ADVANCES IN COMPUTER VISION

6.8300

Prof. Vincent Sitzmann

About Me

I want to build algorithms that can learn to perceive the world the way you can!

My research group @ MIT:

SCENE REPRESENTATION GROUP

Ulm -> TU Munich -> Stanford -> MIT

Vincent Sitzmann

TA Team

Tianyuan Zhang

Ariba Khan

Chenyu Wang

Jane Millward

Benjamin Cohen-Wang

Adriano Hernandez

Juan Atehortua

Isabella Yu

Vivek Gopalakrishnan

Christian Arnold

Administrativa

Communication

Communication

- Generally**, website is source of truth:**
  - https://www.scenerepresentations.org/courses/2025/spring/advances-in-cv/
  - We will link all course materials there (in addition to canvas)

Communication

- **Lectures**
  - Tuesdays and Thursdays between 1:00pm and 2:30pm @ 26-100.
  - Slides will be posted on the course website shortly before each lecture
  - All lectures will be recorded and uploaded after lecture.

Communication

- **Canvas**
  - **https://canvas.mit.edu/courses/31251**

Communication

- **Piazza**
  - **https://piazza.com/class/m5fhhrt0jzc6hx**
  - Please post any/all questions - also helps other students in this class.
  - Create private posts to instructors instead of emailing (really helps keep track!)

Communication

- **Office Hours**
  - One office hour per TA per week. Schedule online soon.
  - Vincent also one office hour per week. **Vincent will not discuss homeworks.**

This course vs. its previous version

This course vs. its previous version

- This course is a **complete remake** of the original “Advances in Computer Vision”, 6.8300 course.

This course vs. its previous version

- The **material is new.**

This course vs. its previous version

- This course is **more advanced.** There is more math that requires you to be **perfectly comfortable with linear algebra and vectorized programming.**

This course vs. its previous version

- That being said, this is not a *rigorous* class - we will not be proving much, the goal is to gain solid intuition.

This course vs. its previous version

- Take the prerequisites seriously!

Prerequisites

Prerequisites

This class is a **graduate-level class**. For the following topics, there will be no explainers and TAs will not be able to help you with them in office hours.

Prerequisites

- **6.7960 Deep Learning:** Proficiency in Python, Numpy, and PyTorch, vectorized programming, CNNs, Transformers, training deep neural networks.

Prerequisites

- **18.06 Linear Algebra:** Vector spaces, change-of-basis, inner products and norms, Eigenvalues, Eigenvectors, Singular Value Decomposition, Convolution.

Evaluation

Evaluation

- Assignments managed via **Gradescope: https://www.gradescope.com/courses/972401**

Evaluation

- **Assignments**: 65%
  - *5* assignments
  - Individual submissions (feel free to discuss, but not share code)
  - The grade on a homework received n days after the deadline (n<=7) will be multiplied by (1-n/14). We will round up to units of full days; submitting 1 hour late counts as using 1 late days
  - Homeworks will not be accepted more than 7 days after the deadline
  - Ten penalty days will be automatically waived for each student.
  - **No incomplete for this course!**

Evaluation

- **Final Projects**: *35%*
  - Research project on perception of your choice written up as a blog post!
  - Graded for clarity and insight as well as novelty and depth of the experiments and analysis. Detailed guidance will be given later in the semester.
  - Teams of **1-3.** We expect the equivalent of ~4 weeks of work per person.
  - Proposal (10%), blog post (90%)
  - **No late days for blog post!**

Honor Code

Discussing ideas & problems is fair game.
But: Write all code & text by yourself!
This includes AI Assistants.

I am running late? Use late days.

What can get me into trouble?
Copying code from anywhere including the internet, classmate's solution, whiteboard, etc.

FAQ

FAQ

- **Q: Is this class a CI-M class?**
  A: No, this is a graduate class.

FAQ

- **Q: Is 6.8301 (the undergraduate version) taught this semester?**
  A: The undergraduate version is taught this semester as well, and it is a CI-M class. For logistical reasons, it had to be renamed to 6.S058. It is taught by Profs. Bill Freeman and Phillip Isola, and does *not* have a Deep Learning prerequisite.

FAQ

- **Q: Is attendance required? Will lectures be recorded?**
  A: Attendance is at your discretion. Yes, lectures will be recorded and uploaded.

Related Courses & Credits

- CMU 16-825: Learning for 3D Vision
  Prof. Shubham Tulsiani
- Berkeley CS294-158: Deep Unsupervised Learning
  Prof. Pieter Abbeel
- Johns Hopkins: FFTs in Graphics and Vision
  Prof. Misha Kazhdan
- Oxford: Computer Vision
  Prof. Christian Rupprecht
- CMU 16-385: Computer Vision
  Prof. Kris Kitani
- MIT 6.7960: Deep Learning
  Prof. Phillip Isola
- MIT 6.8300/6.8301: Advances in Computer Vision,
  Profs. Bill Freeman, Phillip Isola, Antonio Torralba
- University of Amsterdam: Deep Learning II,
  Geometric Deep Learning
  Prof. Erik Bekkers
- Stanford CS348I: Computer Graphics in the Era of AI
  Profs. C. Karen Liu and Jiajun Wu
- University of Tübingen: Computer Vision
  Prof. Andreas Geiger
- **Ben Mildenhall:** Volume Rendering Slides

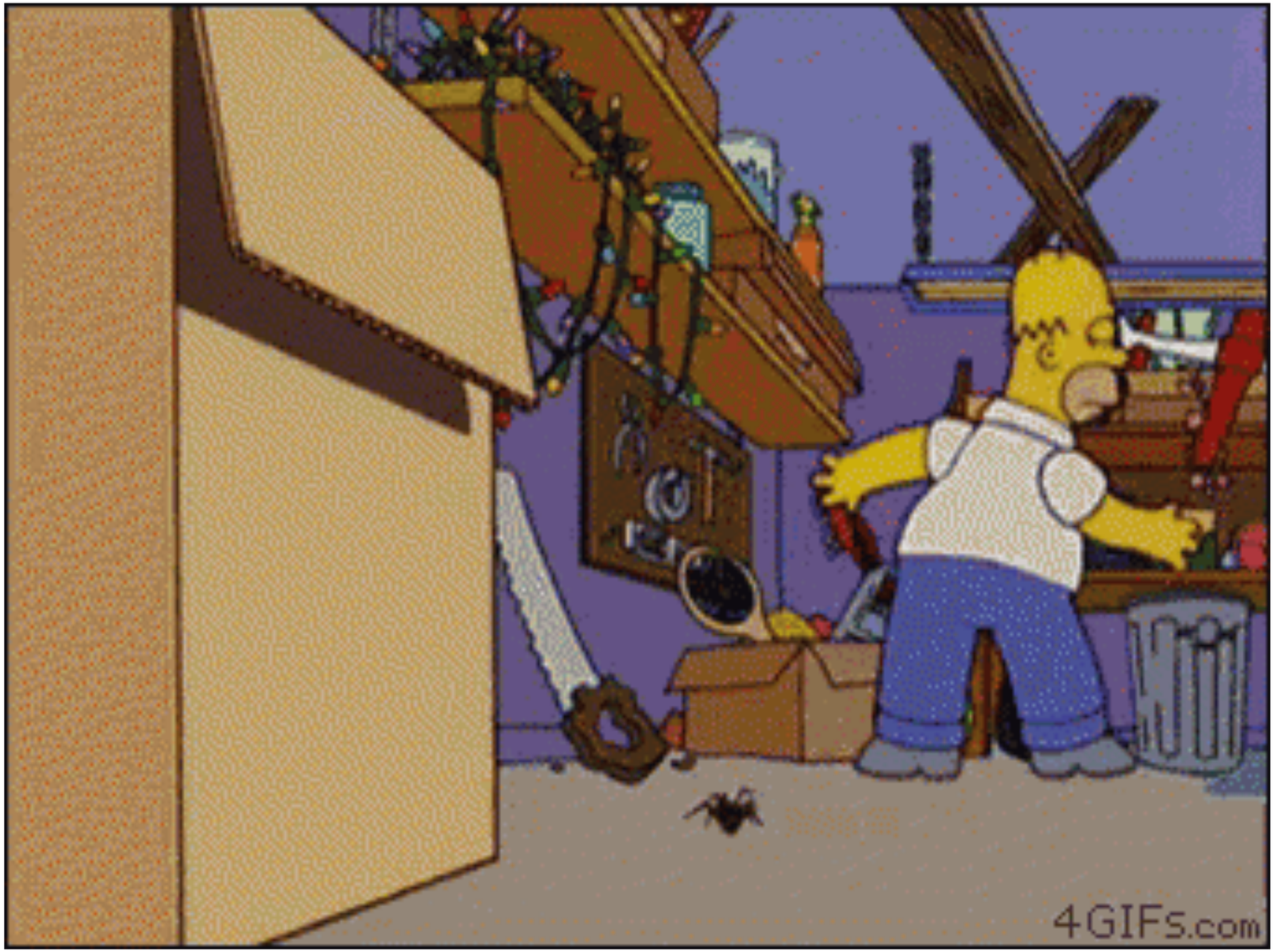
Related Courses & Credits

- CMU 16-825: Learning for 3D Vision
  Pr
- Be
  Ur
  Pr
- Jo
  ar
  Pr
- Ox
  Pr
- Cl
  Prof. Kris Kitani
- MIT 6.7960: Deep Learning
  Prof. Phillip Isola

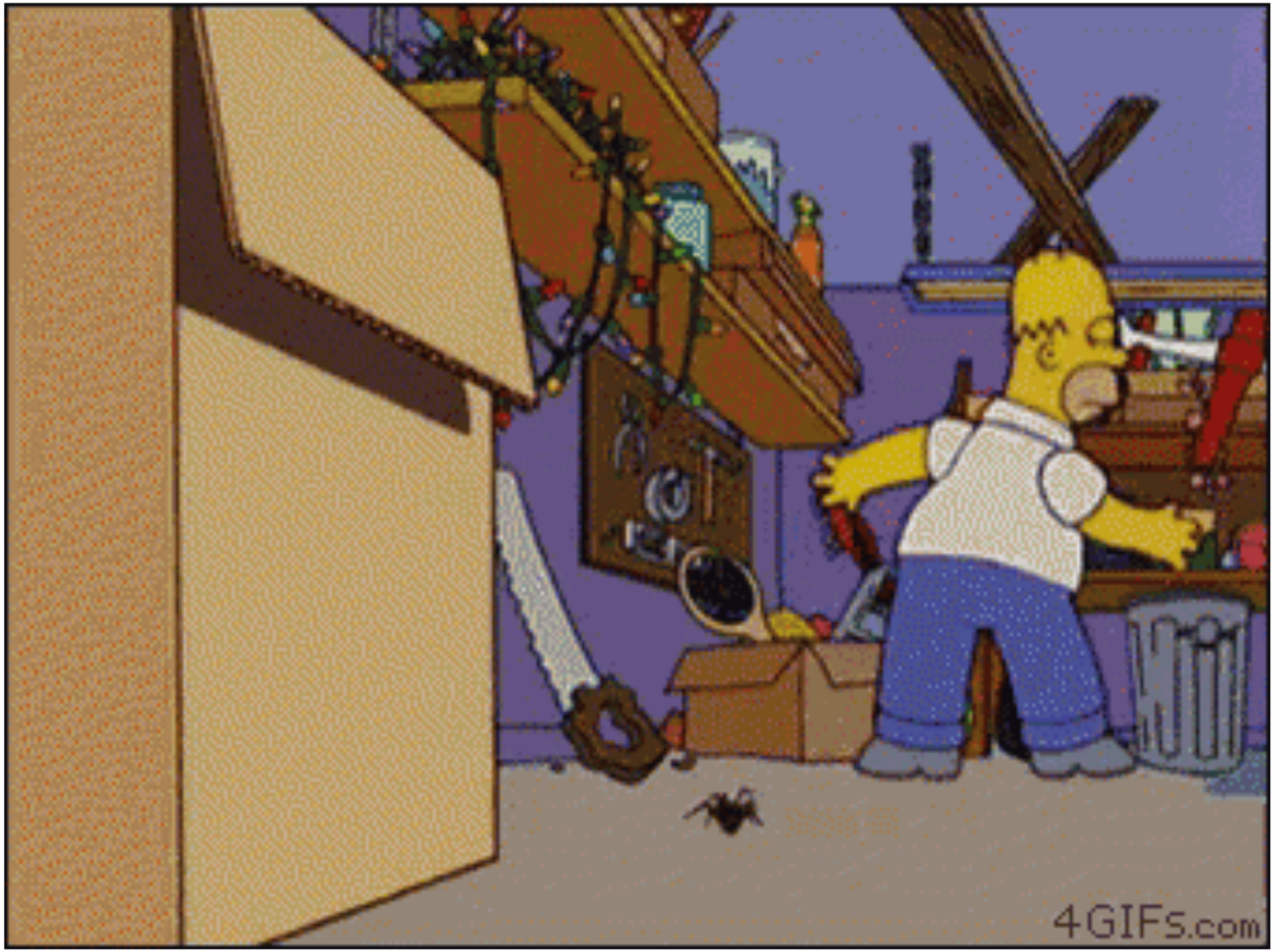
- MIT 6.8300/6.8301: Advances in
  la,
  phics
  Wu
  r
  Vision
  Prof. Andreas Geiger
- **Ben Mildenhall:** Volume Rendering Slides

All of these folks let me use / adapt / build on top of their courses.
Look out for their mentions in the slide credits!

Bugs…

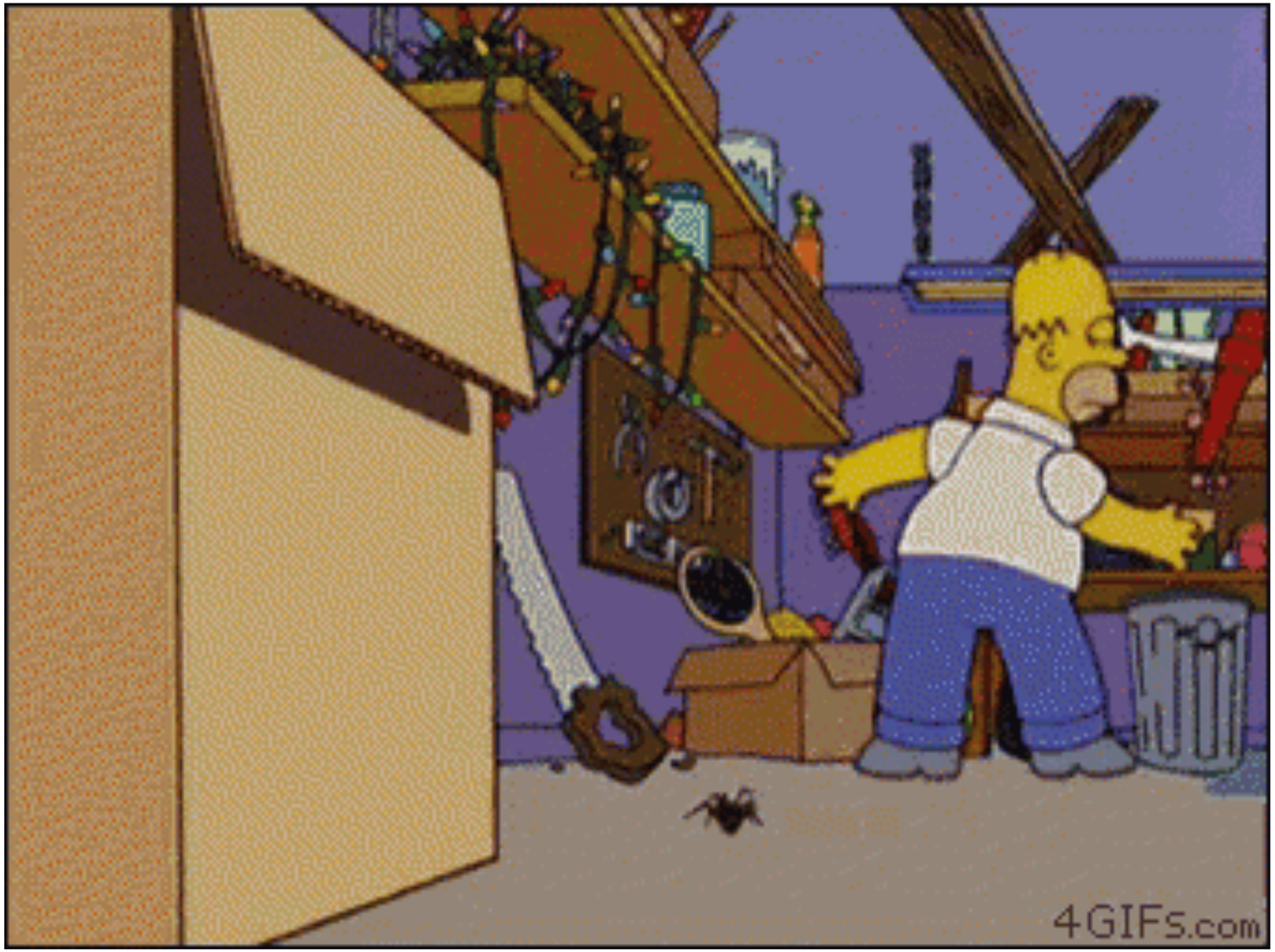


Bugs…



Source: The Simpsons (D’oh!)

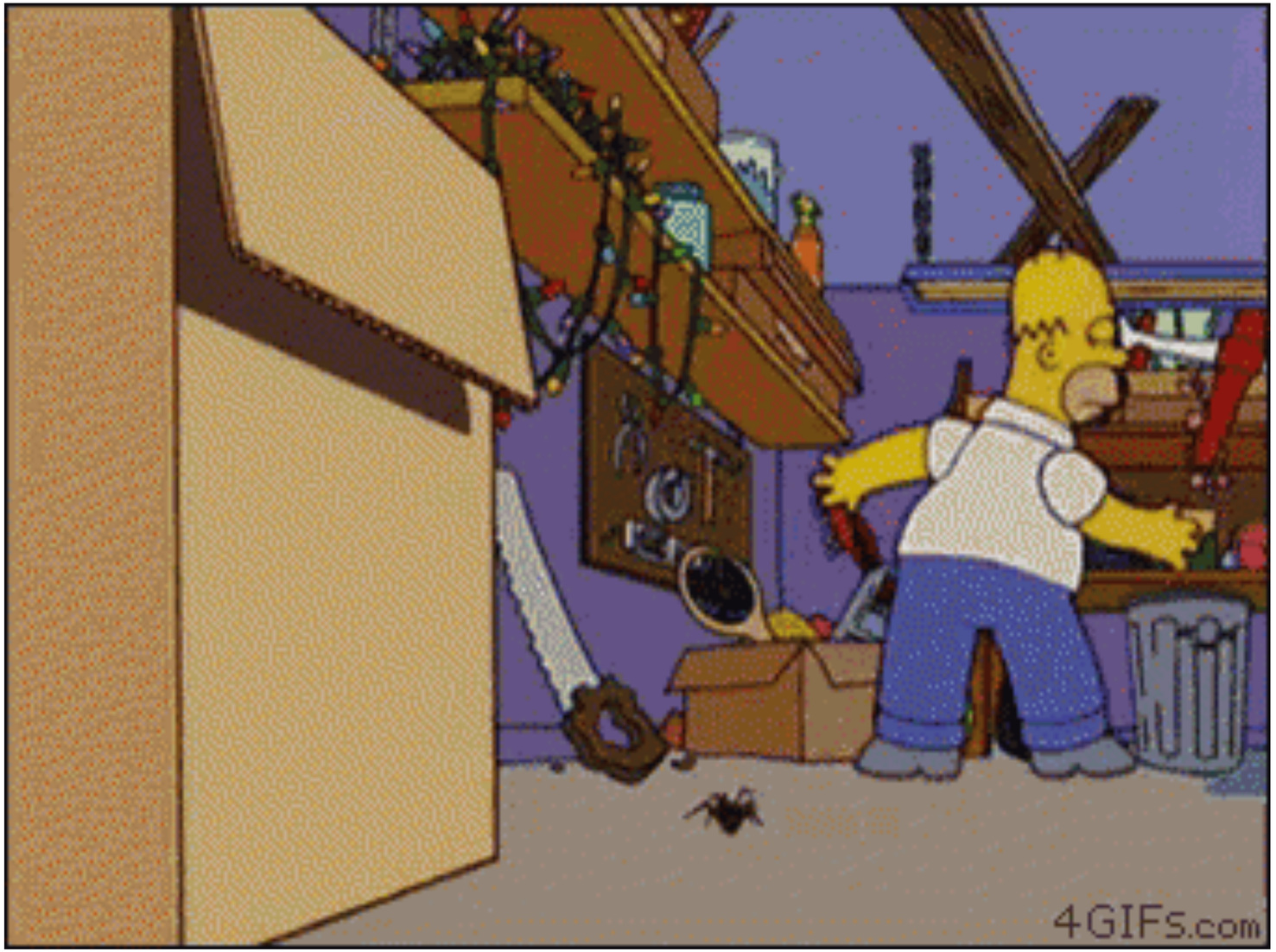
Bugs…



Source: The Simpsons (D’oh!)

- This is the first time I am offering this course.

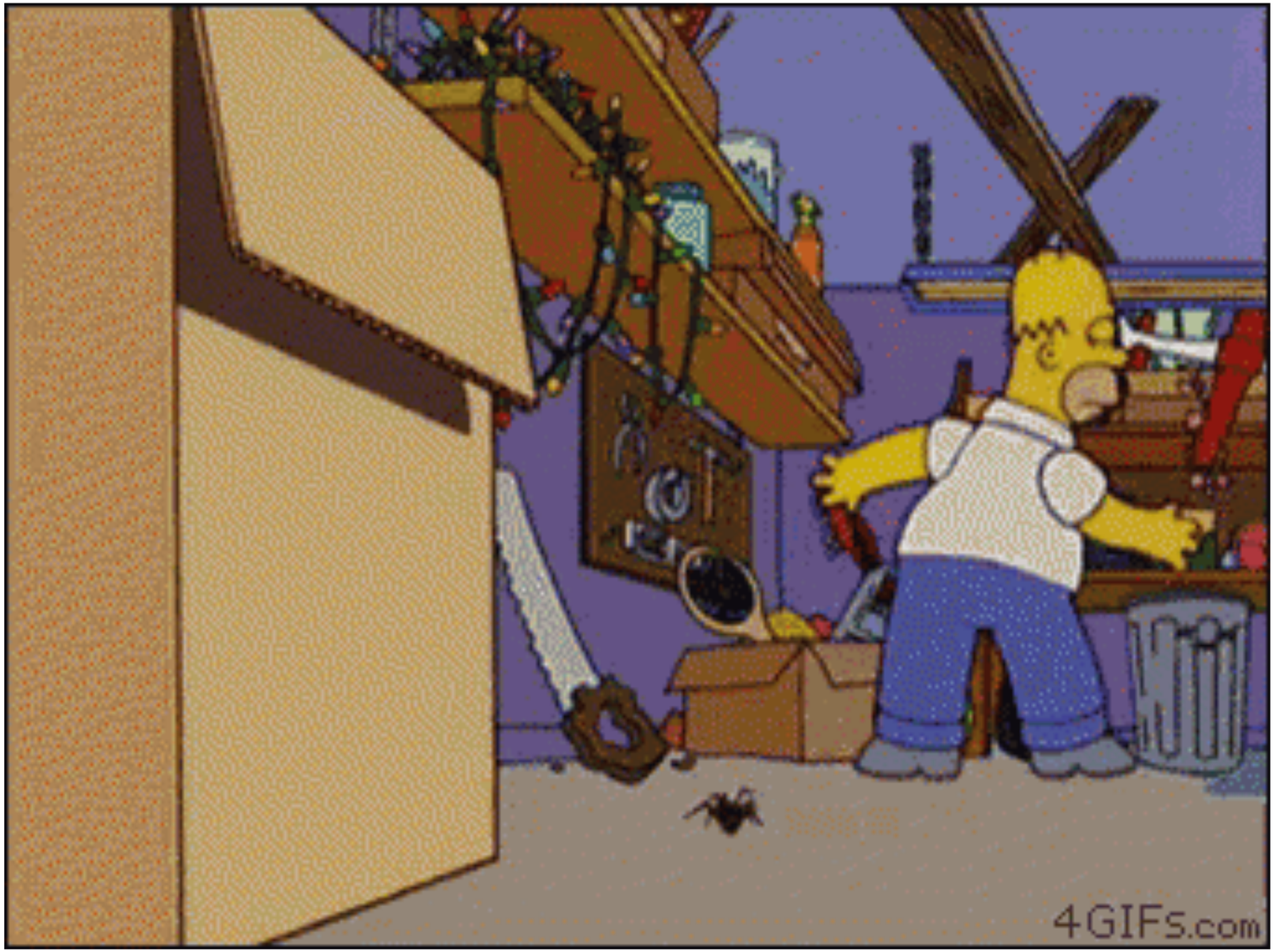
Bugs…



Source: The Simpsons (D’oh!)

- Expect lots of bugs. Most common: Timing bug, lecture too long, too short, changes in lecture order…

Bugs…



Source: The Simpsons (D’oh!)

- Pls give me feedback at the end of the semester! I’ll send around a form :)

ADVANCES IN
MPUTER
N

LECTURE 0:

ntroduction — what appened so far

(Recent) CV History Overview

History slides credits: Andreas Geiger - History of Computer Vision 2021 & Svetlana Lazebnik – Computer Vision: Looking Back to Look Forward

(Recent) CV History Overview

- 1960-1970: Blocks, Edges and Model Fitting

History slides credits: Andreas Geiger - History of Computer Vision 2021 & Svetlana Lazebnik – Computer Vision: Looking Back to Look Forward

(Recent) CV History Overview

- 1960-1970: Blocks, Edges and Model Fitting
- 1970-1981: Low-level vision: stereo, flow, shape-from-x

History slides credits: Andreas Geiger - History of Computer Vision 2021 & Svetlana Lazebnik – Computer Vision: Looking Back to Look Forward

(Recent) CV History Overview

- 1960-1970: Blocks, Edges and Model Fitting
- 1970-1981: Low-level vision: stereo, flow, shape-from-x
- 1985-1988: Neural Networks, backprop., self-driving

History slides credits: Andreas Geiger - History of Computer Vision 2021 & Svetlana Lazebnik – Computer Vision: Looking Back to Look Forward

(Recent) CV History Overview

- 1960-1970: Blocks, Edges and Model Fitting
- 1970-1981: Low-level vision: stereo, flow, shape-from-x
- 1985-1988: Neural Networks, backprop., self-driving
- 1990-2000: Dense stereo and multi-view stereo, MRFs

History slides credits: Andreas Geiger - History of Computer Vision 2021 & Svetlana Lazebnik – Computer Vision: Looking Back to Look Forward

(Recent) CV History Overview

- 1960-1970: Blocks, Edges and Model Fitting
- 1970-1981: Low-level vision: stereo, flow, shape-from-x
- 1985-1988: Neural Networks, backprop., self-driving
- 1990-2000: Dense stereo and multi-view stereo, MRFs
- 2000-2010: Features, descriptors, structure-from-motion.

History slides credits: Andreas Geiger - History of Computer Vision 2021 & Svetlana Lazebnik – Computer Vision: Looking Back to Look Forward

(Recent) CV History Overview

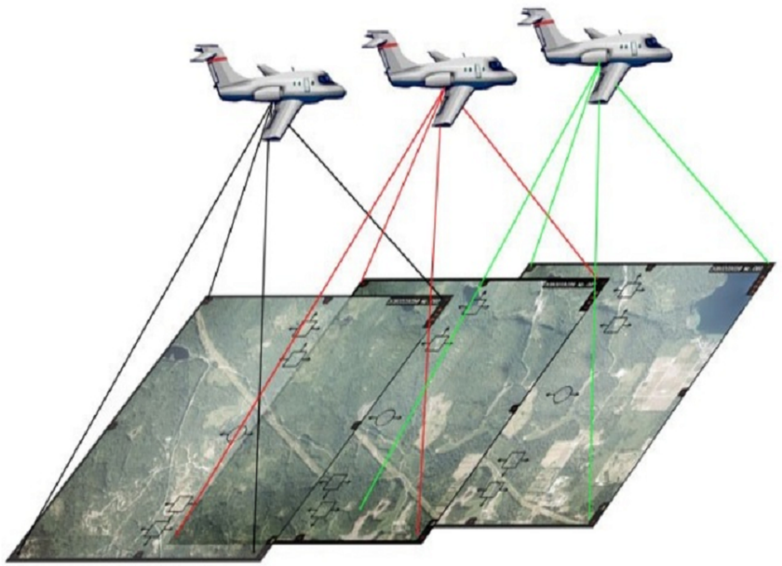
- 1960-1970: Blocks, Edges and Model Fitting
- 1970-1981: Low-level vision: stereo, flow, shape-from-x
- 1985-1988: Neural Networks, backprop., self-driving
- 1990-2000: Dense stereo and multi-view stereo, MRFs
- 2000-2010: Features, descriptors, structure-from-motion.
- 2010-20??: Learning, deep learning, large datasets, rapid growth

History slides credits: Andreas Geiger - History of Computer Vision 2021 & Svetlana Lazebnik – Computer Vision: Looking Back to Look Forward

1957: Stereo(photogrammetry)

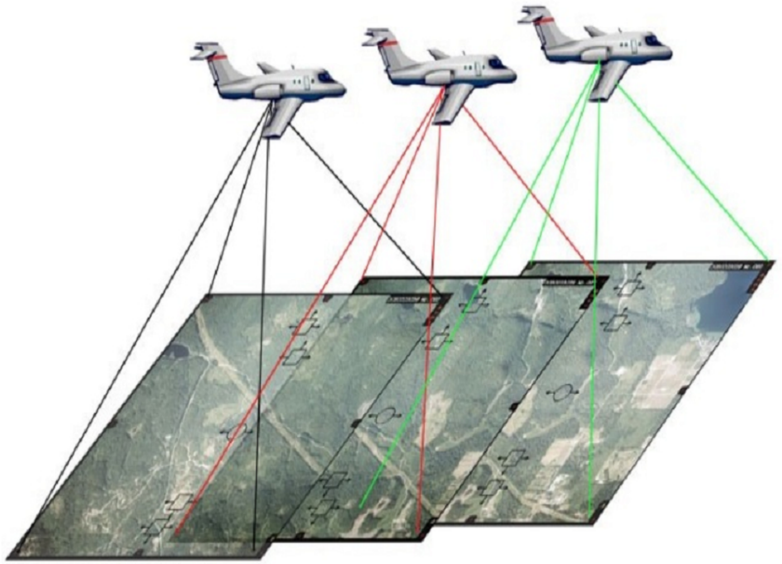
Louis, Hobrough Gilbert. "Methods and apparatus for correlating corresponding points in two images." U.S. Patent No. 2,964,642. 13 Dec. 1960
Slide credit: Christian Rupprecht, Oxford

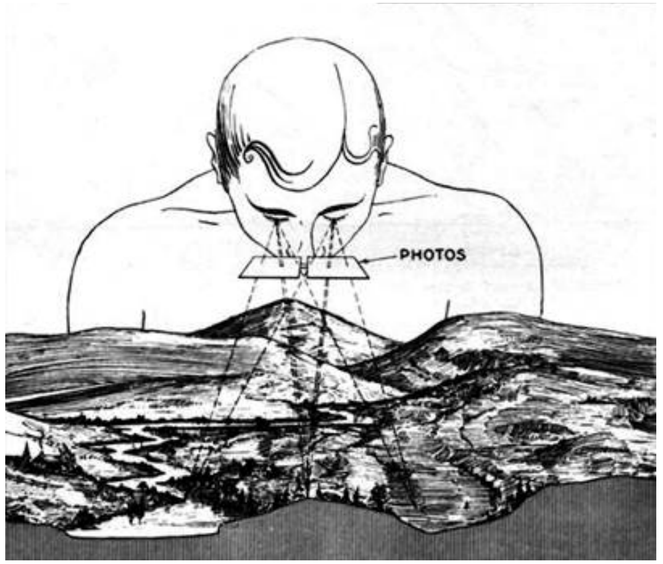
1957: Stereo(photogrammetry)



Louis, Hobrough Gilbert. "Methods and apparatus for correlating corresponding points in two images." U.S. Patent No. 2,964,642. 13 Dec. 1960
Slide credit: Christian Rupprecht, Oxford

1957: Stereo(photogrammetry)

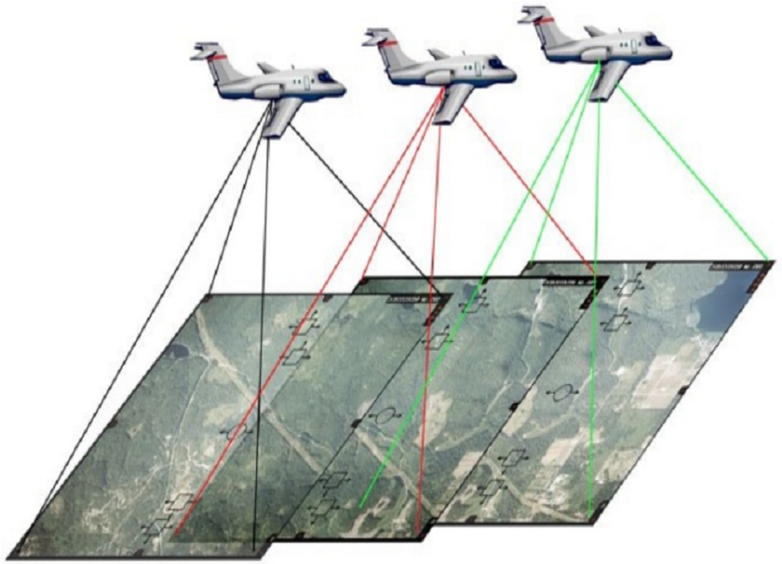


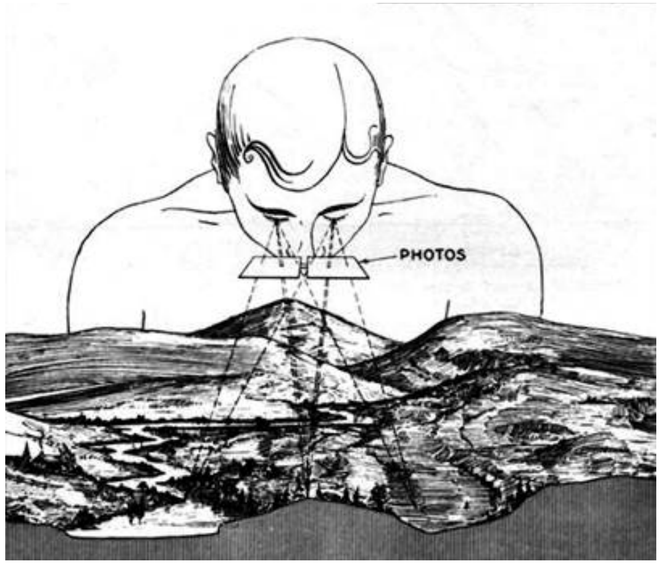


Louis, Hobrough Gilbert. "Methods and apparatus for correlating corresponding points in two images." U.S. Patent No. 2,964,642. 13 Dec. 1960
Slide credit: Christian Rupprecht, Oxford

1957: Stereo(photogrammetry)

- Gilbert Hobrough: analog implementation of stereo image correlation

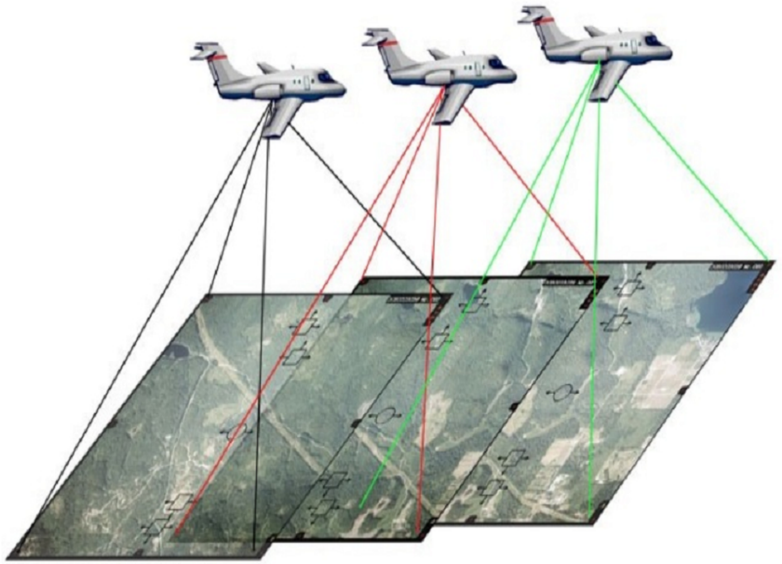


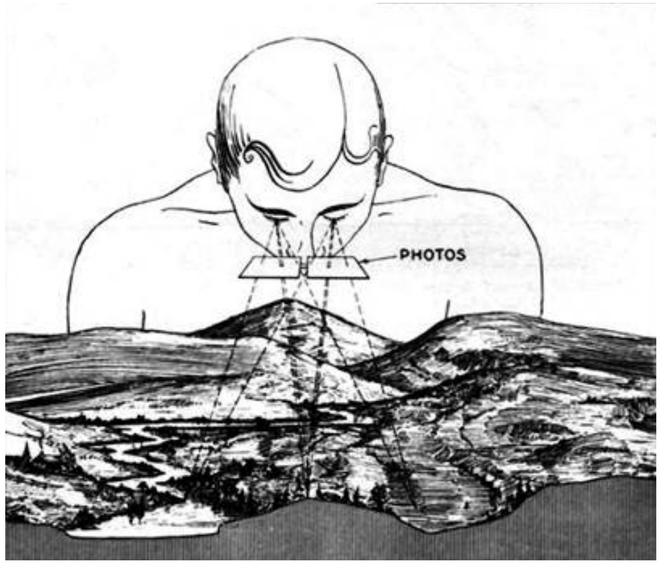


Louis, Hobrough Gilbert. "Methods and apparatus for correlating corresponding points in two images." U.S. Patent No. 2,964,642. 13 Dec. 1960
Slide credit: Christian Rupprecht, Oxford

1957: Stereo(photogrammetry)

- Gilbert Hobrough: analog implementation of stereo image correlation
- Used to create elevation maps (Photogrammetry, since 1840)

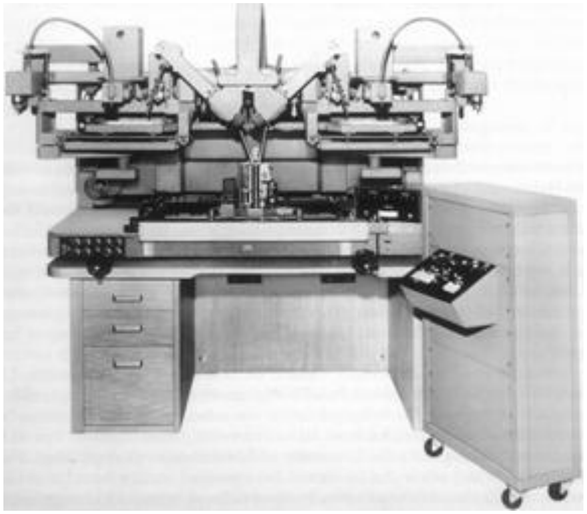




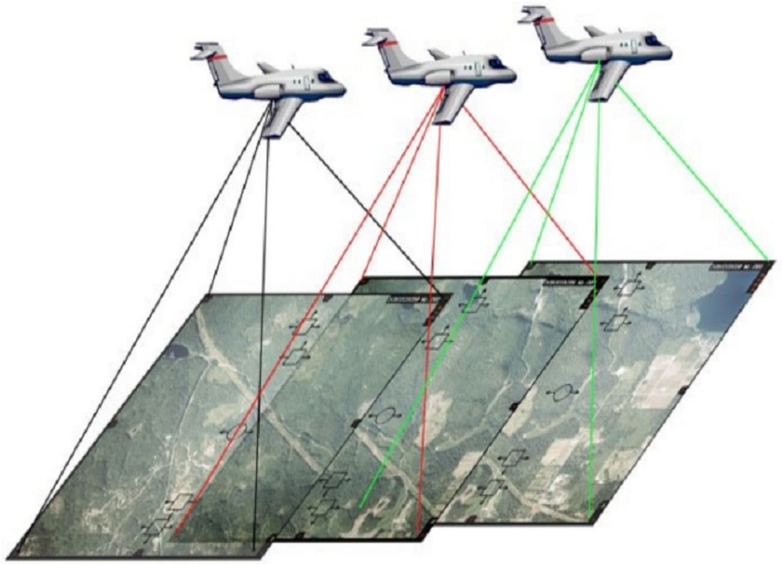
Louis, Hobrough Gilbert. "Methods and apparatus for correlating corresponding points in two images." U.S. Patent No. 2,964,642. 13 Dec. 1960
Slide credit: Christian Rupprecht, Oxford

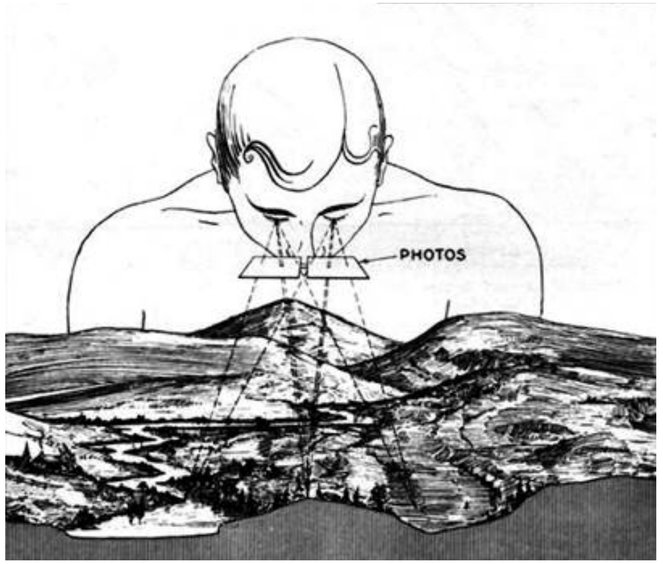
1957: Stereo(photogrammetry)

- Gilbert Hobrough: analog implementation of stereo image correlation
- Used to create elevation maps (Photogrammetry, since 1840)



Wild B8 (721x produced 1961 -1972)



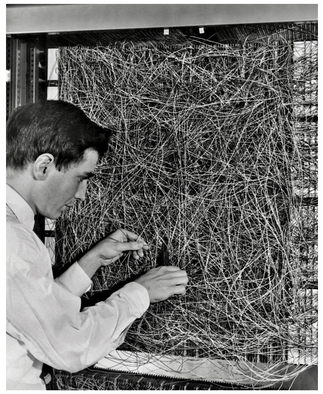


Louis, Hobrough Gilbert. "Methods and apparatus for correlating corresponding points in two images." U.S. Patent No. 2,964,642. 13 Dec. 1960
Slide credit: Christian Rupprecht, Oxford

1958-1962: Rosenblatt’s Perceptron

Rosenblatt, Frank. "The perceptron: a probabilistic model for information storage and organization in the brain." *Psychological review* 65.6 (1958): 386.
Slide credit: Christian Rupprecht, Oxford

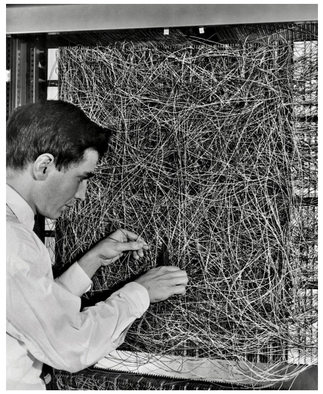
1958-1962: Rosenblatt’s Perceptron



Rosenblatt, Frank. "The perceptron: a probabilistic model for information storage and organization in the brain." *Psychological review* 65.6 (1958): 386.

