CS195: Computer Vision

Introduction to Computer Vision

August 26, 2024

Md Alimoor Reza
Assistant Professor of Computer Science
Drake University

Road Map

- Brief introduction
- Course logistics
- Topics
    - What is computer vision?
    - What makes vision hard?
    - How does human vision work?

Introduction

- Alimoor Reza
  - Assistant Professor of Computer Science, Dept. of Mathematics and Computer Science, Drake University
  - **Office: 323** Collier-Scripps
  - **Email:** md.reza@drake.edu
  - **Phone:** 515-271-1972
  - **Office hours:** Tues/Thurs: 12:00–2:30pm CDT
    additionally by appointment

Zoom link

https://analytics.drake.edu/~reza

Course mechanics

- Syllabus, labs, assignments, quizzes, announcements, etc. on Drake Blackboard
  - http://blackboard.drake.edu/
- Readings from papers and online textbooks

Course mechanics

- Syllabus, schedule, assignments, announcements, etc. on Drake Blackboard
  - https://drake.blackboard.com
  - Syllabus link

- Class meeting times and locations:

  - Location: Collier-Scripps Room#335
  - Time: Monday/Wednesday 3:30 pm - 04:45 pm

Course mechanics

- Office hour times and location:

  - Office hour#1: Tuesday: 12:00 pm - 2:30 pm
  - Office hour#2: Thursday 12:00 pm - 2:30 pm
  - Location: Collier-Scripps#323

Additionally by appointment

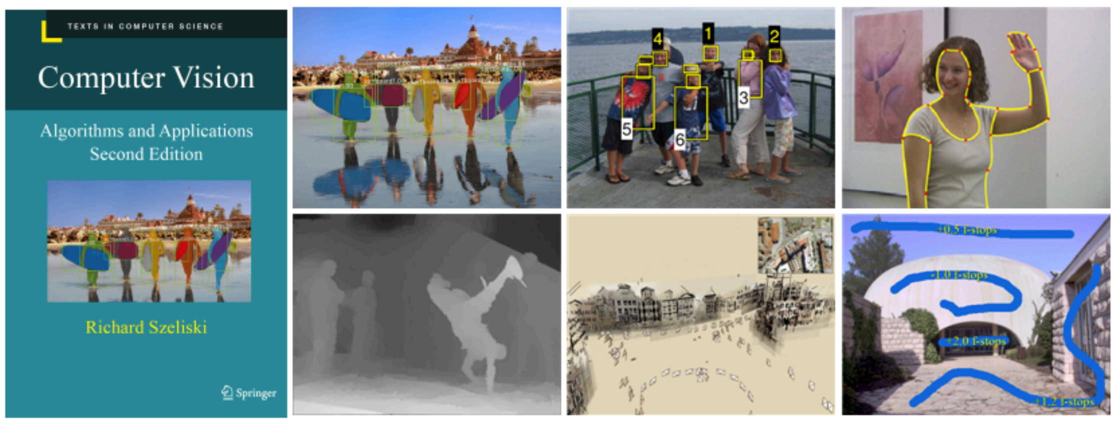
Zoom link

Course mechanics

- Recommended textbook has a free PDF version

Computer Vision: Algorithms and Applications, 2nd ed.

© 2022 Richard Szeliski, The University of Washington



Welcome to the website (https://szeliski.org/Book) for the second edition of my computer vision textbook, which is now available for purchase at Amazon, Springer, and other booksellers.

To download an electronic version of the book, please fill in your information on this page. You are welcome to download the PDF website for personal use, but **not** to repost it on any other website; please post a link to this URL instead.

Note that while the content of this electronic version and the hardcopy versions are the same, the page layout is different, since the electronic version is optimized for online reading. The PDF should be enabled for commenting in your viewer. Also, hyper-links to sections, equations, and references are enabled. To get back to where you were, use the Previous View (Alt-Left-Arrow) command in Acrobat.

The current download count is 190 (since 1/23/2022).

This book is largely based on the computer vision courses that I have co-taught at the University of Washington (2020, 2008, 2005, 2001) with Steve Seitz and Harpreet Sawhney and at Stanford (2003) with David Fleet.

Grading and Requirement

- Notebook assignments (30%): *take home assignments (submit on CodePost).*

Notebooks Assignments (30%) We will explore various computer vision techniques using Python programming language. You'll regularly submit notebook files (.ipynb) to showcase your skills. Since I won't typically run your code due to long computation times, please ensure your results are saved in the notebook. You can expect to submit **6-7 assignments** throughout the course.

Grading and Requirement

- In-class activities (15%): *10-12 simple coding activities or paper-based tasks (submit on Blackboard).*

In-class Activities (15%) You will also submit notebook files with simple coding or paper-based activities. These are easier, as they follow the lecture and are usually done during or right after class. Since I won't run your code, be sure to save the results in the notebook. You can expect to submit about **12-13 in-class activities** throughout the course.

Grading and Requirement

- Quizzes (30%): *3 quizzes, 10% each*

Quizzes (30%) There will be **3 quizzes** that will be administered via Blackboard. They will not be timed, and you will have a few days to complete them. Quizzes should be completed individually. There is no time limit on these quizzes. As in the real world, you will be allowed to use external resources like the class notes and the internet. You will be required to cite any sources that you used while completing these quizzes other than the class notes.

Grading and Requirement

- Final Project (20%): *preferably a group project*

Final project: **1 final project** will consist of research or implementation on a topic of your choice (in consultation with the instructor). Projects may be done individually or in groups of up to 3 students. There will be three deliverables:

- *Project proposal:* a brief (1-2 pages) proposal, due approximately 9^{th} week into the semester.
- *Project report:* a summary report along with an electronic copy of the source code developed during the project.
- *Project presentation:* a class presentation. students are encouraged but not required to submit their work to a regional conference (eg, CCSC) or workshop.

Final project

- On a topic of your choice
- Individually or in small groups (up to 3 students)
- Three deliverables: a brief proposal, a final presentation, and a final report (and source code)
- Wide range of possible projects, e.g.
  - Apply existing technique to new application X
  - Implement and compare techniques U and V
  - Something else broadly related to vision
  - Or develop new technique for problem Y (not necessary but you are welcome)

Grading and Requirement

- Attendance/Participation (05%): *Participation in polls, not based on correctness*

Attendance/Participation (05%): This class is highly interactive, meaning that active participation is both expected and the norm. You will receive credit for your participation, and it will be counted towards your final grade. I will keep track of your involvement using a signature sheet. Throughout the course, I will pose questions using polling software and conduct in-class Q&A sessions to better understand how the class is grasping the content. These responses will not be evaluated for correctness but rather for completion.

Grading and Requirement

- Notebook assignments (30%): *take home assignments (submit on CodePost)*
- In-class activities (15%): *simple coding activities or paper-based tasks (submit on Blackboard)*
- Quizzes (30%): *3 quizzes, 10% each*
- Final Project (20%): *preferably group project*
- Attendance/Participation (05%): *Participation in polls, not based on correctness, physical attendance during lecture time*

Grading Scale

- The tentative grading scale for this course would be as follows:

A (93%-100%)	A- (90%-92.9%)	B+ (87%-89.9%)
B (84%-86.9%)	B- (80%-83.9%)	C+ (77%-79.9%)
C (74%-76.9%)	C- (70%-73.9%)	D (60%-69.9%)
F (0%-59.9%)		

Course overview

- Bias towards algorithms and mathematical models that are widely applicable (more CS than engineering)
  - Many of these are widely applicable outside of computer vision

- Three broad topics:
  1. Recognition
  2. Low-level vision
  3. Geometry

- Three broad technical approaches:
  1. Machine Learning especially deep learning
  2. Probabilistic modeling eg, maximum likelihood (ML) estimation
  3. Expert-designed features (SIFT, SURF) and reasoning

1. Recognition

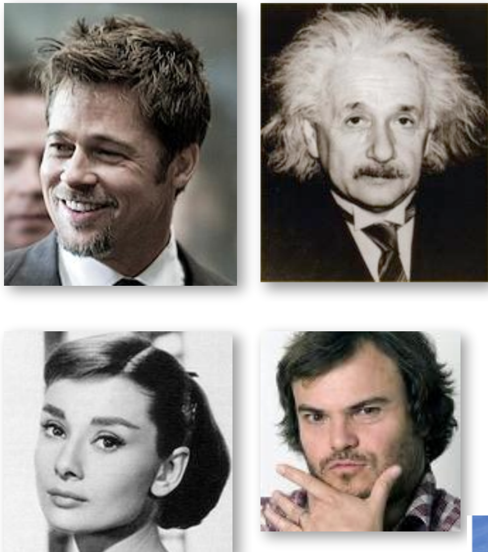
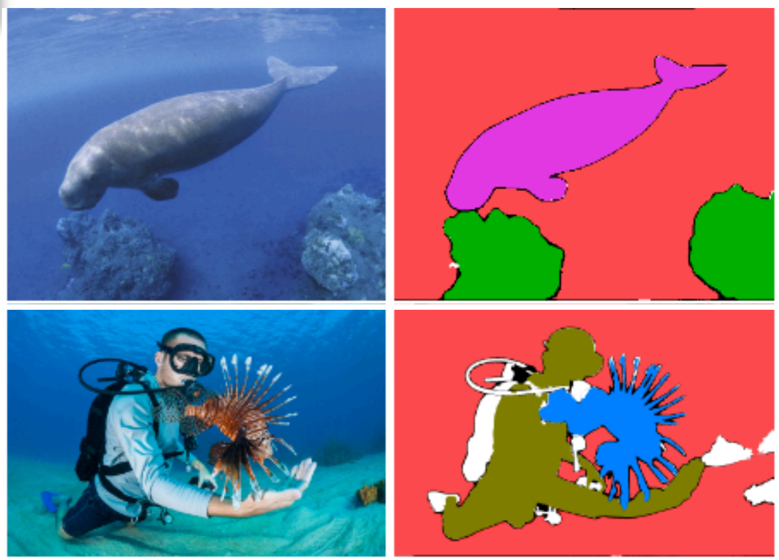


