for individual help.

I will not require that you participate in any of the zoom activities. If you want to just take the materials that I give you and learn them on your own and do the assignments on your own, that's totally fine.



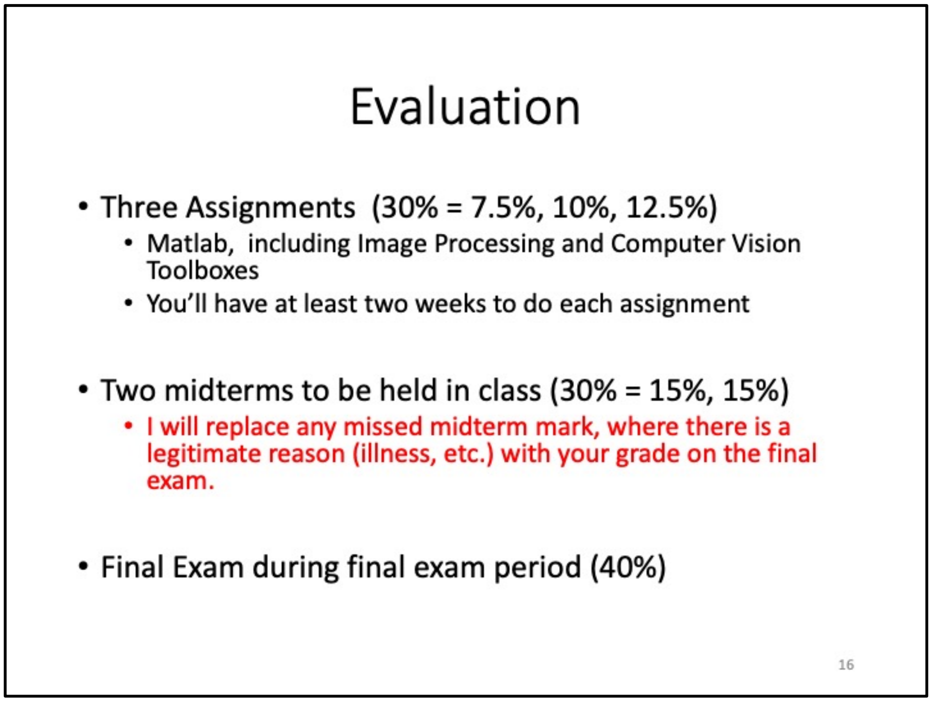
What will I provide for you? I will follow a lecture schedule. Each week I will provide you either with slides with text and sometimes typeset notes, sometimes both. I will also try to make up some exercises for you. And I’ll provide you with matlab code as examples. These materials will be enough for you to do the course on your own if that’s how you want to do it.

The resources that I will make for you are enough. But suppose you want more. You can certainly use Szeliski's book which I've given you. But suppose you want to watch videos. What video resources are there for you? Here are a few pointers. First is a nice course by Aaron Bobick and colleagues from Georgia Tech which was produced by Udacity – and its free: https://www.udacity.com/course/introduction-to-computer-vision--ud810. Another is Mubarak Shah's computer vision playlist from his course at U of Florida.
https://www.youtube.com/playlist?list=PLmyoWnoyCKo8epWKGHAm4m_SyzoYhslk5
I think he explains the concepts very well. In general, I wouldn't recommend studying only by watching videos. They are difficult to search and annotate. But I admit they there are very good for introductory ideas and for a high level understanding.