Slide credit: Christian Rupprecht, Oxford

1958-1962: Rosenblatt's Perceptron



1958 New York Times…

NEW NAVY DEVICE LEARNS BY DOING

Psychologist Shows Embryo of Computer Designed to Read and Grow Wiser

WASHINGTON, July 7 (UPI) —The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.

The embryo—the Weather Bureau's $2,000,000 "704" computer—learned to differentiate between right and left after fifty attempts in the Navy's demonstration for newsmen.

The service said it would use this principle to build the first of its Perceptron thinking machines that will be able to read and write. It is expected to be finished in about a year at a cost of $100,000.

Dr. Frank Rosenblatt, designer of the Perceptron, conducted the demonstration. He said the machine would be the first device to think as the human brain. As do human beings, Perceptron will make mistakes at first, but will grow wiser as it gains experience, he said.

Dr. Rosenblatt, a research psychologist at the Cornell Aeronautical Laboratory, Buffalo, said Perceptrons might be fired to the planets as mechanical space explorers.

Without Human Controls

The Navy said the perceptron would be the first non-living mechanism "capable of receiving, recognizing and identifying its surroundings without any human training or control."

The "brain" is designed to remember images and information it has perceived itself. Ordinary computers remember only what is fed into them on punch cards or magnetic tape.

Later Perceptrons will be able to recognize people and call out their names and instantly translate speech in one language to speech or writing in another language, it was predicted.

Mr. Rosenblatt said in principle it would be possible to build brains that could reproduce themselves on an assembly line and which would be conscious of their existence.

In today's demonstration, the "704" was fed two cards, one with squares marked on the left side and the other with squares on the right side.

Learns by Doing

In the first fifty trials, the machine made no distinction between them. It then started registering a "Q" for the left squares and "O" for the right squares.

Dr. Rosenblatt said he could explain why the machine learned only in highly technical terms. But he said the computer had undergone a "self-induced change in the wiring diagram."

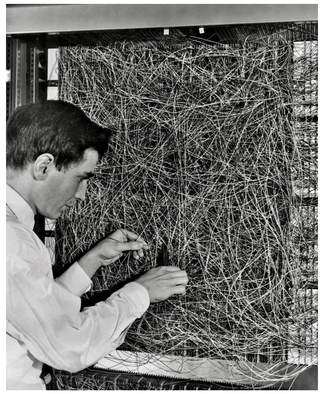
The first Perceptron will have about 1,000 electronic "association cells" receiving electrical impulses from an eye-like scanning device with 400 photo-cells. The human brain has 10,000,000,000 responsive cells, including 100,000,000 connections with the eyes.

Rosenblatt, Frank. "The perceptron: a probabilistic model for information storage and organization in the brain." *Psychological review* 65.6 (1958): 386.

Slide credit: Christian Rupprecht, Oxford

1958-1962: Rosenblatt's Perceptron

- First algorithm and implementation for training single linear "threshold neuron"



NEW NAVY DEVICE LEARNS BY DOING

Psychologist Shows Embryo of Computer Designed to Read and Grow Wiser

WASHINGTON, July 7 (UPI) —The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.

The embryo—the Weather Bureau's $2,000,000 "704" computer—learned to differentiate between right and left after fifty attempts in the Navy's demonstration for newsmen.

The service said it would use this principle to build the first of its Perceptron thinking machines that will be able to read and write. It is expected to be finished in about a year at a cost of $100,000.

Dr. Frank Rosenblatt, designer of the Perceptron, conducted the demonstration. He said the machine would be the first device to think as the human brain. As do human beings, Perceptron will make mistakes at first, but will grow wiser as it gains experience, he said.

Dr. Rosenblatt, a research psychologist at the Cornell Aeronautical Laboratory, Buffalo, said Perceptrons might be fired to the planets as mechanical space explorers.

Without Human Controls

The Navy said the perceptron would be the first non-living mechanism "capable of receiving, recognizing and identifying its surroundings without any human training or control."

The "brain" is designed to remember images and information it has perceived itself. Ordinary computers remember only what is fed into them on punch cards or magnetic tape.

Later Perceptrons will be able to recognize people and call out their names and instantly translate speech in one language to speech or writing in another language, it was predicted.

Mr. Rosenblatt said in principle it would be possible to build brains that could reproduce themselves on an assembly line and which would be conscious of their existence.

1958 New York Times…

In today's demonstration, the "704" was fed two cards, one with squares marked on the left side and the other with squares on the right side.

Learns by Doing

In the first fifty trials, the machine made no distinction between them. It then started registering a "Q" for the left squares and "O" for the right squares.

Dr. Rosenblatt said he could explain why the machine learned only in highly technical terms. But he said the computer had undergone a "self-induced change in the wiring diagram."

The first Perceptron will have about 1,000 electronic "association cells" receiving electrical impulses from an eye-like scanning device with 400 photo-cells. The human brain has 10,000,000,000 responsive cells, including 100,000,000 connections with the eyes.

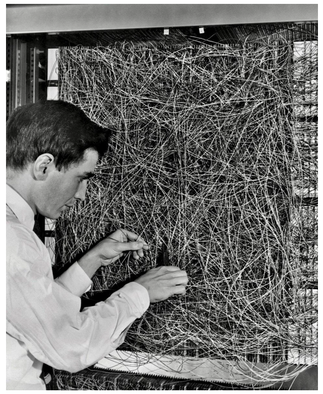
Rosenblatt, Frank. "The perceptron: a probabilistic model for information storage and organization in the brain." *Psychological review* 65.6 (1958): 386.

Slide credit: Christian Rupprecht, Oxford

1958-1962: Rosenblatt's Perceptron

- First algorithm and implementation for training single linear "threshold neuron"
- Perceptron criterion:

$$L(w) = -\sum_{n \in M} w^T x_n y_n$$



1958 New York Times…

NEW NAVY DEVICE LEARNS BY DOING

Psychologist Shows Embryo of Computer Designed to Read and Grow Wiser

WASHINGTON, July 7 (UPI) —The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.

The embryo—the Weather Bureau's $2,000,000 "704" computer—learned to differentiate between right and left after fifty attempts in the Navy's demonstration for newsmen.

The service said it would use this principle to build the first of its Perceptron thinking machines that will be able to read and write. It is expected to be finished in about a year at a cost of $100,000.

Dr. Frank Rosenblatt, designer of the Perceptron, conducted the demonstration. He said the machine would be the first device to think as the human brain. As do human beings, Perceptron will make mistakes at first, but will grow wiser as it gains experience, he said.

Dr. Rosenblatt, a research psychologist at the Cornell Aeronautical Laboratory, Buffalo, said Perceptrons might be fired to the planets as mechanical space explorers.

Without Human Controls

The Navy said the perceptron would be the first non-living mechanism "capable of receiving, recognizing and identifying its surroundings without any human training or control."

The "brain" is designed to remember images and information it has perceived itself. Ordinary computers remember only what is fed into them on punch cards or magnetic tape.

Later Perceptrons will be able to recognize people and call out their names and instantly translate speech in one language to speech or writing in another language, it was predicted.

Mr. Rosenblatt said in principle it would be possible to build brains that could reproduce themselves on an assembly line and which would be conscious of their existence.

In today's demonstration, the "704" was fed two cards, one with squares marked on the left side and the other with squares on the right side.

Learns by Doing

In the first fifty trials, the machine made no distinction between them. It then started registering a "Q" for the left squares and "O" for the right squares.

Dr. Rosenblatt said he could explain why the machine learned only in highly technical terms. But he said the computer had undergone a "self-induced change in the wiring diagram."

The first Perceptron will have about 1,000 electronic "association cells" receiving electrical impulses from an eye-like scanning device with 400 photo-cells. The human brain has 10,000,000,000 responsive cells, including 100,000,000 connections with the eyes.

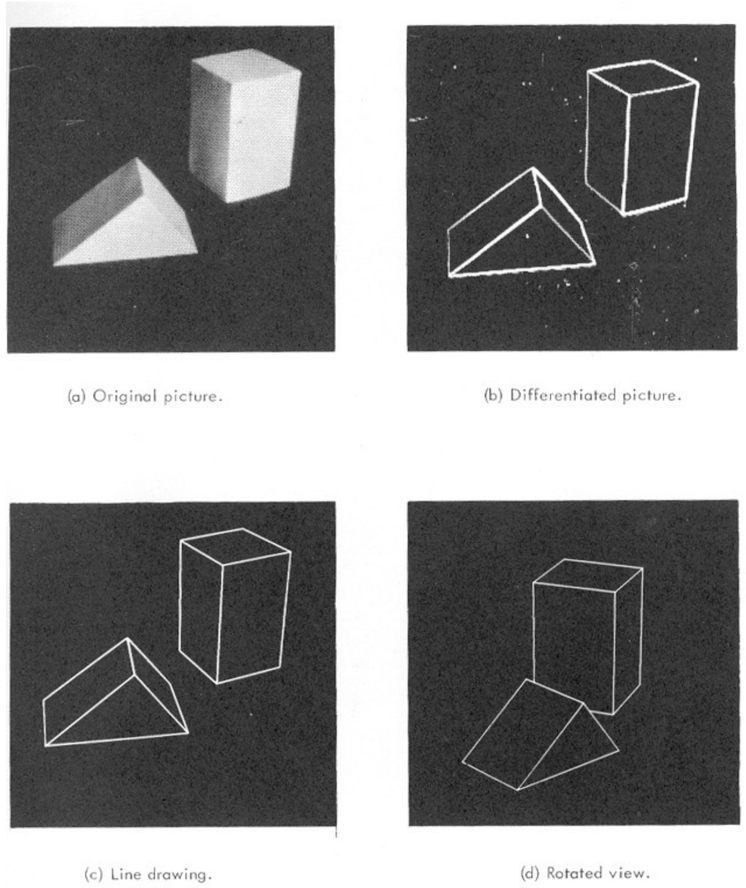
Rosenblatt, Frank. "The perceptron: a probabilistic model for information storage and organization in the brain." *Psychological review* 65.6 (1958): 386.

Slide credit: Christian Rupprecht, Oxford

1963: Larry Roberts – Blocks World

Roberts, Lawrence G. *Machine perception of three-dimensional solids*. Diss. Massachusetts Institute of Technology, 1963.
Slide credit: Christian Rupprecht, Oxford

1963: Larry Roberts – Blocks World



(a) Original picture.

(b) Differentiated picture.

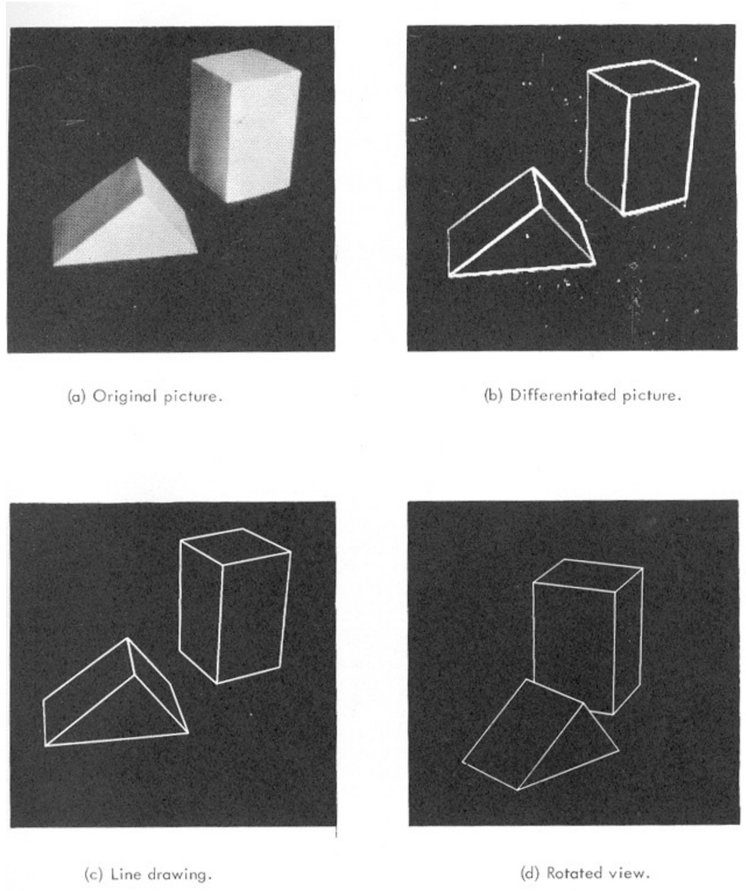
(c) Line drawing.

(d) Rotated view.

Roberts, Lawrence G. *Machine perception of three-dimensional solids*. Diss. Massachusetts Institute of Technology, 1963.
Slide credit: Christian Rupprecht, Oxford

1963: Larry Roberts – Blocks World

- Scene Understanding for Robotics



(a) Original picture.

(b) Differentiated picture.

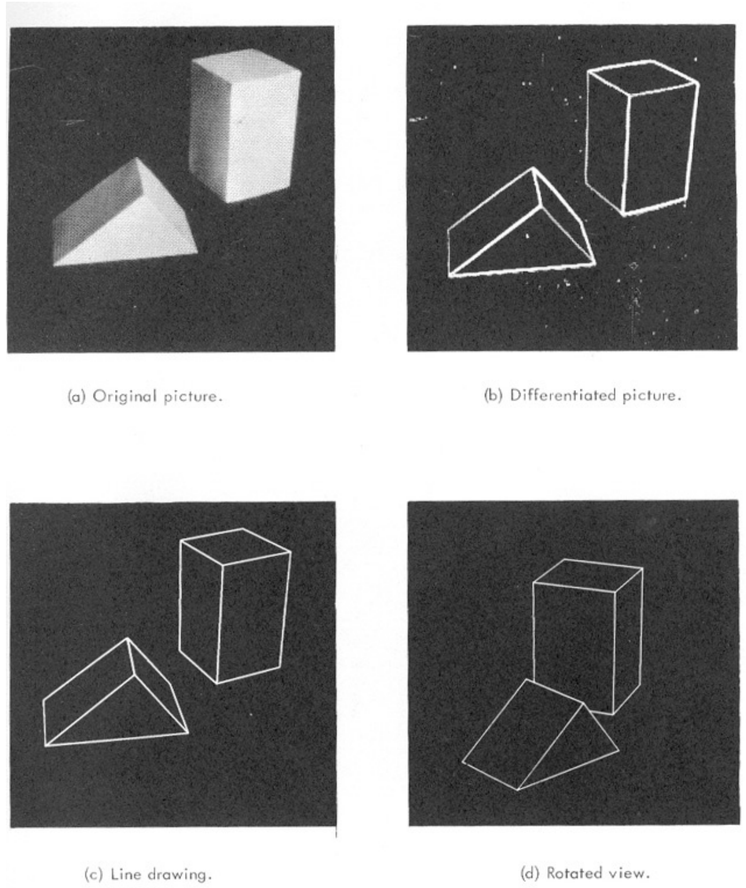
(c) Line drawing.

(d) Rotated view.

Roberts, Lawrence G. *Machine perception of three-dimensional solids*. Diss. Massachusetts Institute of Technology, 1963.
Slide credit: Christian Rupprecht, Oxford

1963: Larry Roberts – Blocks World

- Scene Understanding for Robotics
- Extracts edges



(a) Original picture.

(b) Differentiated picture.

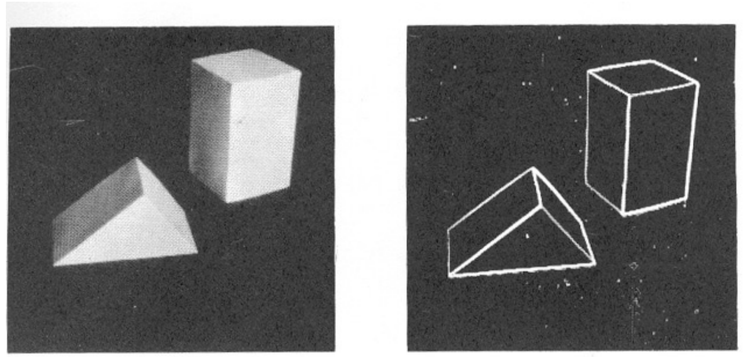
(c) Line drawing.

(d) Rotated view.

Roberts, Lawrence G. *Machine perception of three-dimensional solids*. Diss. Massachusetts Institute of Technology, 1963.
Slide credit: Christian Rupprecht, Oxford

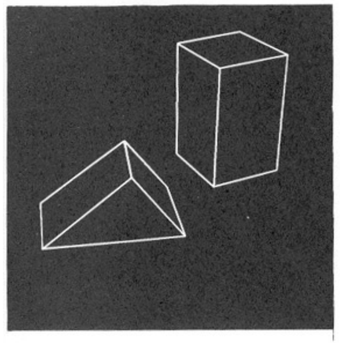
1963: Larry Roberts – Blocks World

- Scene Understanding for Robotics
- Extracts edges
- Infers 3D structure from topological structure of edges

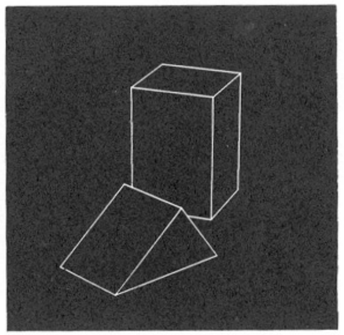


(a) Original picture.

(b) Differentiated picture.



(c) Line drawing.



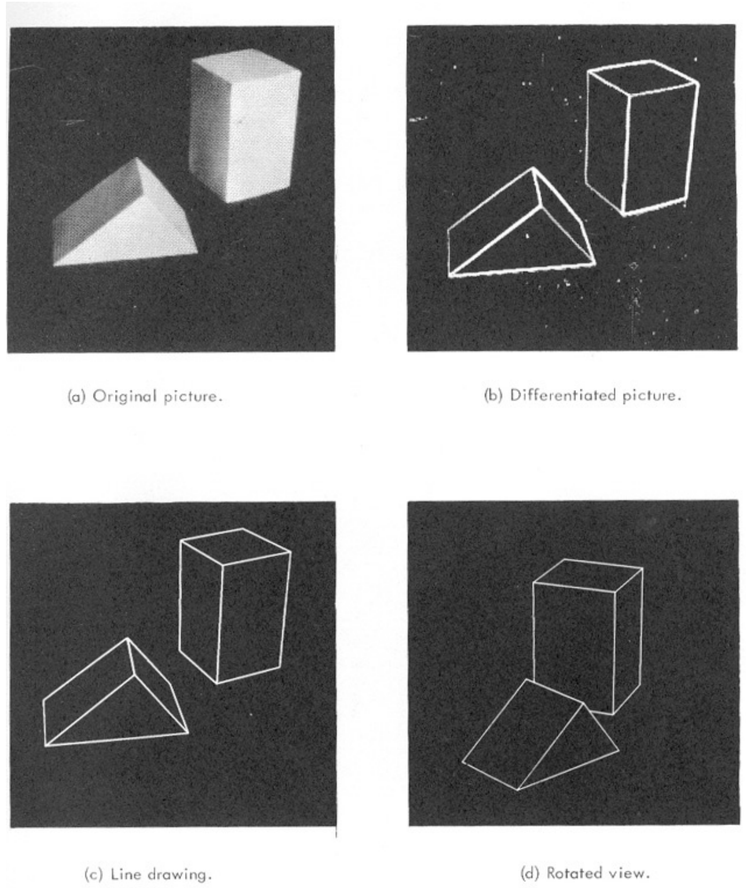
(d) Rotated view.

Roberts, Lawrence G. *Machine perception of three-dimensional solids*. Diss. Massachusetts Institute of Technology, 1963.
Slide credit: Christian Rupprecht, Oxford

1963: Larry Roberts – Blocks World

- Scene Understanding for Robotics
- Extracts edges
- Infers 3D structure from topological structure of edges

- “It is assumed that a photograph is a projection of… **known three-dimensional models**… These assumptions enable a computer to obtain a reasonable, three-dimensional description from the edge information in a photograph by means of a topological, mathematical process.”



(a) Original picture. (b) Differentiated picture.

(c) Line drawing. (d) Rotated view.

Roberts, Lawrence G. *Machine perception of three-dimensional solids*. Diss. Massachusetts Institute of Technology, 1963.
Slide credit: Christian Rupprecht, Oxford

1966: MIT Summer Vision Project

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group — July 7, 1966
Vision Memo. No. 100.

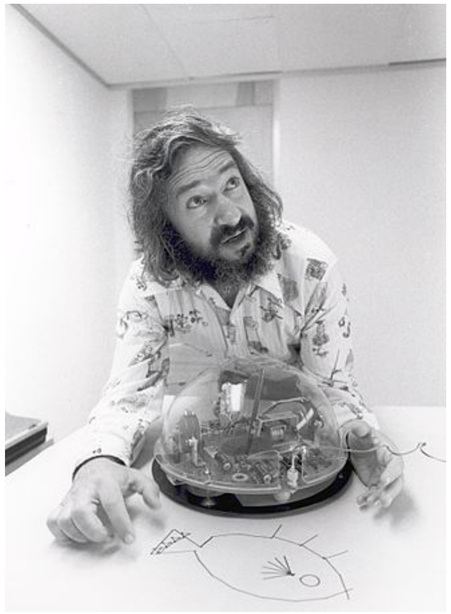
THE SUMMER VISION PROJECT

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 sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

Slide credit: Christian Rupprecht, Oxford

1966: MIT Summer Vision Project



MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group July 7, 1966
Vision Memo. No. 100.

THE SUMMER VISION PROJECT

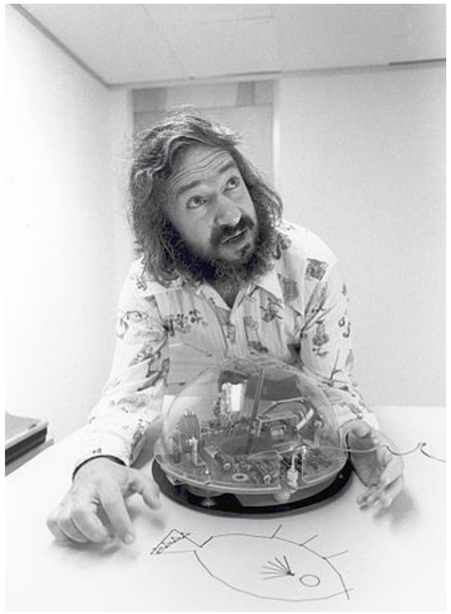
Seymour Papert

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 sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

Slide credit: Christian Rupprecht, Oxford

1966: MIT Summer Vision Project

- Solve computer vision as a summer project



MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group July 7, 1966
Vision Memo. No. 100.

THE SUMMER VISION PROJECT

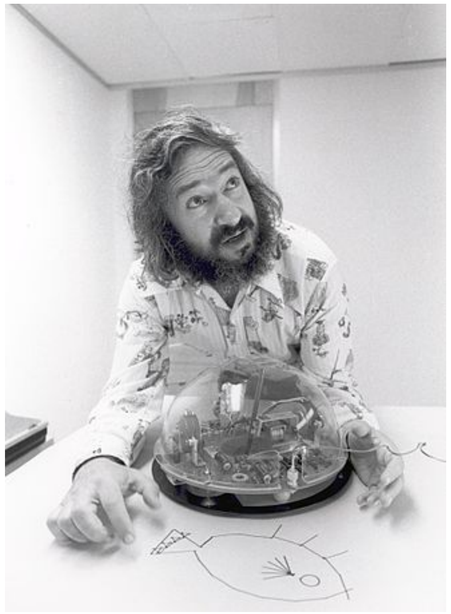
Seymour Papert

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 sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

Slide credit: Christian Rupprecht, Oxford

1966: MIT Summer Vision Project

- Solve computer vision as a summer project
- Committed to block world ideas



MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group — July 7, 1966
Vision Memo. No. 100.

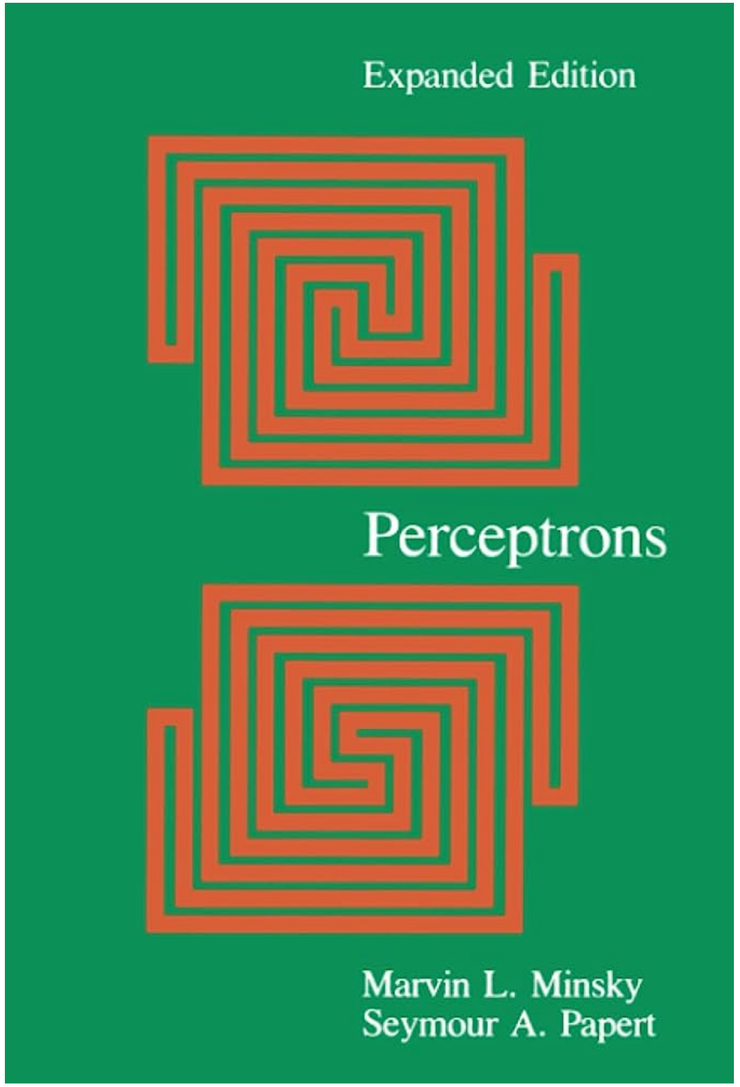
THE SUMMER VISION PROJECT

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 sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

Slide credit: Christian Rupprecht, Oxford

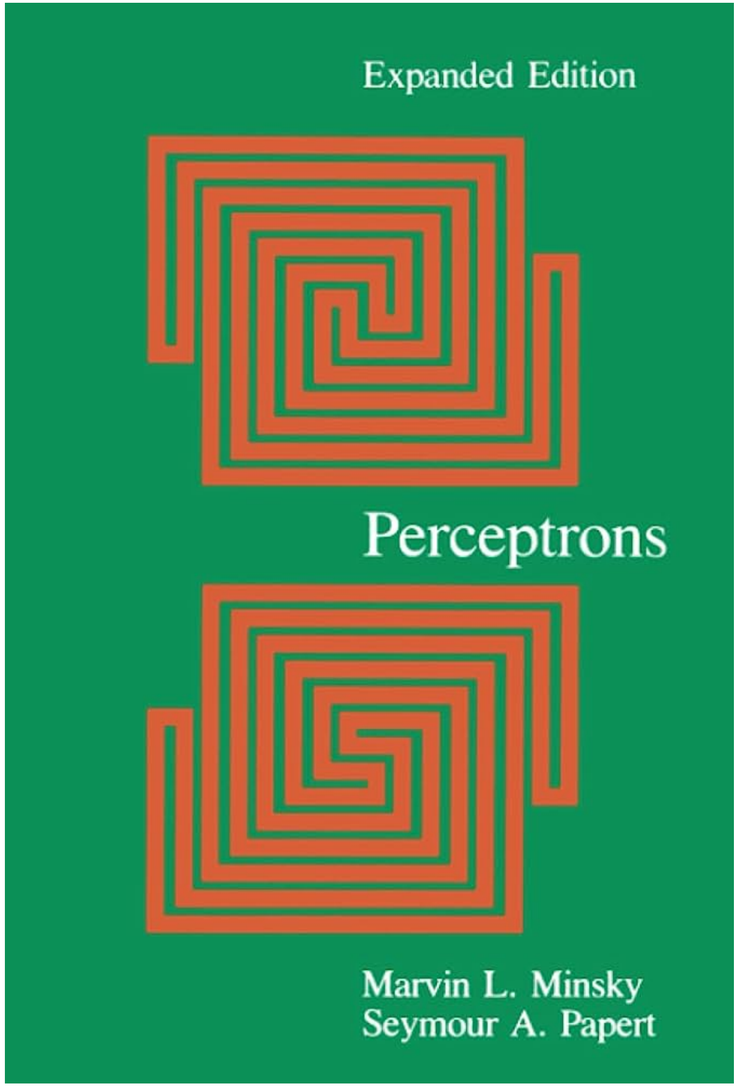
1969: Perceptrons book



Slide credit: Christian Rupprecht, Oxford

1969: Perceptrons book

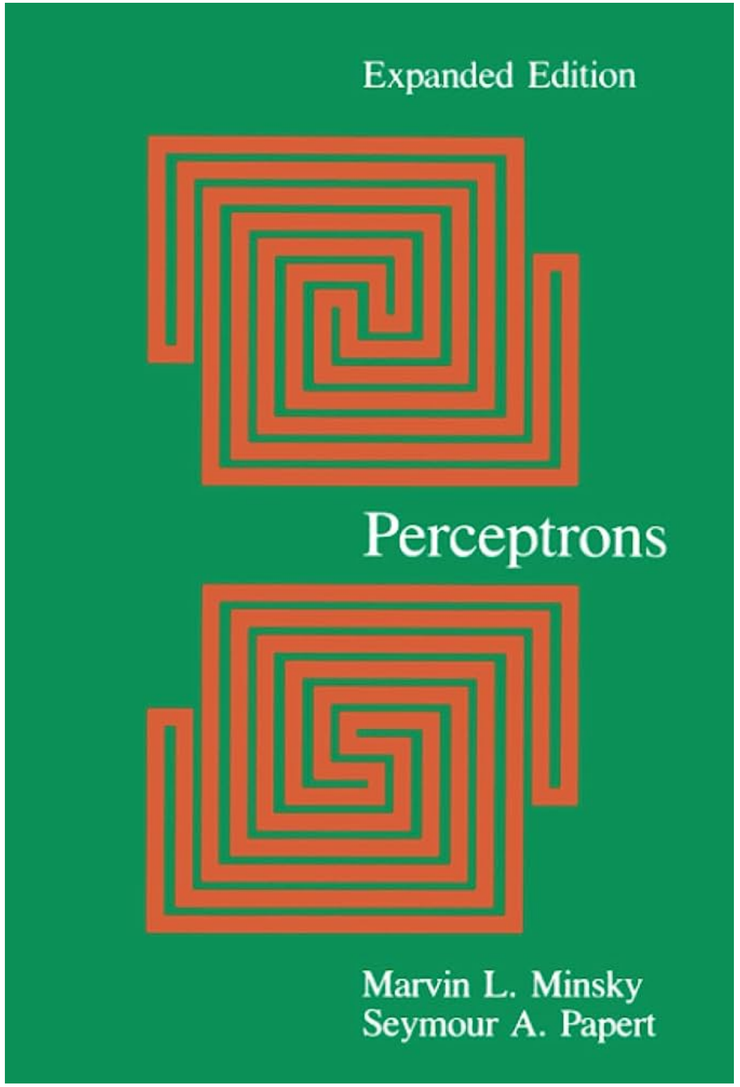
- Minsky and Papert



Slide credit: Christian Rupprecht, Oxford

1969: Perceptrons book

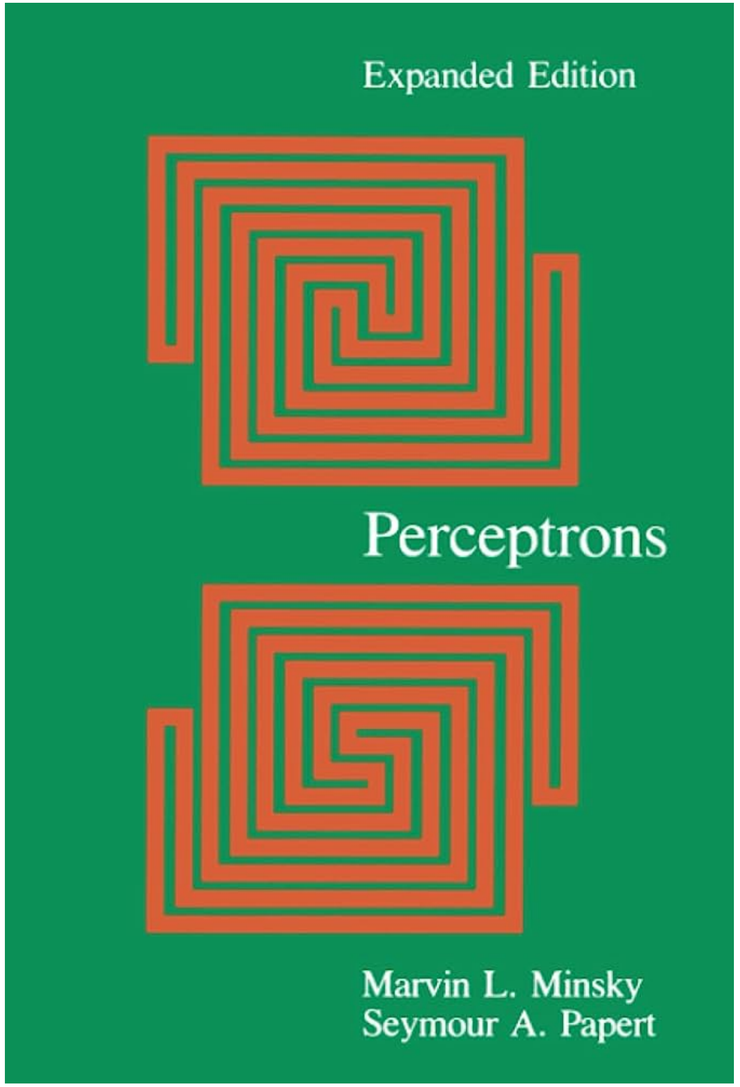
- Minsky and Papert
- Several discouraging results



Slide credit: Christian Rupprecht, Oxford

1969: Perceptrons book

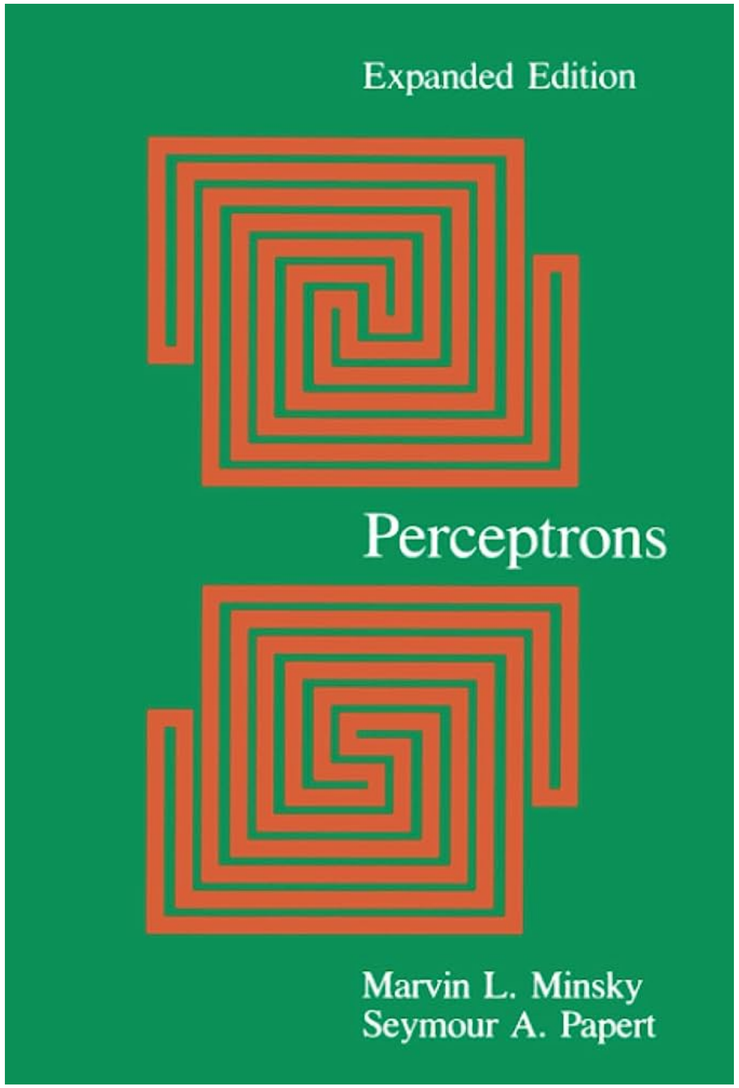
- Minsky and Papert
- Several discouraging results
- *Perceptrons cannot solve the XOR problem*



Slide credit: Christian Rupprecht, Oxford

1969: Perceptrons book

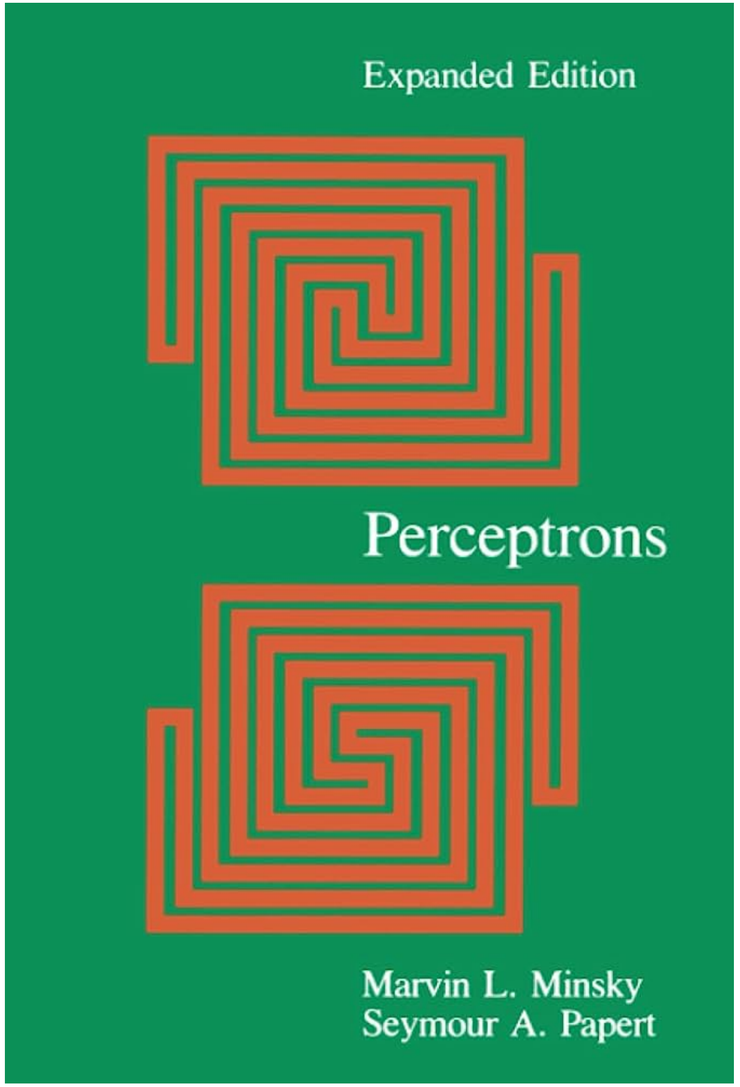
- Minsky and Papert
- Several discouraging results
- *Perceptrons cannot solve the XOR problem*
- Largely contributed to the following “AI winter”



Slide credit: Christian Rupprecht, Oxford

1969: Perceptrons book

- Minsky and Papert
- Several discouraging results
- *Perceptrons cannot solve the XOR problem*
- Largely contributed to the following “AI winter”
- 70s: mostly symbolic AI

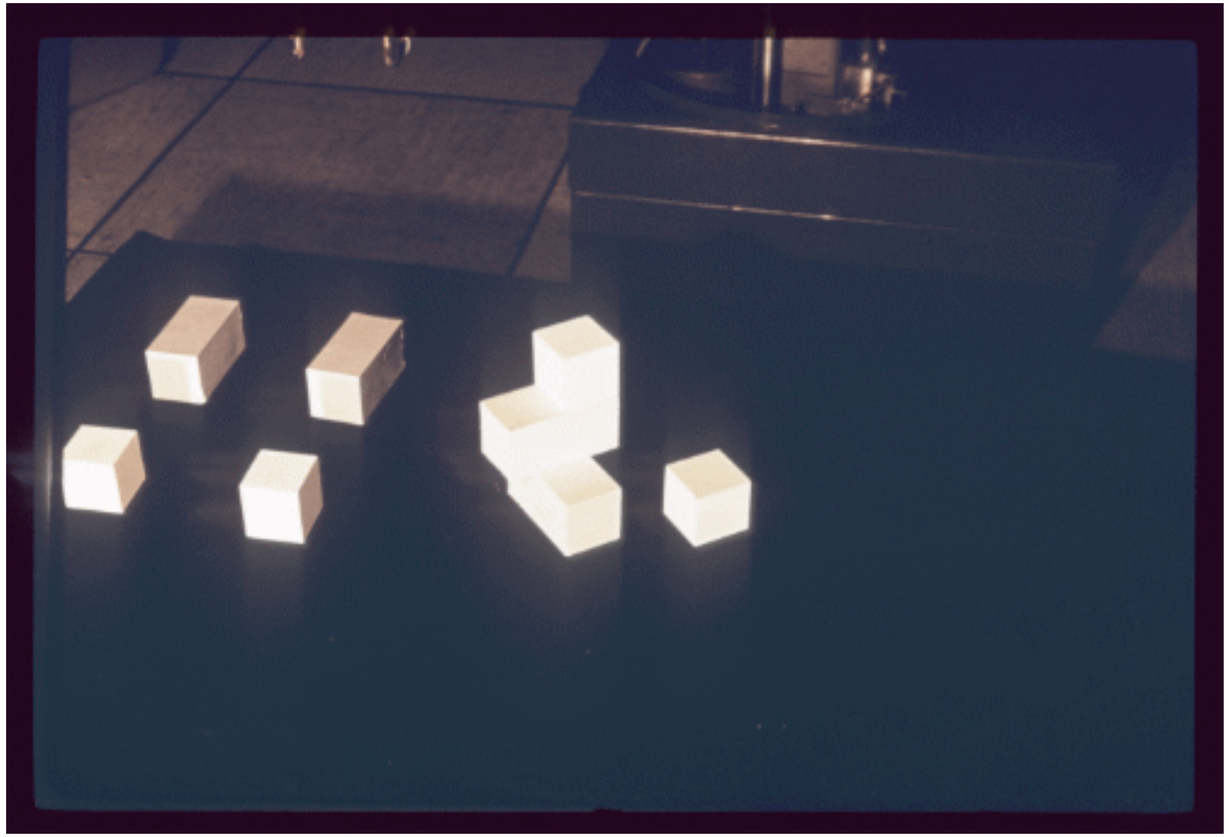


Slide credit: Christian Rupprecht, Oxford

1970: MIT Copy Demo

Patrick Winston and the MIT AI Lab Copy Demo https://people.csail.mit.edu/bkph/phw_copy_demo.shtml
Slide credit: Christian Rupprecht, Oxford