Image classification

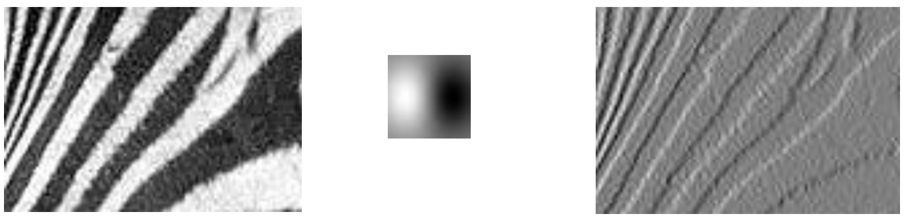
Object Detection



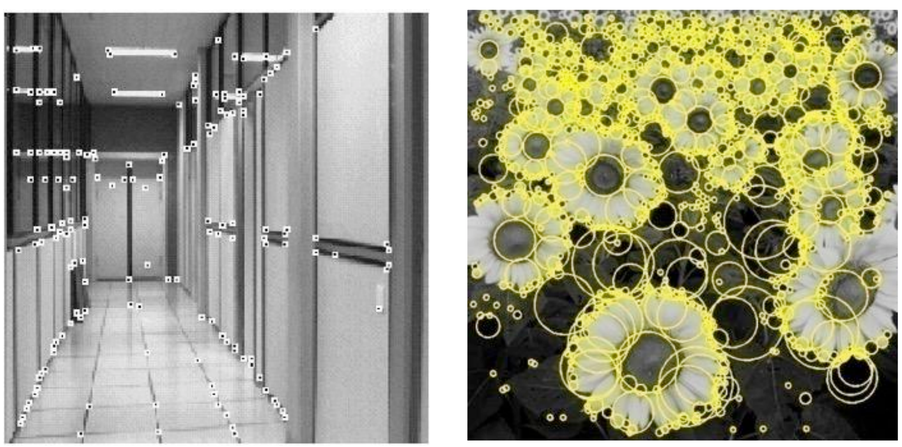
Semantic segmentation

2. Low-level vision

- Basic image processing and image formation



Filtering, edge detection



Feature extraction

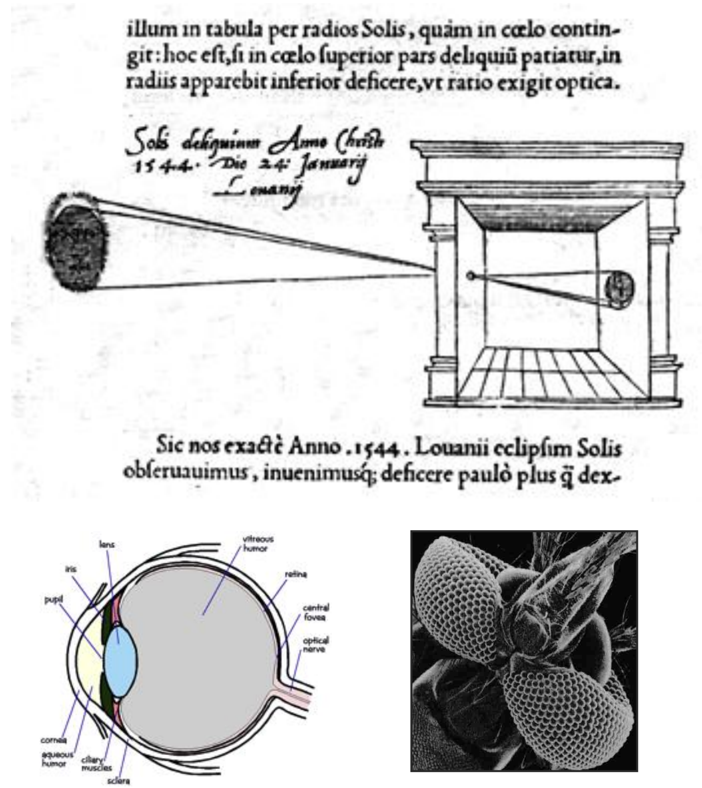
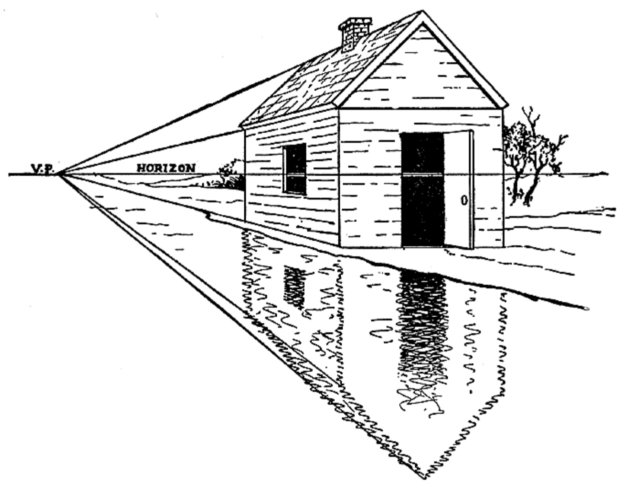
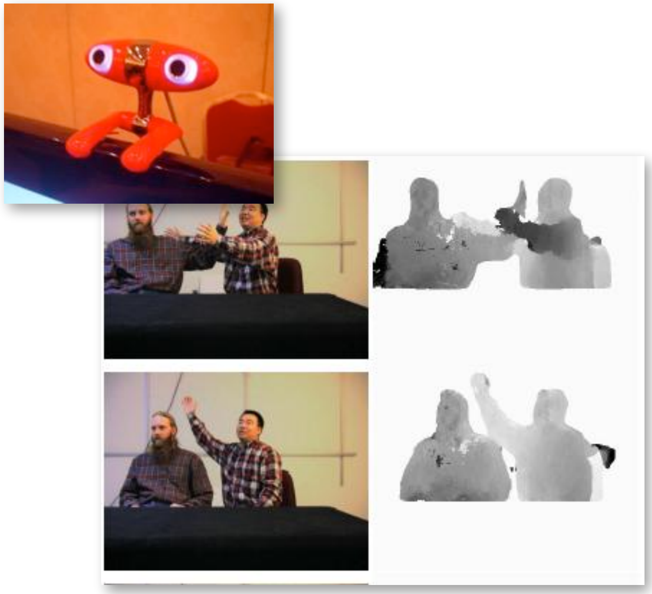


Image formation

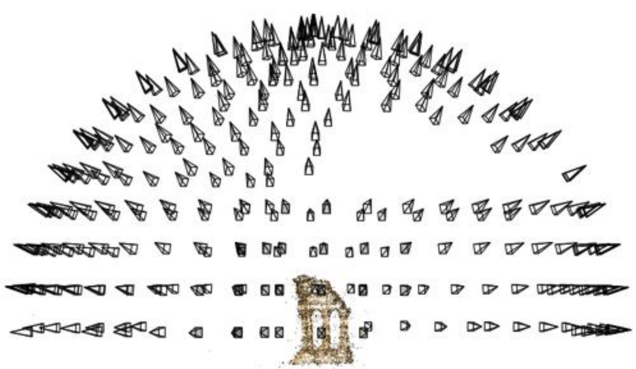
3. Geometry



Projective geometry



Stereo



Multi-view stereo

Structure from motion

Course overview

- Bias towards algorithms and mathematical models that are widely applicable (more CS than engineering)
  - Many of these are widely applicable outside of computer vision

- Some of the techniques we'll encounter:
  - Deep learning e.g., neural networks (CNN, Vision Transformer)
  - Probabilistic learning and inference e.g., maximum likelihood (ML) estimate
  - Probabilistic models e.g., Variational Autoencoder (VAE), diffusion model
  - Projective geometry, homogeneous coordinates
  - Singular value decomposition (SVD)

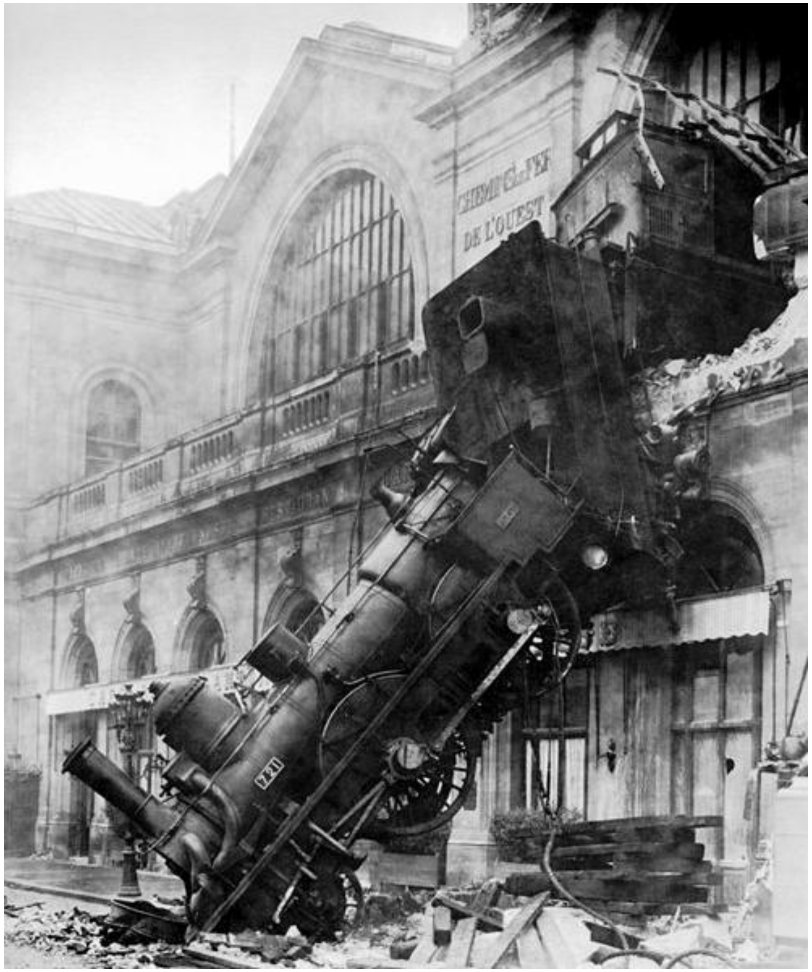
Prerequisites

- Practically:
  - CS167 (ML) or CS143 (AI) or equivalent is recommended but not necessary
  - Proficiency in a programming language, preferably Python or Matlab
  - Some level of mathematical maturity, esp. with linear algebra and statistics
  - Willingness to learn some programming and/or math on your own if necessary