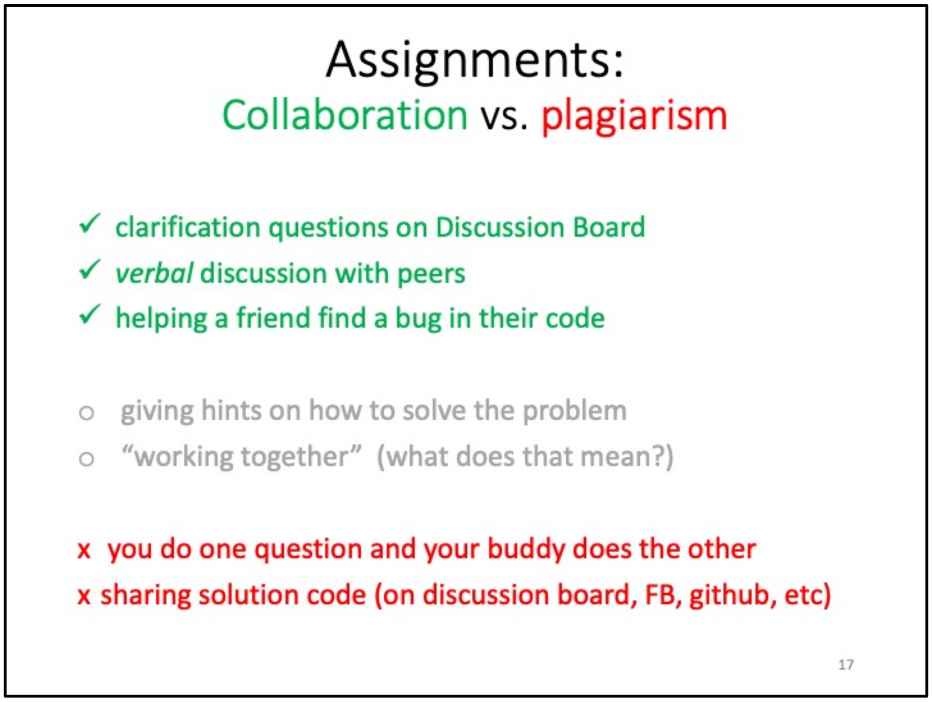


Here are some more video resources. The first https://www.youtube.com/channel/UC0n76gicaarsN_Y9YShWwhw/channels gives a link to talks from ICCV, CVPR, ECCV for last few years.

Also, if you want to know more about very modern computer vision, there are some very good videos from a course at Stanford.
https://www.youtube.com/results?search_query=cs+231+stanford



How will you be evaluated? There will be three assignments worth a total of 40%. You'll have two weeks to do the first two assignments and maybe three weeks to do the third one. The third one will be worth a bit more. The assignments will use Matlab, including the image processing and computer vision toolboxes.
There will be five quizzes. They will be worth 5% each. They will be every two weeks. The final exam will be 35%. If you can't write one of the quizzes because you are sick, or too busy, or you forget, then no problem. You have the option of dropping your worst quiz and replacing it with your final exam grade.



The assignments are to be done individually, not in groups. But what does that mean in practice?
I am tempted to say that you will learn best from the assignments if you do them entirely on own. But I don't actually believe that. I would say it is best to do them MOSTLY on your own, but to also speak with your peers about them and to exchange ideas on how you are thinking about problems. For example, if you are spending a few hours trying to answer a question on your own but getting nowhere, then its time to discuss the problem with either a peer or a TA. I am ok with this. I don't want you spending too much time spinning your wheels and getting nowhere. I give you the green light for this. Go ahead!

But there is a grey zone. If you are helping someone out, then please don't just give away the solution. The best learning comes from figuring stuff out, so it defeats the purpose of the assignment if you just give someone the solution. If you find yourself just giving the answer to your partner, or your friend say "I'll do question 1 and you do question 2" then that's collaboration but its not what we want, and that where the grey zone becomes a red zone!

Also in the red zone is that you share your code. This is absolutely forbidden!

whether it is posting it on the discussion board or facebook or github or whatever.

Midterm and Final Exams

- ✓ do it by yourself
- ✓ closed book
- ✓ email me if you are concerned about a question

X contacting other students during midterm/exam

18

What about the midterms and the final exam? These will be multiple choice, or multiselect, or short answer.

Email Policy

Please email me only for urgent matters, and for questions that are not answered on the Course Outline and Discussion Board.

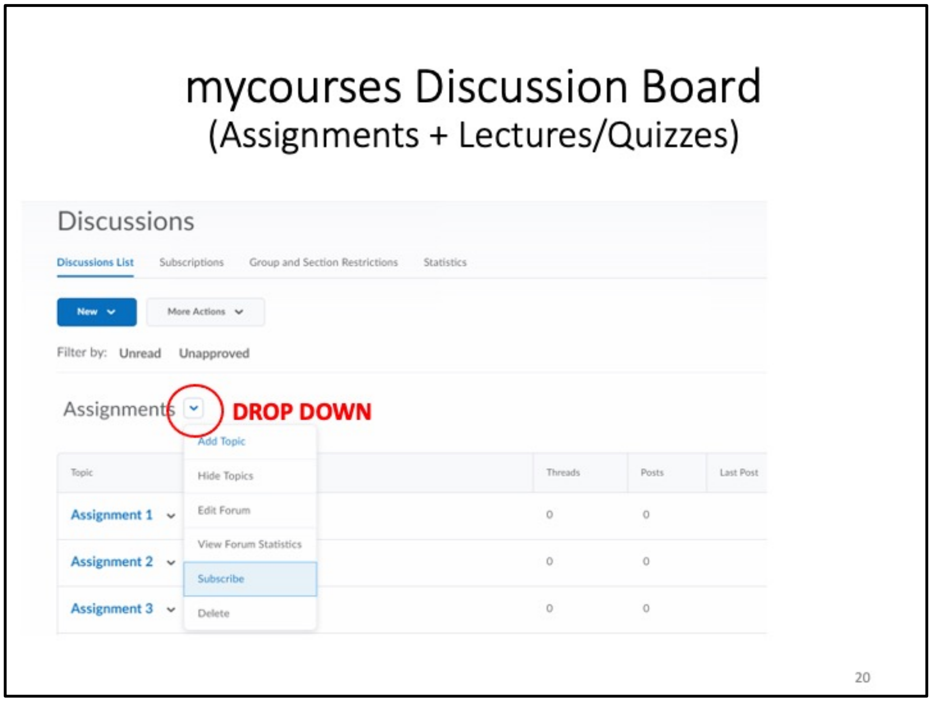
- ✓ requests for assignment extension because of *significant* illness or other *significant unforeseen* personal challenge
- ✓ broken URL links for course materials
- ✓ real-time problems with exams