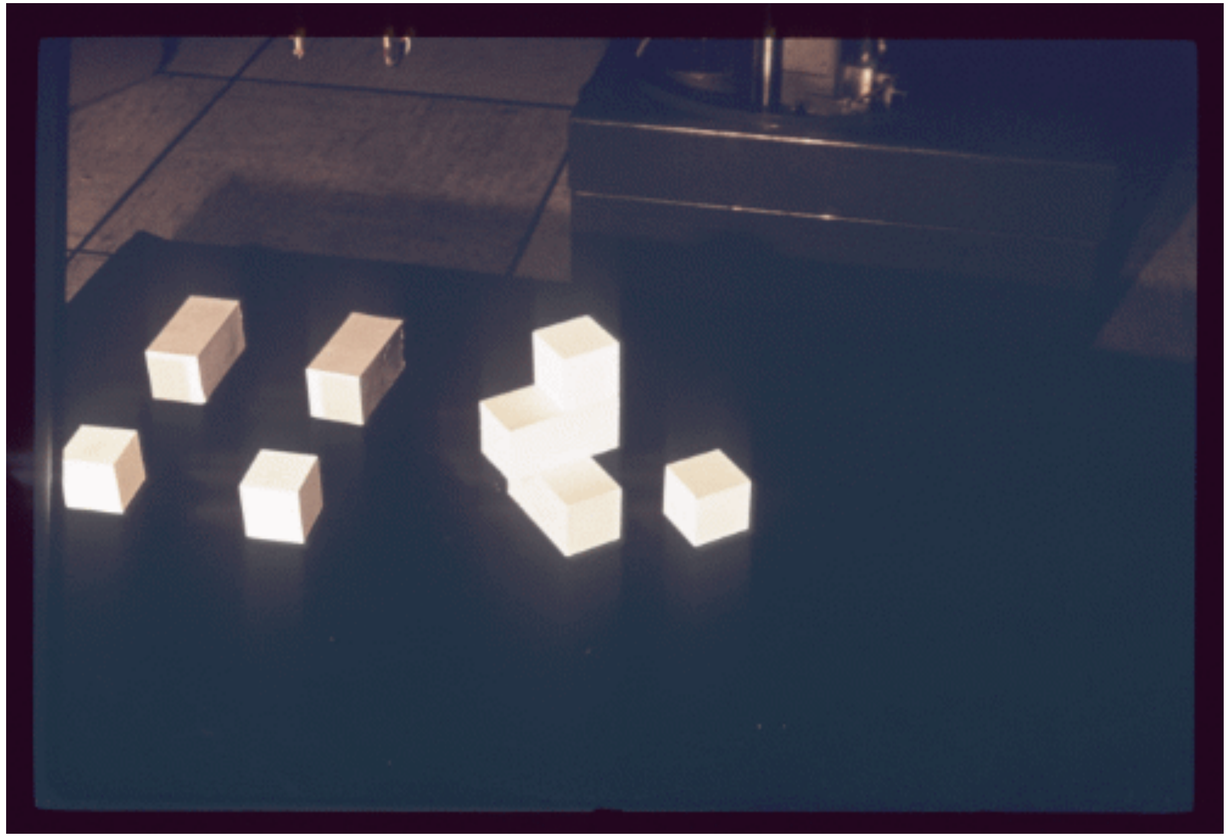
1970: MIT Copy Demo



Patrick Winston and the MIT AI Lab Copy Demo https://people.csail.mit.edu/bkph/phw_copy_demo.shtml
Slide credit: Christian Rupprecht, Oxford

1970: MIT Copy Demo

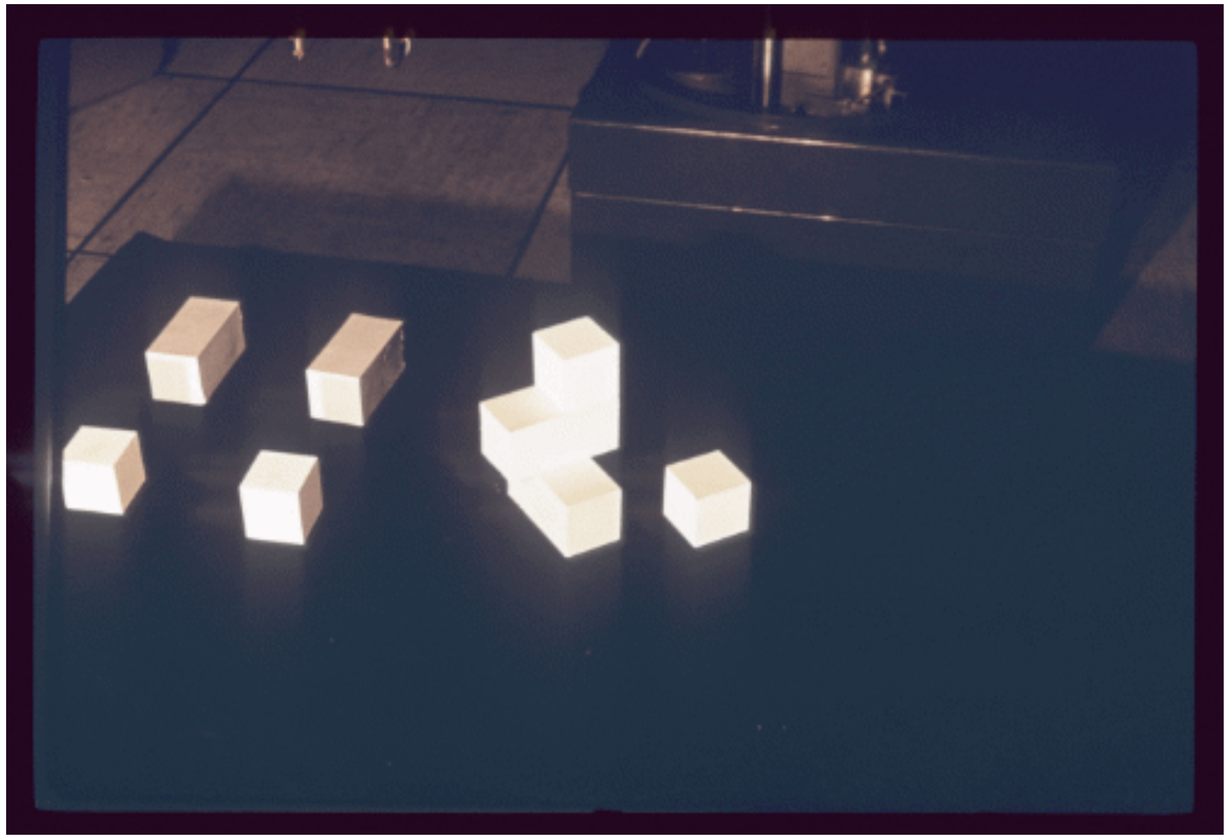
- Vision + Robotics



Patrick Winston and the MIT AI Lab Copy Demo https://people.csail.mit.edu/bkph/phw_copy_demo.shtml
Slide credit: Christian Rupprecht, Oxford

1970: MIT Copy Demo

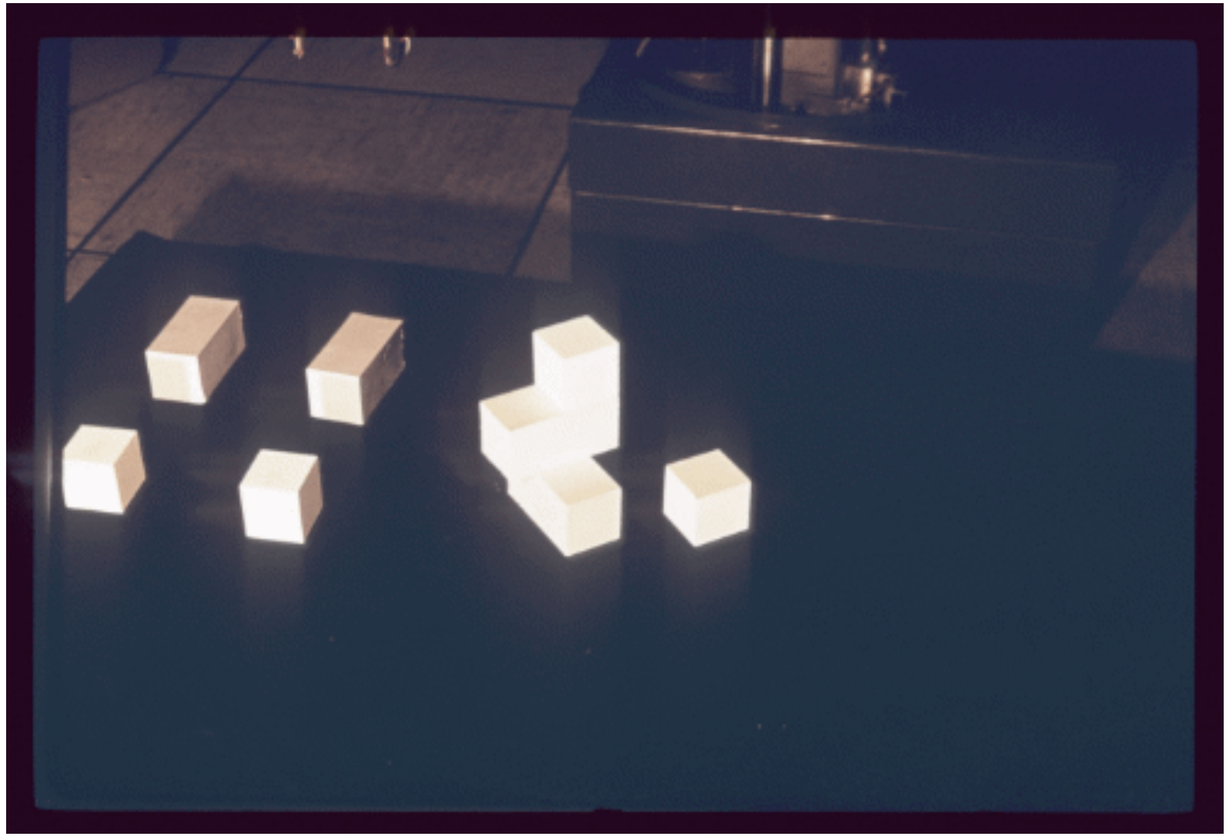
- Vision + Robotics
- Recover the structure of a scene



Patrick Winston and the MIT AI Lab Copy Demo https://people.csail.mit.edu/bkph/phw_copy_demo.shtml
Slide credit: Christian Rupprecht, Oxford

1970: MIT Copy Demo

- Vision + Robotics
- Recover the structure of a scene
- Plan robot movement to copy the block arrangement
- Only works in ideal conditions
- Causes attention to robustness for low level vision tasks



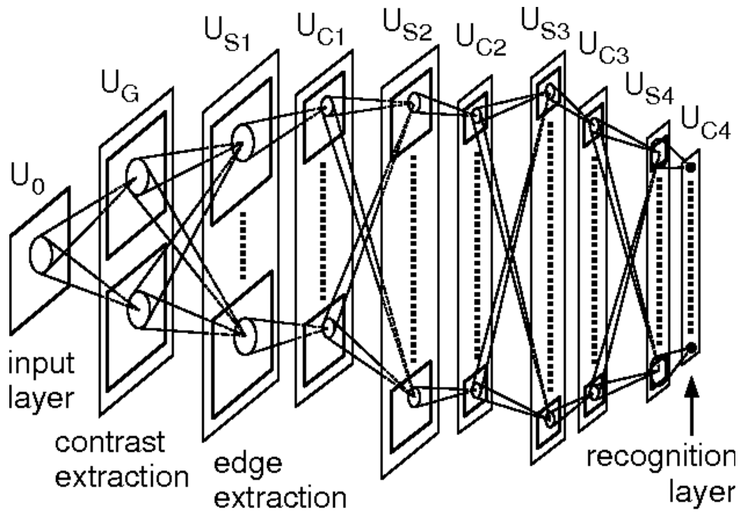
Patrick Winston and the MIT AI Lab Copy Demo https://people.csail.mit.edu/bkph/phw_copy_demo.shtml

Slide credit: Christian Rupprecht, Oxford

1980s: Advances in ML

Slide credit: Christian Rupprecht, Oxford

1980s: Advances in ML

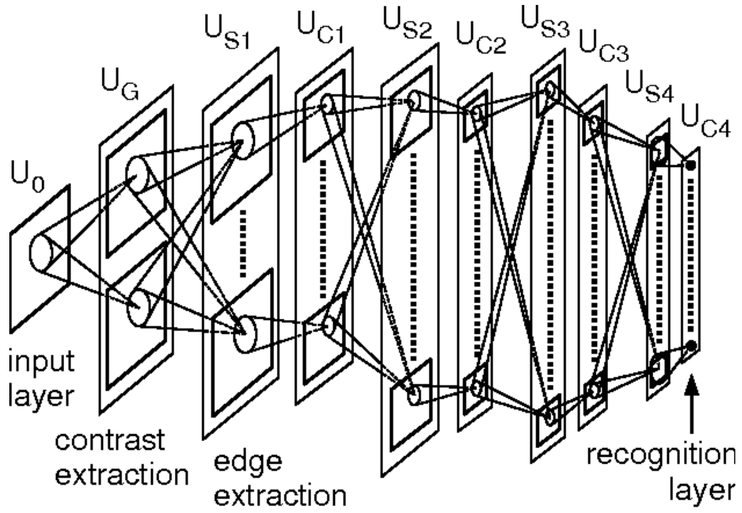


Fukushima (1980)

Slide credit: Christian Rupprecht, Oxford

1980s: Advances in ML

- Neocognitron: Fukushima (1980)

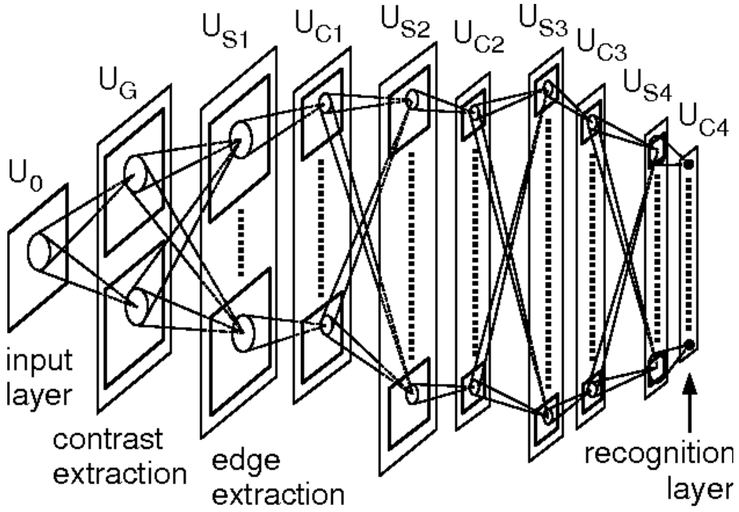


Fukushima (1980)

Slide credit: Christian Rupprecht, Oxford

1980s: Advances in ML

- Neocognitron: Fukushima (1980)
- Back-propagation: Rumelhart, Hinton & Williams (1986)
  - Origins in control theory and optimization: Kelley (1960), Dreyfus (1962), Bryson & Ho (1969), Linnainmaa (1970)
  - Application to neural networks: Werbos (1974)

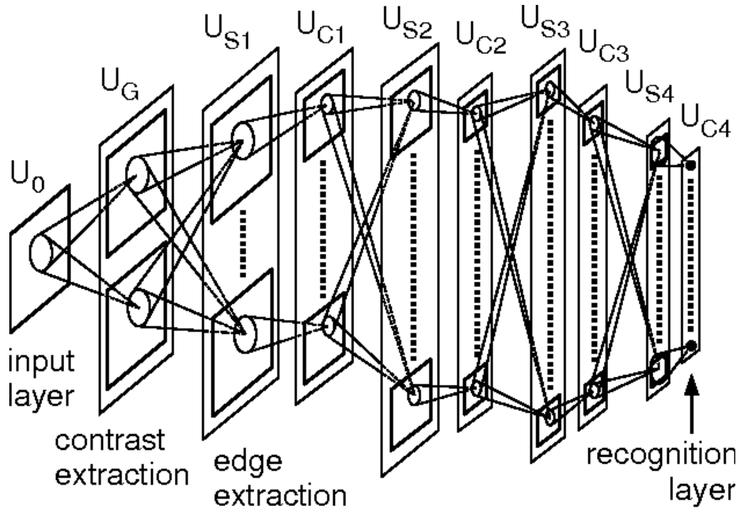


Fukushima (1980)

Slide credit: Christian Rupprecht, Oxford

1980s: Advances in ML

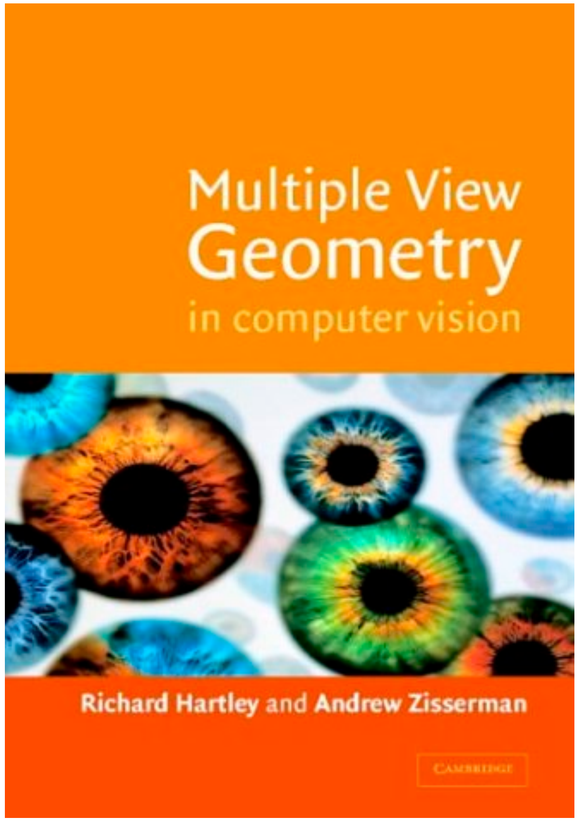
- Neocognitron: Fukushima (1980)
- Back-propagation: Rumelhart, Hinton & Williams (1986)
  - Origins in control theory and optimization: Kelley (1960), Dreyfus (1962), Bryson & Ho (1969), Linnainmaa (1970)
  - Application to neural networks: Werbos (1974)
- Parallel Distributed Processing: Rumelhart et al. (1987)
- Neural networks for digit recognition: LeCun et al. (1989)



Fukushima (1980)

Slide credit: Christian Rupprecht, Oxford

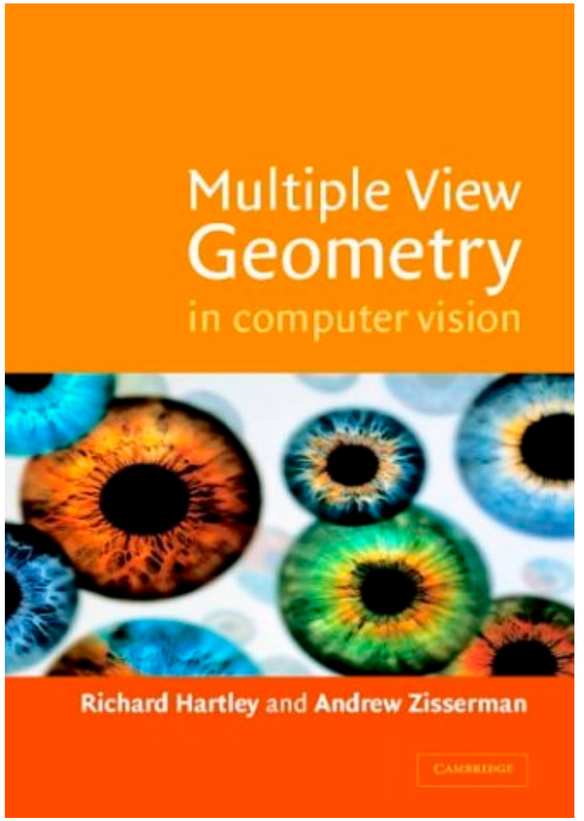
1990s Theme: Geometry



Slide credit: Christian Rupprecht, Oxford

1990s Theme: Geometry

- Fundamental matrix: Faugeras (1992)
- Normalized 8-point algorithm: Hartley (1997)
- RANSAC for robust fundamental matrix estimation: Torr & Murray (1997)
- Bundle adjustment: Triggs et al. (1999)
- Hartley & Zisserman book (2000)
- Projective structure from motion: Faugeras and Luong (2001)



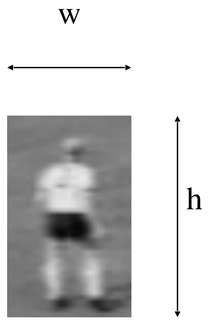
Slide credit: Christian Rupprecht, Oxford

Tracking, Optical Flow and Correspondence

Template Tracking

Slide credit: Christian Rupprecht, Oxford

Template Tracking



Template T

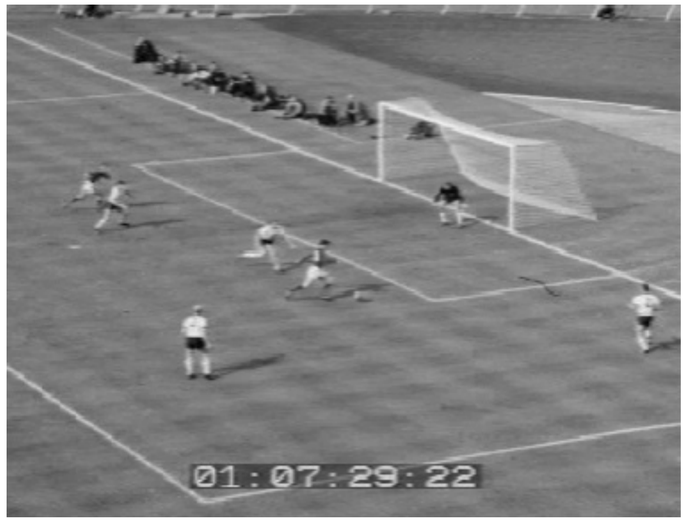
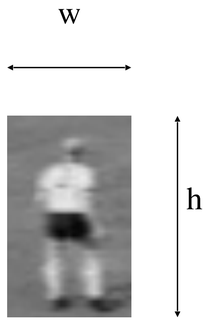


Image I

Slide credit: Christian Rupprecht, Oxford

Template Tracking



Template T

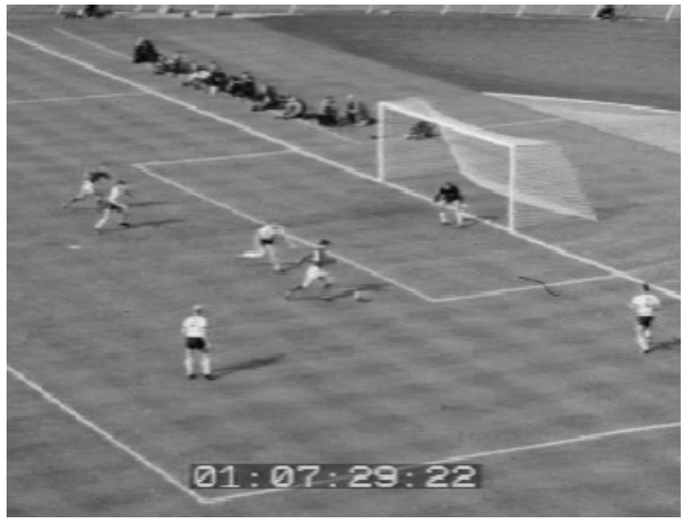
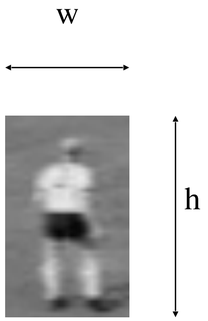


Image I

Sum of squared differences between image I and template T at every location (u, v)

Slide credit: Christian Rupprecht, Oxford

Template Tracking



Template T

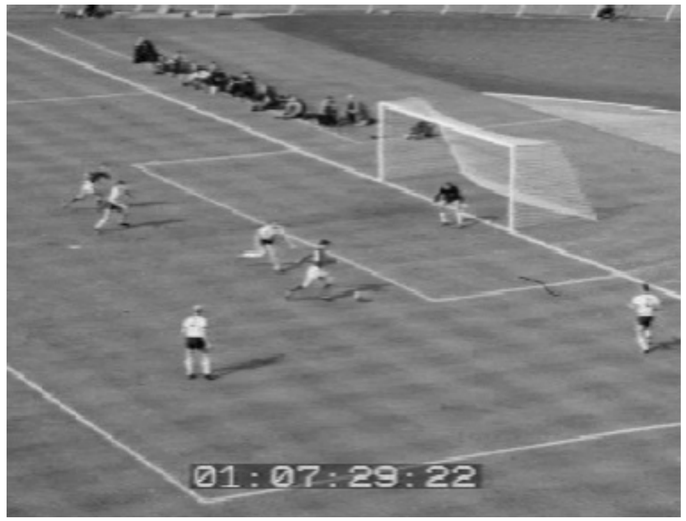
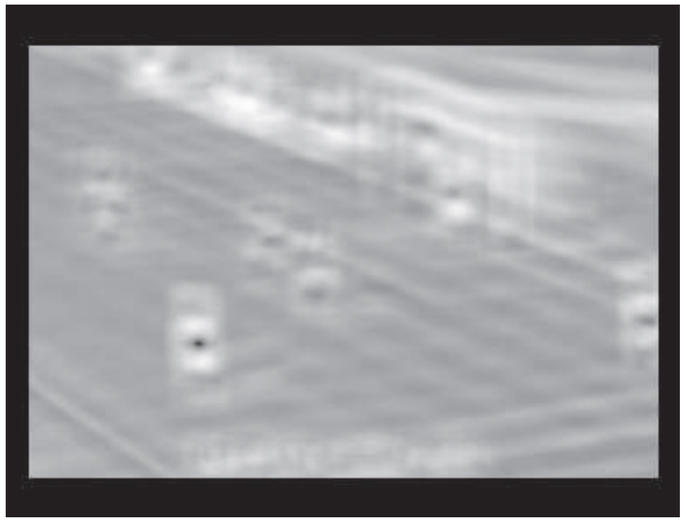


Image I

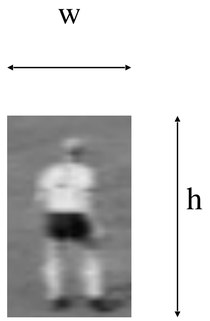
Sum of squared differences between image I and template T at every location (u, v)



Energy E

Slide credit: Christian Rupprecht, Oxford

Template Tracking



Template T

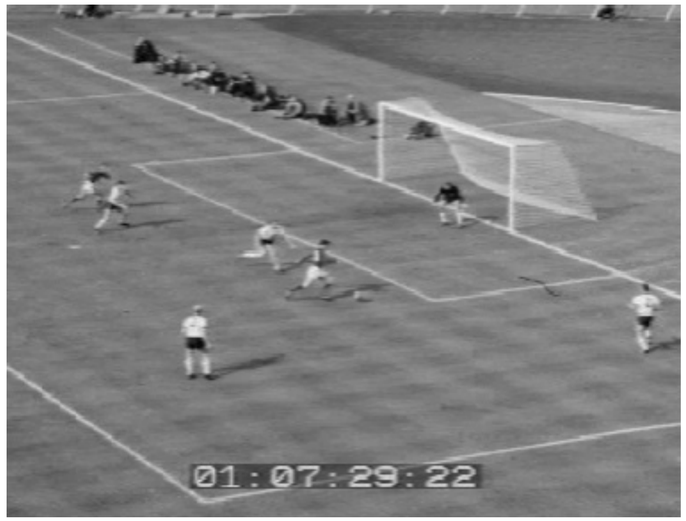
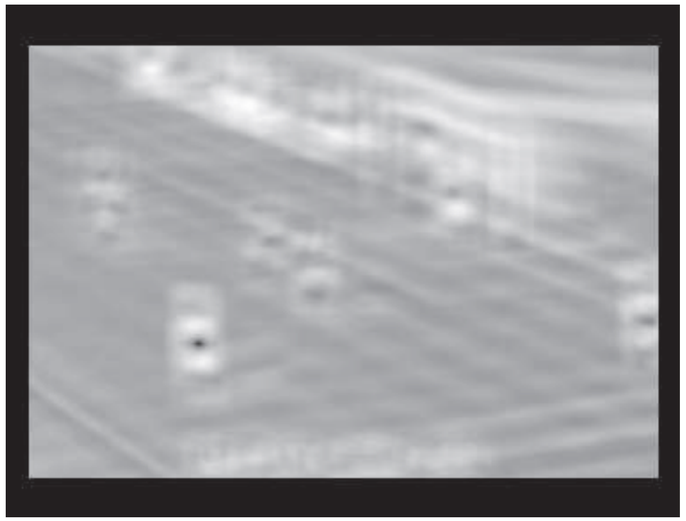


Image I

Sum of squared differences between image I and template T at every location (u, v)

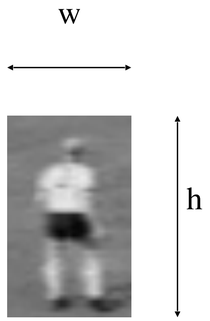


Energy E

$$E(u,v) = \sum_{(x,y)\in\left[-\frac{w}{2},\frac{w}{2}\right]\times\left[-\frac{h}{2},\frac{h}{2}\right]} \Big(I\big(u+x,\ v+y\big) - T\big(x,y\big)\Big)^2$$

Slide credit: Christian Rupprecht, Oxford

Template Tracking



Template T

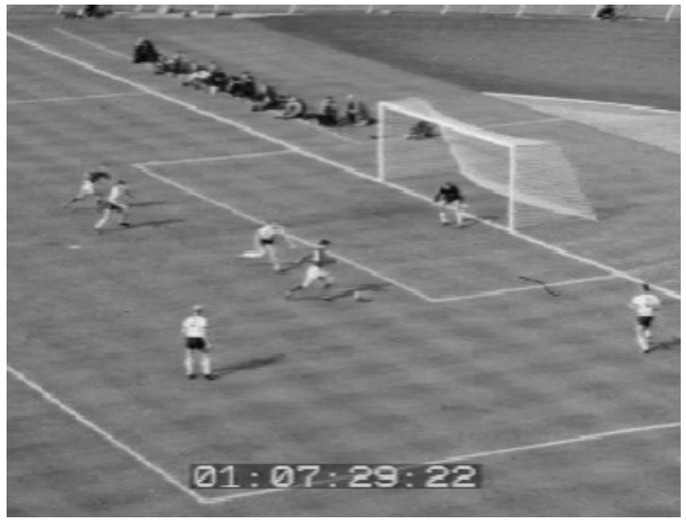
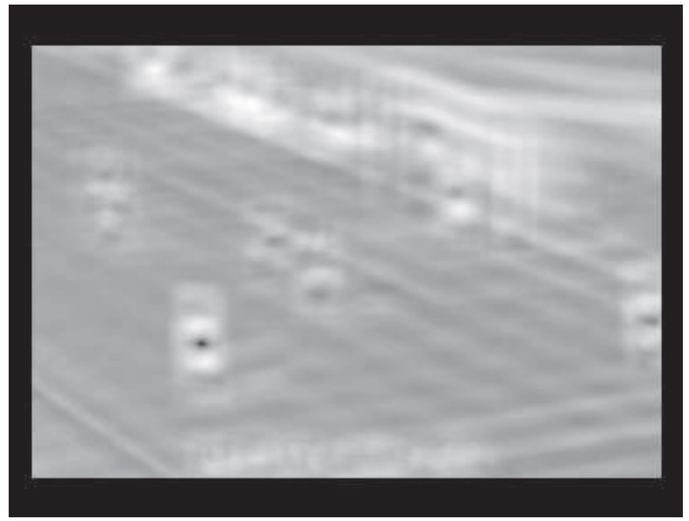


Image I

Sum of squared differences between image I and template T at every location (u, v)



Energy E

$$E(u,v) = \sum_{(x,y)\in\left[-\frac{w}{2},\frac{w}{2}\right]\times\left[-\frac{h}{2},\frac{h}{2}\right]} \Big(I\big(u+x,\ v+y\big) - T\big(x,y\big)\Big)^2$$

- Tracking by Detection (each frame processed individually)
- Very slow: (#pixels_image * #pixels_template) comparisons

Slide credit: Christian Rupprecht, Oxford

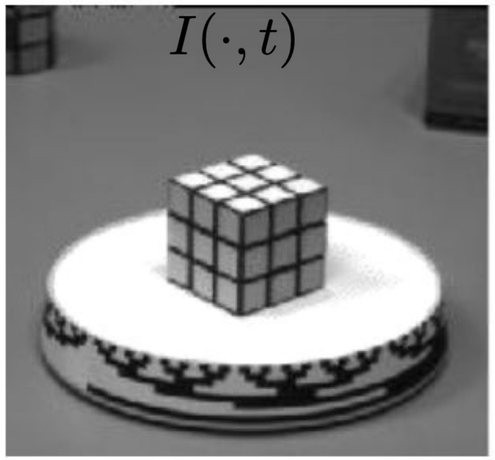
1981: Optical Flow – The Beginnings

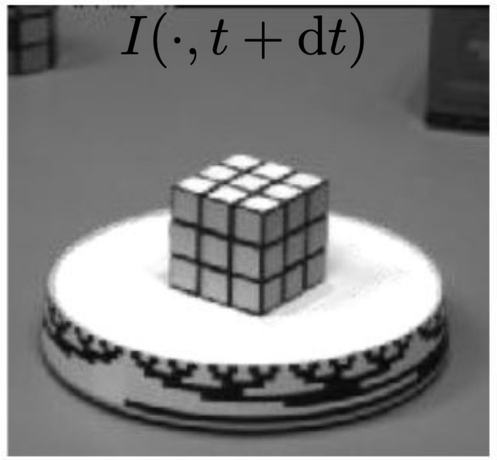
- Pattern of apparent motion: densely tracking pixels between frames
- Horn-Schunck algorithm
- 2D correspondence search: more difficult than stereo

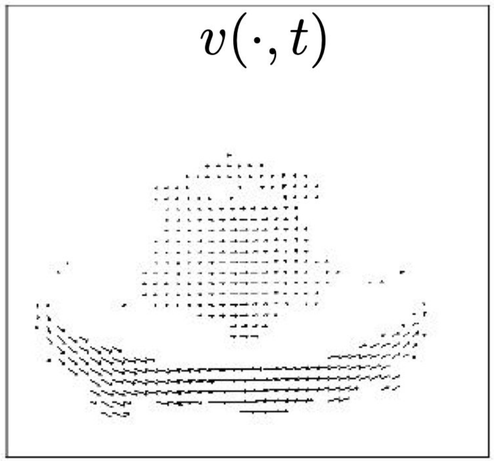
Determining Optical Flow

Berthold K.P. Horn and Brian G. Schunck
Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

ABSTRACT
Optical flow cannot be computed locally, since only one independent measurement is available from the image sequence at a point, while the flow velocity has two components. A second constraint is needed. A method for finding the optical flow pattern is presented which assumes that the apparent velocity of the brightness pattern varies smoothly almost everywhere in the image. An iterative implementation is shown which successfully computes the optical flow for a number of synthetic image sequences. The algorithm is robust in that it can handle image sequences that are quantized rather coarsely in space and time. It is also insensitive to quantization of brightness levels and additive noise. Examples are included where the assumption of smoothness is violated at singular points or along lines in the image.







Slide credit: Christian Rupprecht, Oxford

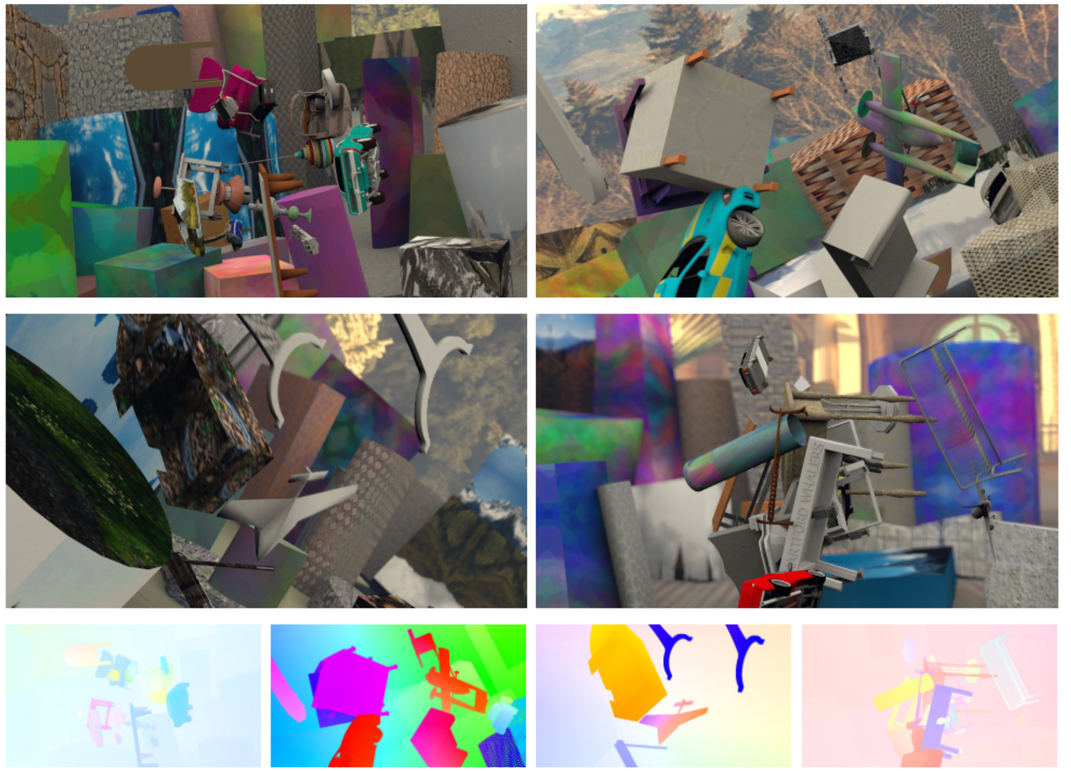
Optical Flow – The Beginnings

Raw estimate

Smoothed estimate

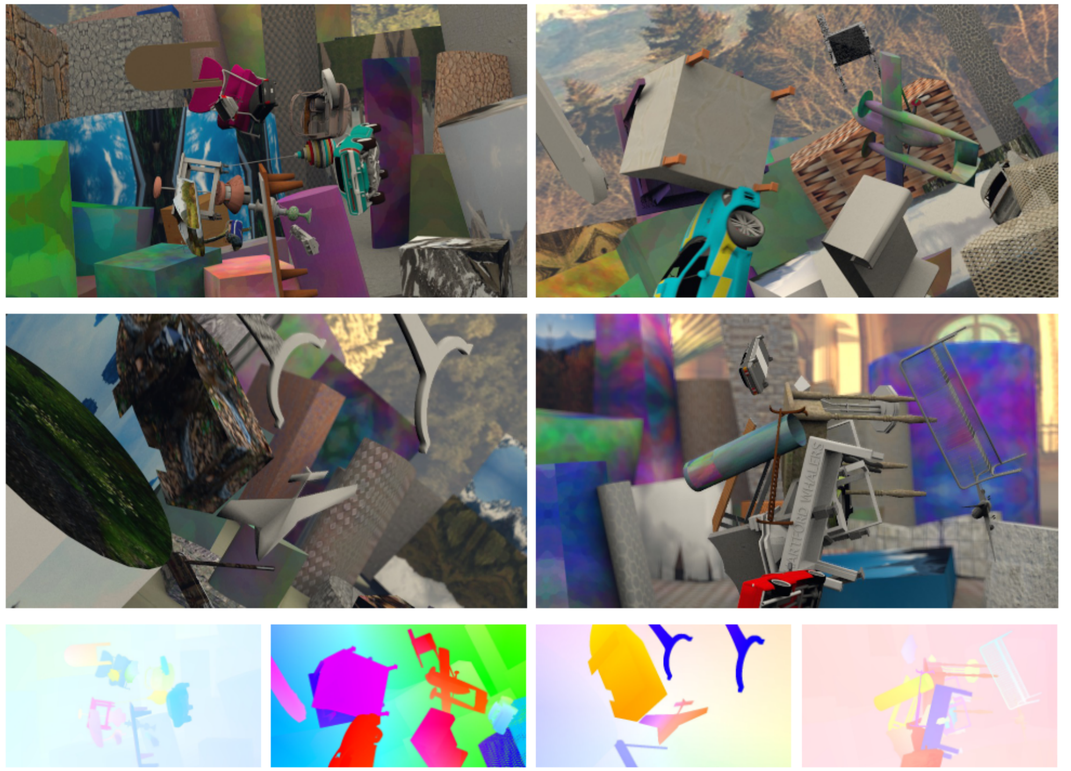
Slide credit: Christian Rupprecht, Oxford