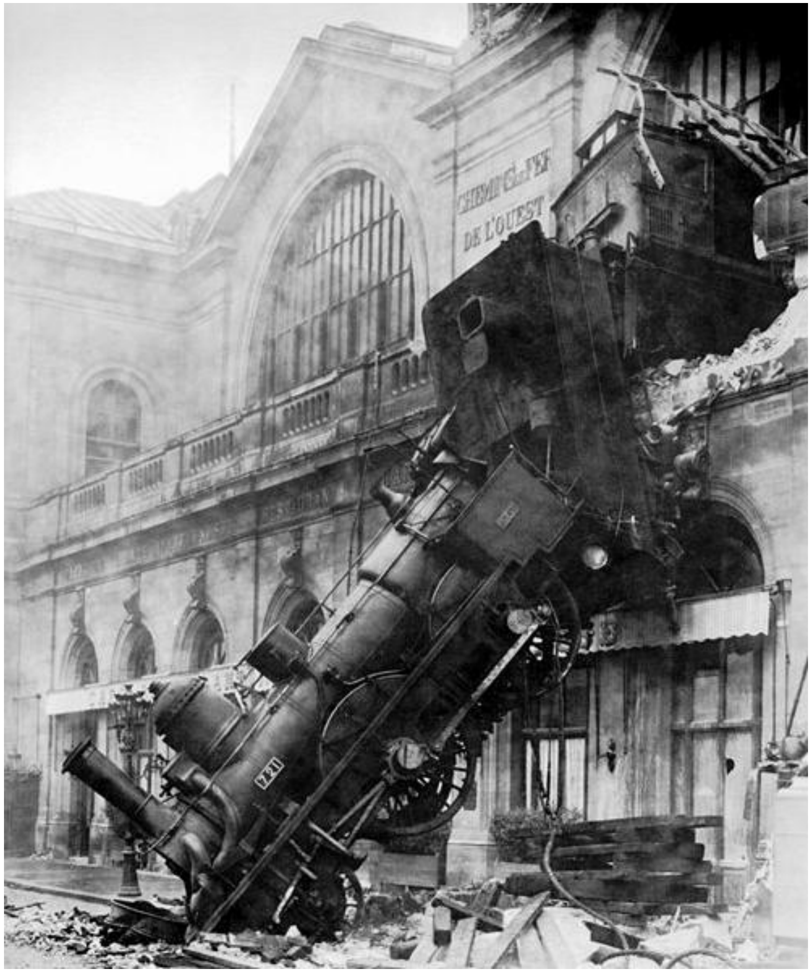
Why study computer vision?

- Compelling applications
- Draws on many different areas
  - Machine learning, image processing, graph theory, optics, geometry, statistics, linear algebra, algorithms, optimization, computer graphics…
- Difficult computational problems
  - Essentially all non-trivial inference with today's computer vision models is NP-hard
- An exciting time for the field

What is computer vision?



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




721

Goal: from images to meaning

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

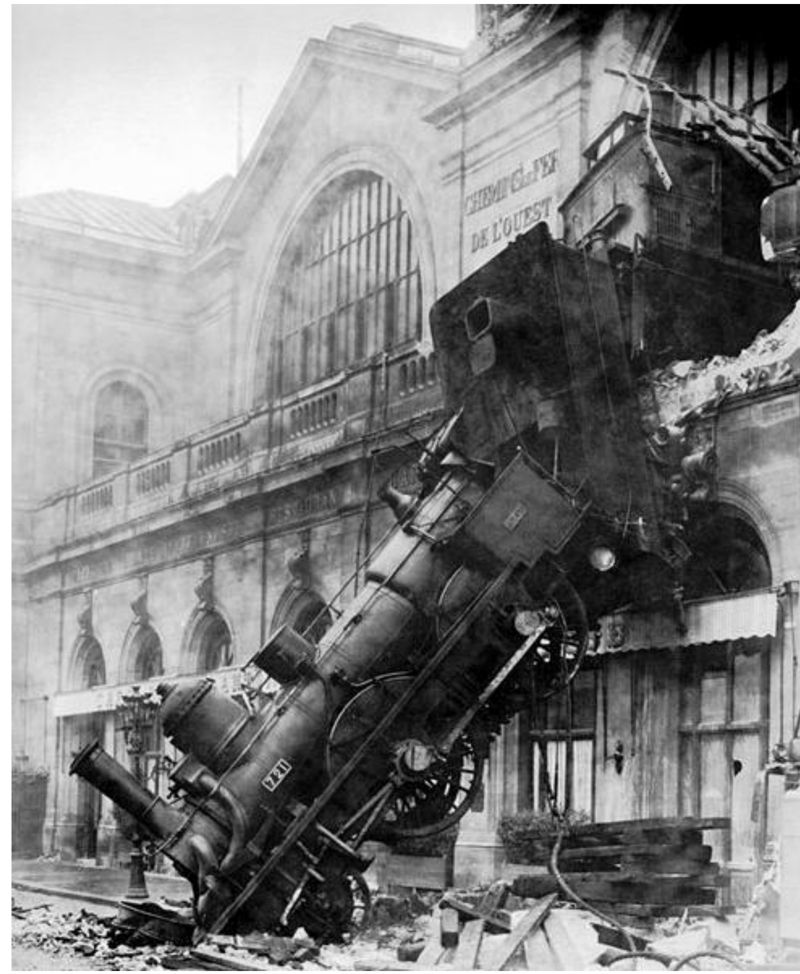


Image processing

- Take an image, produce another image

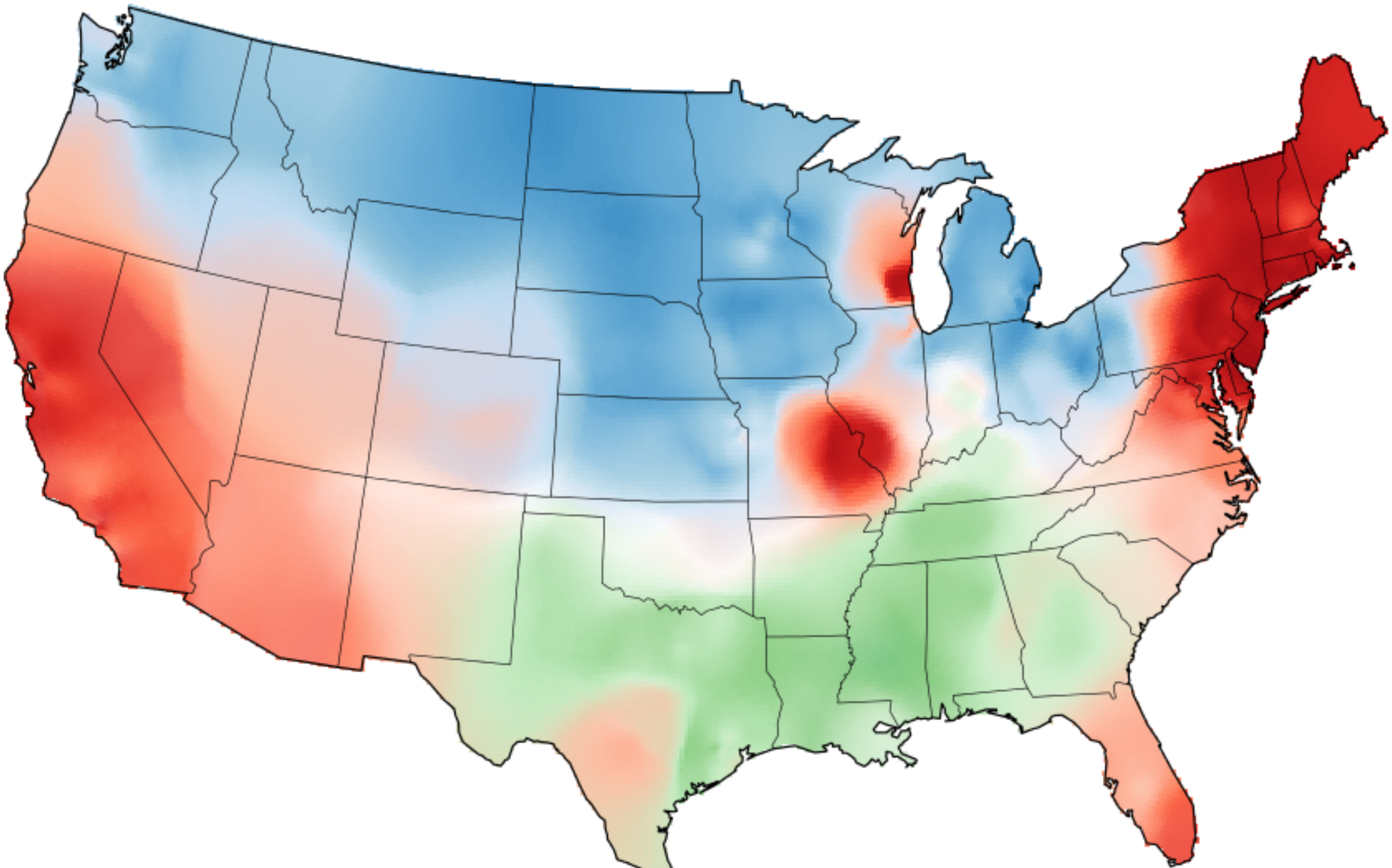
Computer graphics

- Take a description, produce an image

From “Final Fantasy”

Visualization

- Take data, produce an image

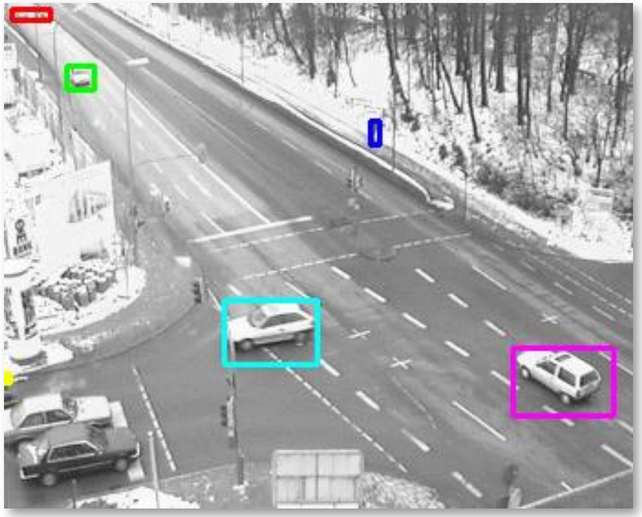


Joshua Katz, Department of Statistics, NC State University

Computer vision

- Takes an image, produces a description

Indoors, office, people, scale, laughing, humor, Obama, mirrors, …



flickr from YAHOO!