- X "I missed Quiz/Assignment n because … Can I replace that 0 grade with …?
- X …

19

A few details on how we will communicate: first my email policy -- please only email me for urgent matters. If something non-urgent comes up in the course, then please use the Discussion Board. That way, I can answer you there, and others in the class will have the same access to the information. Like in a lecture when you don't understand something, chances are pretty darn high that someone else doesn't understand either. Please see the course outline for how I expect you to use the discussion board.

That said, there are things you SHOULD email me for. If you have a significant illness or accident that prevents you from doing an assignment on time, then you should let me know. I'll accommodate you as best I can.
Also, if you notice a broken link on mycourses please, email me and I'll quickly take care of it. Or if there is a real time problem with a quiz, let me know that too of course.



Please subscribe to the various discussion boards. Definitely subscribe to announcements and to assignments.

Who are you (by degree) ?

B. Sci./B.A	80
M. Sc.	20
Other	10
	110

21

Hopefully we'll get to know each other as the semester goes along. We're a big group now. There are about 110 of you currently registered. There are typically about 60-80 students in this course, so I'm guessing the numbers will drop down, but who knows what will happen this year?

Who am I ?

- studied ECE (Lafayette & Brown)
- did postdocs at McGill and then moved to Yale
- prof here since 1998!
- taught this course many times
- I also teach an advanced course in shape analysis (very occasionally)

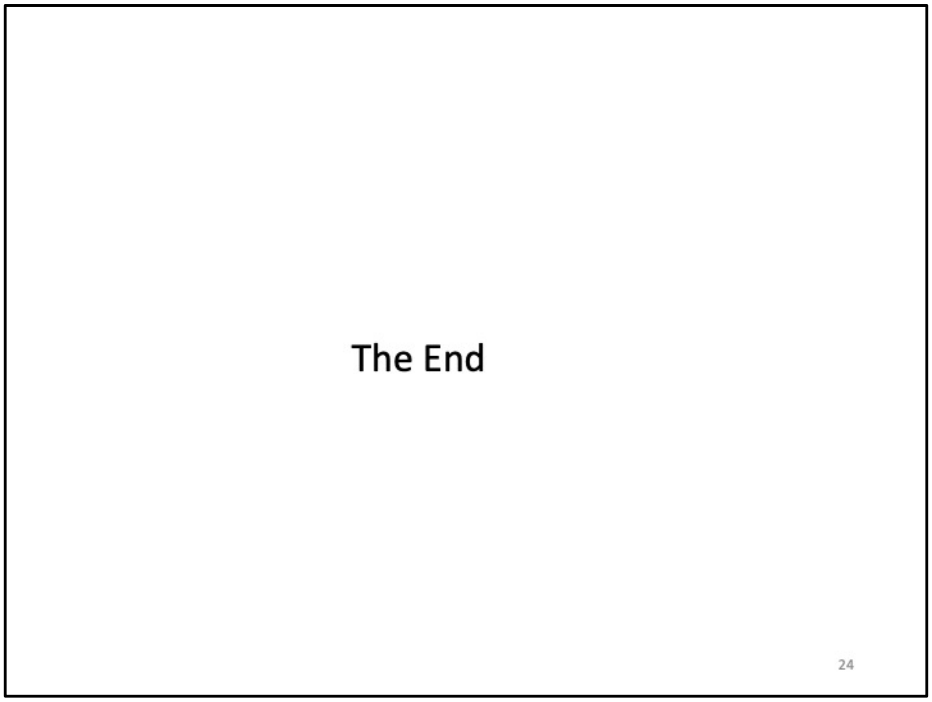
22

My Research

(see my web page: http://www.cim.mcgill.ca/~siddiqi

23

I am interested in shape and in visual form. This has taken on many different directions and meanings. In my doctoral work I came up with the notion of a shock graph. I later worked on a variety of geometric flows. Then I also contributed to medial representations and wrote a book on the subject. For the past decade or so I've been very interested in biological patterns and how to model them using CS and applied math. We have started to write papers in general science journals. My view is that computer science can help biologists go well beyond where they presently are, to tackle fundamental questions related to form and function. I am an associate member of the math and stats dept. at McGil, of MILA and of the Goodman Cancer Institute.



That's all for now, folks. Hopefully that gives you a sense of what this course is about. Please read the Course Outline for more details on the logistics and expectations.