Supervised Learning age: Optical Flow – Flying Things



Slide credit: Christian Rupprecht, Oxford

Supervised Learning age: Optical Flow – Flying Things

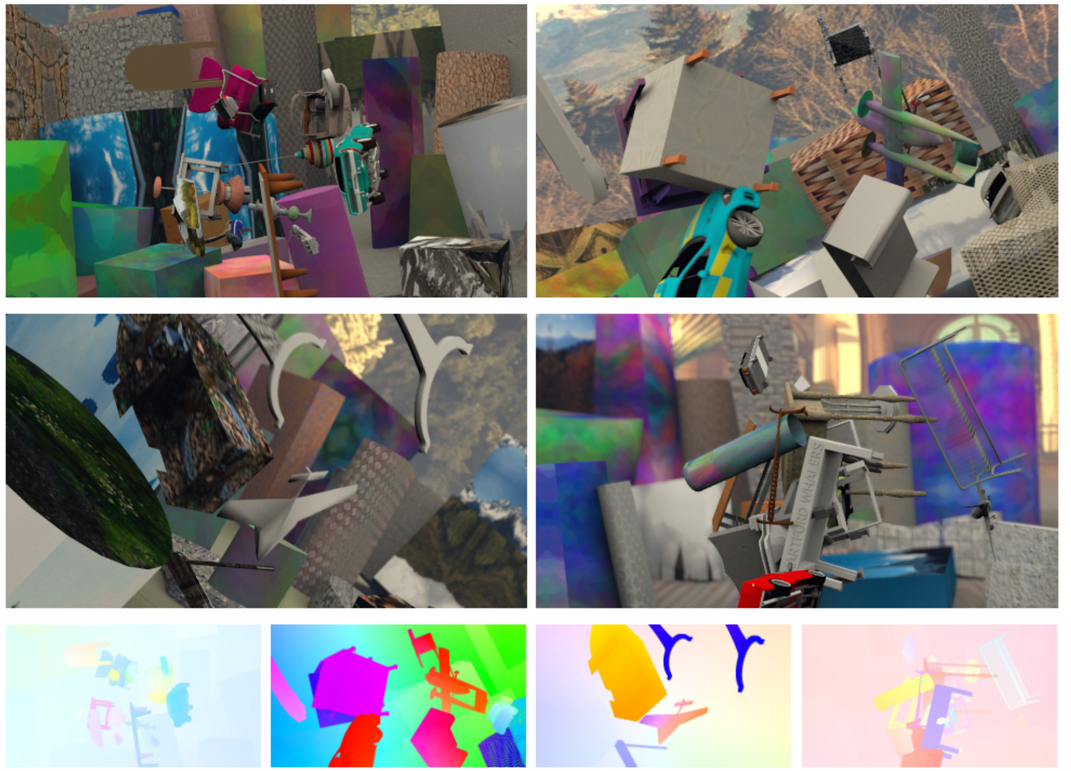


Mayer, Nikolaus, Eddy Ilg, Philip Hausser, Philipp Fischer, Daniel Cremers, Alexey Dosovitskiy, and Thomas Brox. "A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation.", CVPR 2016

Slide credit: Christian Rupprecht, Oxford

Supervised Learning age: Optical Flow – Flying Things

- Easier to create – automatic pipeline

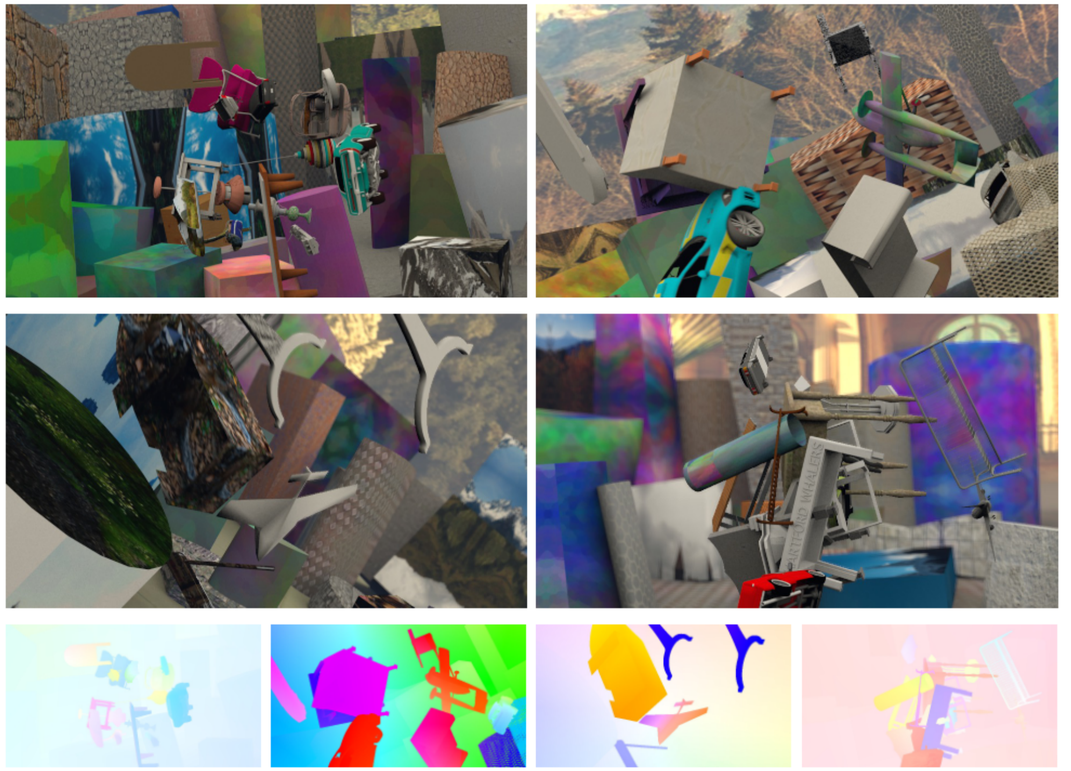


Mayer, Nikolaus, Eddy Ilg, Philip Hausser, Philipp Fischer, Daniel Cremers, Alexey Dosovitskiy, and Thomas Brox. "A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation.", CVPR 2016

Slide credit: Christian Rupprecht, Oxford

Supervised Learning age: Optical Flow – Flying Things

- Easier to create – automatic pipeline
- Generalises well to real data



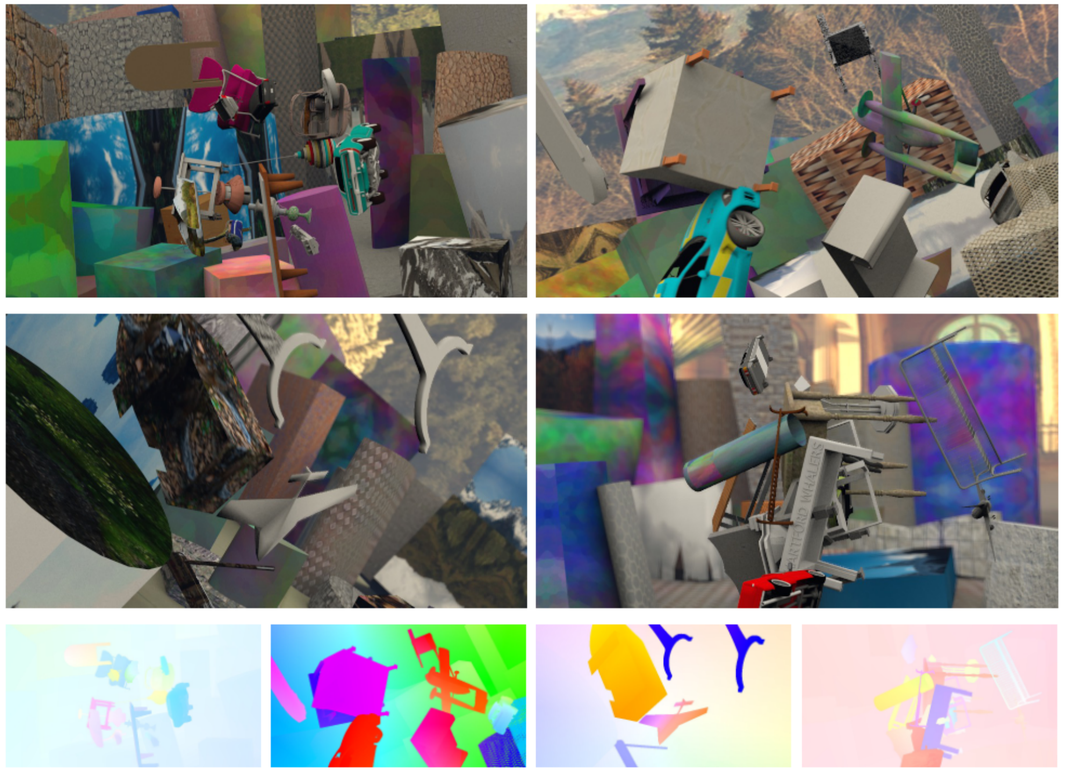
Mayer, Nikolaus, Eddy Ilg, Philip Hausser, Philipp Fischer, Daniel Cremers, Alexey Dosovitskiy, and Thomas Brox. "A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation.", CVPR 2016

Slide credit: Christian Rupprecht, Oxford

Supervised Learning age: Optical Flow – Flying Things

- Easier to create – automatic pipeline
- Generalises well to real data

-> optical flow is a low level vision problem and thus sim2real transfer works well



Mayer, Nikolaus, Eddy Ilg, Philip Hausser, Philipp Fischer, Daniel Cremers, Alexey Dosovitskiy, and Thomas Brox. "A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation.", CVPR 2016

Slide credit: Christian Rupprecht, Oxford

Optical Flow - Sintel

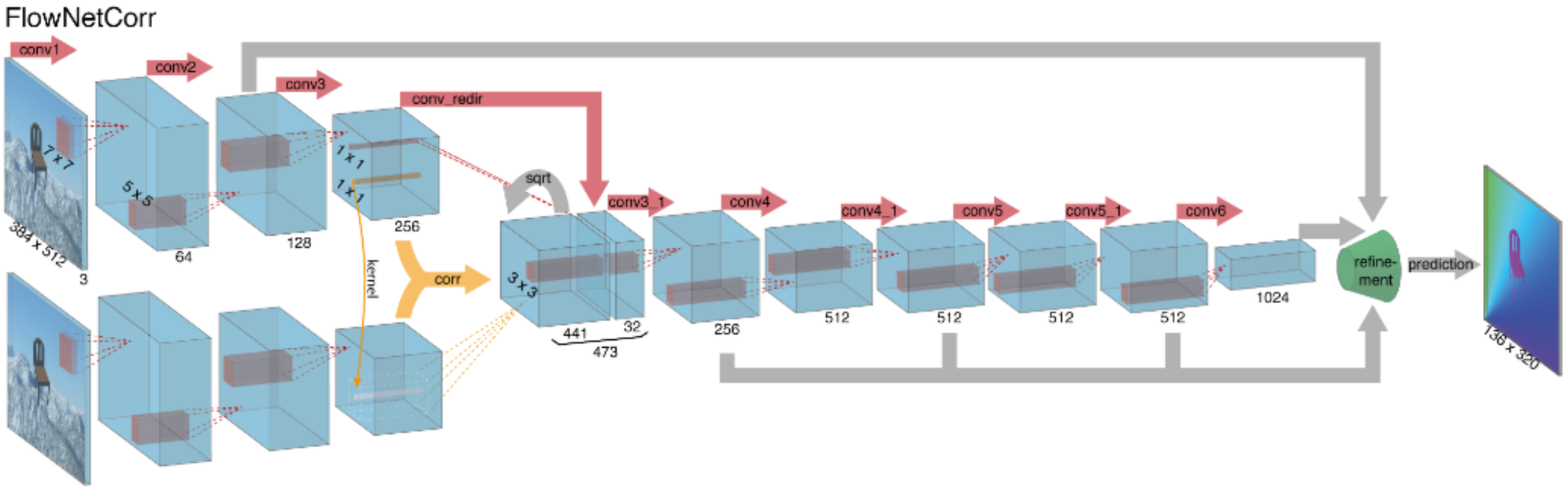
Slide credit: Christian Rupprecht, Oxford

Optical Flow - Sintel

Butler, Daniel J., Jonas Wulff, Garrett B. Stanley, and Michael J. Black. "A naturalistic open source movie for optical flow evaluation.“ ECCV 2012

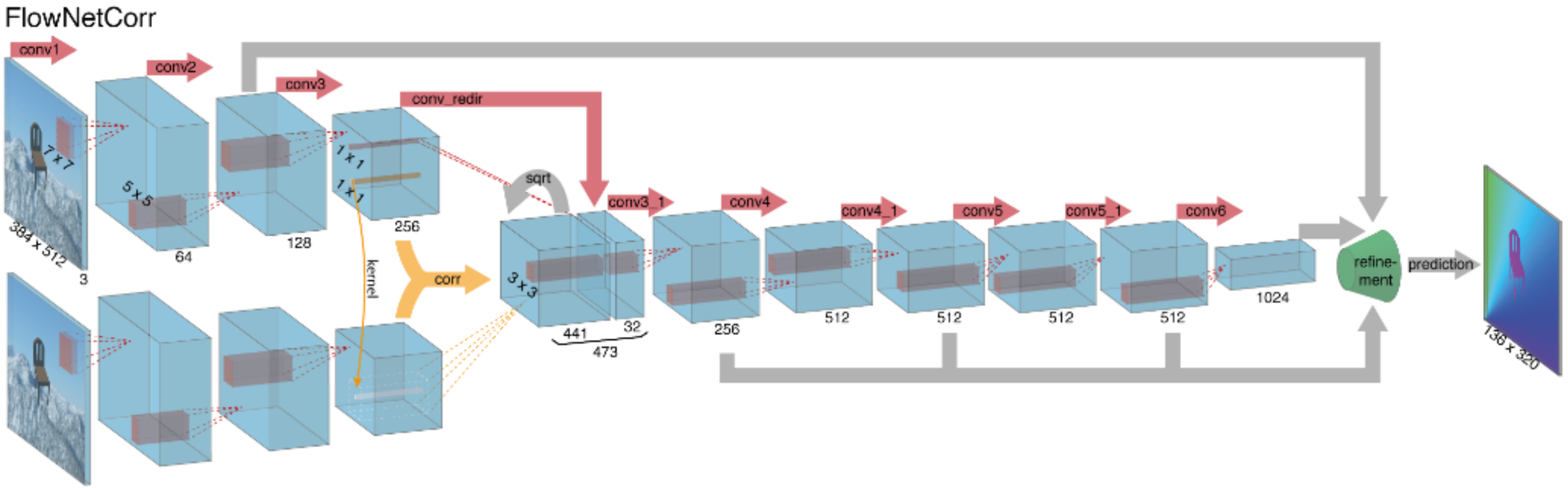
Slide credit: Christian Rupprecht, Oxford

Learning Optical fLow



Slide credit: Christian Rupprecht, Oxford

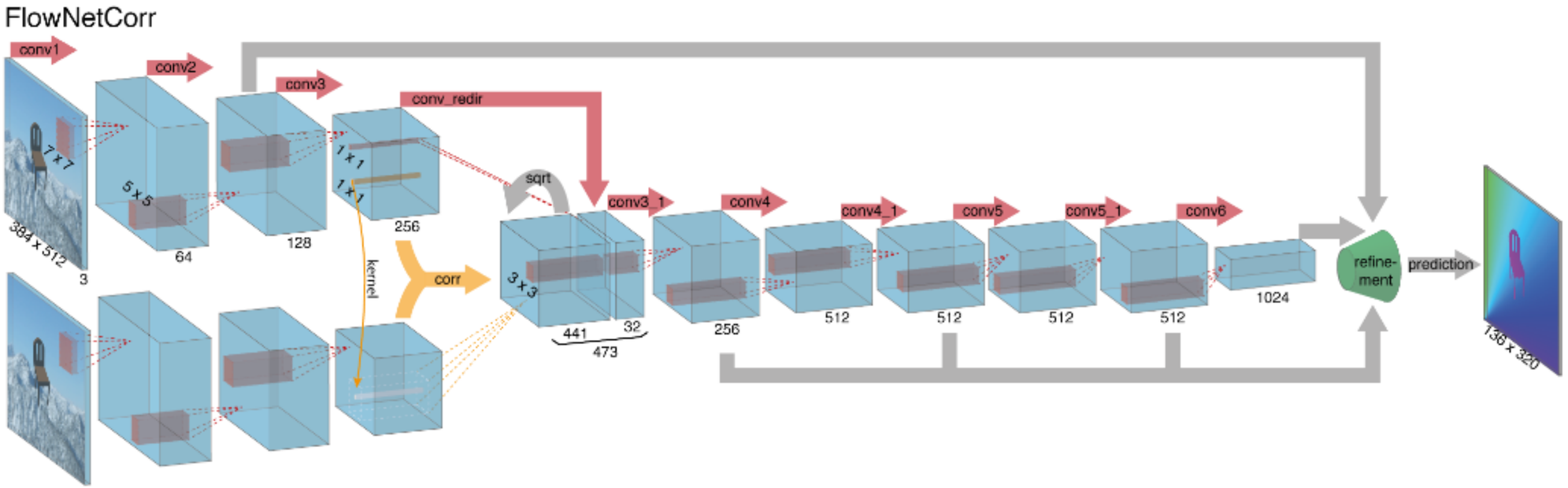
Learning Optical fLow



- Siamese architecture

Slide credit: Christian Rupprecht, Oxford

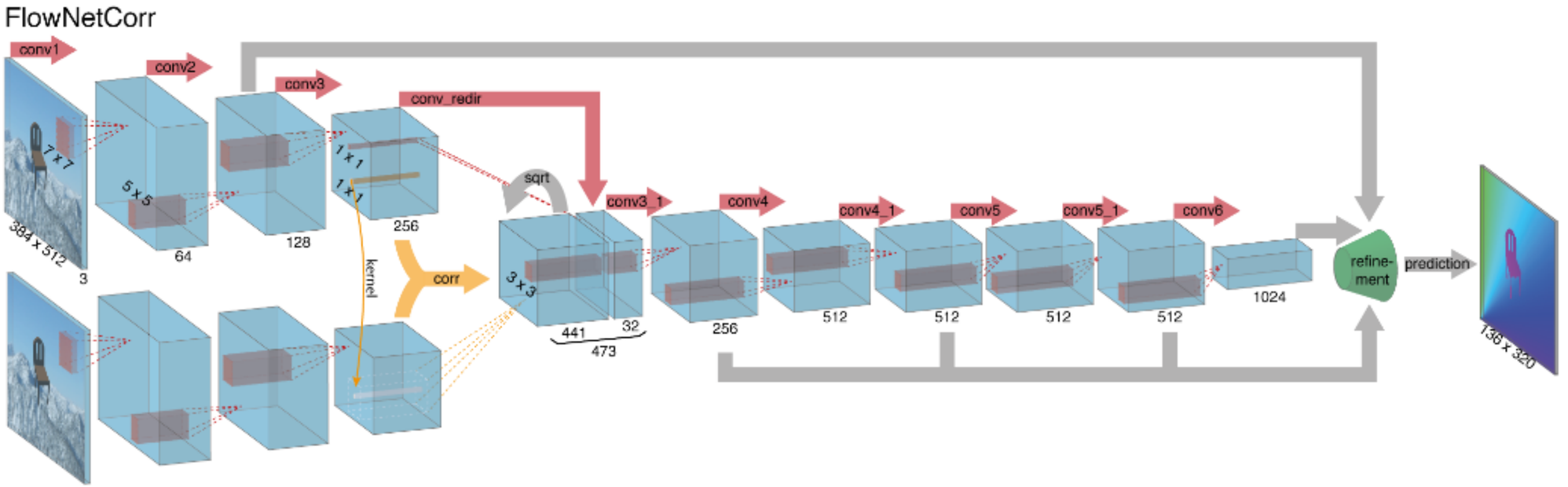
Learning Optical fLow



- Siamese architecture
- Compute correlation volume inside network

Slide credit: Christian Rupprecht, Oxford

Learning Optical fLow

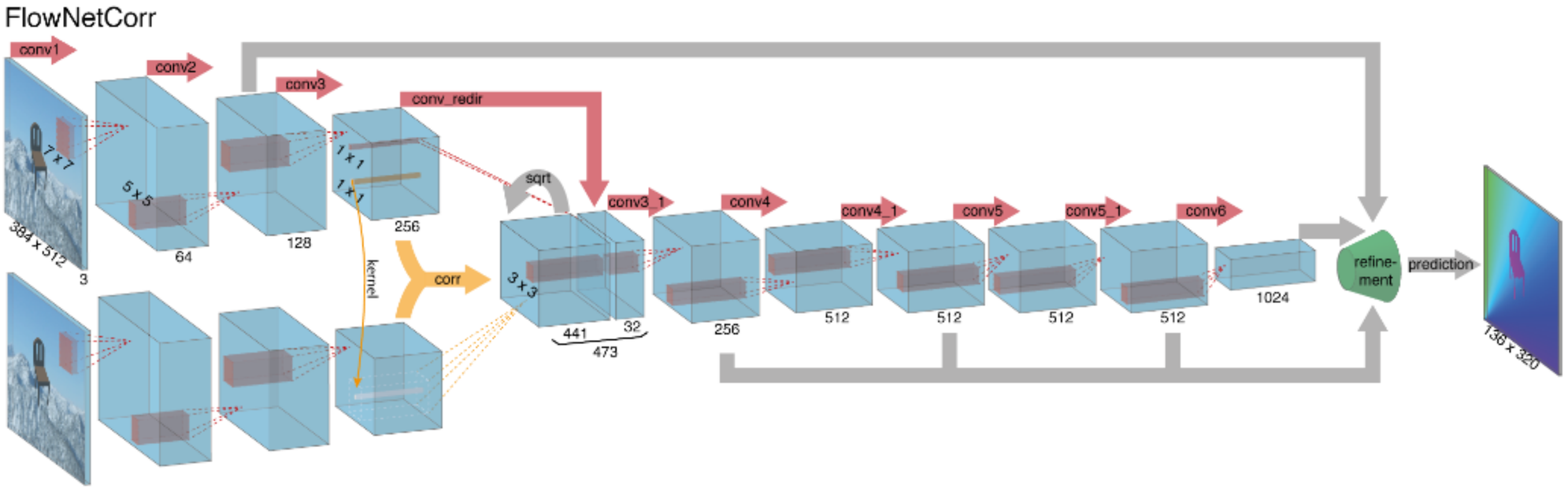


- Siamese architecture
- Compute correlation volume inside network

$$c(\mathbf{x}_1, \mathbf{x}_2) = \sum_{\mathbf{o} \in [-k,k] \times [-k,k]} \langle \mathbf{f}_1(\mathbf{x}_1 + \mathbf{o}), \mathbf{f}_2(\mathbf{x}_2 + \mathbf{o}) \rangle$$

Slide credit: Christian Rupprecht, Oxford

Learning Optical fLow

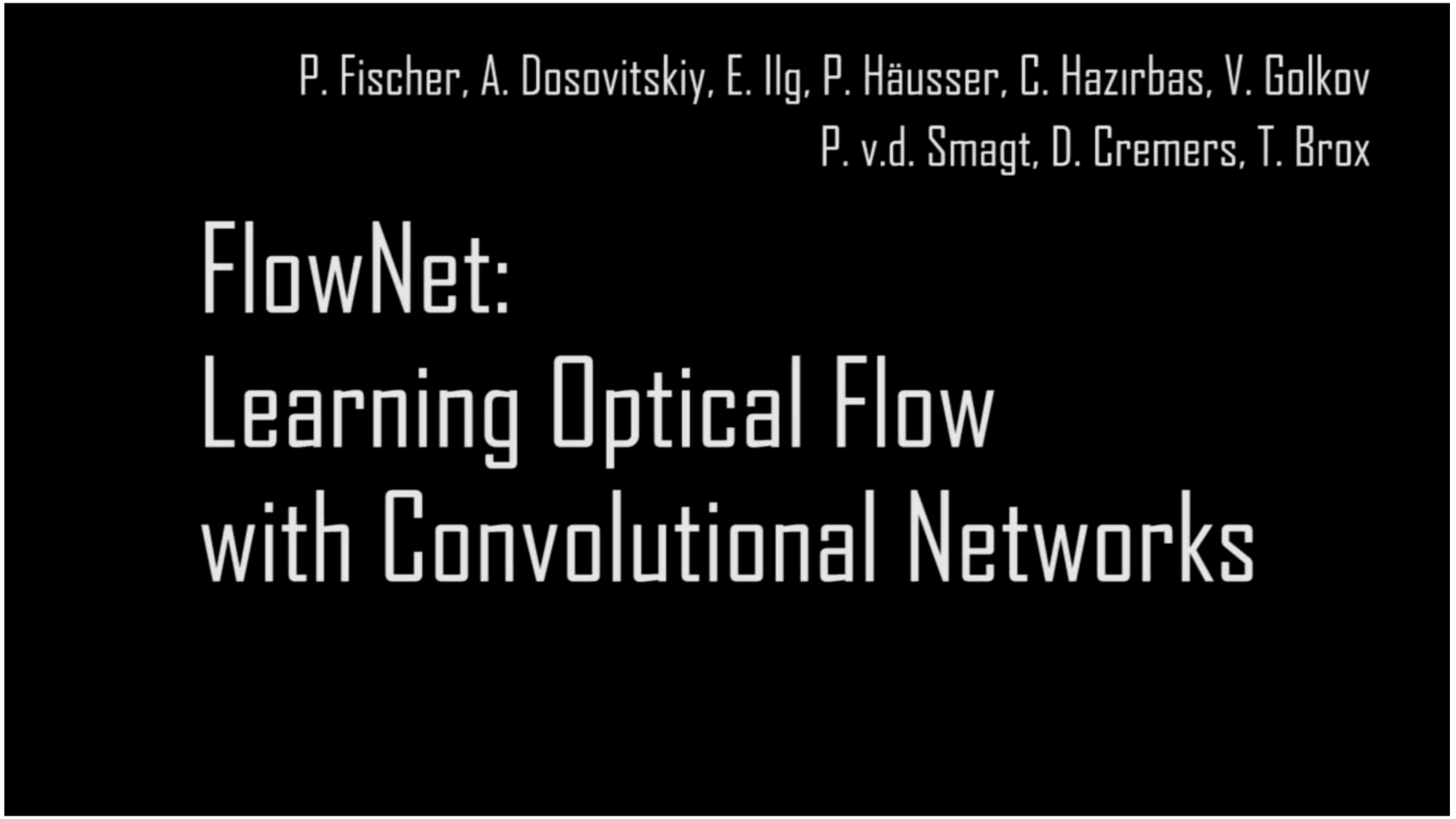


- Siamese architecture
- Compute correlation volume inside network
- Trained on FlyingThings3D

$$c(\mathbf{x}_1, \mathbf{x}_2) = \sum_{\mathbf{o} \in [-k,k] \times [-k,k]} \langle \mathbf{f}_1(\mathbf{x}_1 + \mathbf{o}), \mathbf{f}_2(\mathbf{x}_2 + \mathbf{o}) \rangle$$

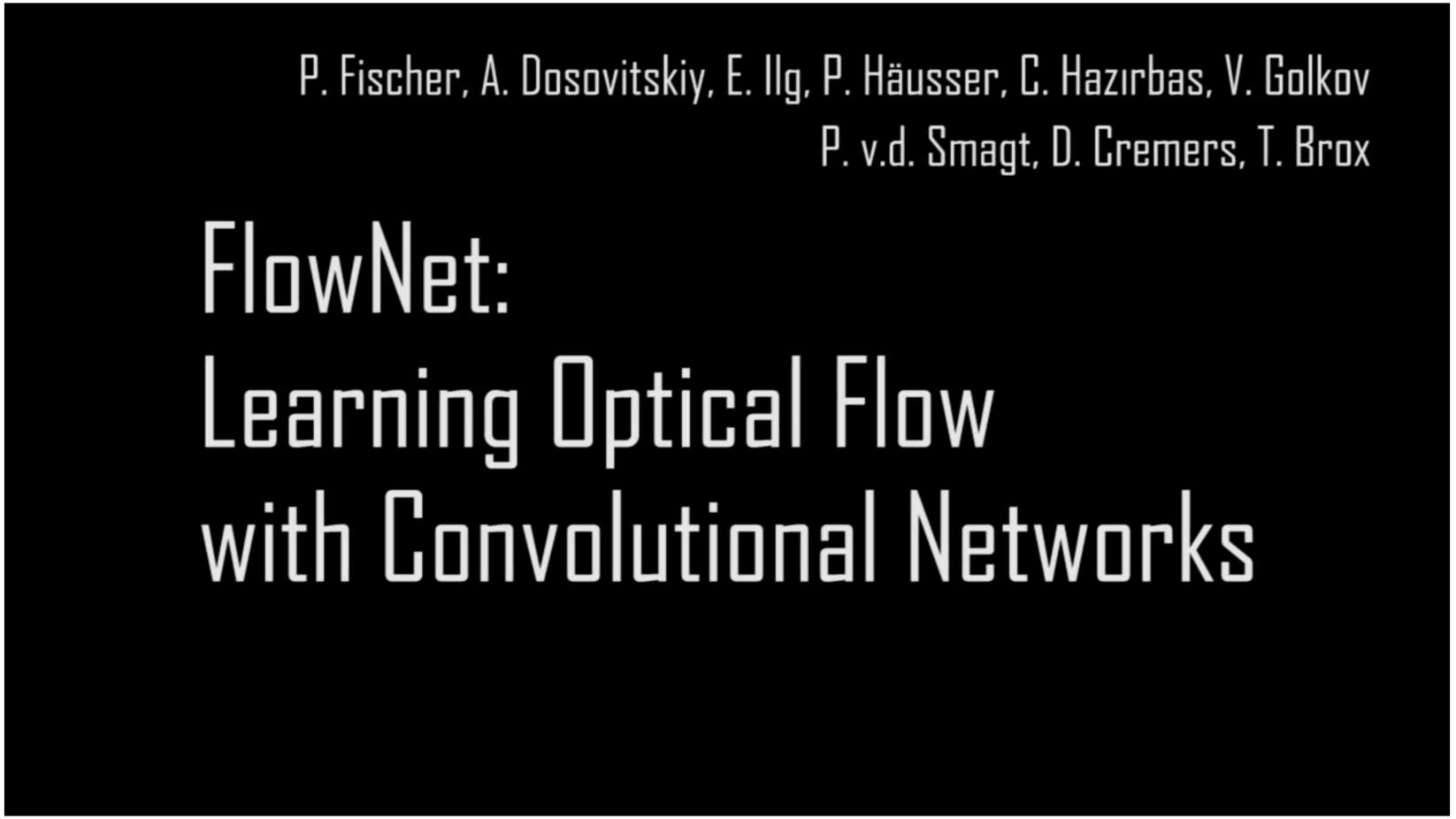
Slide credit: Christian Rupprecht, Oxford

Optical Flow –Learned: FlowNet



Slide credit: Christian Rupprecht, Oxford

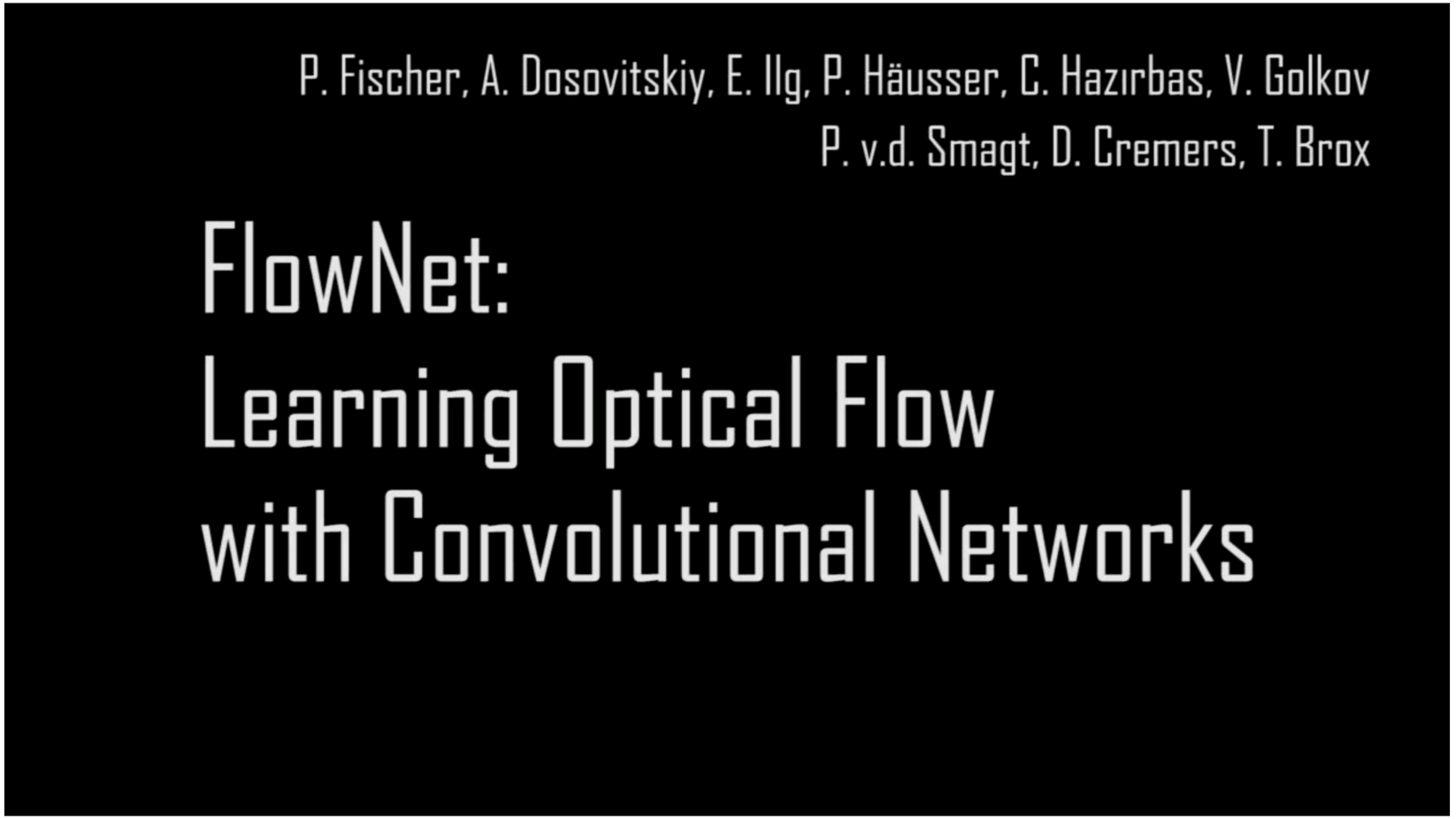
Optical Flow –Learned: FlowNet



A. Dosovitskiy, P. Fischer, E. Ilg, P. Haeusser, C. Hazirbas, V. Golkov, P. van der Smagt, D. Cremers and T. Brox
“FlowNet: Learning Optical Flow ith Convolutional Networks), ICCV 2015

Slide credit: Christian Rupprecht, Oxford

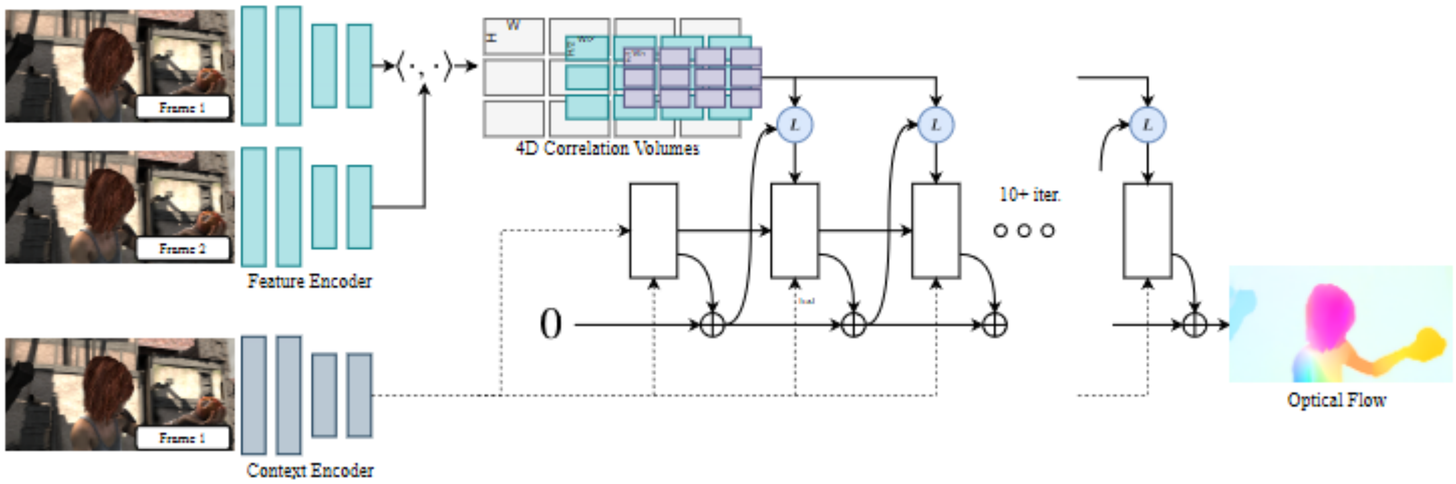
Optical Flow –Learned: FlowNet



A. Dosovitskiy, P. Fischer, E. Ilg, P. Haeusser, C. Hazirbas, V. Golkov, P. van der Smagt, D. Cremers and T. Brox
“FlowNet: Learning Optical Flow ith Convolutional Networks), ICCV 2015

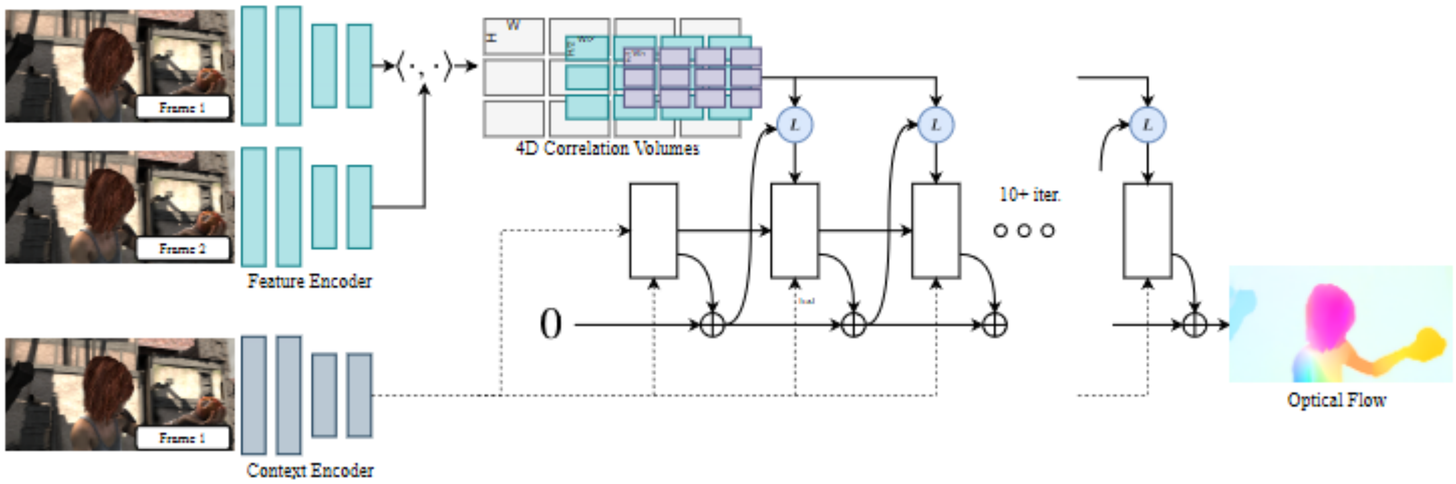
Slide credit: Christian Rupprecht, Oxford

Optical Flow – Now(ish): RAFT



Slide credit: Christian Rupprecht, Oxford

Optical Flow – Now(ish): RAFT



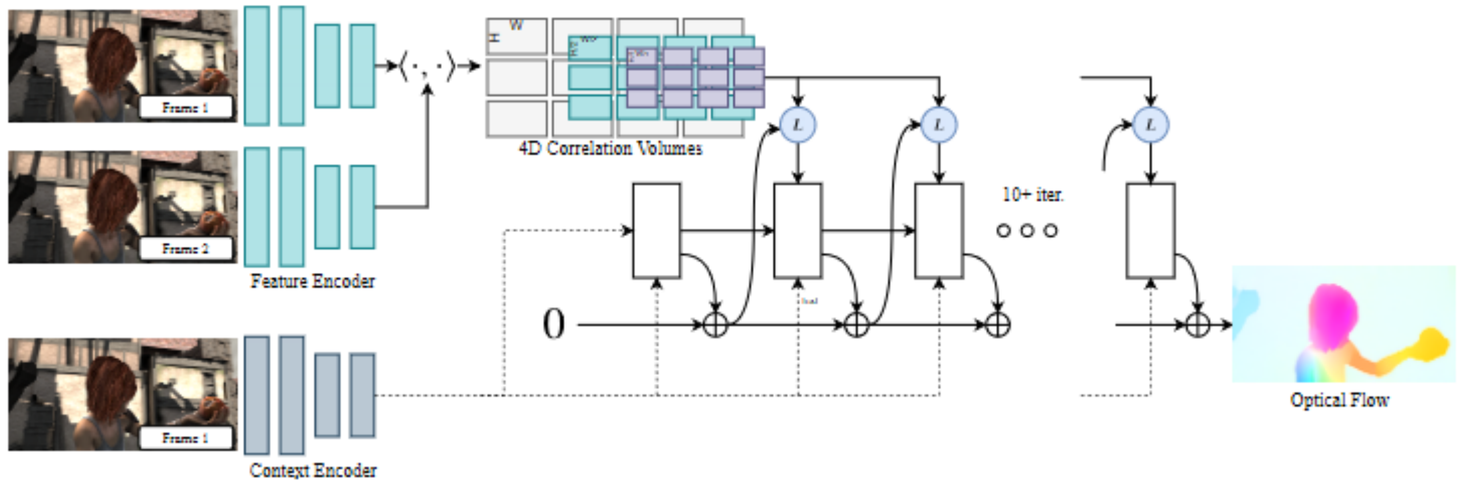
Teed, Zachary, and Jia Deng. "Raft: Recurrent all-pairs field transforms for optical flow.", ECCV 2020

Slide credit: Christian Rupprecht, Oxford

Optical Flow – Now(ish): RAFT

Improvements (by RAFT and other papers):

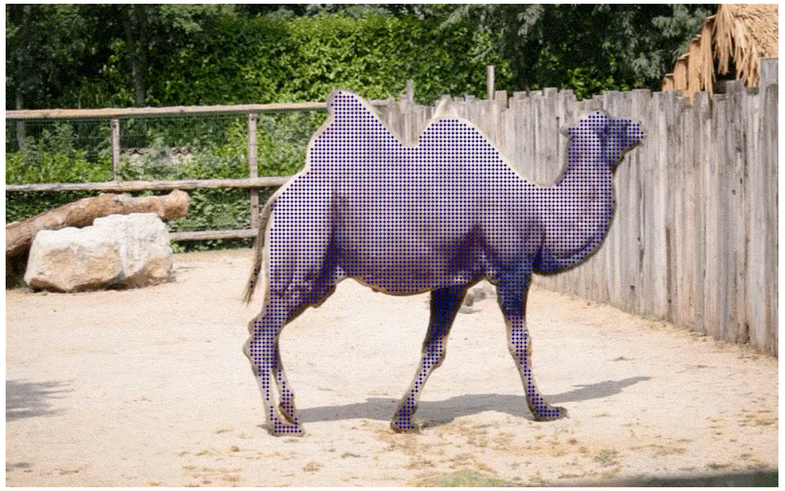
- Better backbone architectures
- Multi-scale correlation volume
- Iterative refinements: predict and update the flow estimate through several iterations

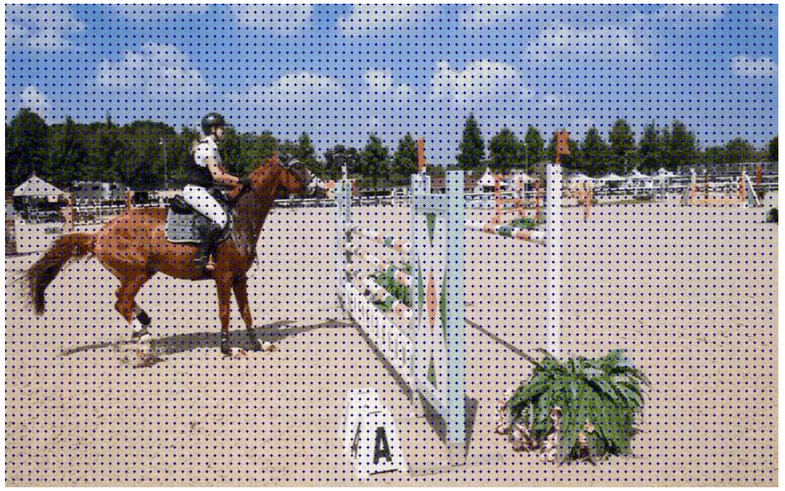


Teed, Zachary, and Jia Deng. "Raft: Recurrent all-pairs field transforms for optical flow.", ECCV 2020

Slide credit: Christian Rupprecht, Oxford

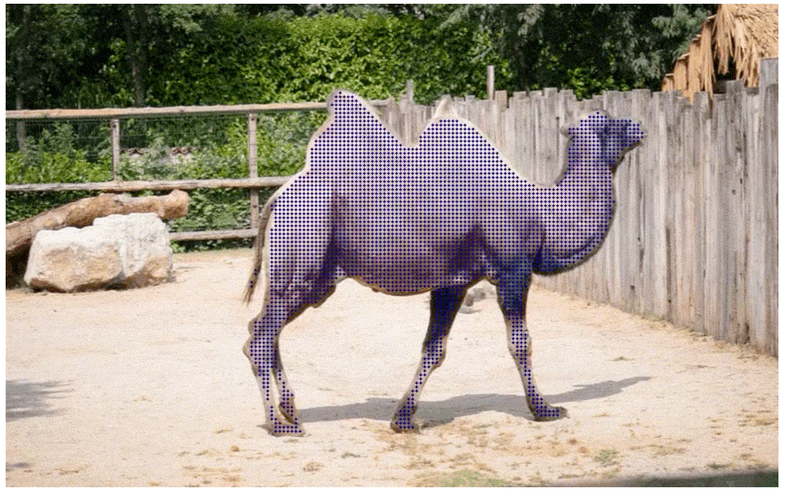
Recently: Point Tracking Revival

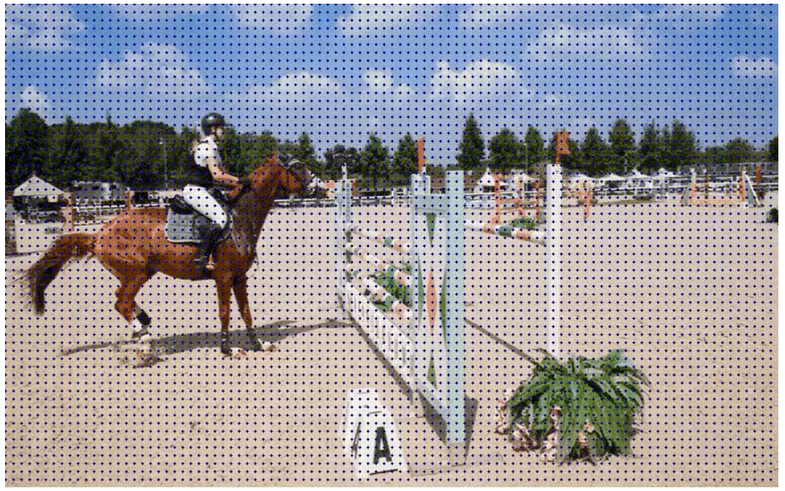




Harley, Adam W., Zhaoyuan Fang, and Katerina Fragkiadaki. "Particle video revisited: Tracking through occlusions using point trajectories." ECCV, 2022.