YouTube

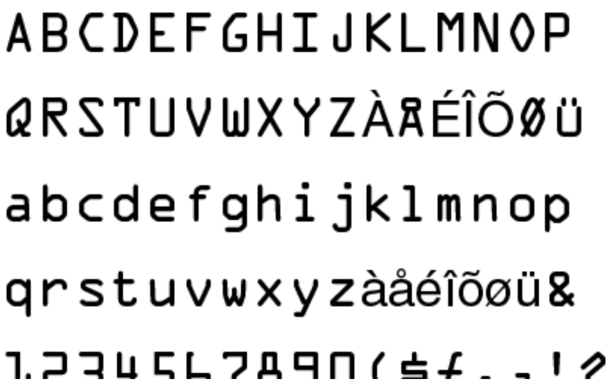
Can computers see as well as humans?

- Yes and no, but mostly no (so far). But it progressing rapidly

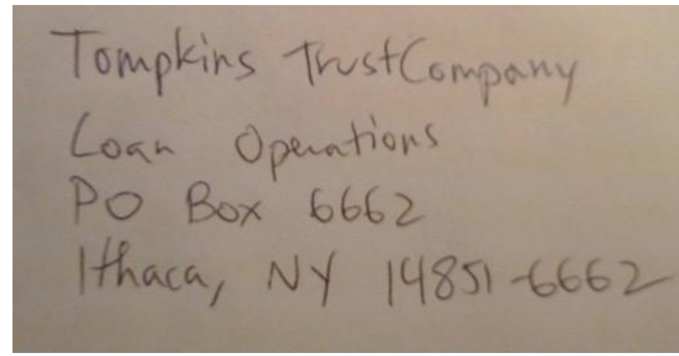
- Current vision technology is useful in select applications, with:
  - Specific, constrained environments, and/or
  - High tolerance for errors

The most successful and ubiquitous application of computer vision … ?

Optical character recognition (OCR)



Document digitization



Postal address recognition

License plate readers

Automatic check processing

Source: S. Seitz

Industrial inspection
(aka Machine Vision)

Facebook's face detection

Facebook's face detection

Facebook's face detection

THE VERGE TECH REVIEWS SCIENCE CREATORS ENTERTAINMENT VIDEO MORE

POLICY \ TECH \ ARTIFICIAL INTELLIGENCE

Facebook is shutting down its Face Recognition tagging program

The program has been opt-in since 2019

By Adi Robertson | @thedextriarchy | Nov 2, 2021, 1:53pm EDT

SHARE

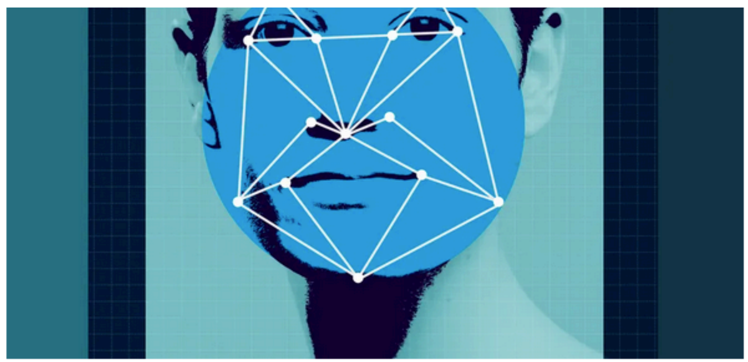


Illustration by James Bareham / The Verge

Meta (formerly known as Facebook) is discontinuing Facebook's Face Recognition feature following a lengthy privacy battle. Meta says the change will roll out in the coming weeks.

[Reference: https://www.theverge.com/2021/11/2/22759613/meta-facebook-face-recognition-automatic-tagging-feature-shutdown](https://www.theverge.com/2021/11/2/22759613/meta-facebook-face-recognition-automatic-tagging-feature-shutdown)

Why is computer vision difficult?

Why is computer vision difficult?

Viewpoint variation

Illumination changes

Scale changes

Why is computer vision difficult?

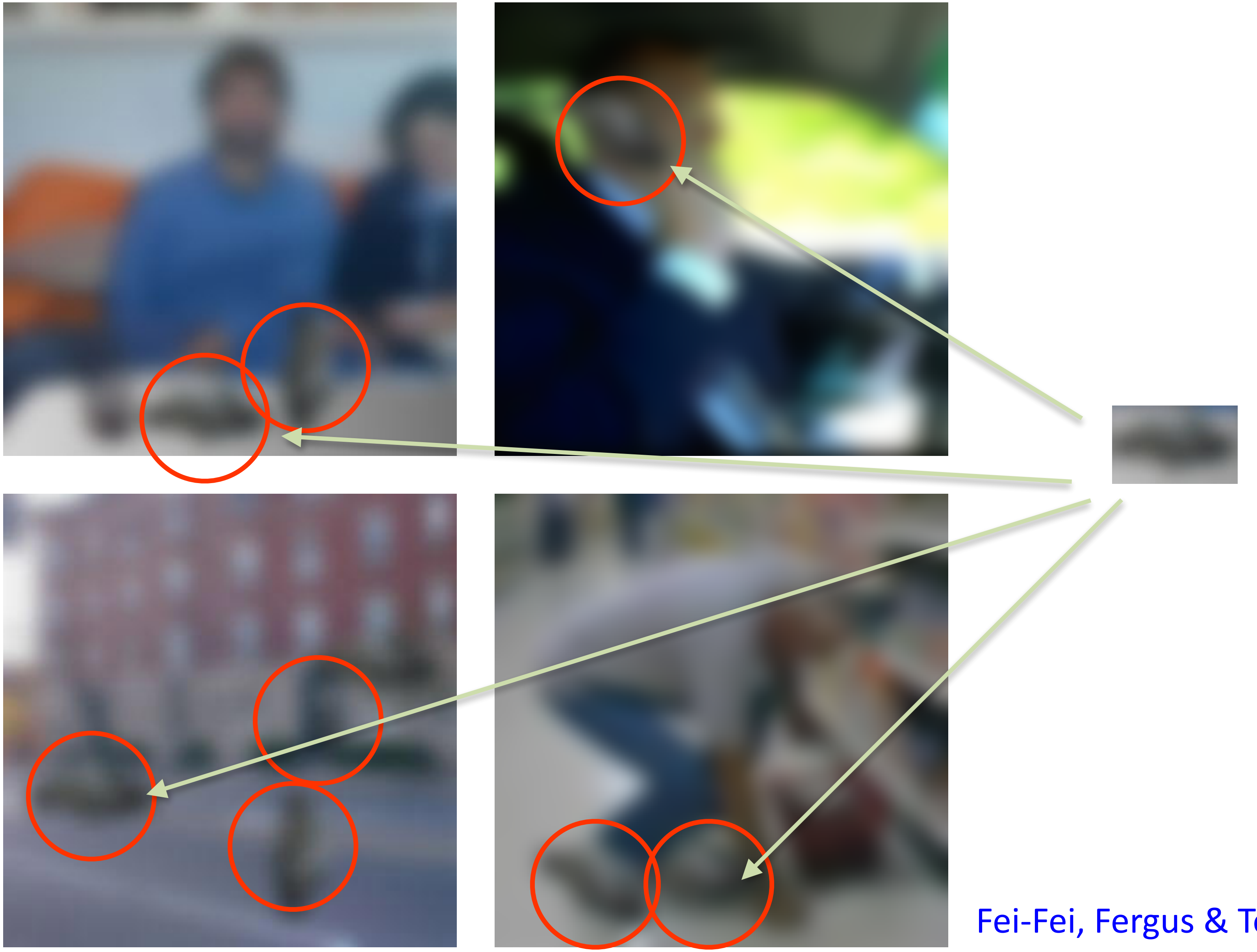
Intra-class variation

Motion (Source: S. Lazebnik)

Background clutter

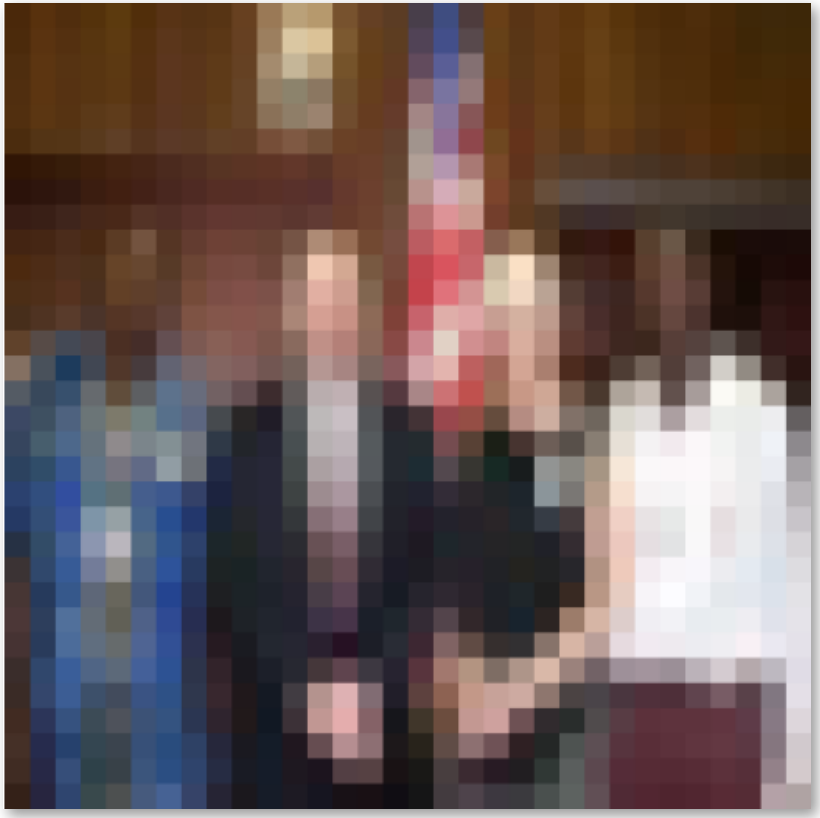
Occlusion

Role of high-level reasoning (using context)



Fei-Fei, Fergus & Torralba

Role of high-level reasoning



Source: “80 million tiny images” by Torralba, et al.

Perception is inherently ambiguous

Many scenes could have created a given 2D image

- People figure out the "most likely" one based on experience, intuition, convention, … ?

Julian Beever Artwork

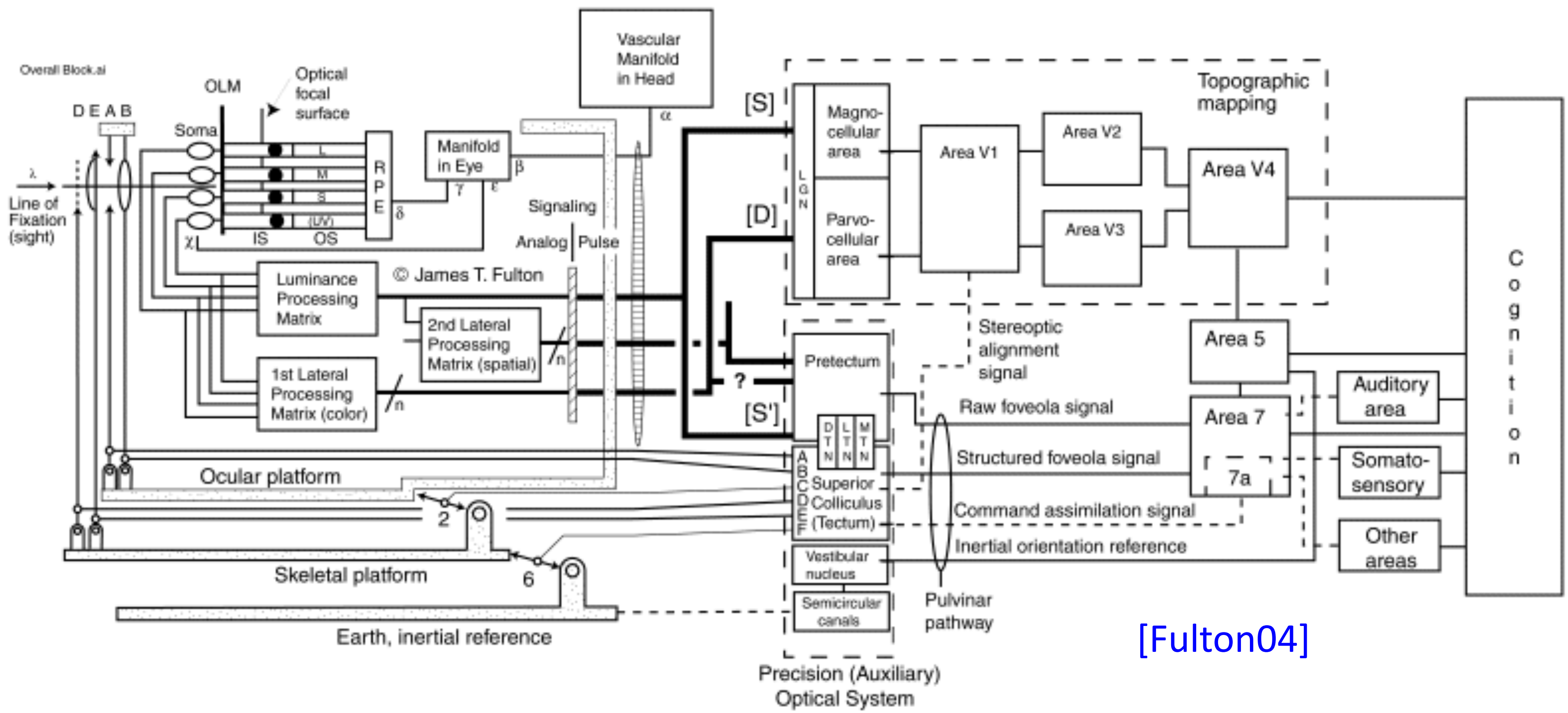
Perception is inherently ambiguous

Julian Beever
Artwork

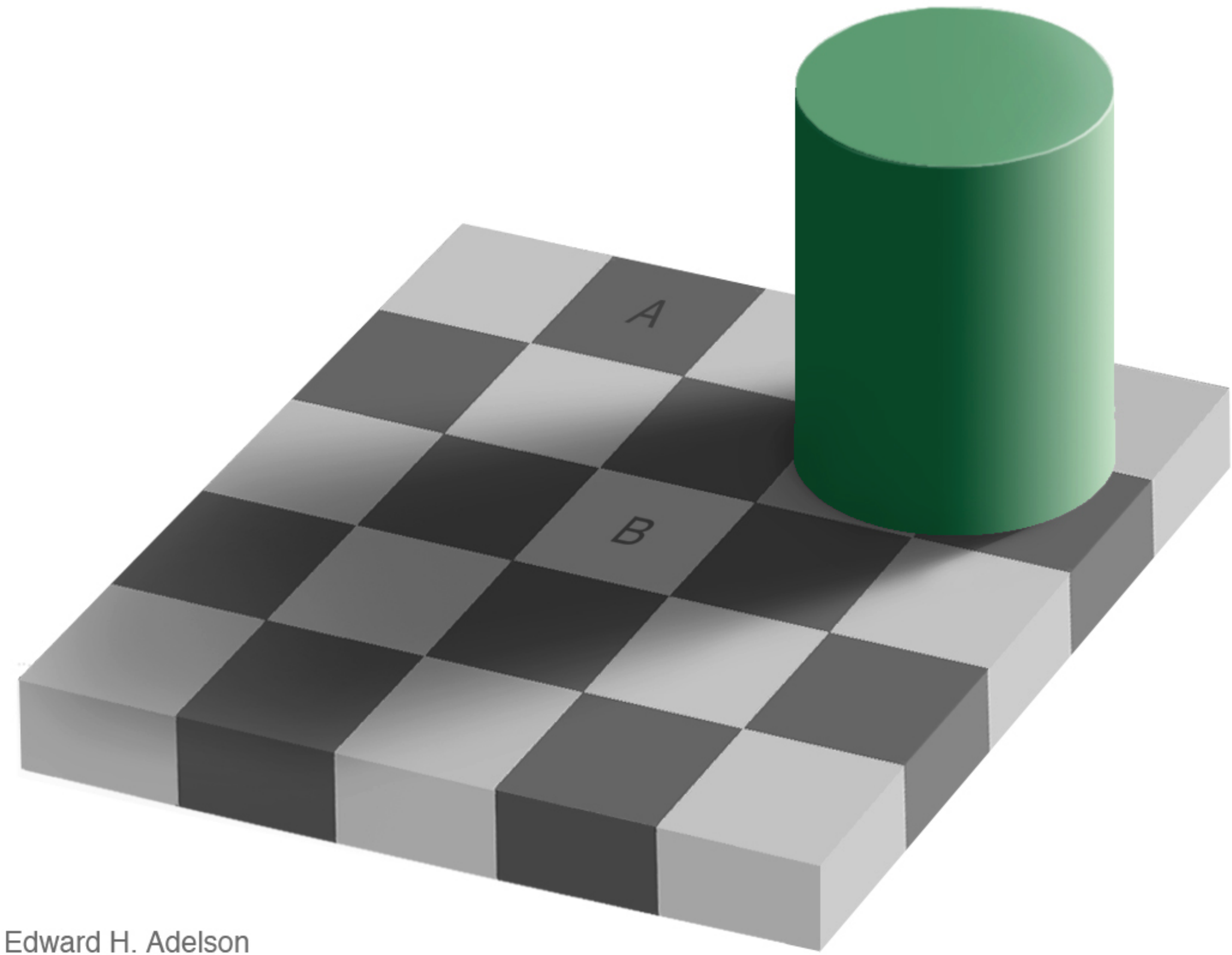
How does human vision work?

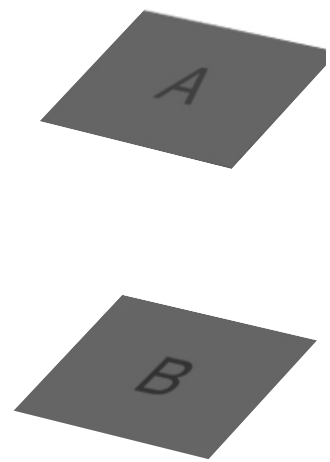
How do people (and animals) see?

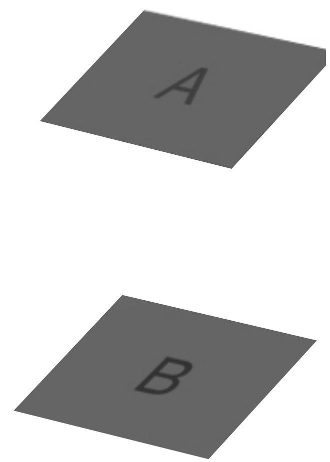
- We don't really know.