Recently: Point Tracking Revival





Harley, Adam W., Zhaoyuan Fang, and Katerina Fragkiadaki. "Particle video revisited: Tracking through occlusions using point trajectories." ECCV, 2022.

Motion Estimation Today: Supervised on large datasets

Slide credit: Christian Rupprecht, Oxford

Motion Estimation Today: Supervised on large datasets

Point Tracking

Long-term tracking of individual points

Slide credit: Christian Rupprecht, Oxford

Motion Estimation Today: Supervised on large datasets

Point Tracking

Long-term tracking of individual points

Optical Flow

Dense correspondences between a pair of frames

Slide credit: Christian Rupprecht, Oxford

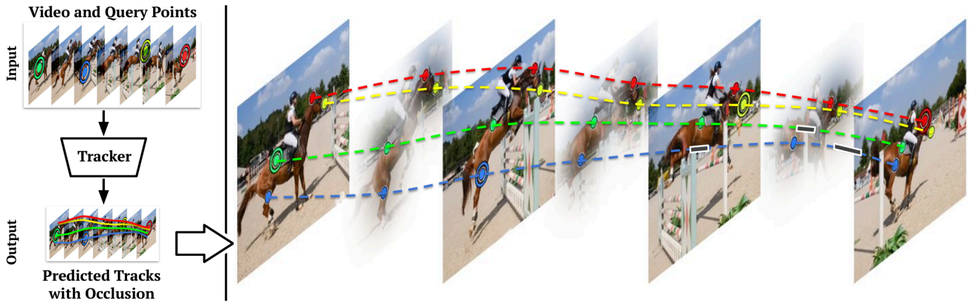
Motion Estimation Today: Supervised on large datasets

Point Tracking

Long-term tracking of individual points

PIPs

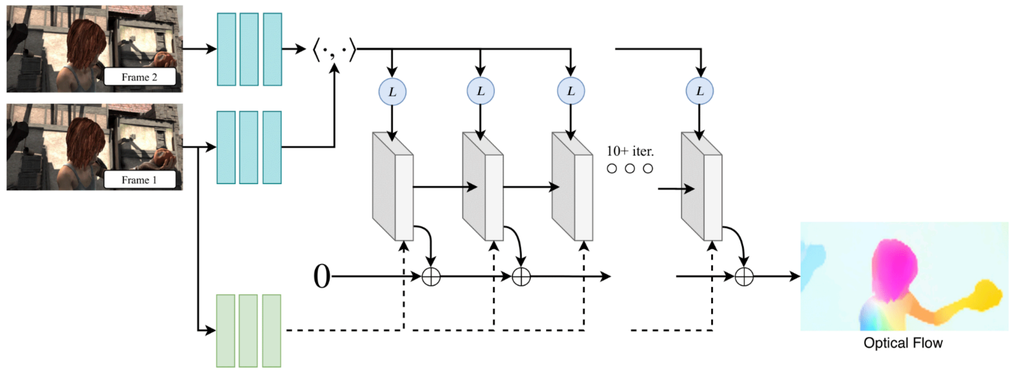
TAP-Net



Optical Flow

Dense correspondences between a pair of frames

RAFT



Particle Video Revisited: Tracking Through Occlusions Using Point Trajectories. Harley et. al. ECCV 2022
TAP-Vid: A Benchmark for Tracking Any Point in a Video. Doersch et al., NeurIPS D&B 2022
RAFT: Recurrent All-Pairs Field Transforms for Optical Flow. Teed et al., ECCV 2020

Slide credit: Christian Rupprecht, Oxford

Shape-from-X

1970: Shape from Shading

Slide credit: Christian Rupprecht, Oxford

1970: Shape from Shading

- Recover 3D from a single 2D image

Slide credit: Christian Rupprecht, Oxford

1970: Shape from Shading

- Recover 3D from a single 2D image
- Assume simple lighting and material (Lambertian with constant albedo)

Slide credit: Christian Rupprecht, Oxford

1970: Shape from Shading

- Recover 3D from a single 2D image
- Assume simple lighting and material (Lambertian with constant albedo)
- Strong smoothness assumptions

Slide credit: Christian Rupprecht, Oxford

1970: Shape from Shading

- Recover 3D from a single 2D image
- Assume simple lighting and material (Lambertian with constant albedo)
- Strong smoothness assumptions

Shape-from-X
- Shading: Horn (1970)
- Contour: Guzman (1971), Waltz (1975), etc.
- Texture: Bajczy & Lieberman (1976)
- Stereo: Marr & Poggio (1976)

Slide credit: Christian Rupprecht, Oxford

1970: Shape from Shading

- Recover 3D from a single 2D image
- Assume simple lighting and material (Lambertian with constant albedo)
- Strong smoothness assumptions

Shape-from-X
- Shading: Horn (1970)
- Contour: Guzman (1971), Waltz (1975), etc.
- Texture: Bajczy & Lieberman (1976)
- Stereo: Marr & Poggio (1976)

Slide credit: Christian Rupprecht, Oxford

1970: Shape from Shading

- Recover 3D from a single 2D image
- Assume simple lighting and material (Lambertian with constant albedo)
- Strong smoothness assumptions

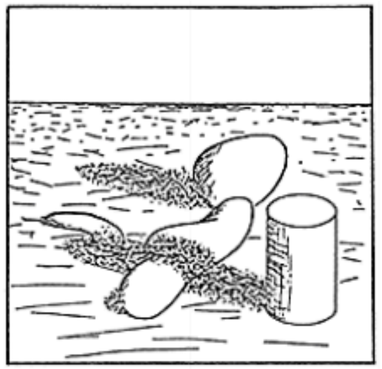
Shape-from-X
- Shading: Horn (1970)
- Contour: Guzman (1971), Waltz (1975), etc.
- Texture: Bajczy & Lieberman (1976)
- Stereo: Marr & Poggio (1976)

Slide credit: Christian Rupprecht, Oxford

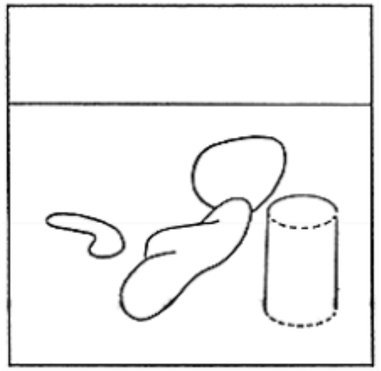
1978: Intrinsic Images

Slide credit: Christian Rupprecht, Oxford

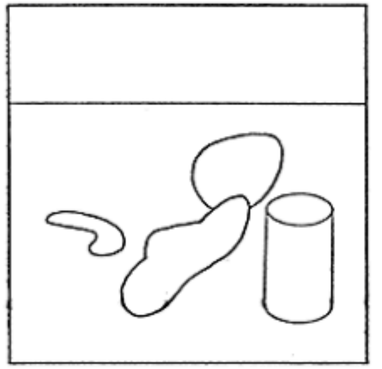
1978: Intrinsic Images



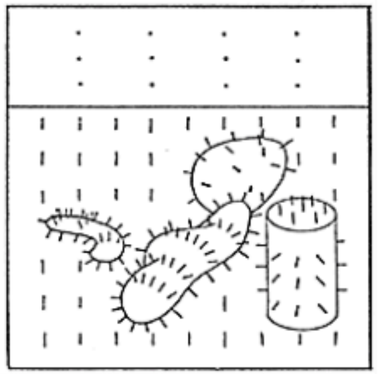
(a) ORIGINAL SCENE



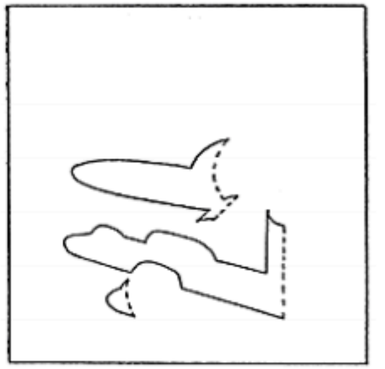
(b) DISTANCE



(c) REFLECTANCE



(d) ORIENTATION (VECTOR)



(e) ILLUMINATION

Figure 3 A set of intrinsic images derived from a single monochrome intensity image

The images are depicted as line drawings, but, in fact, would contain values at every point. The solid lines in the intrinsic images represent discontinuities in the scene characteristic; the dashed lines represent discontinuities in its derivative.

Slide credit: Christian Rupprecht, Oxford

1978: Intrinsic Images

- Decompose images into its intrinsic 2D layers
  - Reflectance
  - Shading
  - Shape
  - Motion, etc.

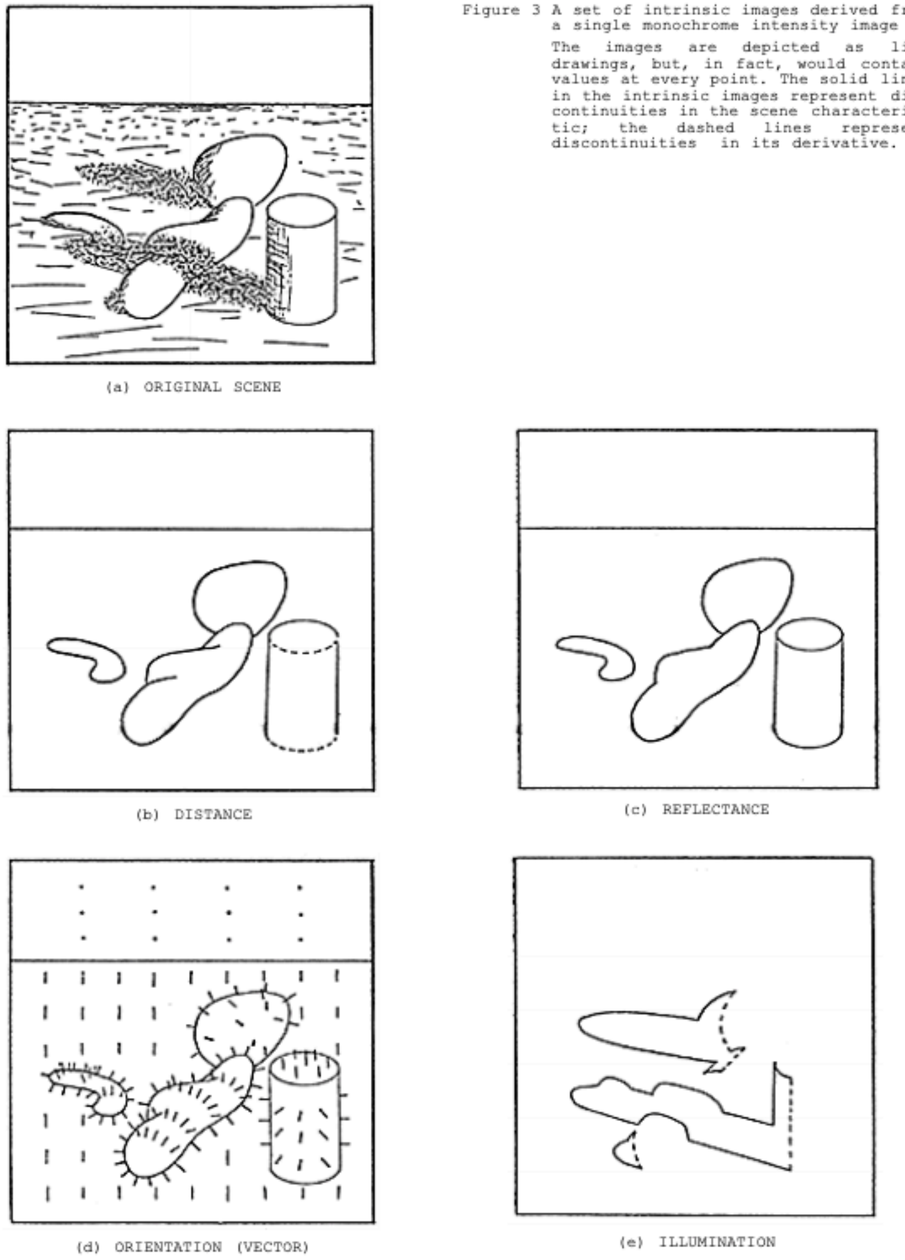


Figure 3 A set of intrinsic images derived from a single monochrome intensity image

The images are depicted as line drawings, but, in fact, would contain values at every point. The solid lines in the intrinsic images represent discontinuities in the scene characteristic; the dashed lines represent discontinuities in its derivative.

(a) ORIGINAL SCENE

(b) DISTANCE

(c) REFLECTANCE

(d) ORIENTATION (VECTOR)

(e) ILLUMINATION

Slide credit: Christian Rupprecht, Oxford

1978: Intrinsic Images

- Decompose images into its intrinsic 2D layers
  - Reflectance
  - Shading
  - Shape
  - Motion, etc.
- Useful for downstream tasks: e.g. removing lighting simplifies object detection

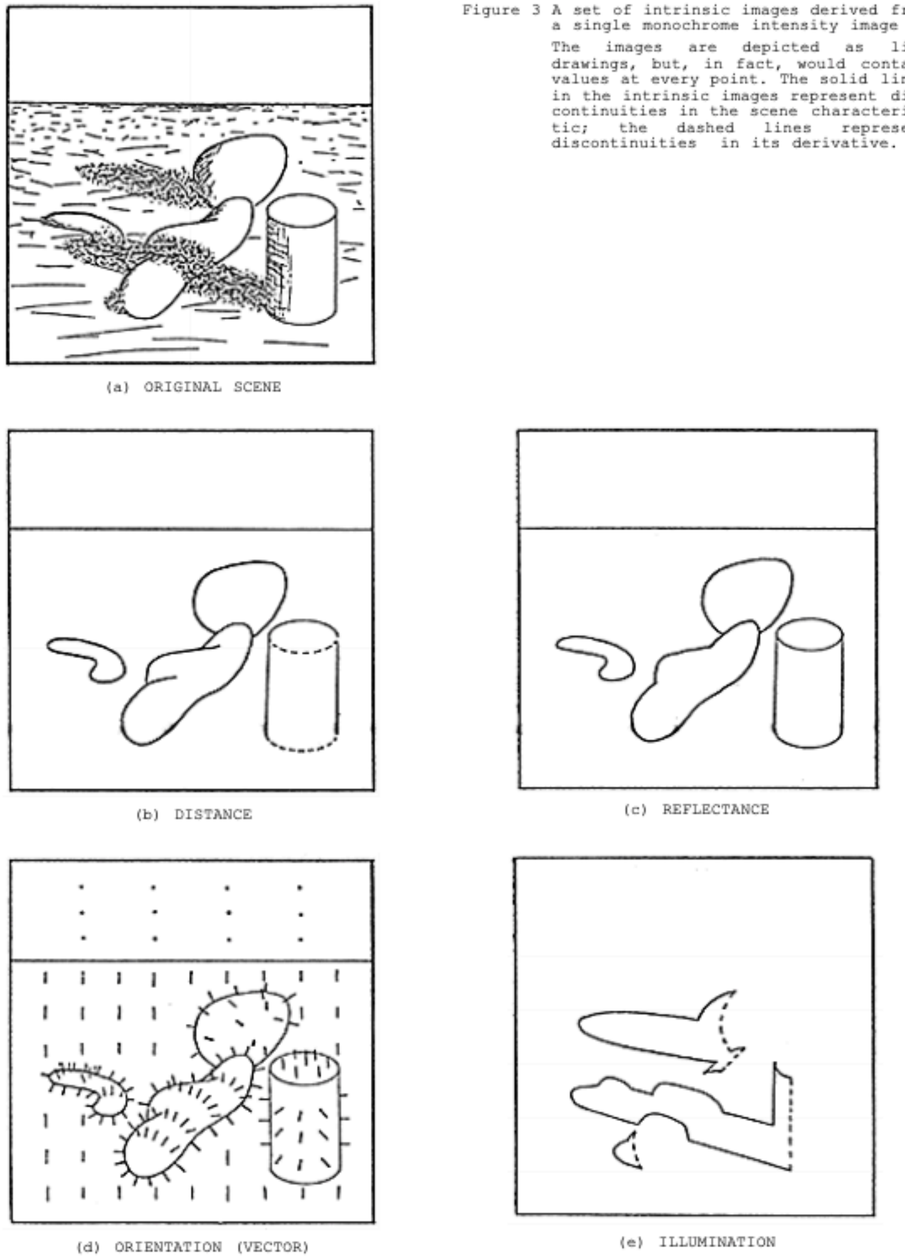


Figure 3 A set of intrinsic images derived from a single monochrome intensity image

The images are depicted as line drawings, but, in fact, would contain values at every point. The solid lines in the intrinsic images represent discontinuities in the scene characteristic; the dashed lines represent discontinuities in its derivative.

(a) ORIGINAL SCENE

(b) DISTANCE

(c) REFLECTANCE

(d) ORIENTATION (VECTOR)

(e) ILLUMINATION

Slide credit: Christian Rupprecht, Oxford

1978: Intrinsic Images

MIT Intrinsic Images dataset (2009)

Wu, Jiaye, et al. "Measured Albedo in the Wild: Filling the Gap in Intrinsics Evaluation." *arXiv preprint arXiv:2306.15662* (2023)

Slide credit: Christian Rupprecht, Oxford

1980: Photometric Stereo

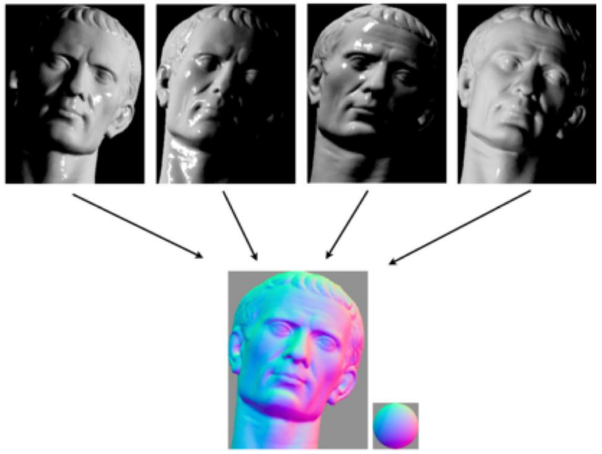
Photometric method for determining surface orientation from multiple images

Robert J. Woodham

Department of Computer Science
University of British Columbia
2075 Wesbrook Mall
Vancouver, B.C., Canada
V6T 1W5

Abstract. A novel technique called photometric stereo is introduced. The idea of photometric stereo is to vary the direction of incident illumination between successive images, while holding the viewing direction constant. It is shown that this provides sufficient information to determine surface orientation at each image point. Since the imaging geometry is not changed, the correspondence between image points is known *a priori.* The technique is photometric because it uses the radiance values recorded at a single image location, in successive views, rather than the relative positions of displaced features.

Photometric stereo is used in computer-based image understanding. It can be applied in two ways. First, it is a general technique for determining surface orientation at each image point. Second, it is a technique for determining object points that have a particular surface orientation. These applications are illustrated using synthesized examples.



Slide credit: Christian Rupprecht, Oxford

1980: Photometric Stereo

- Recover 3D from multiple (>2) 2D images with varying lighting

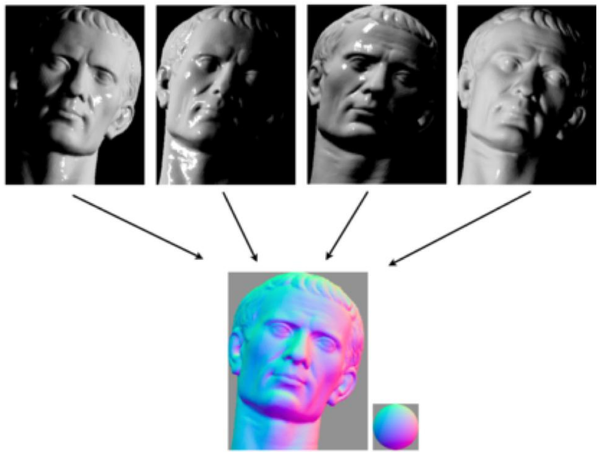
Photometric method for determining surface orientation from multiple images

Robert J. Woodham

Department of Computer Science
University of British Columbia
2075 Wesbrook Mall
Vancouver, B.C., Canada
V6T 1W5

Abstract. A novel technique called photometric stereo is introduced. The idea of photometric stereo is to vary the direction of incident illumination between successive images, while holding the viewing direction constant. It is shown that this provides sufficient information to determine surface orientation at each image point. Since the imaging geometry is not changed, the correspondence between image points is known *a priori.* The technique is photometric because it uses the radiance values recorded at a single image location, in successive views, rather than the relative positions of displaced features.

Photometric stereo is used in computer-based image understanding. It can be applied in two ways. First, it is a general technique for determining surface orientation at each image point. Second, it is a technique for determining object points that have a particular surface orientation. These applications are illustrated using synthesized examples.



Slide credit: Christian Rupprecht, Oxford

1980: Photometric Stereo

- Recover 3D from multiple (>2) 2D images with varying lighting
- Highly detailed and accurate

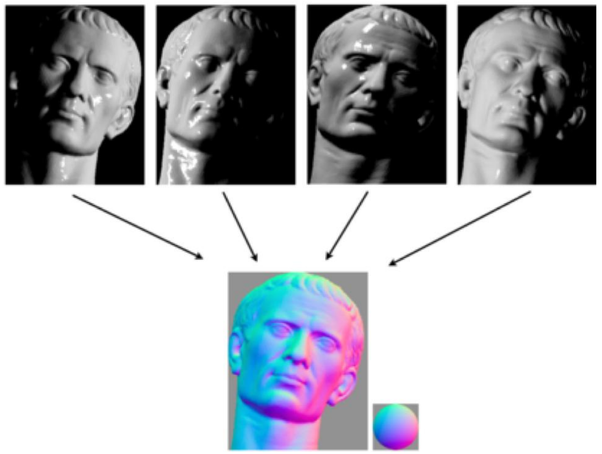
Photometric method for determining surface orientation from multiple images

Robert J. Woodham

Department of Computer Science
University of British Columbia
2075 Wesbrook Mall
Vancouver, B.C., Canada
V6T 1W5

Abstract. A novel technique called photometric stereo is introduced. The idea of photometric stereo is to vary the direction of incident illumination between successive images, while holding the viewing direction constant. It is shown that this provides sufficient information to determine surface orientation at each image point. Since the imaging geometry is not changed, the correspondence between image points is known *a priori.* The technique is photometric because it uses the radiance values recorded at a single image location, in successive views, rather than the relative positions of displaced features.

Photometric stereo is used in computer-based image understanding. It can be applied in two ways. First, it is a general technique for determining surface orientation at each image point. Second, it is a technique for determining object points that have a particular surface orientation. These applications are illustrated using synthesized examples.



Slide credit: Christian Rupprecht, Oxford

1980: Photometric Stereo

- Recover 3D from multiple (>2) 2D images with varying lighting
- Highly detailed and accurate
- Still Lambertian lighting assumption – relaxed later

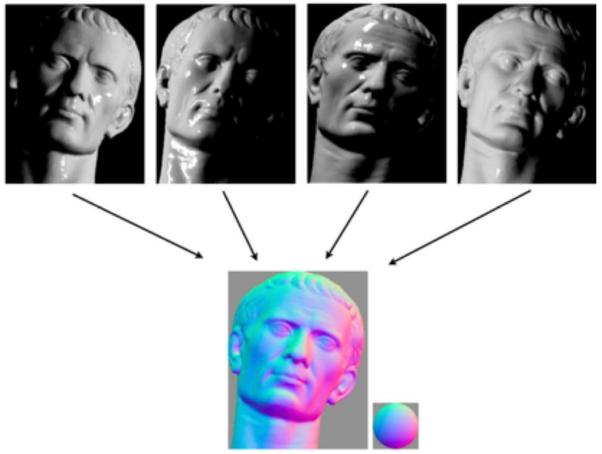
Photometric method for determining surface orientation from multiple images

Robert J. Woodham

Department of Computer Science
University of British Columbia
2075 Wesbrook Mall
Vancouver, B.C., Canada
V6T 1W5

Abstract. A novel technique called photometric stereo is introduced. The idea of photometric stereo is to vary the direction of incident illumination between successive images, while holding the viewing direction constant. It is shown that this provides sufficient information to determine surface orientation at each image point. Since the imaging geometry is not changed, the correspondence between image points is known *a priori.* The technique is photometric because it uses the radiance values recorded at a single image location, in successive views, rather than the relative positions of displaced features.

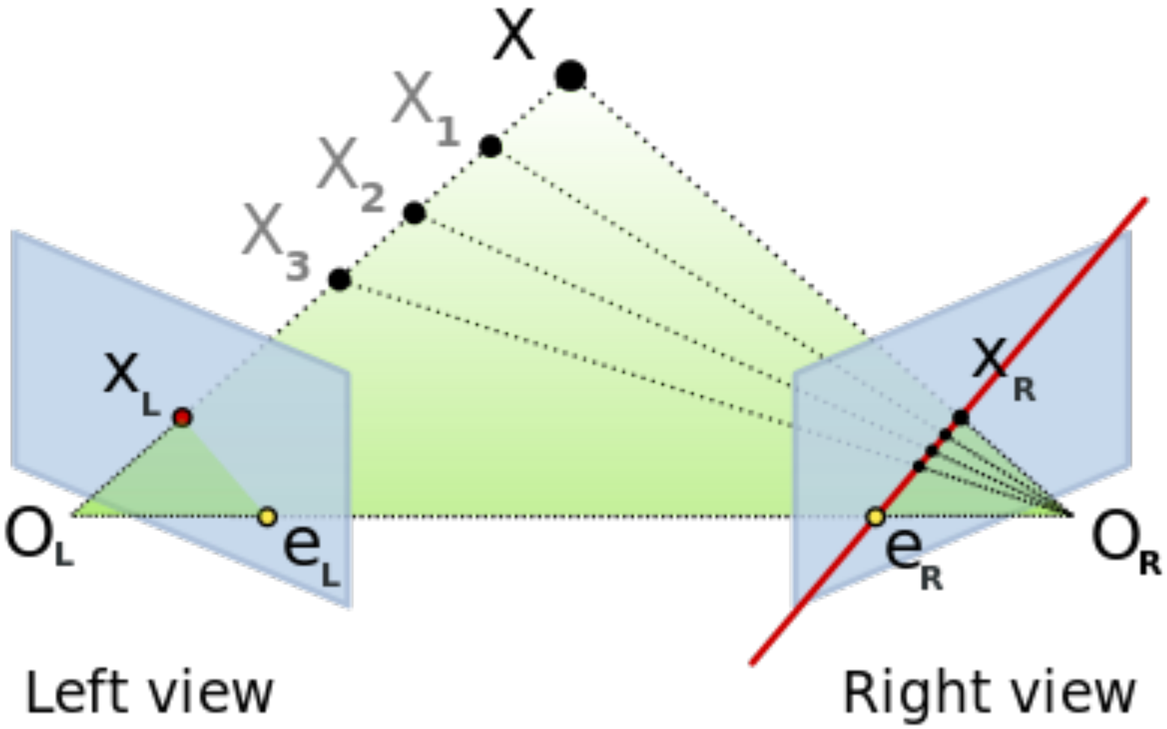
Photometric stereo is used in computer-based image understanding. It can be applied in two ways. First, it is a general technique for determining surface orientation at each image point. Second, it is a technique for determining object points that have a particular surface orientation. These applications are illustrated using synthesized examples.



Slide credit: Christian Rupprecht, Oxford

Stereo

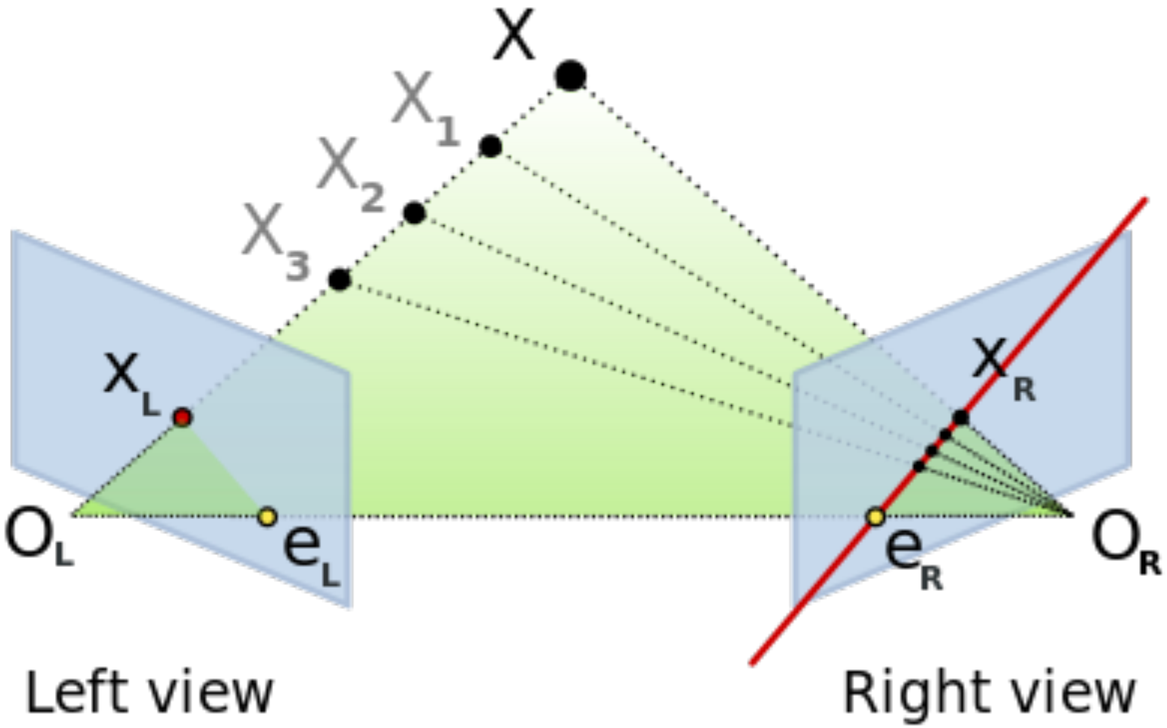
1981: Essential Matrix



H. Christopher Longuet-Higgins (September 1981). "A computer algorithm for reconstructing a scene from two projections". Nature.
Slide credit: Christian Rupprecht, Oxford

1981: Essential Matrix

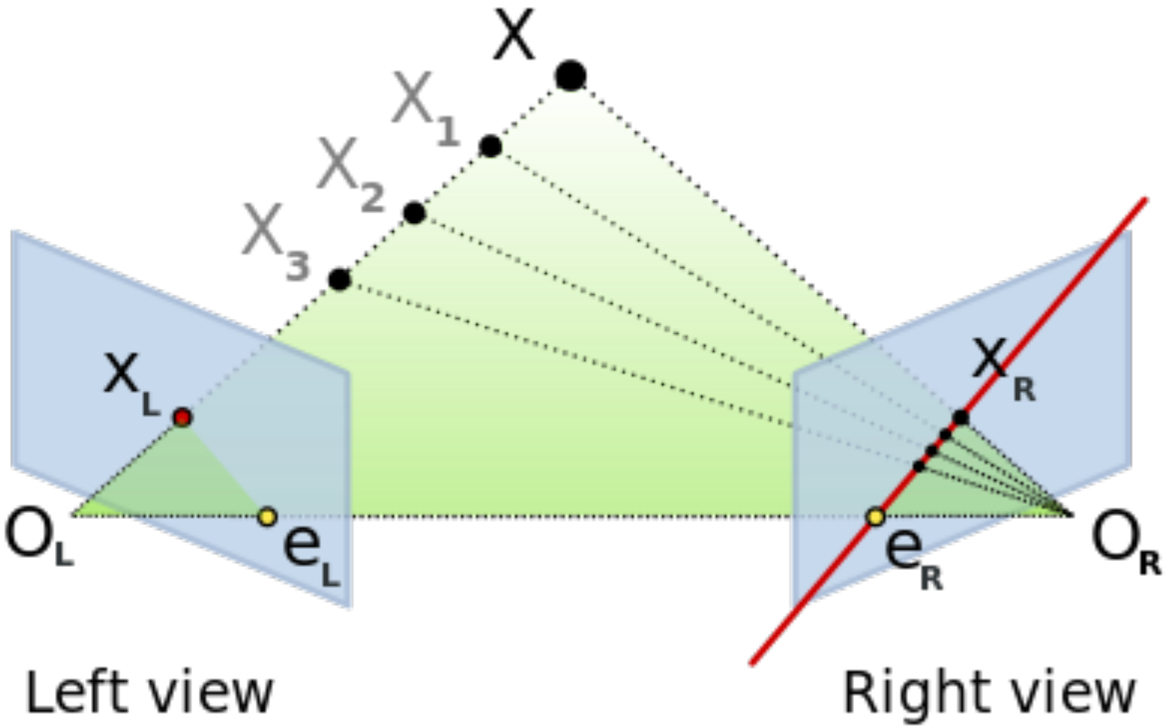
- 2-view Geometry: a matrix that maps points to epipolar lines



H. Christopher Longuet-Higgins (September 1981). "A computer algorithm for reconstructing a scene from two projections". Nature.
Slide credit: Christian Rupprecht, Oxford

1981: Essential Matrix

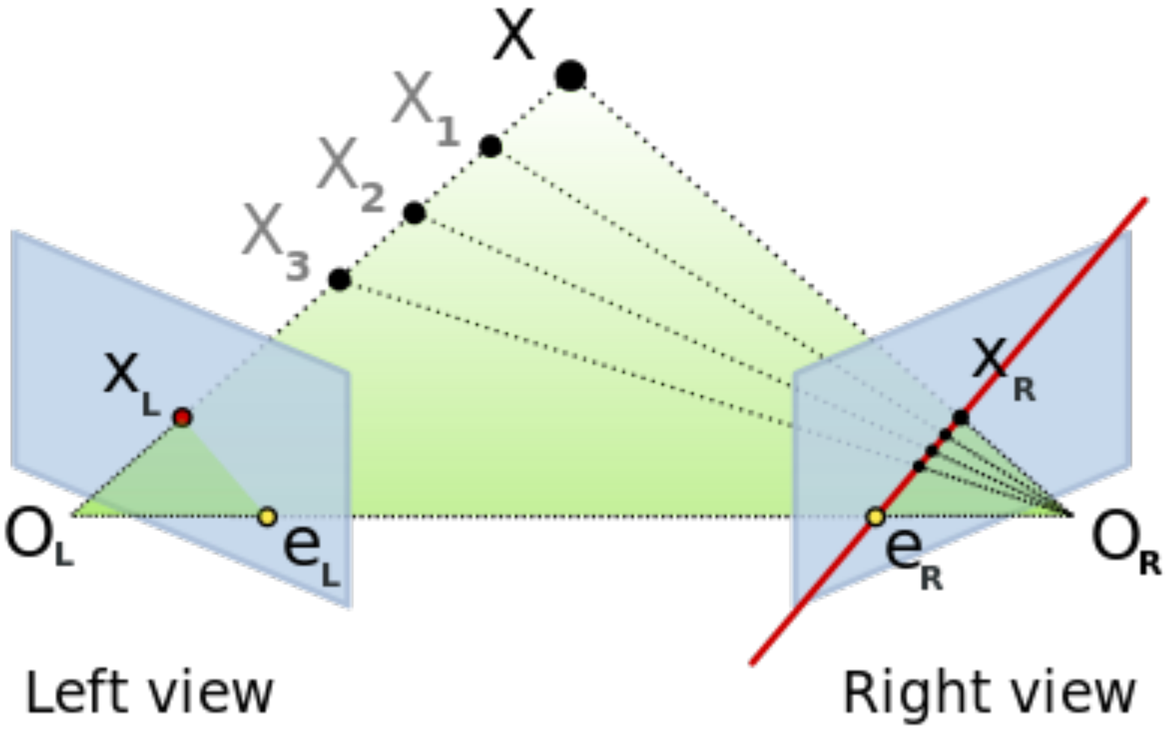
- 2-view Geometry: a matrix that maps points to epipolar lines
- Correspondence search becomes a 1D problem



H. Christopher Longuet-Higgins (September 1981). "A computer algorithm for reconstructing a scene from two projections". Nature.
Slide credit: Christian Rupprecht, Oxford

1981: Essential Matrix

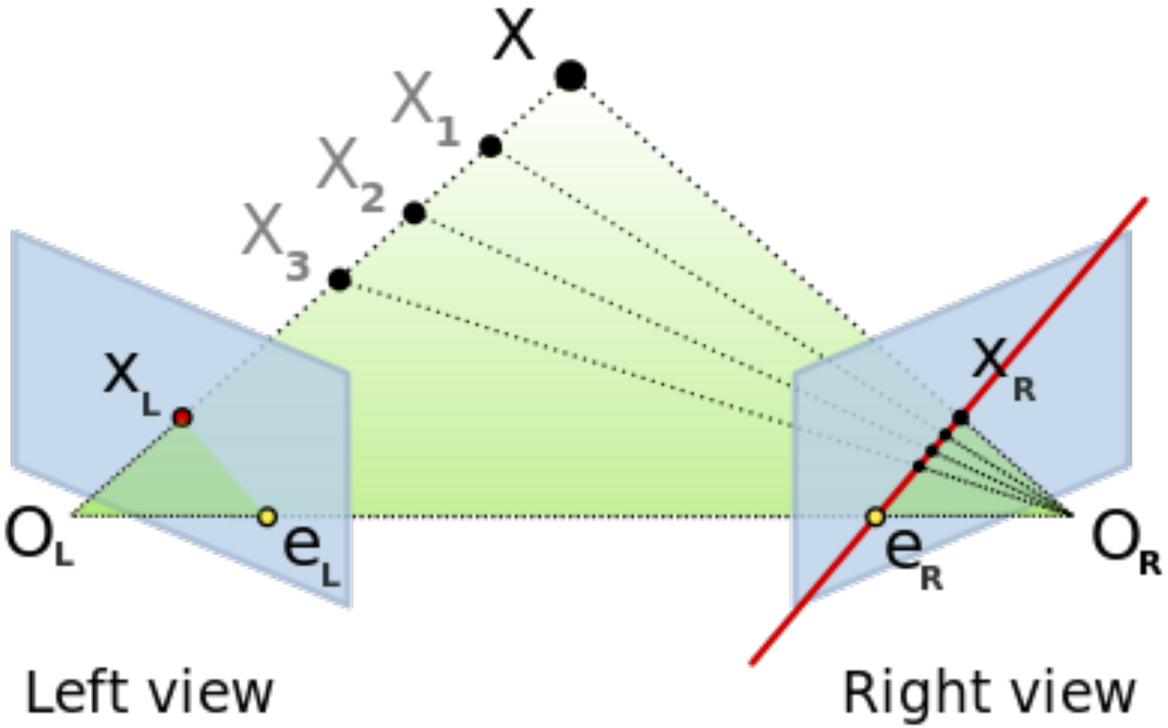
- 2-view Geometry: a matrix that maps points to epipolar lines
- Correspondence search becomes a 1D problem
- Essential Matrix can be computed from 2D correspondences



H. Christopher Longuet-Higgins (September 1981). "A computer algorithm for reconstructing a scene from two projections". Nature.
Slide credit: Christian Rupprecht, Oxford

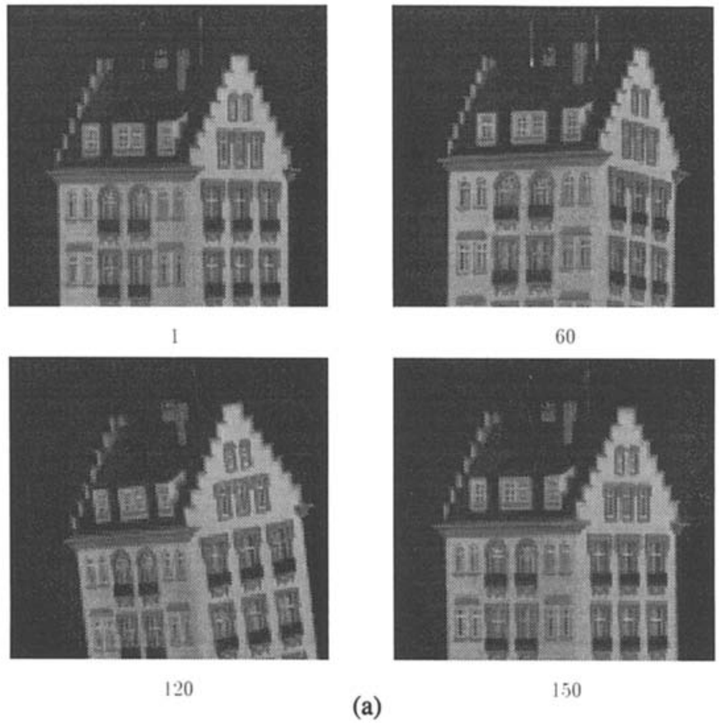
1981: Essential Matrix

- 2-view Geometry: a matrix that maps points to epipolar lines
- Correspondence search becomes a 1D problem
- Essential Matrix can be computed from 2D correspondences
- Rediscovery of known ideas >100 years old



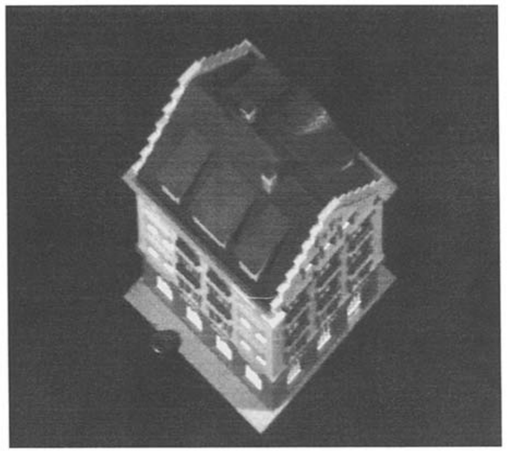
H. Christopher Longuet-Higgins (September 1981). "A computer algorithm for reconstructing a scene from two projections". Nature.
Slide credit: Christian Rupprecht, Oxford

1992: Structure-from-motion



(a)

Fig. 2a. The "Hotel" stream: four of the 150 frames.