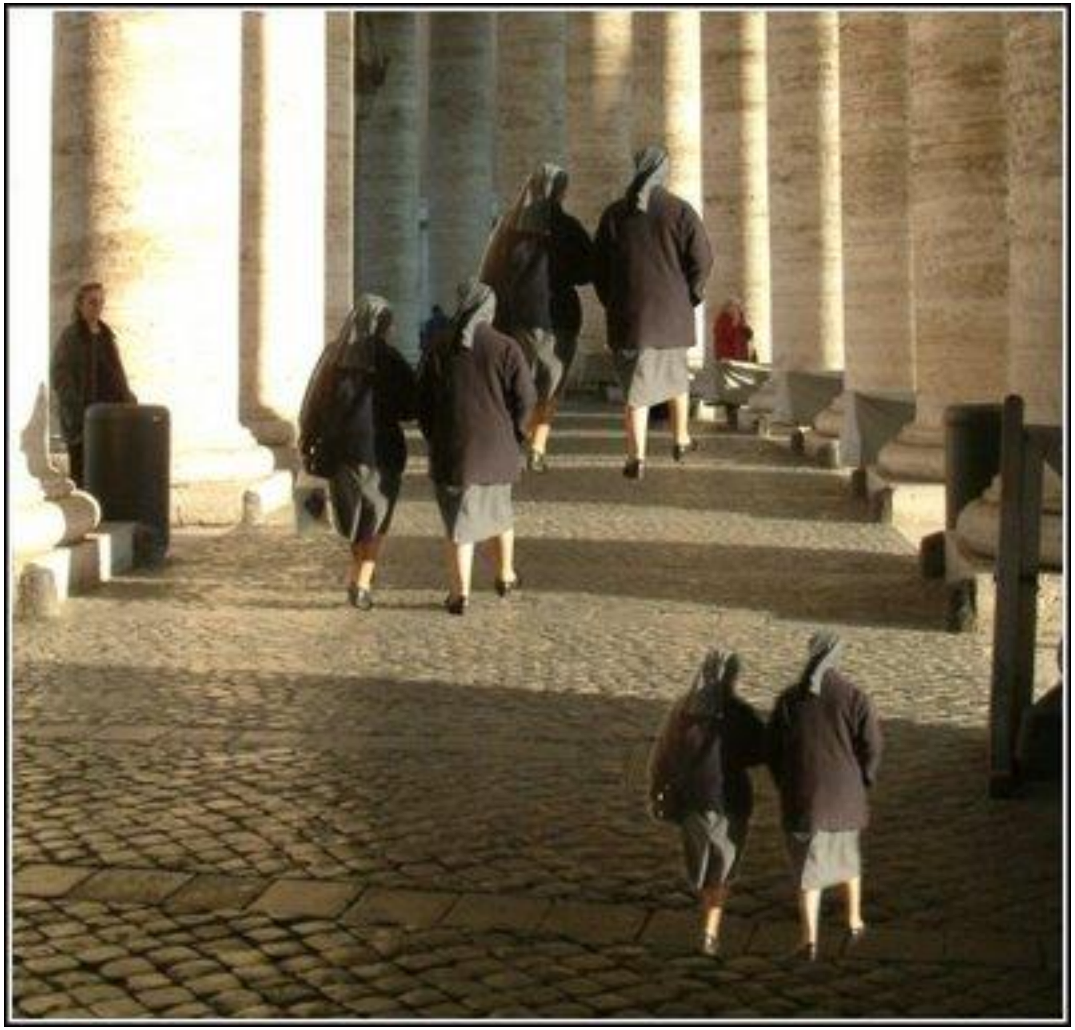
[Fulton04]


A
B
Edward H. Adelson


A
B


A
B

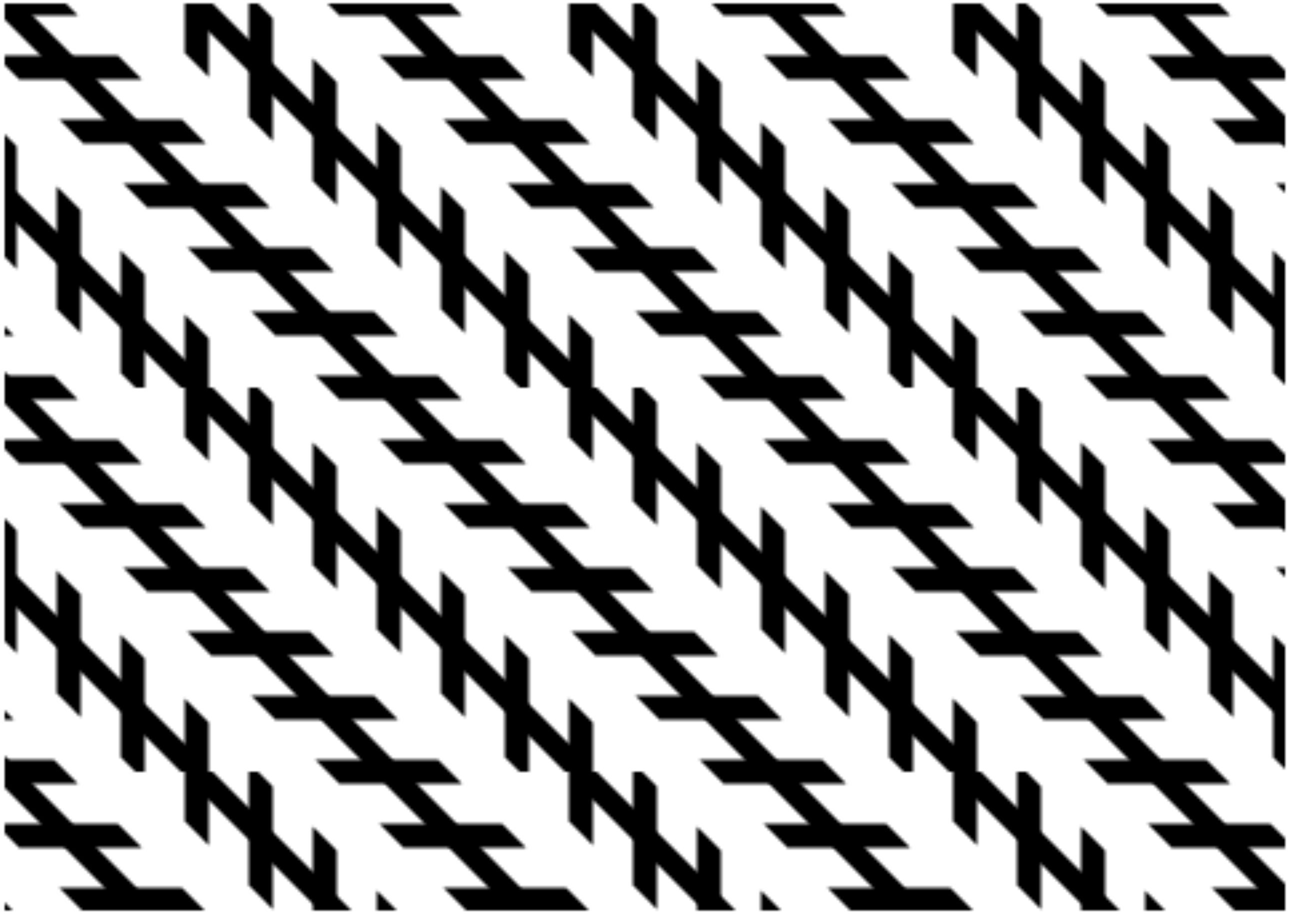
How does human vision work?



How does human vision work?