Tomasi, Carlo, and Takeo Kanade. "Shape and motion from image streams under orthography: a factorization method." *IJCV* (1992)

Slide credit: Christian Rupprecht, Oxford

1992: Structure-from-motion

- Estimating the 3D structure from image collections of static scenes

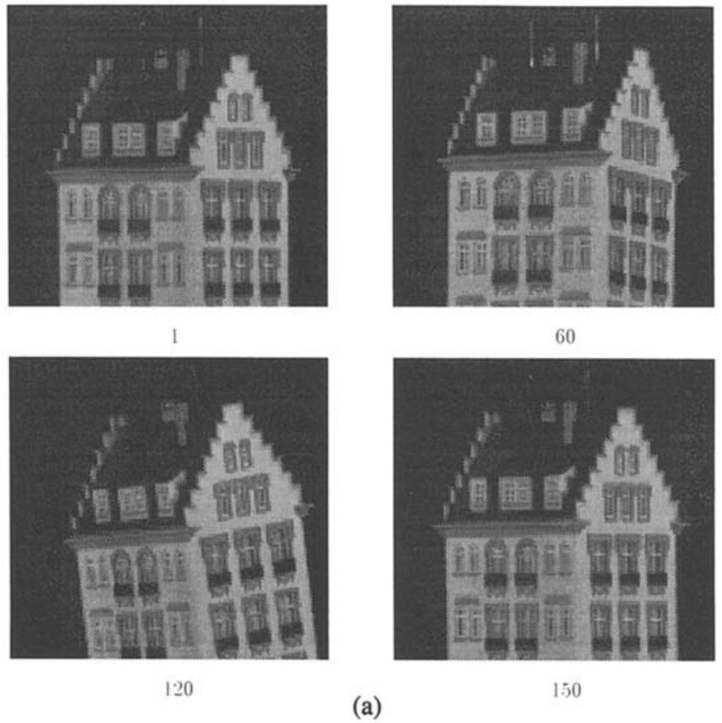
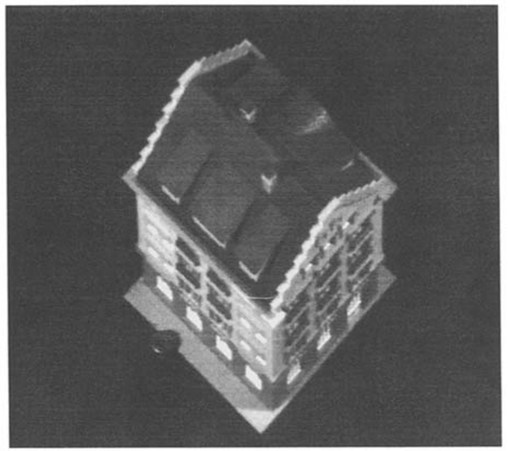


Fig. 2a. The "Hotel" stream: four of the 150 frames.



Tomasi, Carlo, and Takeo Kanade. "Shape and motion from image streams under orthography: a factorization method." *IJCV* (1992)

Slide credit: Christian Rupprecht, Oxford

1992: Structure-from-motion

- Estimating the 3D structure from image collections of static scenes
- Static scenes require only a single (moving) camera

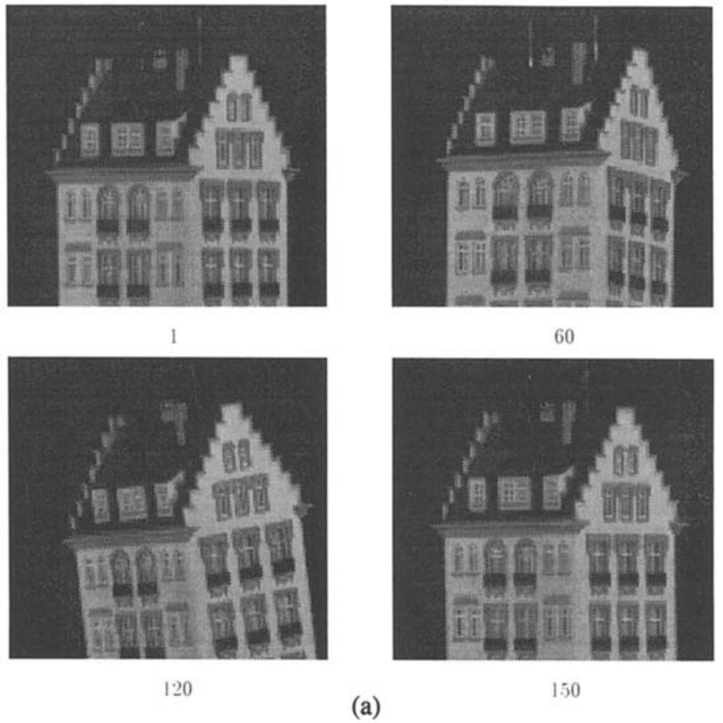
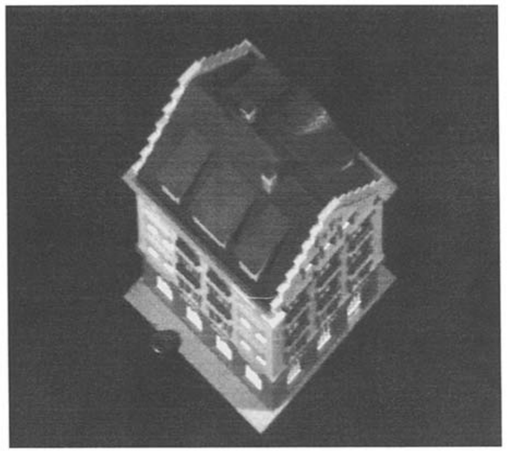


Fig. 2a. The "Hotel" stream: four of the 150 frames.

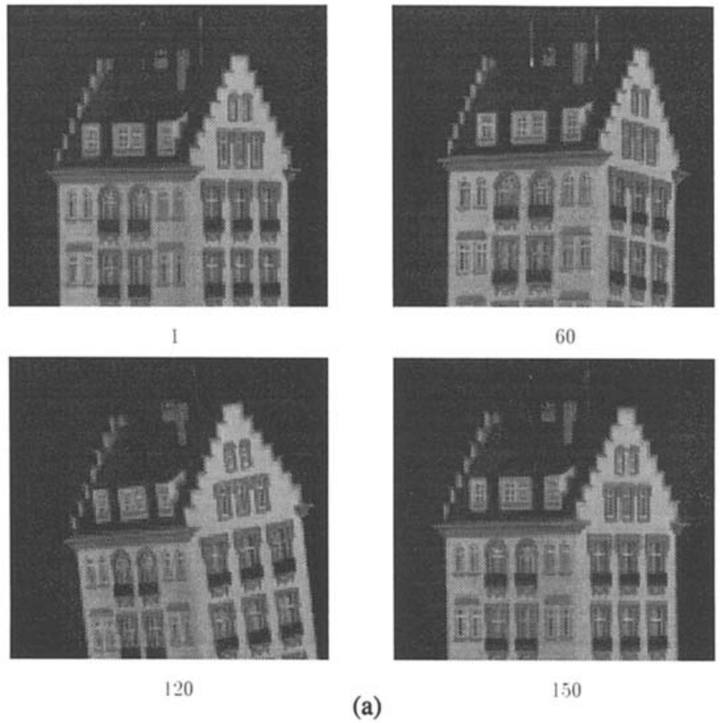


Tomasi, Carlo, and Takeo Kanade. "Shape and motion from image streams under orthography: a factorization method." *IJCV* (1992)

Slide credit: Christian Rupprecht, Oxford

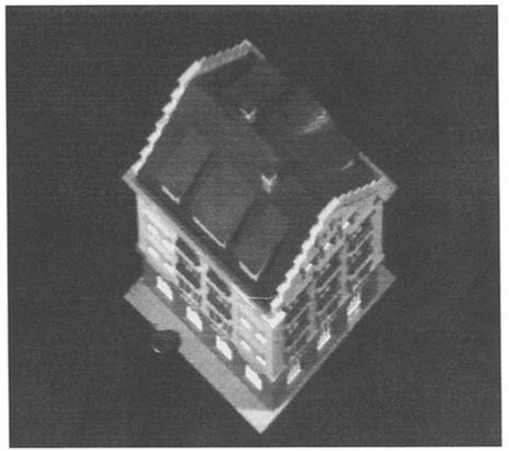
1992: Structure-from-motion

- Estimating the 3D structure from image collections of static scenes
- Static scenes require only a single (moving) camera
- Closed-form SVD solution: Tomasi-Kanade factorisation for orthographic projection.



(a)

Fig. 2a. The "Hotel" stream: four of the 150 frames.

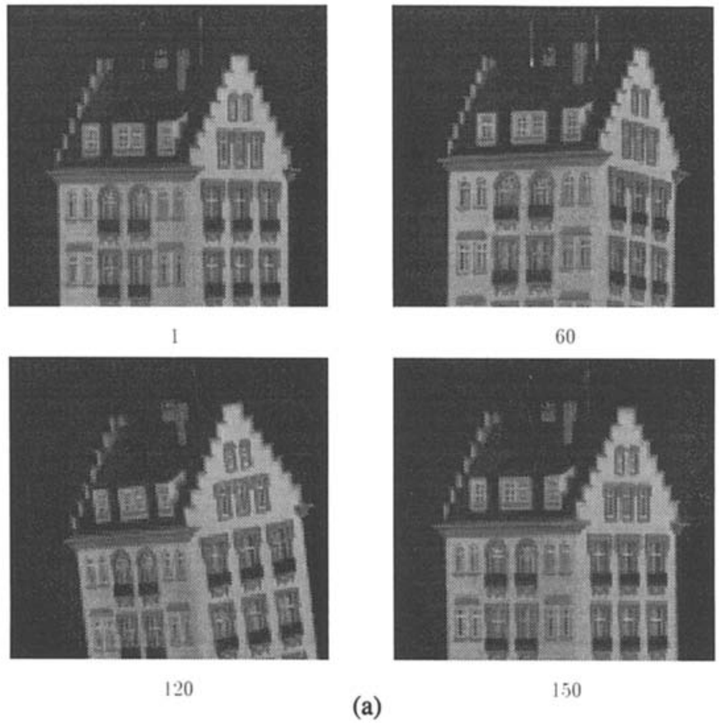


Tomasi, Carlo, and Takeo Kanade. "Shape and motion from image streams under orthography: a factorization method." *IJCV* (1992)

Slide credit: Christian Rupprecht, Oxford

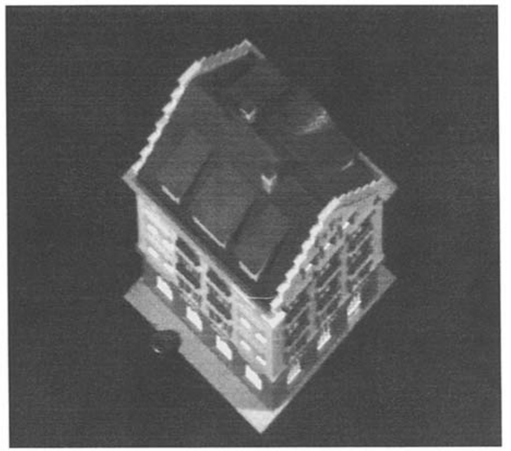
1992: Structure-from-motion

- Estimating the 3D structure from image collections of static scenes
- Static scenes require only a single (moving) camera
- Closed-form SVD solution: Tomasi-Kanade factorisation for orthographic projection.
- Later: non-linear least squares for projective cameras



(a)

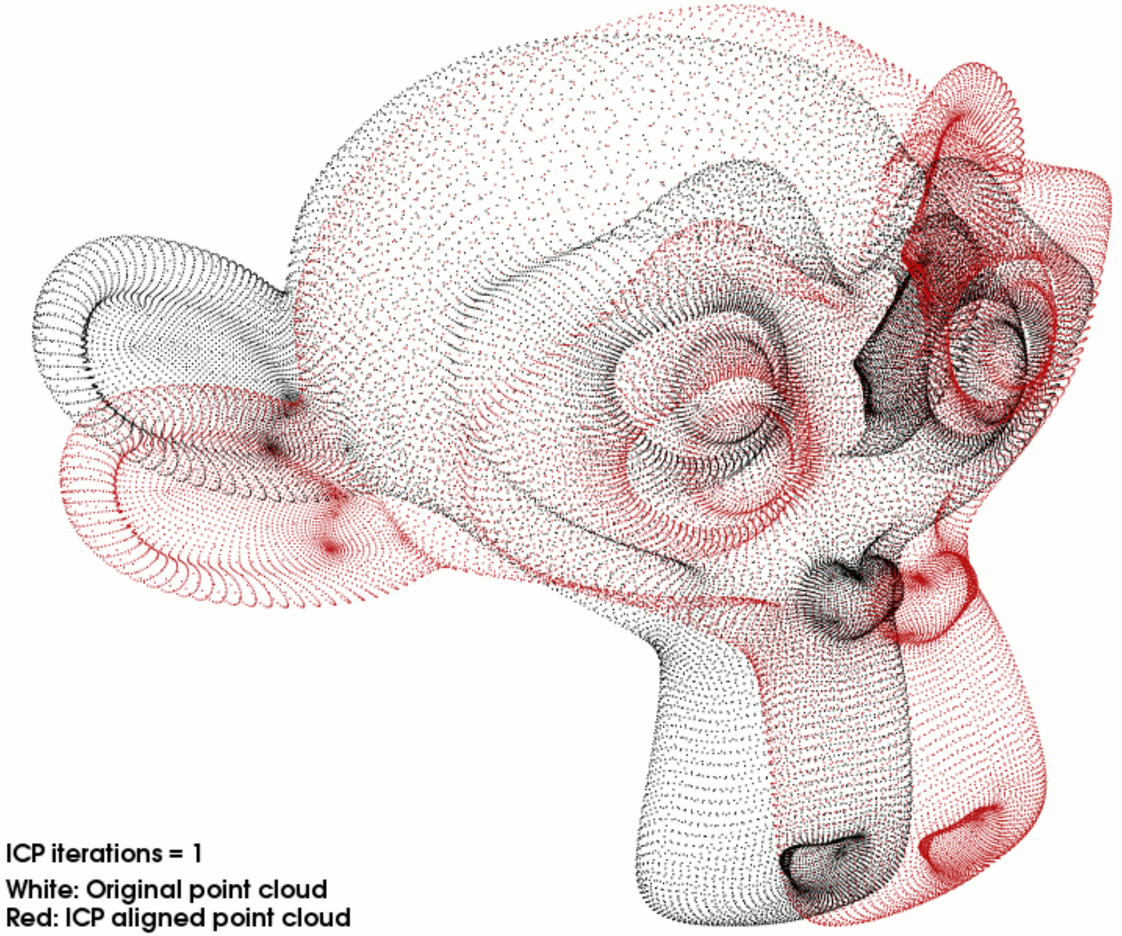
Fig. 2a. The "Hotel" stream: four of the 150 frames.



Tomasi, Carlo, and Takeo Kanade. "Shape and motion from image streams under orthography: a factorization method." *IJCV* (1992)

Slide credit: Christian Rupprecht, Oxford

1992: Iterative Closest Points

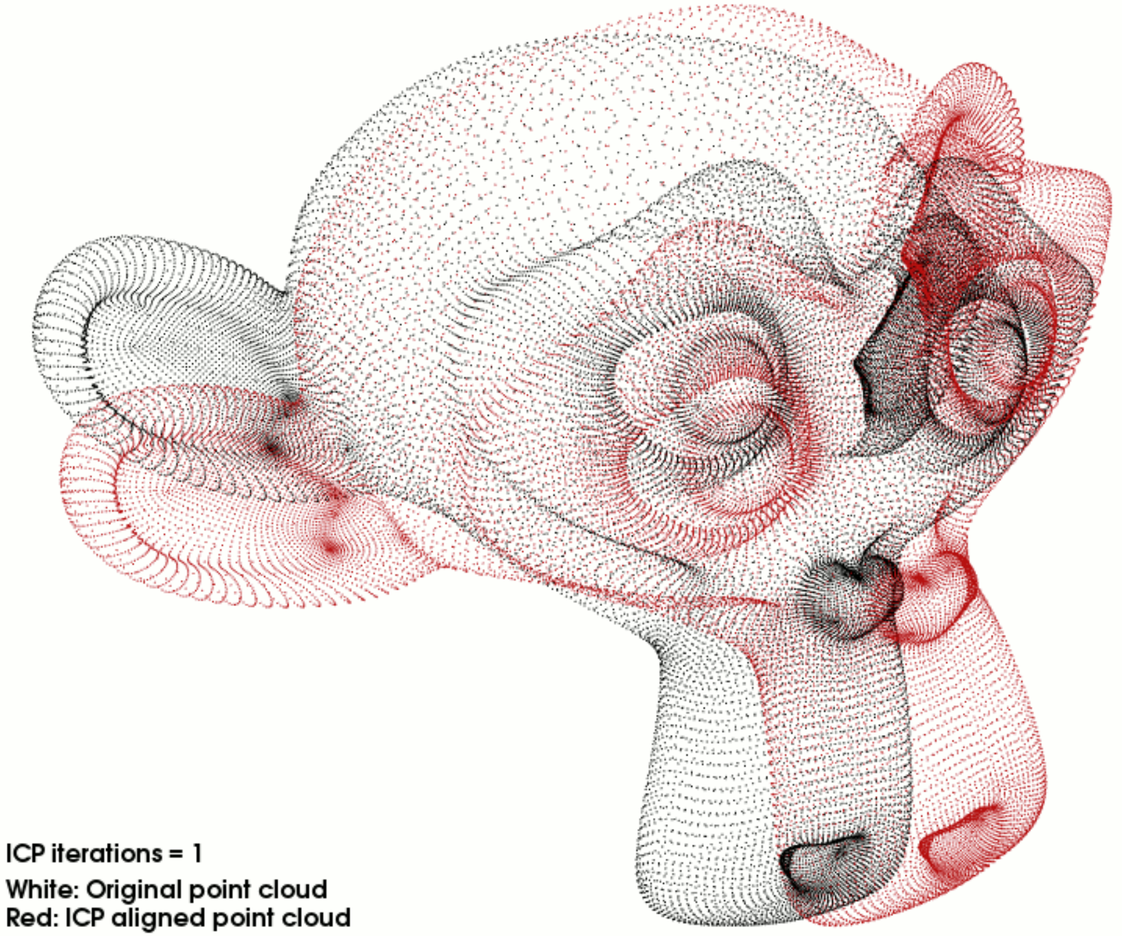


https://github.com/yassram/iterative-closest-point

Besl, Paul J., and Neil D. McKay. "Method for registration of 3-D shapes." *Sensor fusion IV: control paradigms and data structures*. Vol. 1611. Spie, 1992.

Slide credit: Christian Rupprecht, Oxford

1992: Iterative Closest Points



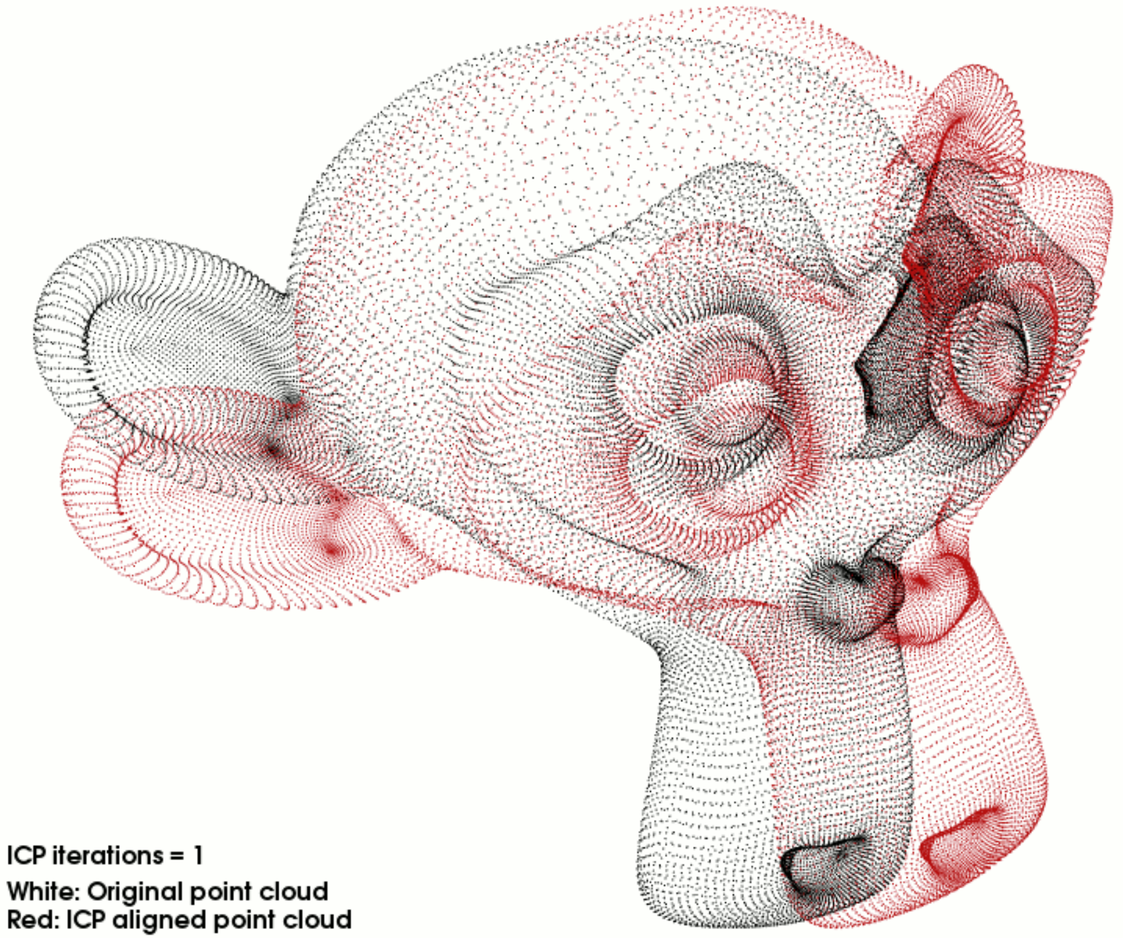
https://github.com/yassram/iterative-closest-point

Besl, Paul J., and Neil D. McKay. "Method for registration of 3-D shapes." *Sensor fusion IV: control paradigms and data structures*. Vol. 1611. Spie, 1992.

Slide credit: Christian Rupprecht, Oxford

1992: Iterative Closest Points

- Register two point clouds by minimizi
  points



https://github.com/yassram/iterative-closest-point

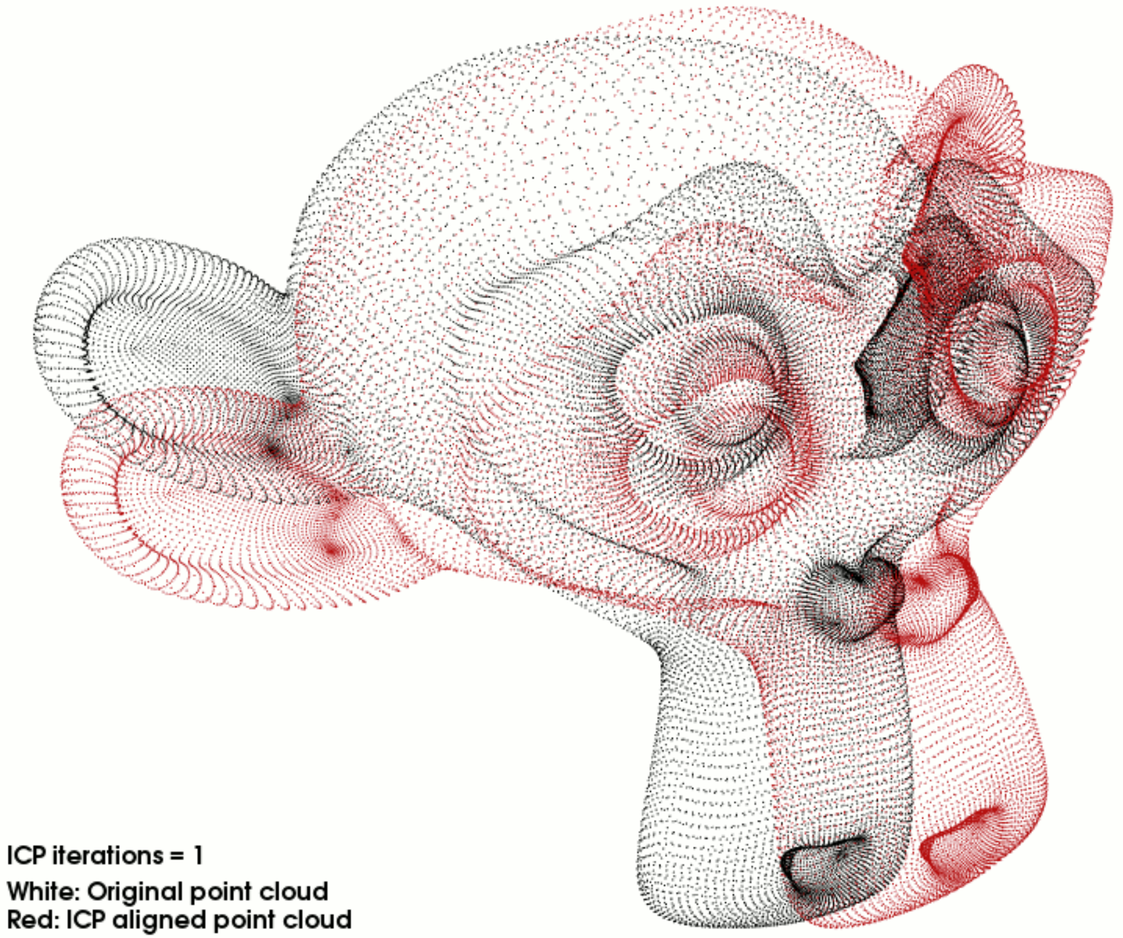
Besl, Paul J., and Neil D. McKay. "Method for registration of 3-D shapes." *Sensor fusion IV: control paradigms and data structures*. Vol. 1611. Spie, 1992.

Slide credit: Christian Rupprecht, Oxford

1992: Iterative Closest Points

- Register two point clouds by minimizi
  points

Uses:



https://github.com/yassram/iterative-closest-point

Besl, Paul J., and Neil D. McKay. "Method for registration of 3-D shapes." *Sensor fusion IV: control paradigms and data structures*. Vol. 1611. Spie, 1992.

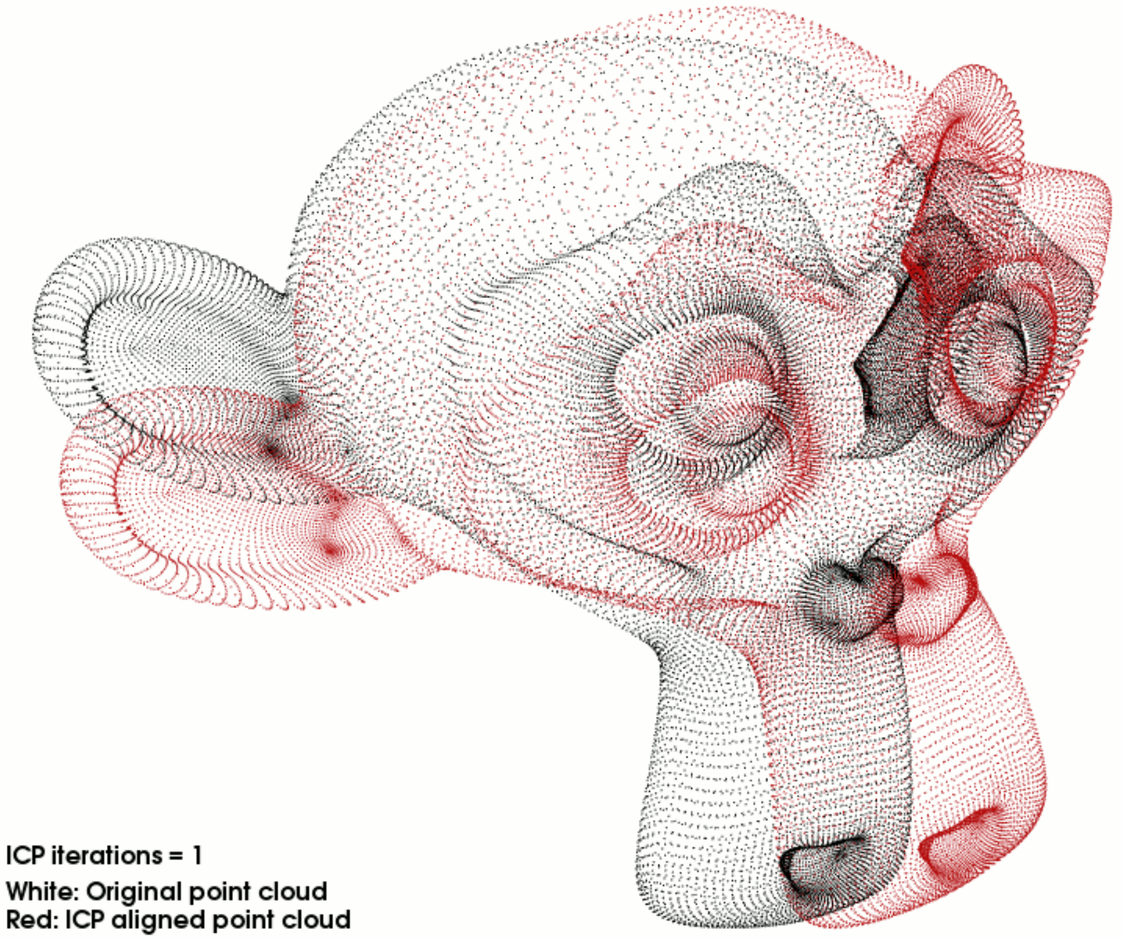
Slide credit: Christian Rupprecht, Oxford

1992: Iterative Closest Points

- Register two point clouds by minimizi
  points

Uses:

- Align partial scans



https://github.com/yassram/iterative-closest-point

Besl, Paul J., and Neil D. McKay. "Method for registration of 3-D shapes." *Sensor fusion IV: control paradigms and data structures*. Vol. 1611. Spie, 1992.

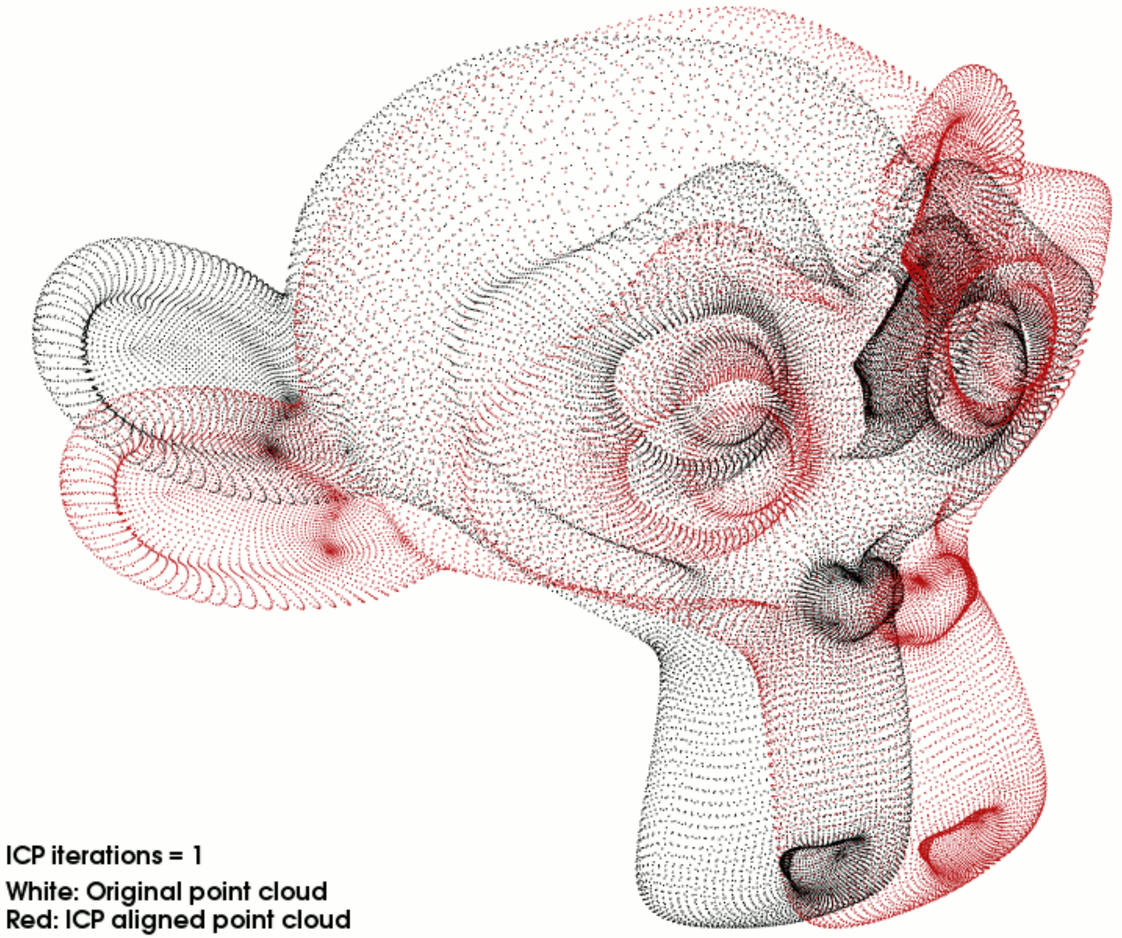
Slide credit: Christian Rupprecht, Oxford

1992: Iterative Closest Points

- Register two point clouds by minimizi
  points

Uses:

- Align partial scans
- Estimate relative camera poses from



https://github.com/yassram/iterative-closest-point

Besl, Paul J., and Neil D. McKay. "Method for registration of 3-D shapes." *Sensor fusion IV: control paradigms and data structures*. Vol. 1611. Spie, 1992.

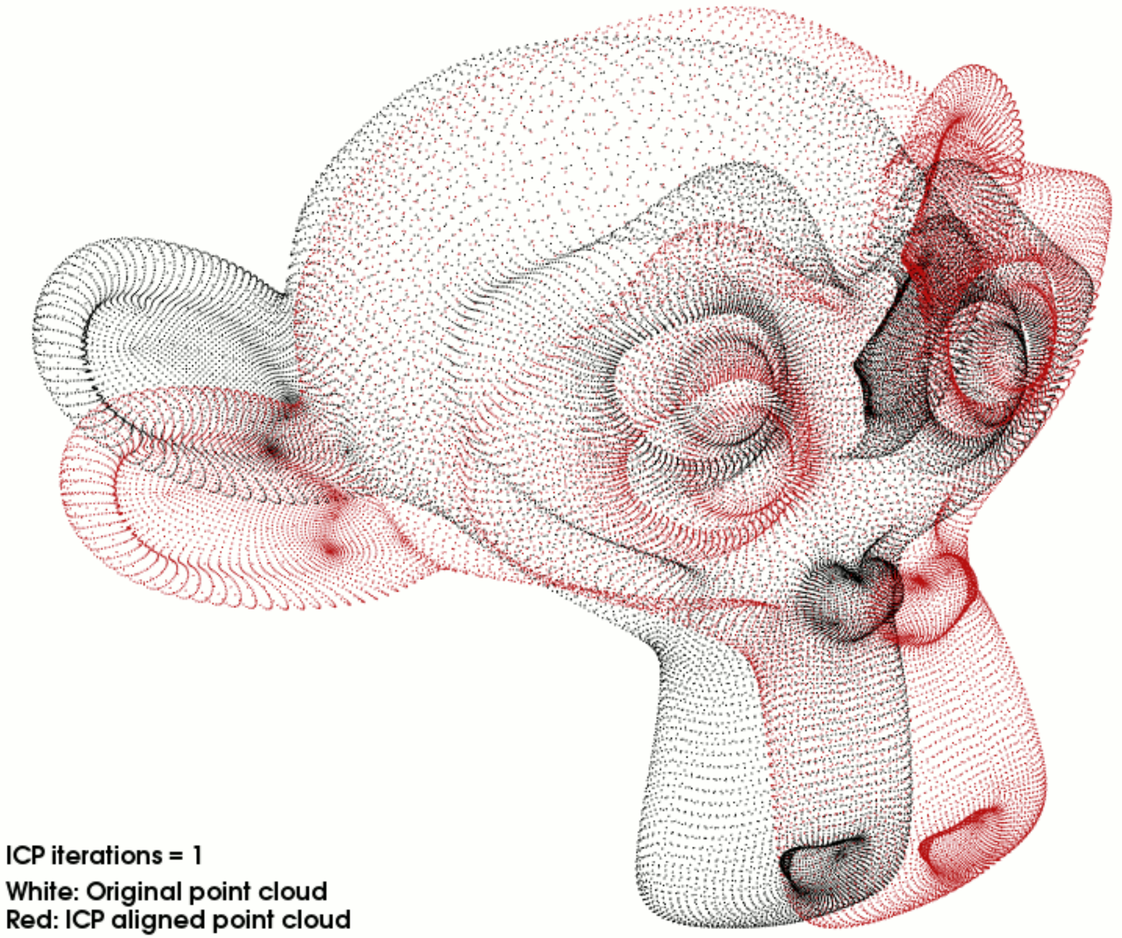
Slide credit: Christian Rupprecht, Oxford

1992: Iterative Closest Points

- Register two point clouds by minimizi
  points

Uses:

- Align partial scans
- Estimate relative camera poses from
- Localization within 3D maps



https://github.com/yassram/iterative-closest-point

Besl, Paul J., and Neil D. McKay. "Method for registration of 3-D shapes." *Sensor fusion IV: control paradigms and data structures*. Vol. 1611. Spie, 1992.

Slide credit: Christian Rupprecht, Oxford

1998: Multi-view stereo

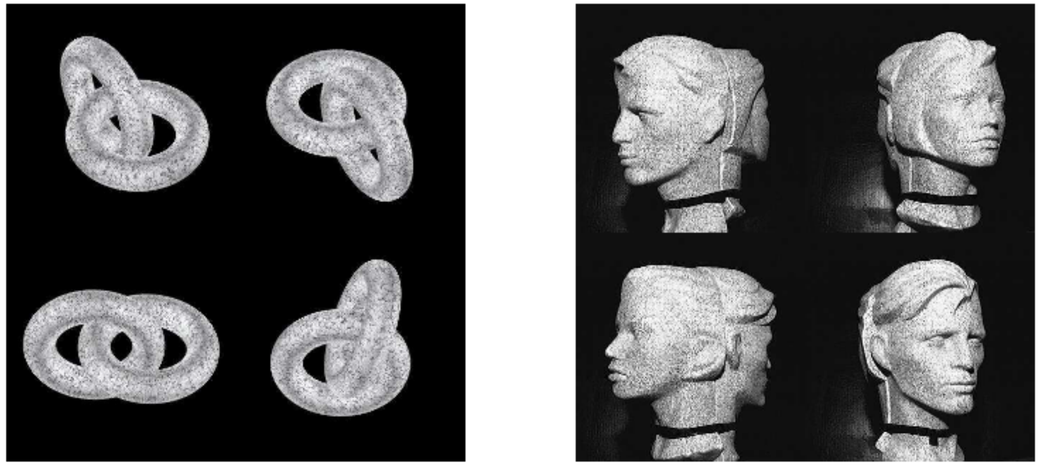


Fig. 3. Multicamera images of 3D objets. On the left hand side, two crossing synthetic tori (24 images). On the right hand side, real images: two human heads (18 images).

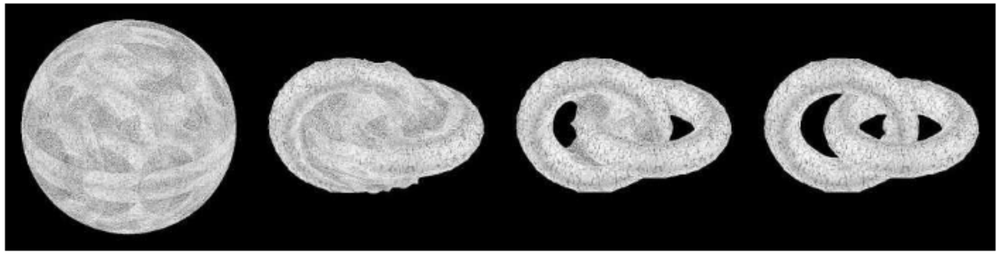


Fig. 4. Evolution of the surface for the two tori.