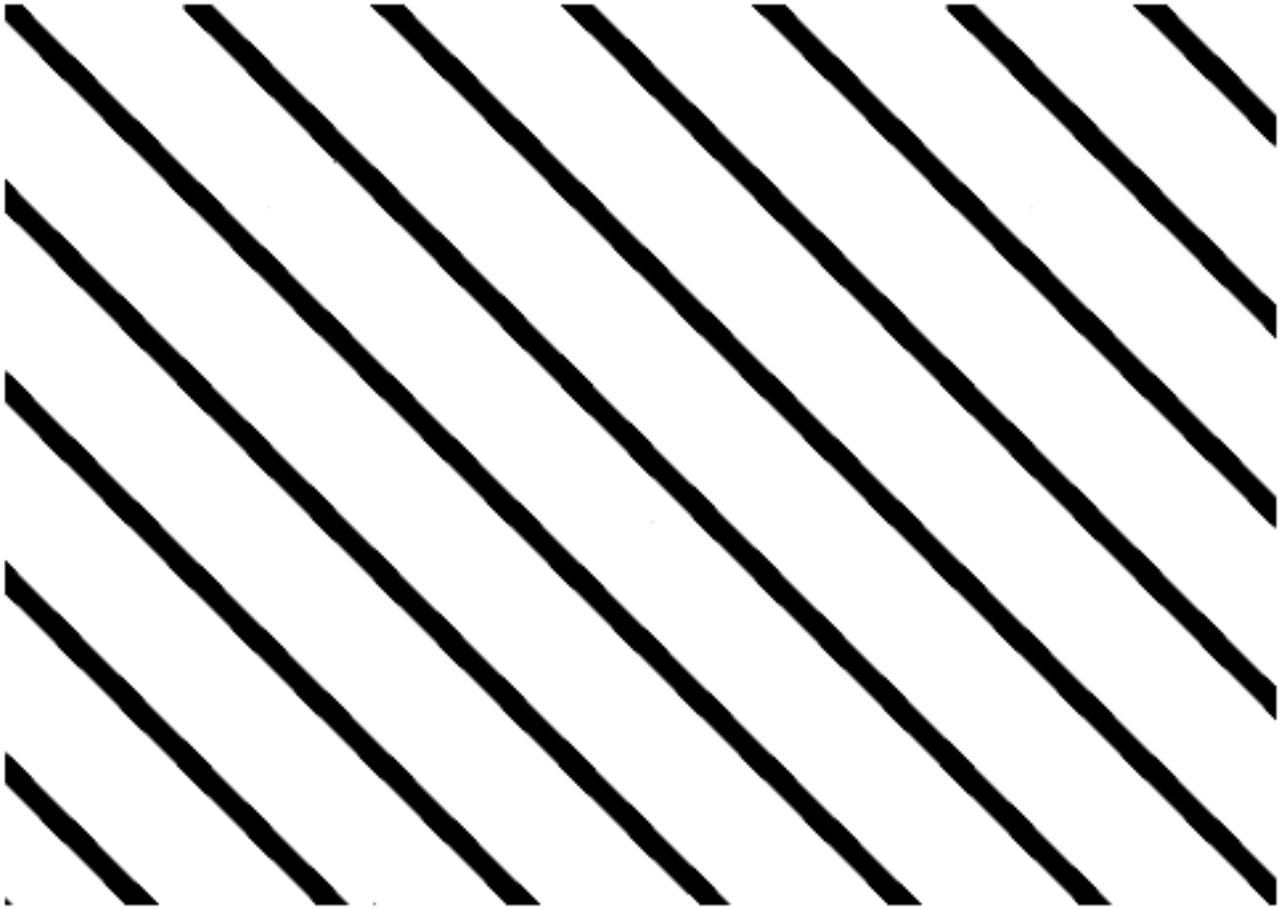
Zollner illusion

- Are these lines parallel?



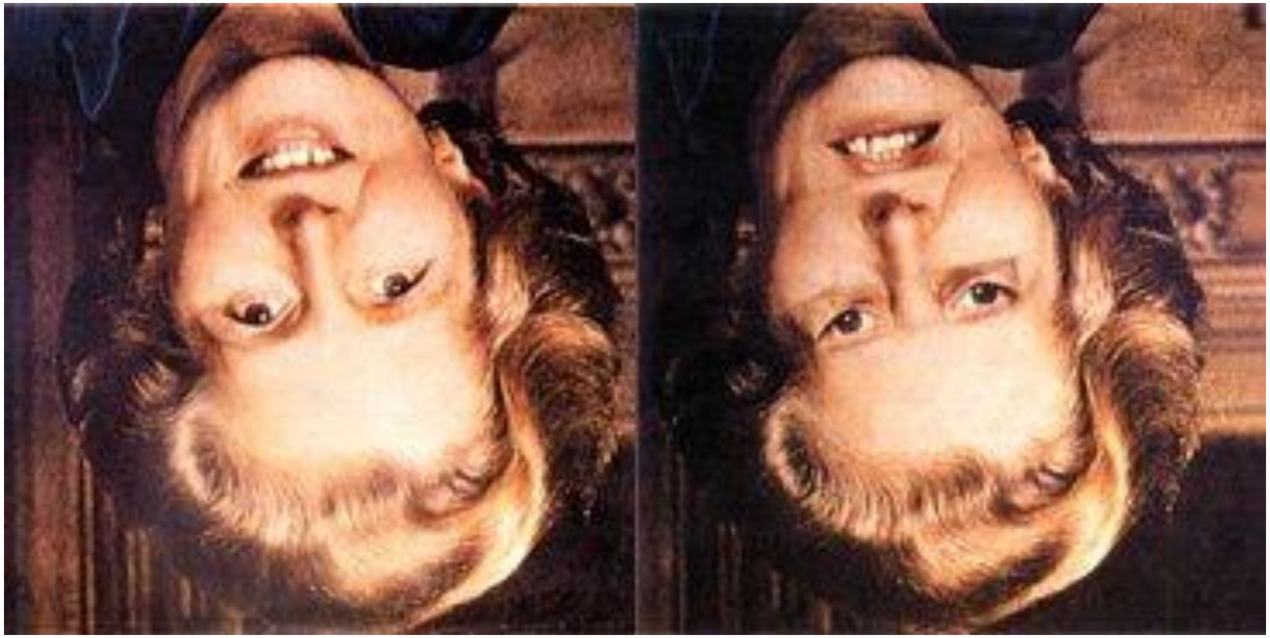
Zollner illusion

- After removing the hatches on these lines, they look parallel



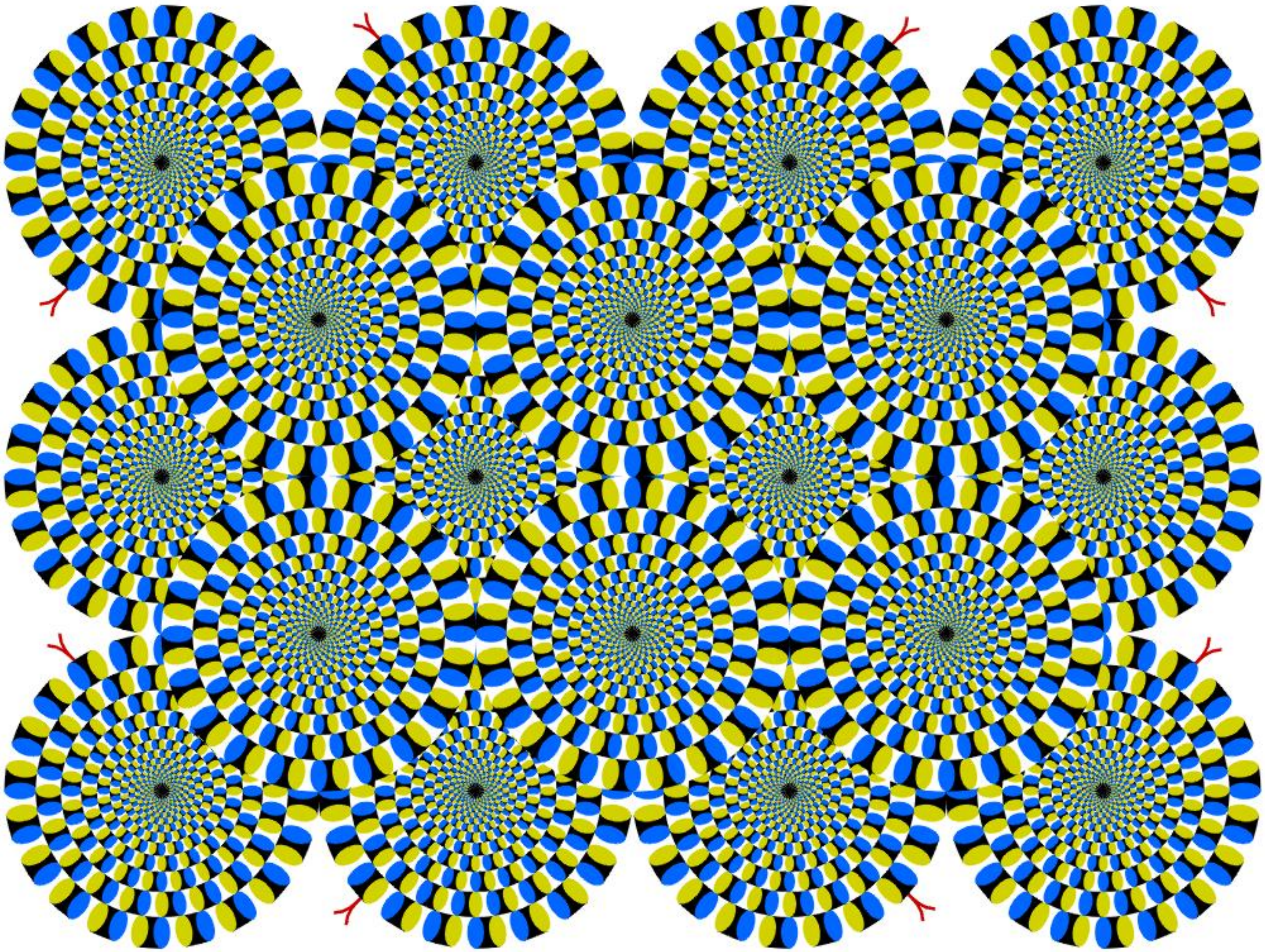
https://www.illusionsindex.org/ir/zoellner-illusion

The Thatcher effect

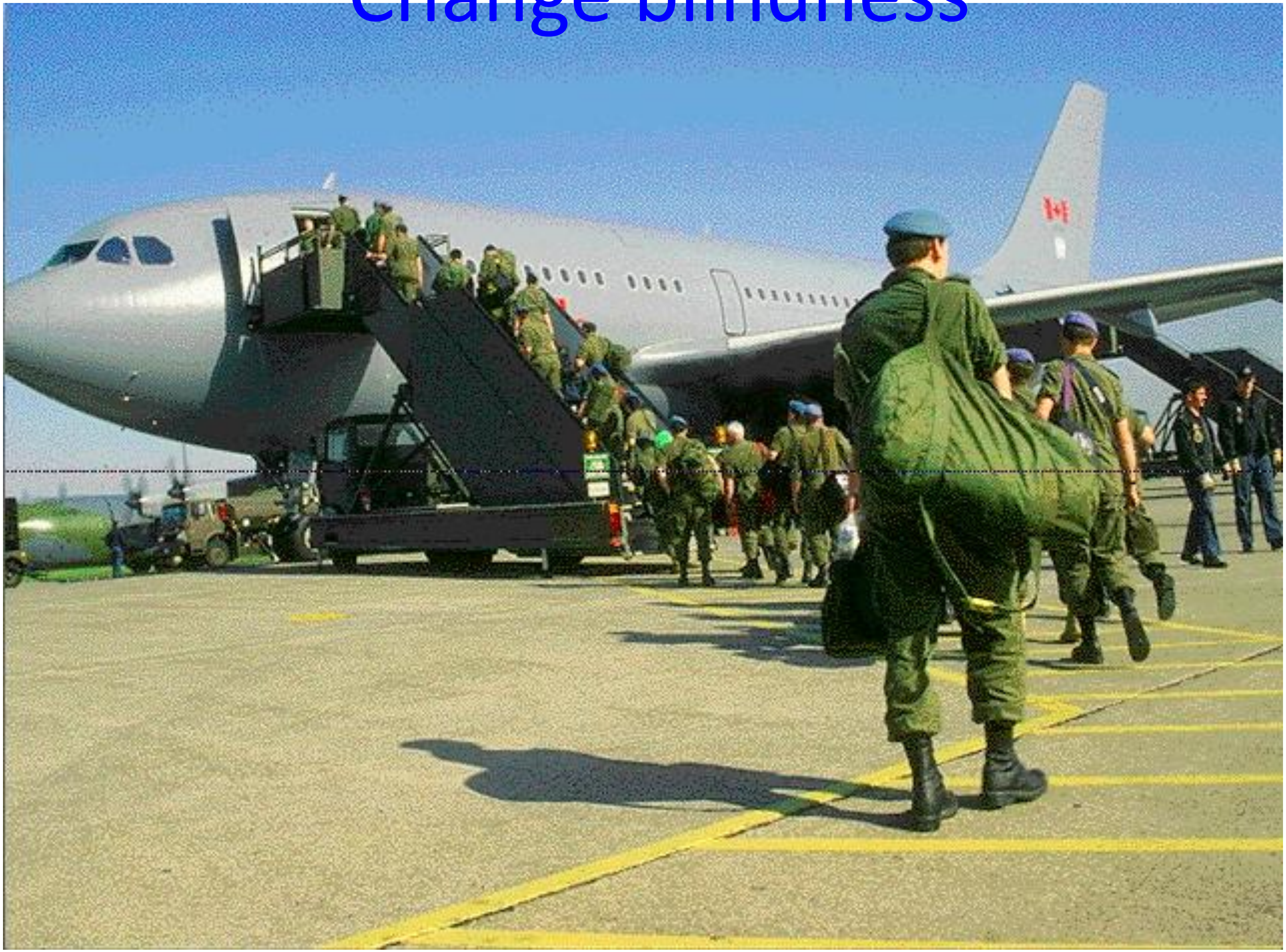


[Thompson 1980]

Rotation illusion

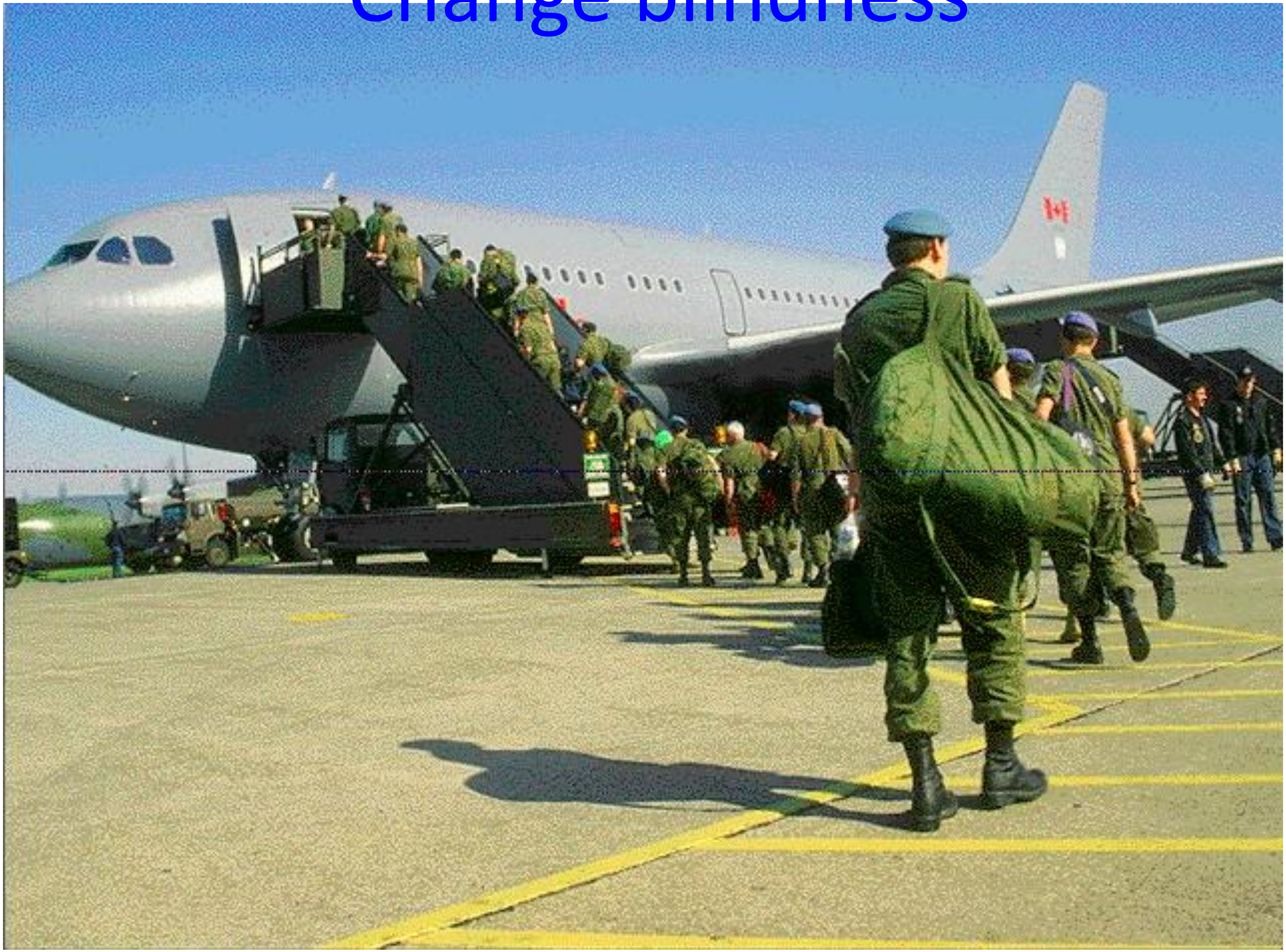


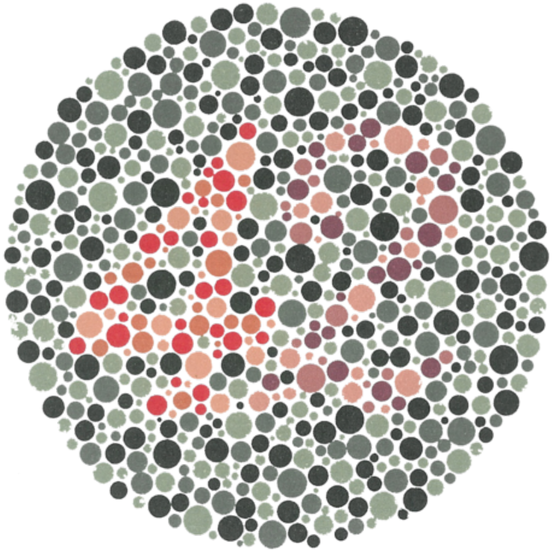
Change blindness

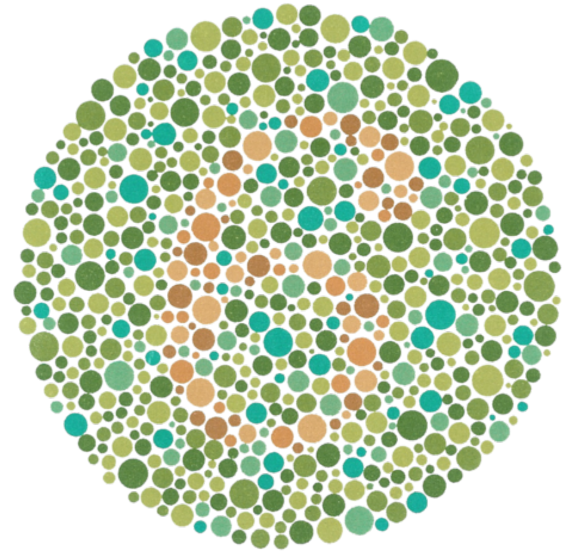


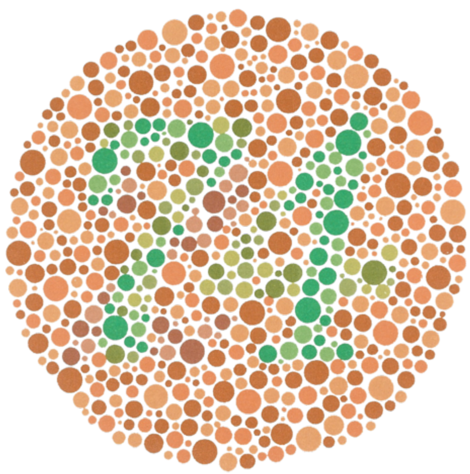
Change blindness

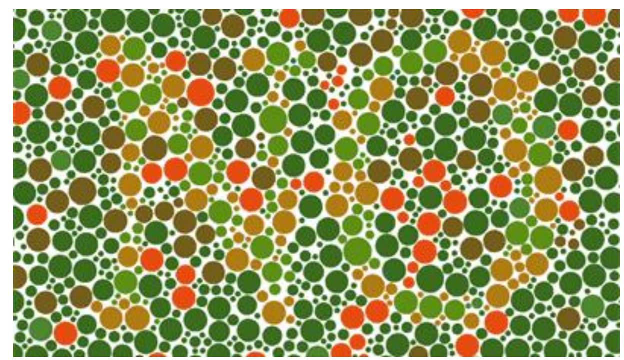
Change blindness











We see the world differently. Even if figure out how to replicate the human vision system the next question would be — Whose vision system do we replicate?

This is yet another example where we experience the world differently

Change blindness

Change blindness

Conclusion: why is computer vision so difficult?

Bad news:

– Computers lack higher-level prior knowledge
– Perception is inherently ambiguous
– We don’t know how the human brain works
– Haven’t found mathematical models that represent human vision well
– The models we do have require intensive (usually intractable) computation

Good news:

– So much progress is being made! Especially in applications where perfect performance isn't needed.