Faugeras, Olivier, and Renaud Keriven. "Complete dense stereovision using level set methods." *Computer Vision—ECCV'98*

Slide credit: Christian Rupprecht, Oxford

1998: Multi-view stereo

- 3D reconstruction from multiple input images – this time with level-set methods

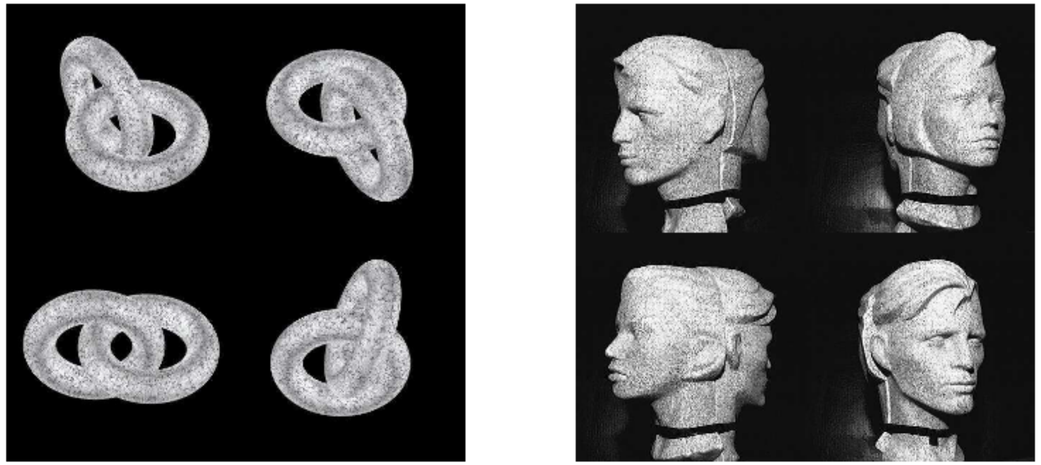


Fig. 3. Multicamera images of 3D objets. On the left hand side, two crossing synthetic tori (24 images). On the right hand side, real images: two human heads (18 images).

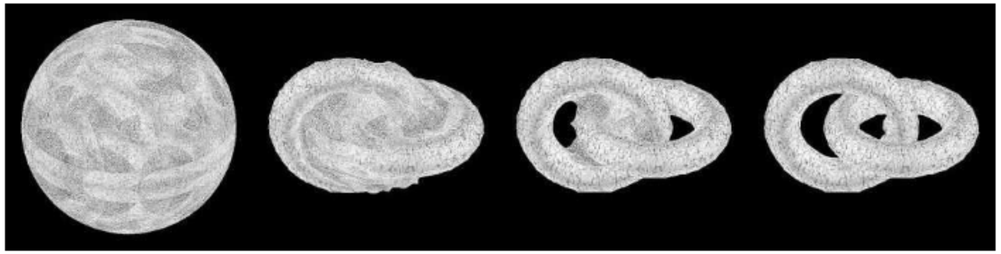


Fig. 4. Evolution of the surface for the two tori.

Faugeras, Olivier, and Renaud Keriven. "Complete dense stereovision using level set methods." *Computer Vision—ECCV'98*

Slide credit: Christian Rupprecht, Oxford

1998: Multi-view stereo

- 3D reconstruction from multiple input images – this time with level-set methods
- Surfaces instead of points

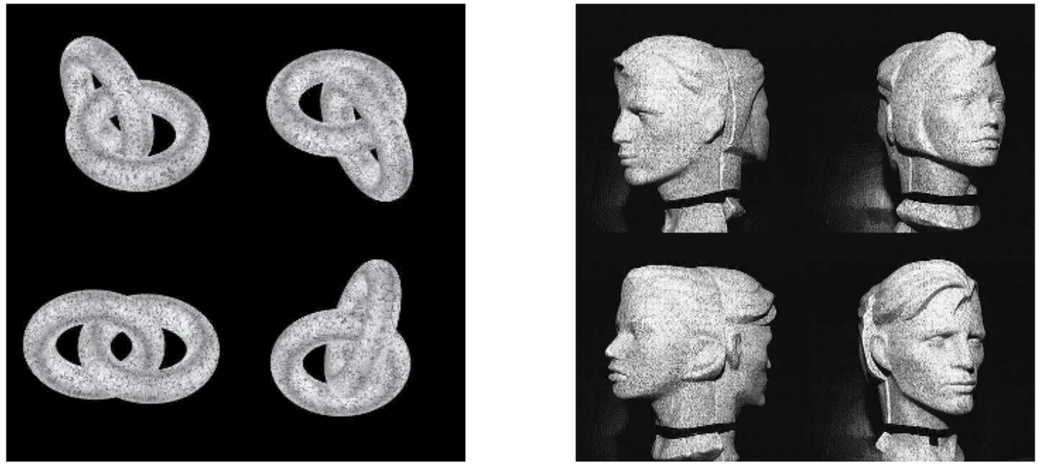


Fig. 3. Multicamera images of 3D objets. On the left hand side, two crossing synthetic tori (24 images). On the right hand side, real images: two human heads (18 images).

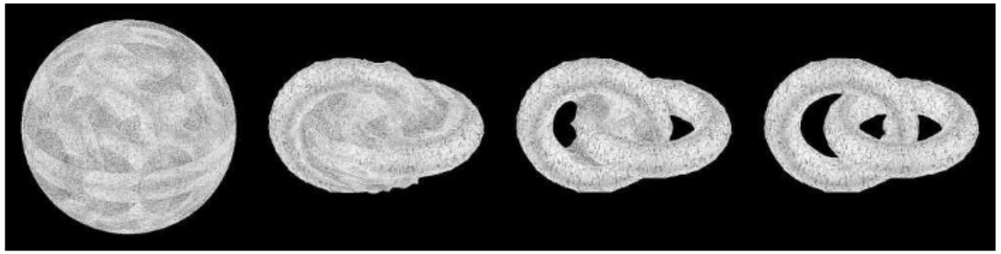


Fig. 4. Evolution of the surface for the two tori.

Faugeras, Olivier, and Renaud Keriven. "Complete dense stereovision using level set methods." *Computer Vision—ECCV'98*

Slide credit: Christian Rupprecht, Oxford

1998: Multi-view stereo

- 3D reconstruction from multiple input images – this time with level-set methods
- Surfaces instead of points
- Modelling visibility

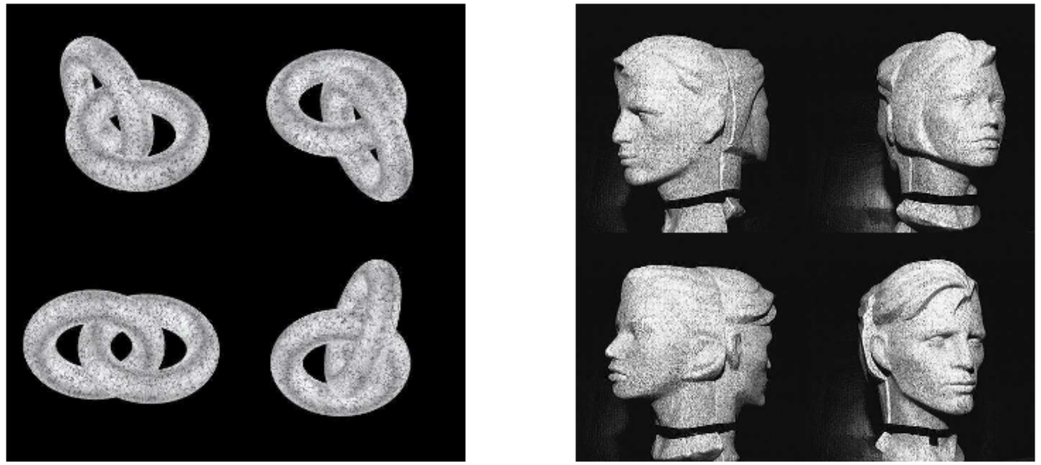


Fig. 3. Multicamera images of 3D objets. On the left hand side, two crossing synthetic tori (24 images). On the right hand side, real images: two human heads (18 images).

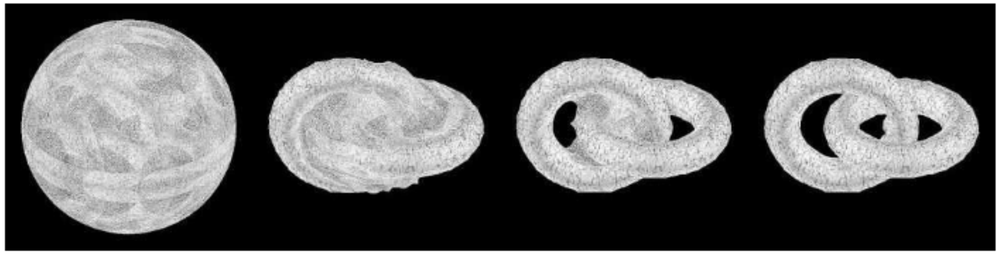


Fig. 4. Evolution of the surface for the two tori.

Faugeras, Olivier, and Renaud Keriven. "Complete dense stereovision using level set methods." *Computer Vision—ECCV'98*

Slide credit: Christian Rupprecht, Oxford

1998: Multi-view stereo

- 3D reconstruction from multiple input images – this time with level-set methods
- Surfaces instead of points

- Modelling visibility
- Convergence proofs

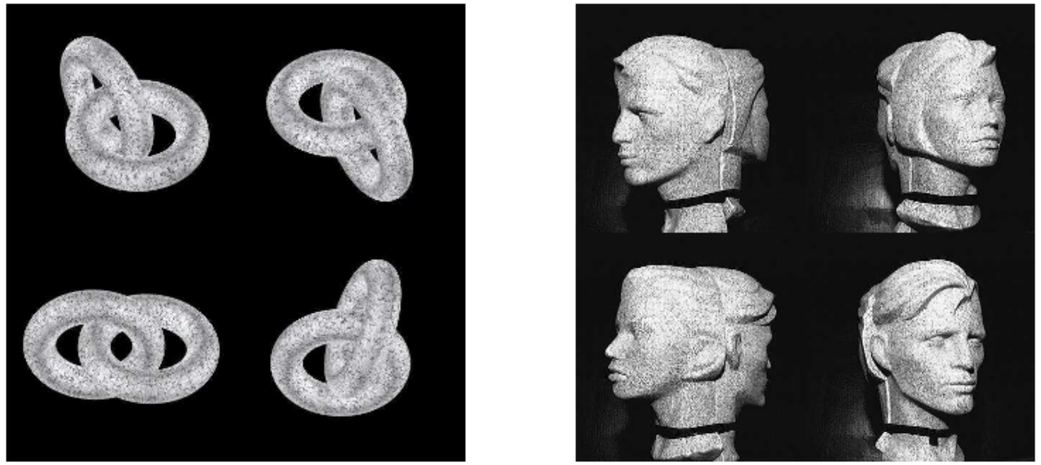


Fig. 3. Multicamera images of 3D objets. On the left hand side, two crossing synthetic tori (24 images). On the right hand side, real images: two human heads (18 images).

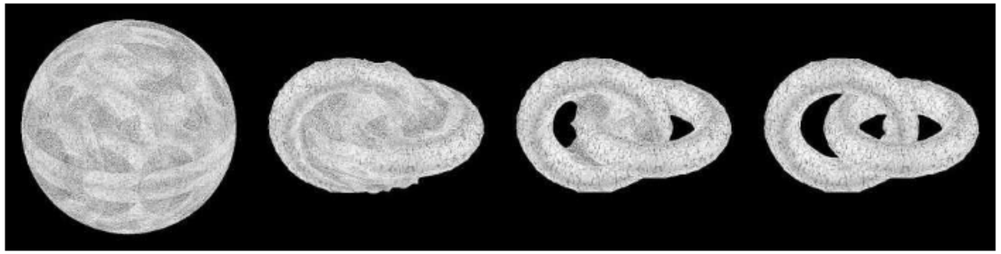


Fig. 4. Evolution of the surface for the two tori.

Faugeras, Olivier, and Renaud Keriven. "Complete dense stereovision using level set methods." *Computer Vision—ECCV'98*

Slide credit: Christian Rupprecht, Oxford

2000s: Large-scale SfM

- 2006: Photo Tourism (Snavely et al,, SIGGAPH'06)
  - 3D reconstruction from internet images
  - Large scale compute
- 2009: Building Rome in a Day (Agarwal et al. ICCV'09)
  - Search "rome" on flickr
  - Reconstruction: 150k images, 21h, 500CPUs

Slide credit: Christian Rupprecht, Oxford

2000s: Large-scale SfM

- 2006: Photo Tourism (Snavely et al,, SIGGAPH'06)
  - 3D reconstruction from internet images
  - Large scale compute
- 2009: Building Rome in a Day (Agarwal et al. ICCV'09)
  - Search "rome" on flickr
  - Reconstruction: 150k images, 21h, 500CPUs

Slide credit: Christian Rupprecht, Oxford

2000s: Large-scale SfM

- 2006: Photo Tourism (Snavely et al,, SIGGAPH'06)
  - 3D reconstruction from internet images
  - Large scale compute
- 2009: Building Rome in a Day (Agarwal et al. ICCV'09)
  - Search "rome" on flickr
  - Reconstruction: 150k images, 21h, 500CPUs

Slide credit: Christian Rupprecht, Oxford

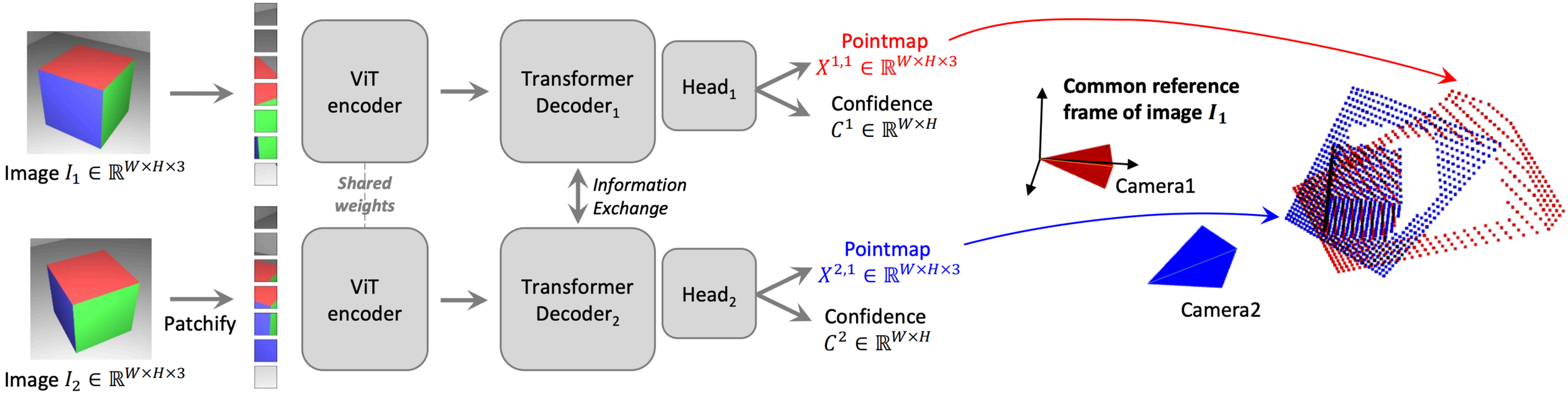
2016: COLMAP

- Open-source C++ framework
- Integrating the best features from prior work
- Defacto standard for SfM

Sparse model of central Rome using 21K photos produced by COLMAP's SfM pipeline

Schonberger, Johannes L., and Jan-Michael Frahm. "Structure-from-motion revisited." CVPR 2016.
Slide credit: Christian Rupprecht, Oxford

2024: DUST3R: 3D Reconstruction as Supervised Learning



View Synthesis

1850: Photosculpture

Slide credit: Christian Rupprecht, Oxford

1850: Photosculpture

- 24 photographs of an object/person

Slide credit: Christian Rupprecht, Oxford

1850: Photosculpture

- 24 photographs of an object/person
- Cut contour from wood

Slide credit: Christian Rupprecht, Oxford

1850: Photosculpture

- 24 photographs of an object/person
- Cut contour from wood
- Assemble radial sculpture

Slide credit: Christian Rupprecht, Oxford

1850: Photosculpture

- 24 photographs of an object/person
- Cut contour from wood
- Assemble radial sculpture

Slide credit: Christian Rupprecht, Oxford

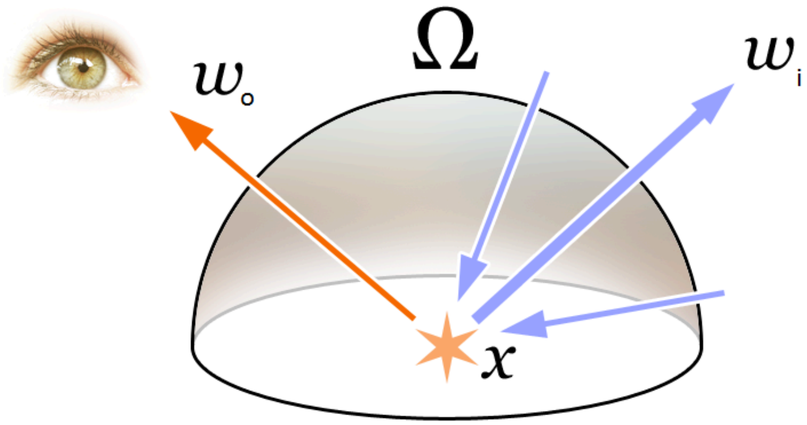
1986: The Rendering Equation

How much light (of wavelength λ) is leaving a point x in the direction of ω_o at time t?

Immel, David S.; Cohen, Michael F.; Greenberg, Donald P. "*A radiosity method for non-diffuse environments”*, SIGGRAPH 1986
Kajiya, James T.*"The rendering equation".* Conference on Computer graphics and interactive techniques 1986
Slide credit: Christian Rupprecht, Oxford

1986: The Rendering Equation

How much light (of wavelength λ) is leaving a point x in the direction of ω_o at time t?

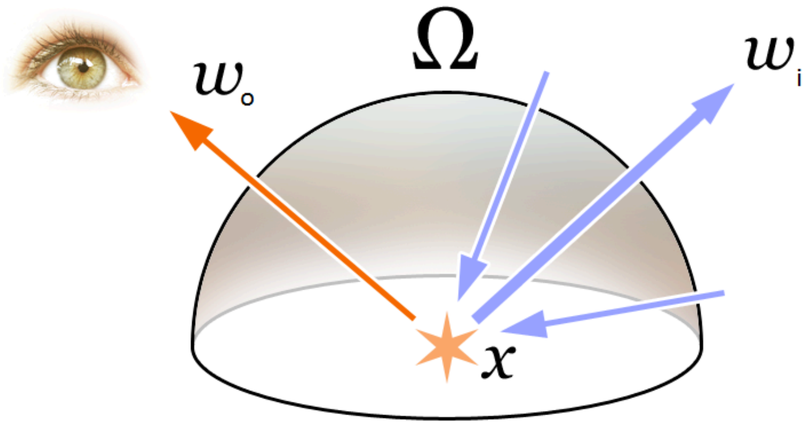


Immel, David S.; Cohen, Michael F.; Greenberg, Donald P. "*A radiosity method for non-diffuse environments*", SIGGRAPH 1986
Kajiya, James T.*"The rendering equation".* Conference on Computer graphics and interactive techniques 1986
Slide credit: Christian Rupprecht, Oxford

1986: The Rendering Equation

How much light (of wavelength λ) is leaving a point x in the direction of ω_o at time t?

$$L_o(x,\ \omega_o, \lambda, t) = L_e(x, \omega_o, \lambda, t) + L_r(x, \omega_o, \lambda, t)$$

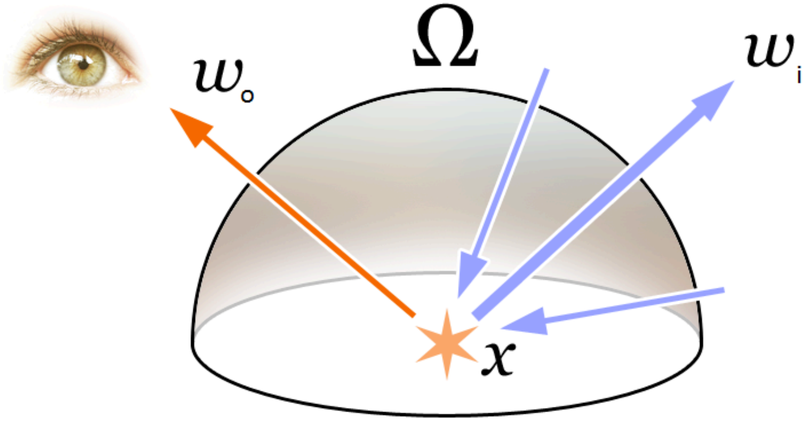


Immel, David S.; Cohen, Michael F.; Greenberg, Donald P. "*A radiosity method for non-diffuse environments*", SIGGRAPH 1986
Kajiya, James T.*"The rendering equation".* Conference on Computer graphics and interactive techniques 1986
Slide credit: Christian Rupprecht, Oxford

1986: The Rendering Equation

How much light (of wavelength λ) is leaving a point x in the direction of ω_o at time t?

$$L_o(x,\ \omega_o, \lambda, t) = \underbrace{L_e(x, \omega_o, \lambda, t)}_{\text{emitted radiance (glowing things)}} + L_r(x, \omega_o, \lambda, t)$$

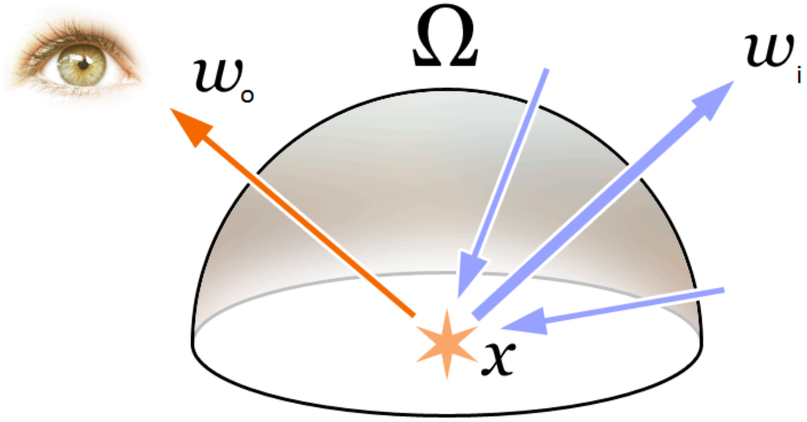


Immel, David S.; Cohen, Michael F.; Greenberg, Donald P. "*A radiosity method for non-diffuse environments*", SIGGRAPH 1986
Kajiya, James T.*"The rendering equation".* Conference on Computer graphics and interactive techniques 1986
Slide credit: Christian Rupprecht, Oxford

1986: The Rendering Equation

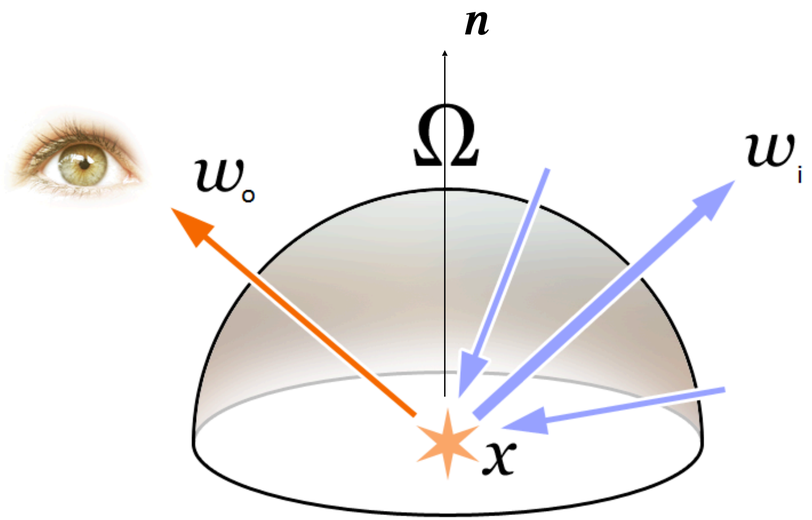
How much light (of wavelength λ) is leaving a point x in the direction of ω_o at time t?

$$L_o(x,\ \omega_o, \lambda, t) = \underbrace{L_e(x, \omega_o, \lambda, t)}_{\text{emitted radiance (glowing things)}} + \underbrace{L_r(x, \omega_o, \lambda, t)}_{\text{reflected radiance}}$$



Immel, David S.; Cohen, Michael F.; Greenberg, Donald P. "*A radiosity method for non-diffuse environments*", SIGGRAPH 1986
Kajiya, James T.*"The rendering equation".* Conference on Computer graphics and interactive techniques 1986
Slide credit: Christian Rupprecht, Oxford

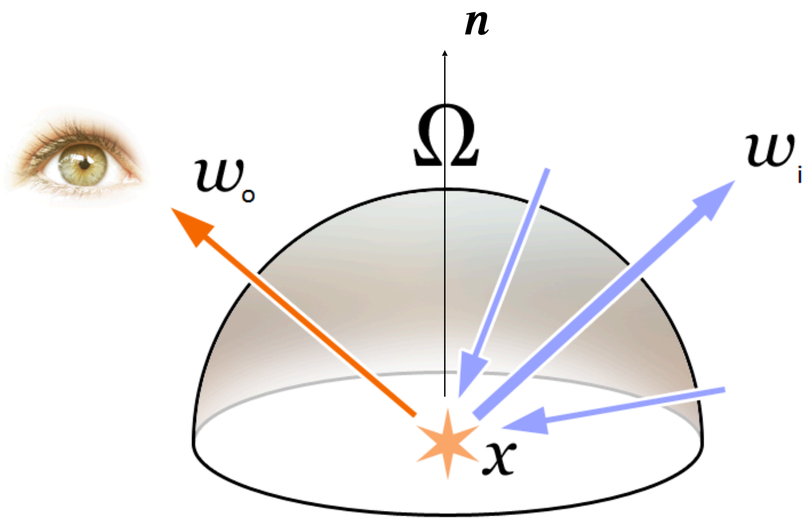
1986: The Rendering Equation



Slide credit: Christian Rupprecht, Oxford

1986: The Rendering Equation

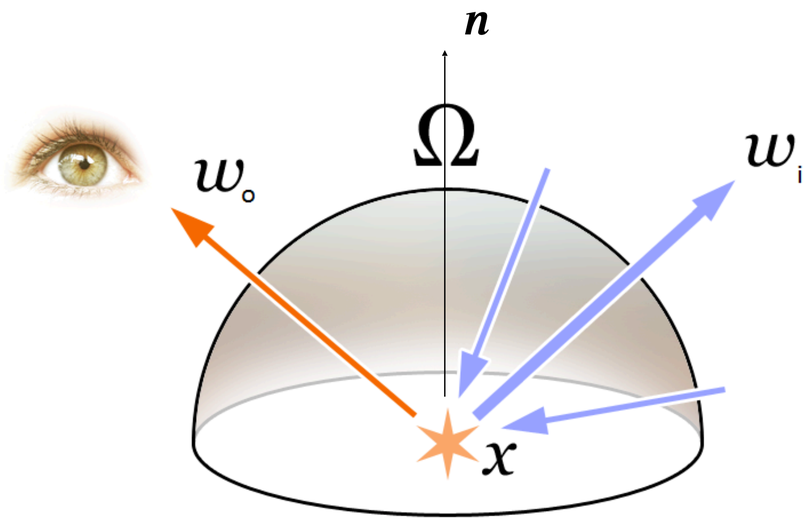
$$L_r(x,\ \omega_o, \lambda, t) = \int_{\Omega} f_i(x,\ \omega_i,\ \omega_o, \lambda, t) L_i(x, \omega_i, \lambda, t)(\omega_i \cdot \boldsymbol{n}) d\omega_i$$



Slide credit: Christian Rupprecht, Oxford

1986: The Rendering Equation

$$L_r(x,\ \omega_o, \lambda, t) = \int_{\Omega} \underbrace{f_i(x,\ \omega_i,\ \omega_o, \lambda, t)}_{\text{bidirectional reflectance distribution function (BRDF)}} L_i(x, \omega_i, \lambda, t)(\omega_i \cdot \boldsymbol{n}) d\omega_i$$



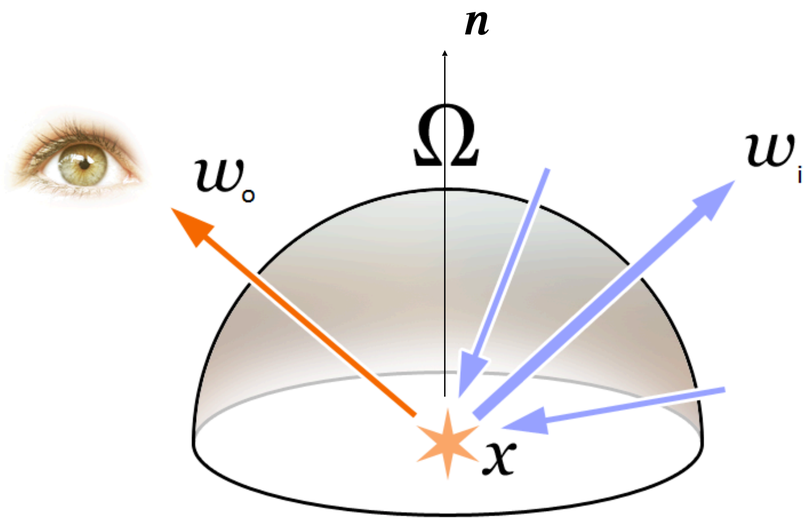
Slide credit: Christian Rupprecht, Oxford

1986: The Rendering Equation

incoming radiance at x from direction ω_i

$$L_r(x,\ \omega_o, \lambda, t) = \int_{\Omega} f_i(x,\ \omega_i,\ \omega_o, \lambda, t) L_i(x, \omega_i, \lambda, t)(\omega_i \cdot \boldsymbol{n}) d\omega_i$$

bidirectional reflectance distribution function (BRDF)



Slide credit: Christian Rupprecht, Oxford

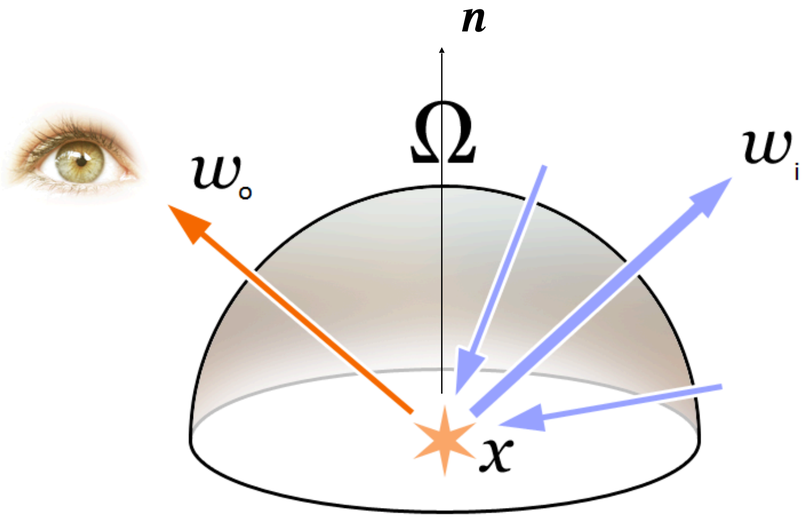
1986: The Rendering Equation

$$L_r(x,\ \omega_o, \lambda, t) = \int_{\Omega} f_i(x,\ \omega_i,\ \omega_o, \lambda, t) L_i(x, \omega_i, \lambda, t) (\omega_i \cdot \boldsymbol{n}) d\omega_i$$

incoming radiance at x from direction ω_i

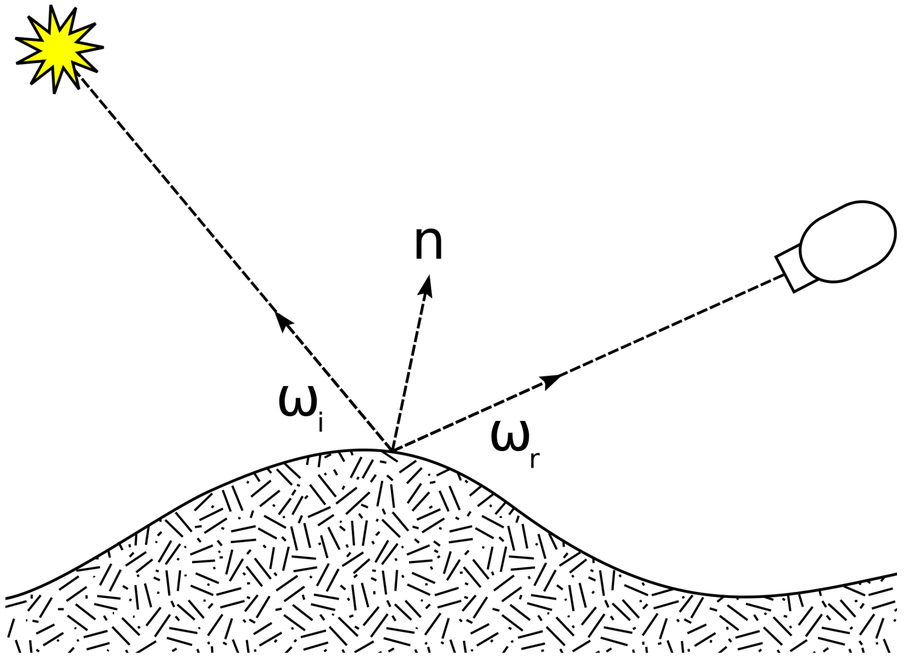
bidirectional reflectance distribution function (BRDF)

surface normal



Slide credit: Christian Rupprecht, Oxford

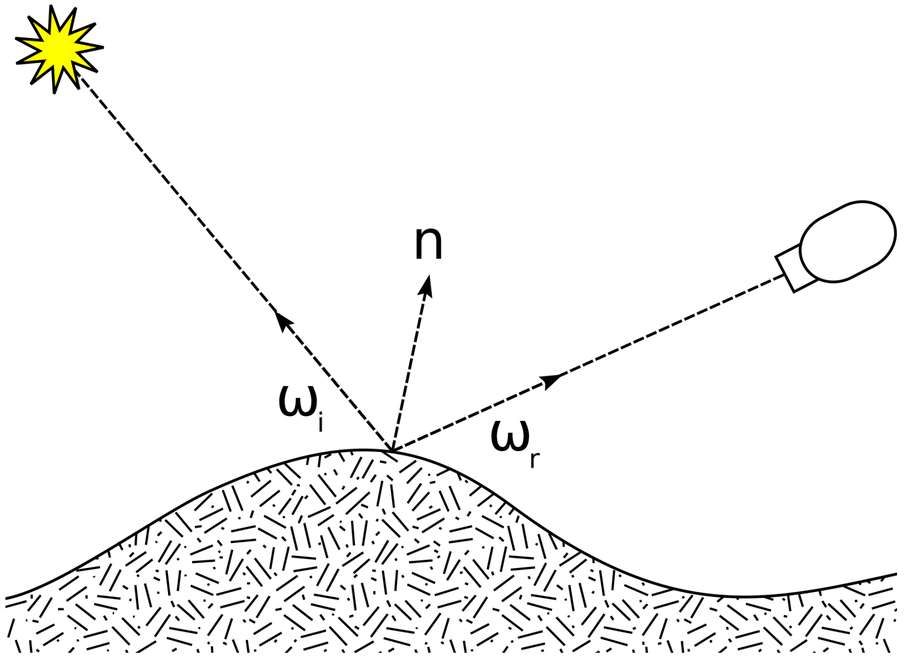
1965: The BRDF



Nicodemus, Fred (1965). "*Directional reflectance and emissivity of an opaque surface*". Applied Optics
Slide credit: Christian Rupprecht, Oxford

1965: The BRDF

$$f_r(\omega_i, \omega_r) = \frac{dL_r(\omega_r)}{L_i(\omega_i)\ (\omega_i \cdot n)\ d\omega_i}$$



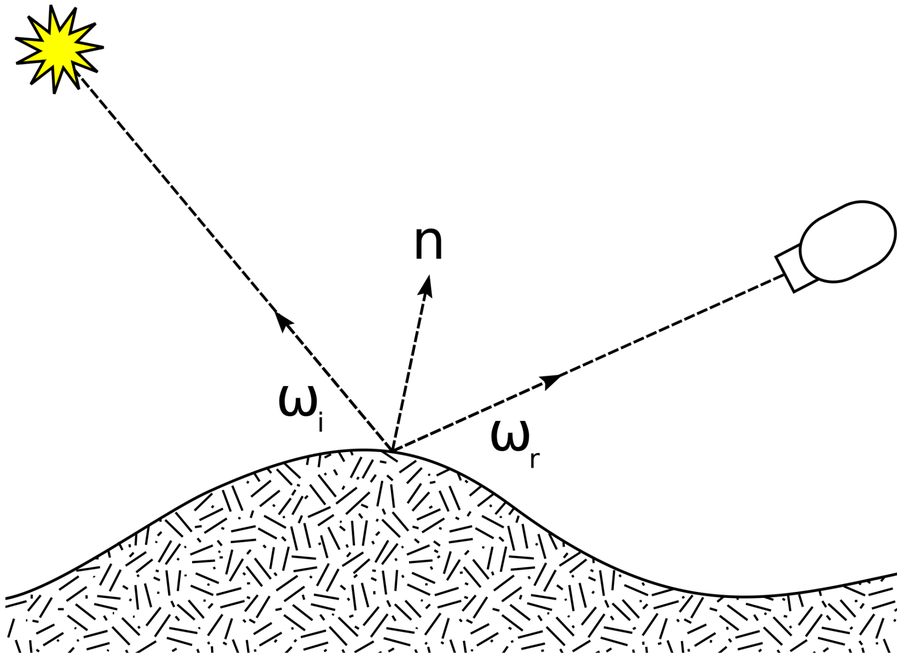
Nicodemus, Fred (1965). "*Directional reflectance and emissivity of an opaque surface*". Applied Optics

Slide credit: Christian Rupprecht, Oxford

1965: The BRDF

$$f_r(\omega_i, \omega_r) = \frac{dL_r(\omega_r)}{L_i(\omega_i)\ (\omega_i \cdot n)\ d\omega_i}$$

- Positivity: $f_r(\omega_i, \omega_r) > 0$

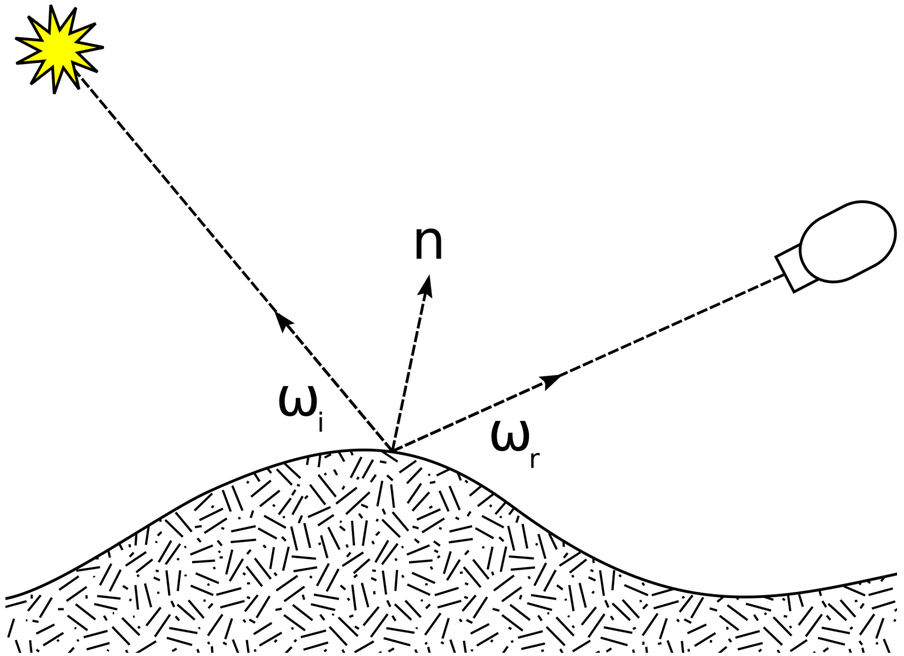


Nicodemus, Fred (1965). "*Directional reflectance and emissivity of an opaque surface*". Applied Optics
Slide credit: Christian Rupprecht, Oxford

1965: The BRDF

$$f_r(\omega_i, \omega_r) = \frac{dL_r(\omega_r)}{L_i(\omega_i)\ (\omega_i \cdot n)\ d\omega_i}$$

- Positivity: $f_r(\omega_i, \omega_r) > 0$
- Reciprocity: $f_r(\omega_i, \omega_r) = f_r(\omega_r, \omega_i)$



Nicodemus, Fred (1965). "*Directional reflectance and emissivity of an opaque surface*". Applied Optics

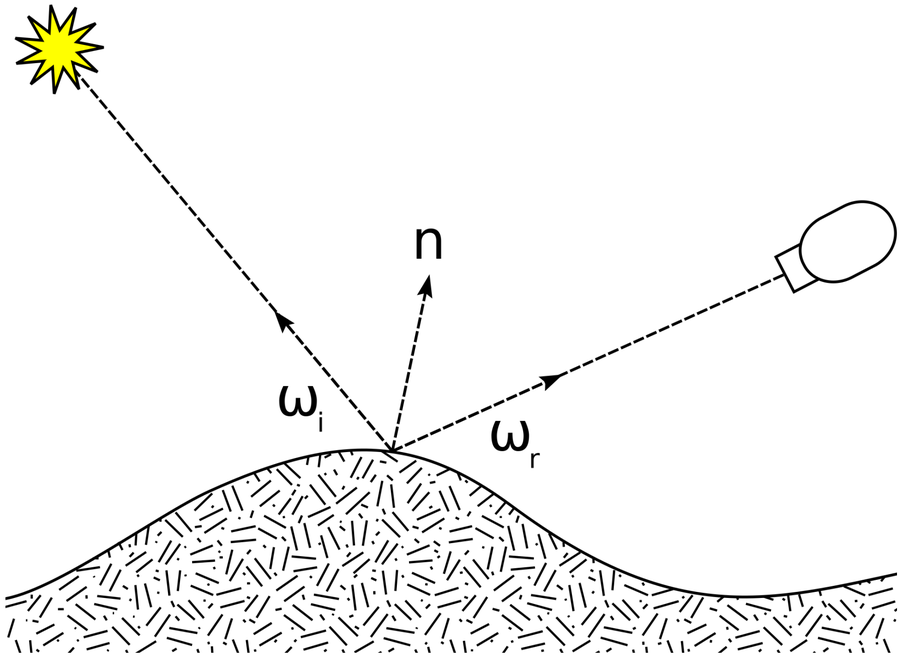
Slide credit: Christian Rupprecht, Oxford

1965: The BRDF

$$f_r(\omega_i, \omega_r) = \frac{dL_r(\omega_r)}{L_i(\omega_i)\ (\omega_i \cdot n)\ d\omega_i}$$

- Positivity: $f_r(\omega_i, \omega_r) > 0$
- Reciprocity: $f_r(\omega_i, \omega_r) = f_r(\omega_r, \omega_i)$
- Energy conservation:

$$\forall \omega_i, \int_\Omega f_r(\omega_i, \omega_r)(\omega_r \cdot n) d\omega_r \leq 1$$



Nicodemus, Fred (1965). "*Directional reflectance and emissivity of an opaque surface*". Applied Optics

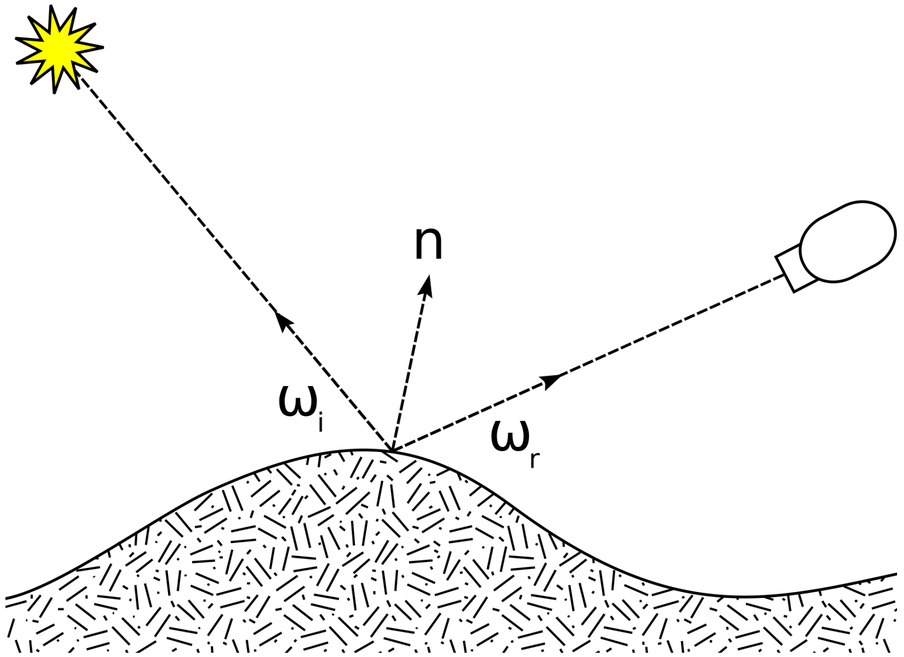
Slide credit: Christian Rupprecht, Oxford

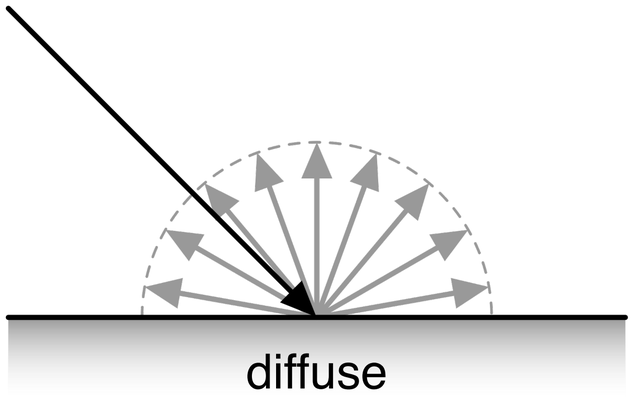
1965: The BRDF

$$f_r(\omega_i, \omega_r) = \frac{dL_r(\omega_r)}{L_i(\omega_i)\ (\omega_i \cdot n)\ d\omega_i}$$

- Positivity: $f_r(\omega_i, \omega_r) > 0$
- Reciprocity: $f_r(\omega_i, \omega_r) = f_r(\omega_r, \omega_i)$
- Energy conservation:

$$\forall \omega_i, \int_\Omega f_r(\omega_i, \omega_r)(\omega_r \cdot n) d\omega_r \leq 1$$





Nicodemus, Fred (1965). "*Directional reflectance and emissivity of an opaque surface*". Applied Optics

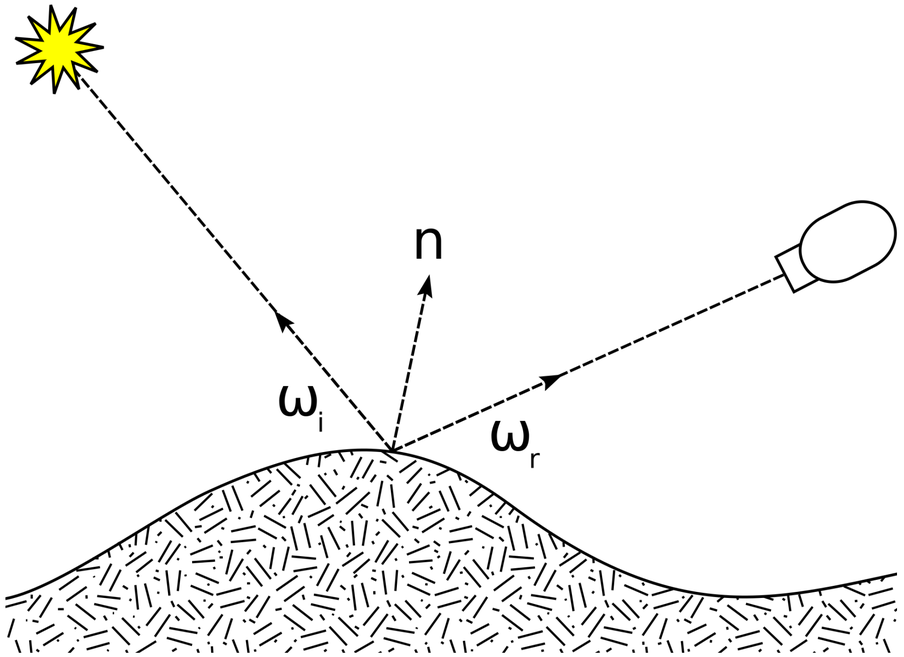
Slide credit: Christian Rupprecht, Oxford

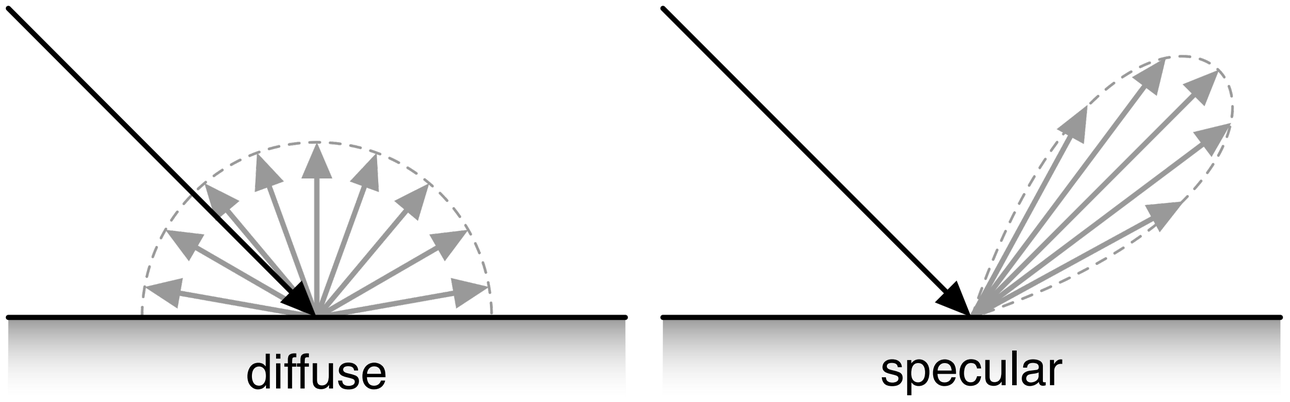
1965: The BRDF

$$f_r(\omega_i, \omega_r) = \frac{dL_r(\omega_r)}{L_i(\omega_i)\ (\omega_i \cdot n)\ d\omega_i}$$

- Positivity: $f_r(\omega_i, \omega_r) > 0$
- Reciprocity: $f_r(\omega_i, \omega_r) = f_r(\omega_r, \omega_i)$
- Energy conservation:

$$\forall \omega_i, \int_\Omega f_r(\omega_i, \omega_r)(\omega_r \cdot n) d\omega_r \leq 1$$





Nicodemus, Fred (1965). "*Directional reflectance and emissivity of an opaque surface*". Applied Optics

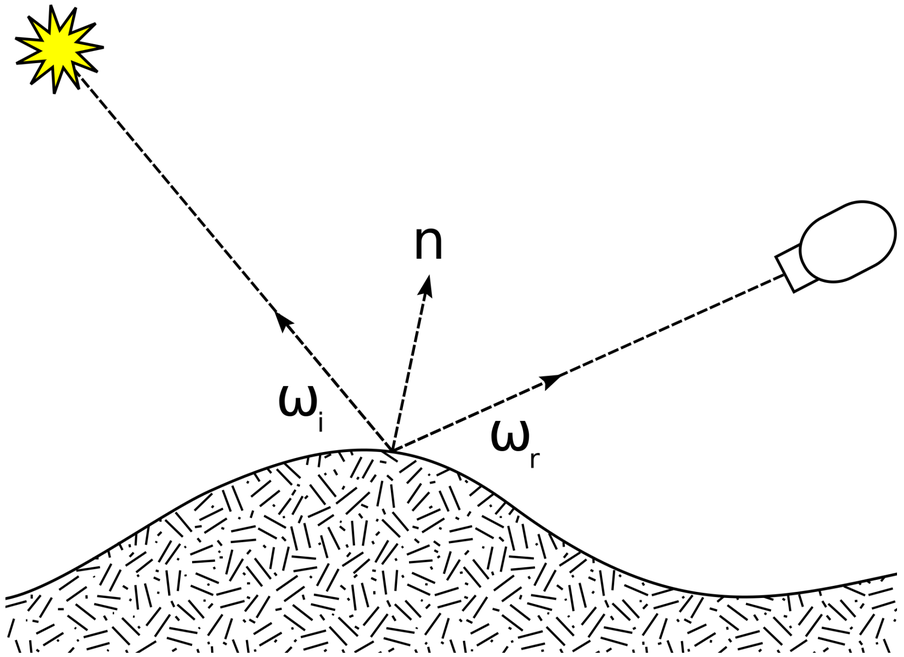
Slide credit: Christian Rupprecht, Oxford

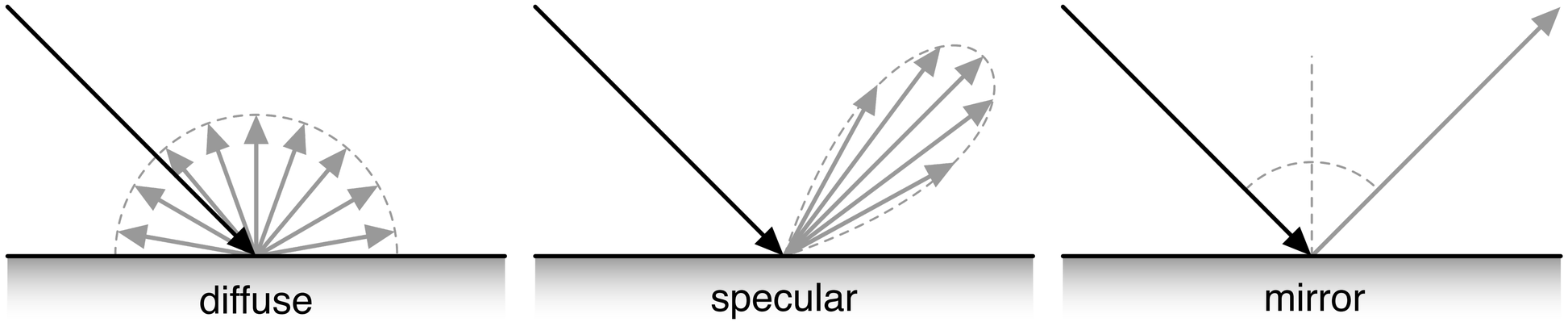
1965: The BRDF

$$f_r(\omega_i, \omega_r) = \frac{dL_r(\omega_r)}{L_i(\omega_i)\ (\omega_i \cdot n)\ d\omega_i}$$

- Positivity: $f_r(\omega_i, \omega_r) > 0$
- Reciprocity: $f_r(\omega_i, \omega_r) = f_r(\omega_r, \omega_i)$
- Energy conservation:

$$\forall \omega_i, \int_\Omega f_r(\omega_i, \omega_r)(\omega_r \cdot n) d\omega_r \leq 1$$





Nicodemus, Fred (1965). "*Directional reflectance and emissivity of an opaque surface*". Applied Optics

Slide credit: Christian Rupprecht, Oxford

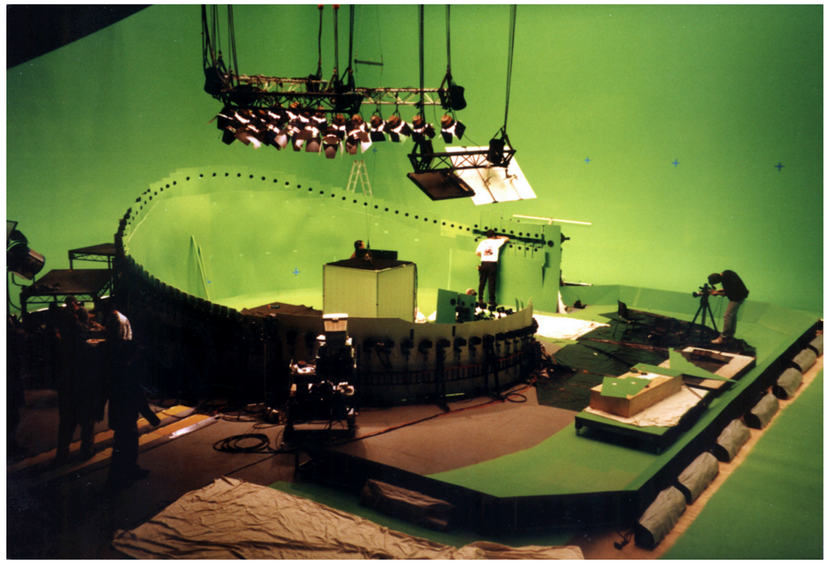
2000s: Lightfield camera arrays

- Use many synchronized cameras to capture a scene from many angle simultaneously
- Film use: The Matrix (1999)

Slide credit: Christian Rupprecht, Oxford

2000s: Lightfield camera arrays

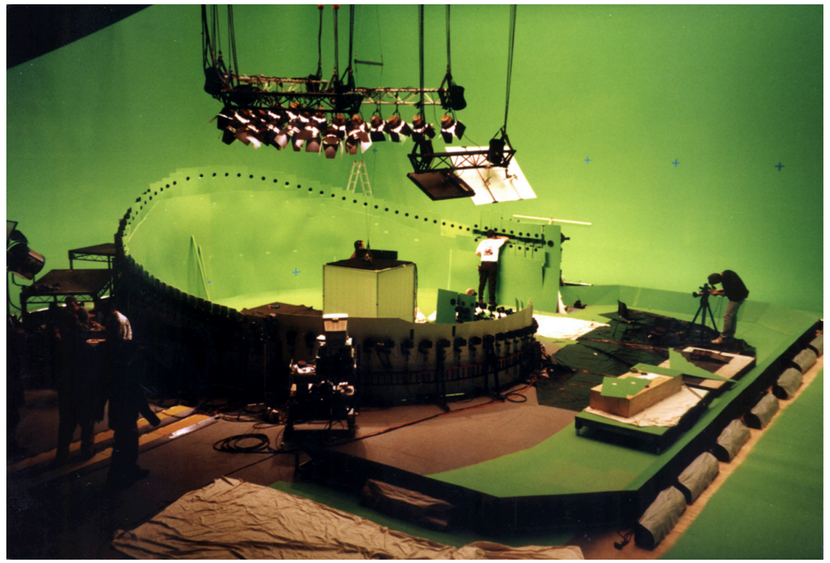
- Use many synchronized cameras to capture a scene from many angle simultaneously
- Film use: The Matrix (1999)



Slide credit: Christian Rupprecht, Oxford

2000s: Lightfield camera arrays

- Use many synchronized cameras to capture a scene from many angle simultaneously
- Film use: The Matrix (1999)



Slide credit: Christian Rupprecht, Oxford

2020: Neural Fields / Neural Implicit Representations

“Implicit Neural Representations with Periodic Activation Functions“, Sitzmann et al., NeurIPS 2020

ReLU

SIREN

2020: Neural Fields / Neural Implicit Representations

“Implicit Neural Representations with Periodic Activation Functions“, Sitzmann et al., NeurIPS 2020

ReLU

SIREN

2020: Neural Fields / Neural Implicit Representations

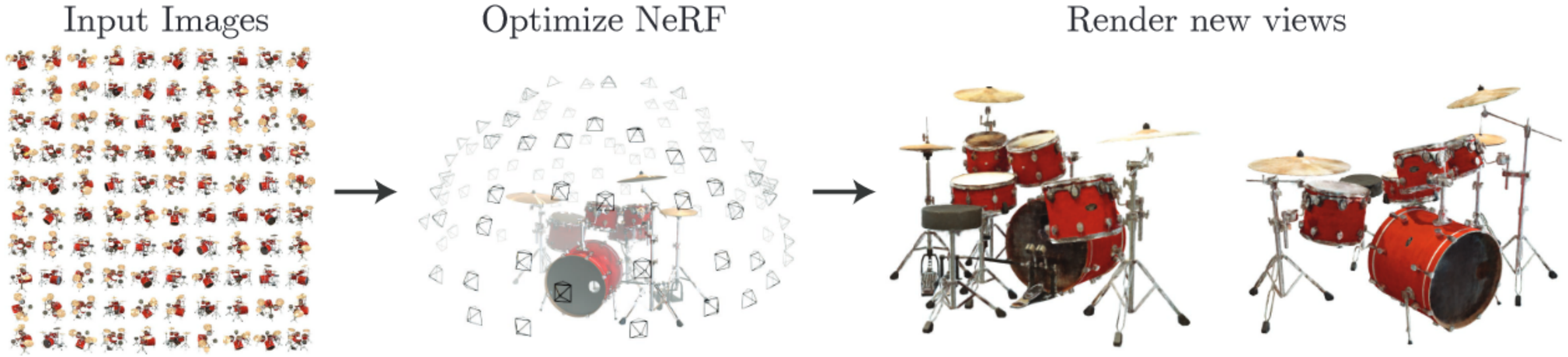
“Implicit Neural Representations with Periodic Activation Functions“, Sitzmann et al., NeurIPS 2020

~1MB compared to
110MB of full mesh!

ReLU

SIREN

2020: Neural Radiance Fields

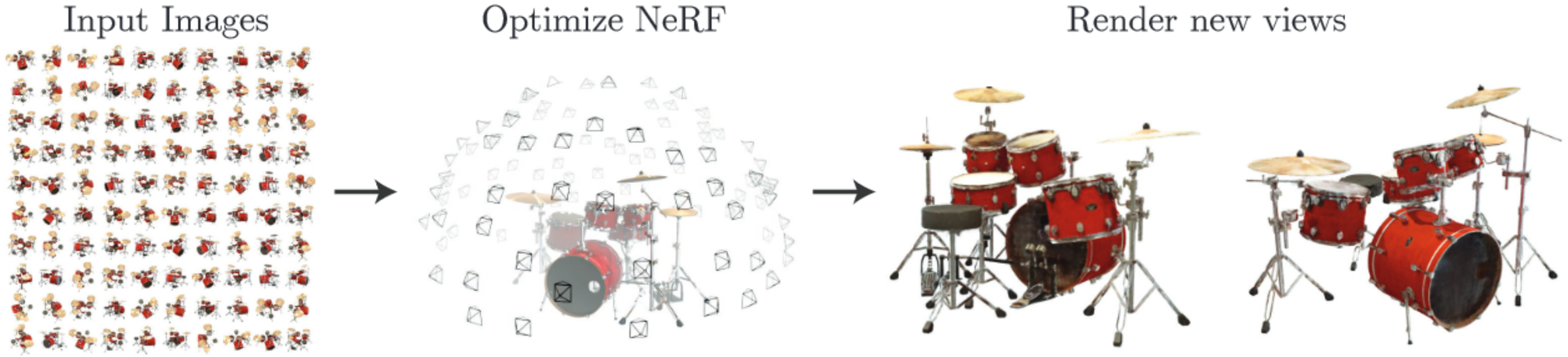


Mildenhall, Ben, et al. "Nerf: Representing scenes as neural radiance fields for view synthesis.“, ECCV 2020

Slide credit: Christian Rupprecht, Oxford

2020: Neural Radiance Fields

- Input: Image collection

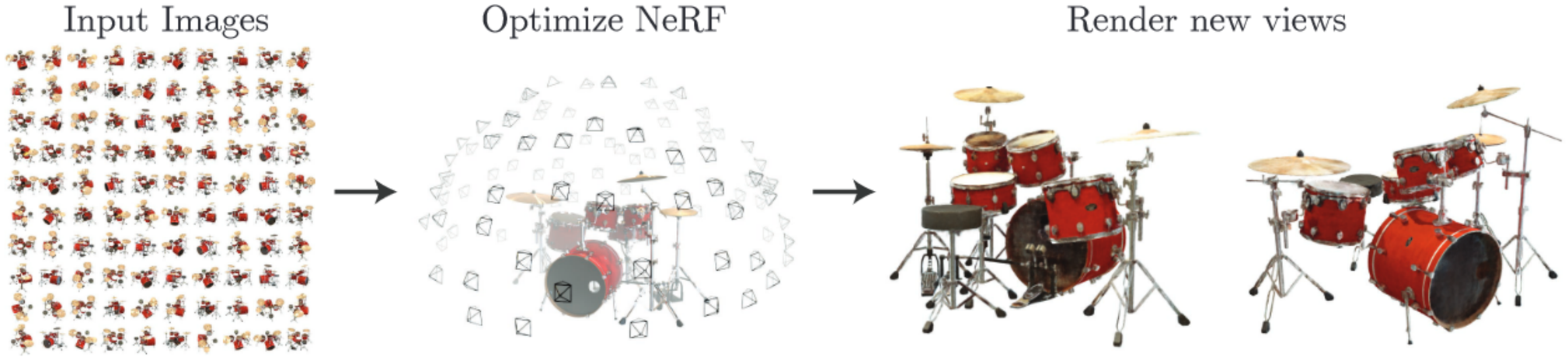


Mildenhall, Ben, et al. "Nerf: Representing scenes as neural radiance fields for view synthesis.“, ECCV 2020

Slide credit: Christian Rupprecht, Oxford

2020: Neural Radiance Fields

- Input: Image collection
- Learning: mapping coordinates (x,y,z) to color and occupancy

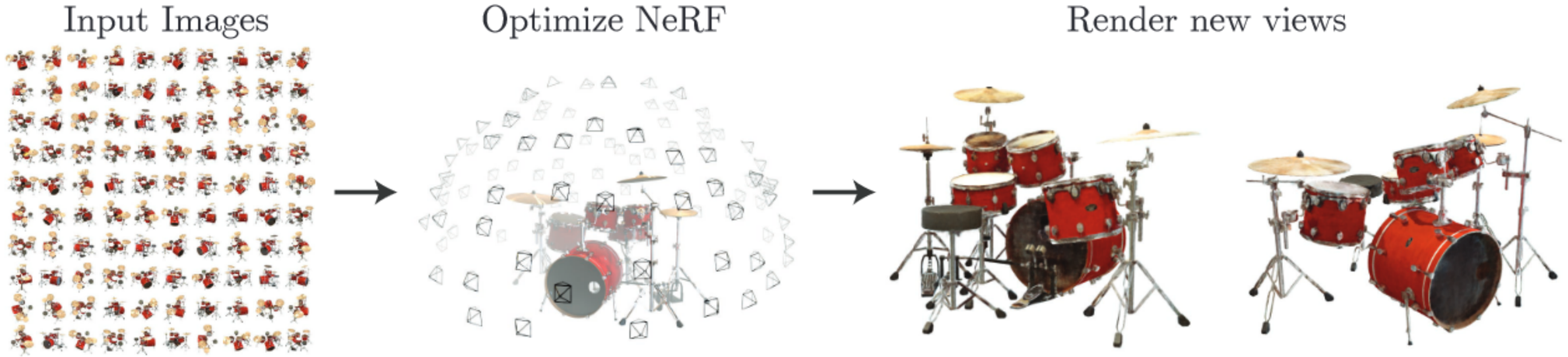


Mildenhall, Ben, et al. "Nerf: Representing scenes as neural radiance fields for view synthesis.“, ECCV 2020

Slide credit: Christian Rupprecht, Oxford

2020: Neural Radiance Fields

- Input: Image collection
- Learning: mapping coordinates (x,y,z) to color and occupancy
- Output: rendering from novel viewpoints

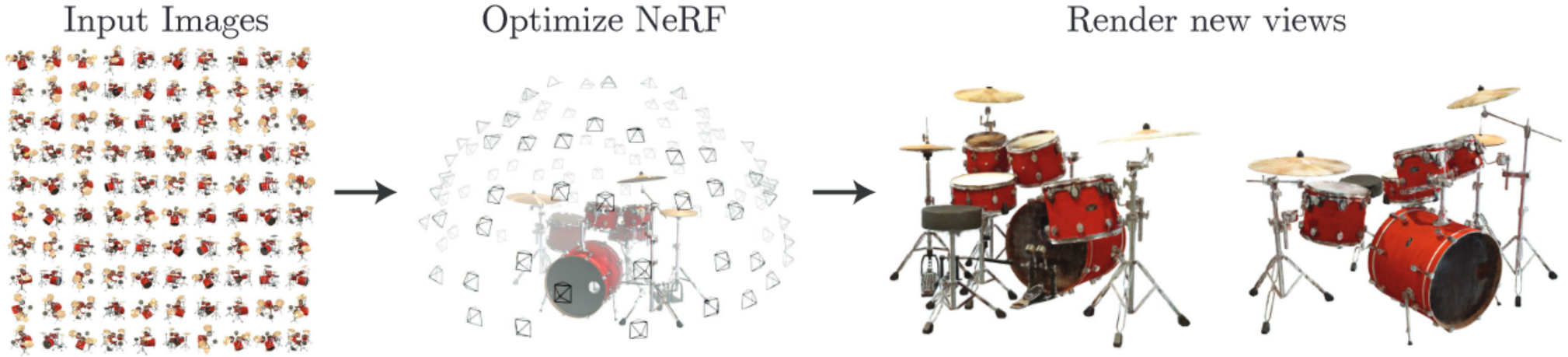


Mildenhall, Ben, et al. "Nerf: Representing scenes as neural radiance fields for view synthesis.“, ECCV 2020

Slide credit: Christian Rupprecht, Oxford

2020: Neural Radiance Fields

- Input: Image collection
- Learning: mapping coordinates (x,y,z) to color and occupancy
- Output: rendering from novel viewpoints



Mildenhall, Ben, et al. "Nerf: Representing scenes as neural radiance fields for view synthesis.“, ECCV 2020

Slide credit: Christian Rupprecht, Oxford

2024: Gaussian Splatting

Kerbl, B., Kopanas, G., Leimkühler, T., & Drettakis, G. (2023). 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics (ToG)*

Slide credit: Christian Rupprecht, Oxford

2024: Gaussian Splatting

- Intuition: ray-casting through every pixel is wasteful (mostly empty space)

Kerbl, B., Kopanas, G., Leimkühler, T., & Drettakis, G. (2023). 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics (ToG)*

Slide credit: Christian Rupprecht, Oxford

2024: Gaussian Splatting

- Intuition: ray-casting through every pixel is wasteful (mostly empty space)
- Represent the scene as a collection of points with size: 3D Gaussians (mean & covariance)

Kerbl, B., Kopanas, G., Leimkühler, T., & Drettakis, G. (2023). 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics (ToG)*

Slide credit: Christian Rupprecht, Oxford

2024: Gaussian Splatting

- Intuition: ray-casting through every pixel is wasteful (mostly empty space)
- Represent the scene as a collection of points with size: 3D Gaussians (mean & covariance)
- Projecting 3D Gaussians to the image plane results in approx. 2D Gaussians (Zwicker et al.,2002)

Kerbl, B., Kopanas, G., Leimkühler, T., & Drettakis, G. (2023). 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics (ToG)*

Slide credit: Christian Rupprecht, Oxford

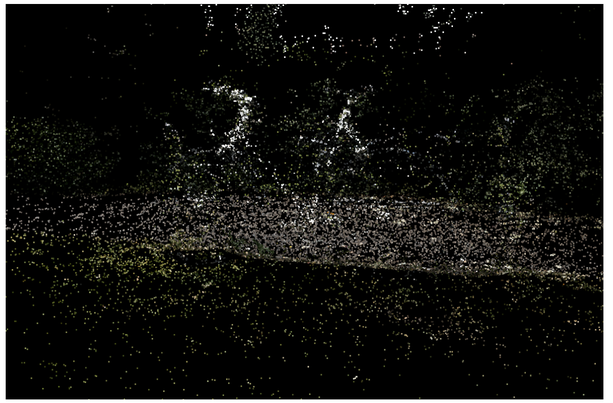
2024: Gaussian Splatting

- Intuition: ray-casting through every pixel is wasteful (mostly empty space)
- Represent the scene as a collection of points with size: 3D Gaussians (mean & covariance)
- Projecting 3D Gaussians to the image plane results in approx. 2D Gaussians (Zwicker et al.,2002)
- Advantage: only spend time/memory on the surface of objects

Kerbl, B., Kopanas, G., Leimkühler, T., & Drettakis, G. (2023). 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics (ToG)*

Slide credit: Christian Rupprecht, Oxford

2024: Gaussian Splatting

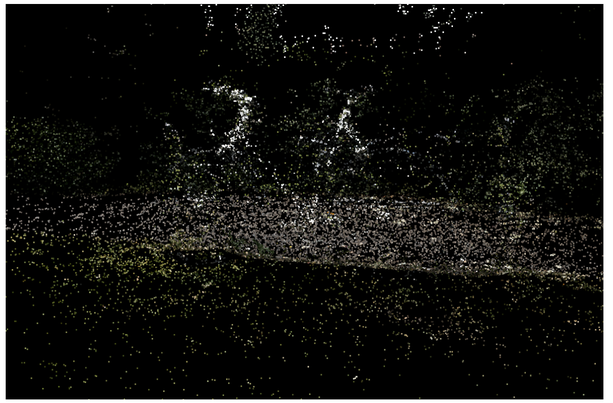


- Intuition: ray-casting through every pixel is wasteful (mostly empty space)
- Represent the scene as a collection of points with size: 3D Gaussians (mean & covariance)
- Projecting 3D Gaussians to the image plane results in approx. 2D Gaussians (Zwicker et al.,2002)
- Advantage: only spend time/memory on the surface of objects

Kerbl, B., Kopanas, G., Leimkühler, T., & Drettakis, G. (2023). 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics (ToG)*

Slide credit: Christian Rupprecht, Oxford

2024: Gaussian Splatting



- Intuition: ray-casting through every pixel is wasteful (mostly empty space)
- Represent the scene as a collection of points with size: 3D Gaussians (mean & covariance)
- Projecting 3D Gaussians to the image plane results in approx. 2D Gaussians (Zwicker et al.,2002)
- Advantage: only spend time/memory on the surface of objects

Kerbl, B., Kopanas, G., Leimkühler, T., & Drettakis, G. (2023). 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics (ToG)*

Slide credit: Christian Rupprecht, Oxford

2024: Gaussian Splatting

- Intuition: ray-casting through every pixel is wasteful (mostly empty space)
- Represent the scene as a collection of points with size: 3D Gaussians (mean & covariance)
- Projecting 3D Gaussians to the image plane results in approx. 2D Gaussians (Zwicker et al.,2002)
- Advantage: only spend time/memory on the surface of objects

Kerbl, B., Kopanas, G., Leimkühler, T., & Drettakis, G. (2023). 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics (ToG)*

Slide credit: Christian Rupprecht, Oxford

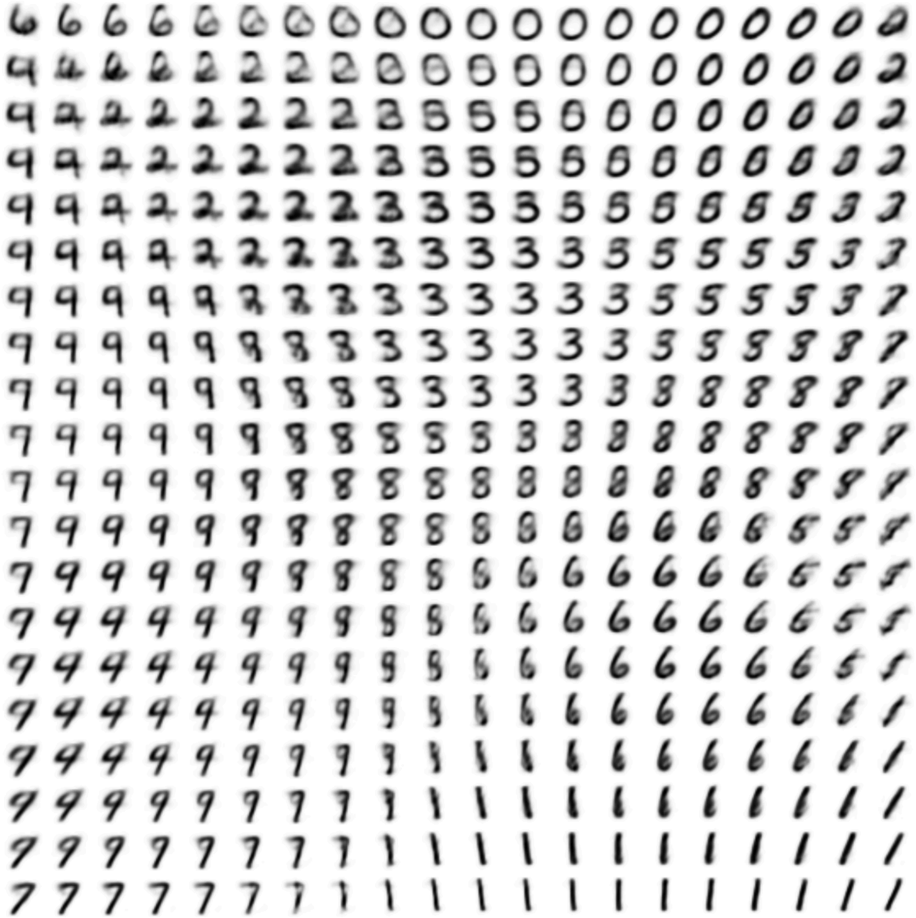

Menu
Views
Capture
3D Gaussians
Camera Point view
Metrics
57.56 (17.37 ms)
VSync On


Menu
Views
Capture
3D Gaussians
Camera Point view
Metrics
57.56 (17.37 ms)
VSync On

Generative Modeling

2013: VAEs

(a) Learned Frey Face manifold

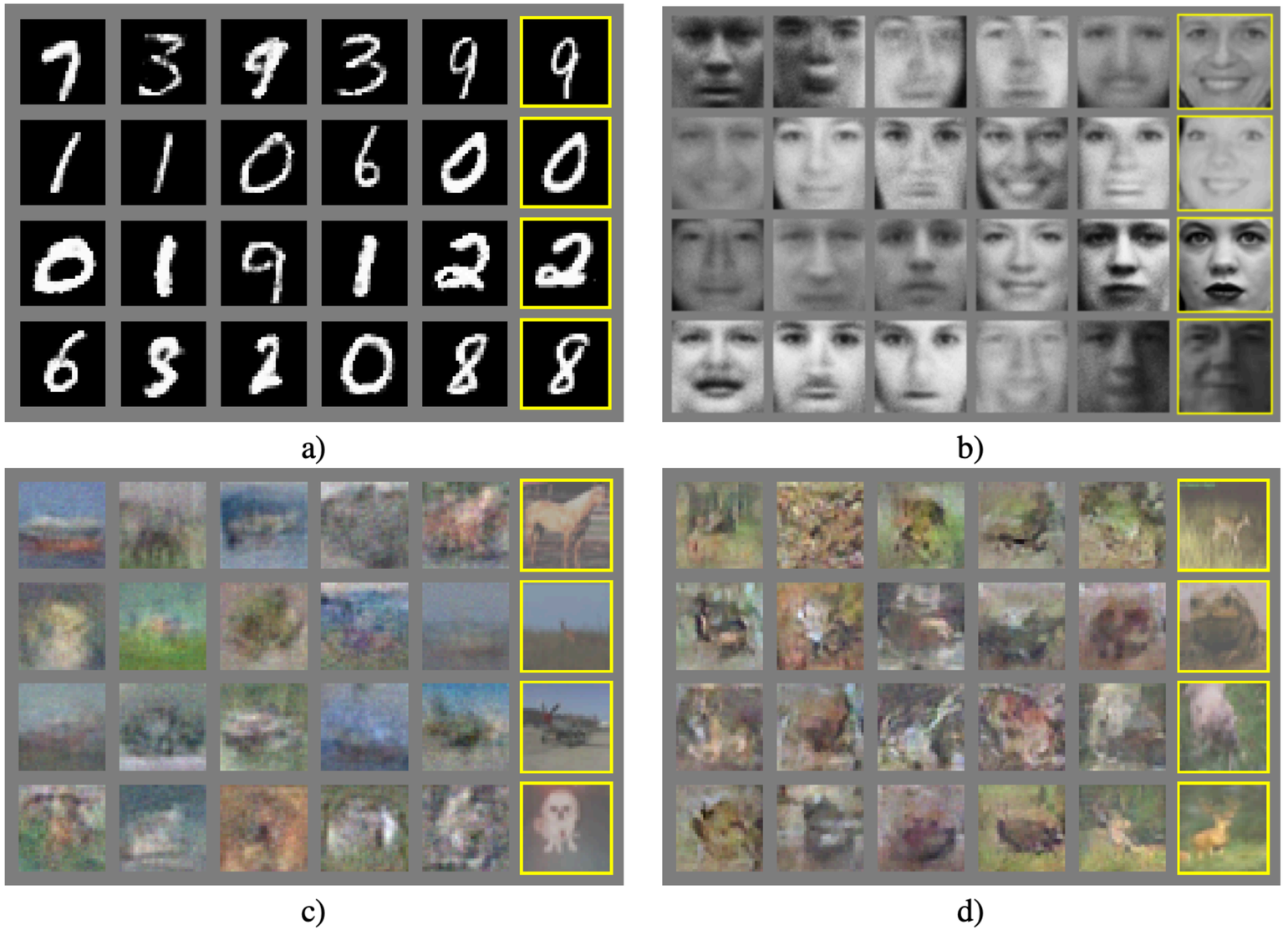


(b) Learned MNIST manifold

Video source: https://x.com/runwayml/status/1807822396415467686

Auto-encoding Variational Bayes, Kingma et al. 2013

2014: GANs



Generative Adversarial Networks, Goodfellow et al. 2014

GANs by 2019

Figure 1: Class-conditional samples generated by our model.

LARGE SCALE GAN TRAINING FOR HIGH FIDELITY NATURAL IMAGE SYNTHESIS, Brock et al. 2019

VAEs by 2020

Figure 1: 256×256-pixel samples generated by NVAE, trained on CelebA HQ [28].

NVAE: A Deep Hierarchical Variational Autoencoder, Vahdat et al. 2020

2020: Diffusion

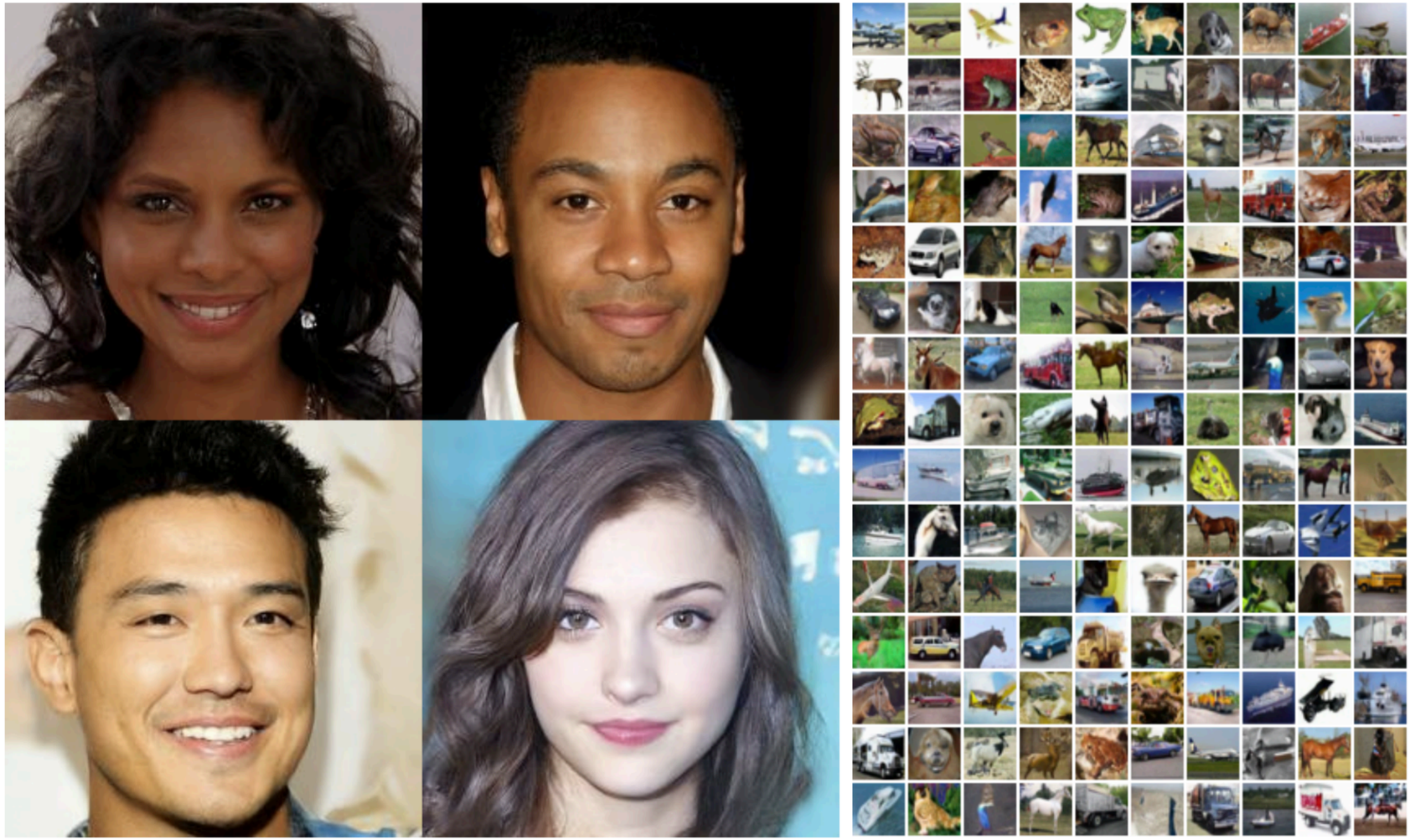


Figure 1: Generated samples on CelebA-HQ 256 × 256 (left) and unconditional CIFAR10 (right)

Deep Unsupervised Learning using Nonequilibrium Thermodynamics, Sohl-Dickstein et al. 2015
Denoising Diffusion Probabilistic Models, Ho et al. 2006

Since 2022: Video Generative Models

Video source: https://x.com/runwayml/status/1807822396415467686

Diffusion Probabilistic Modeling for Video Generation, Yang et al. 2022

Video Diffusion Models, Ho et al. 2022

Since 2022: Video Generative Models

Video source: https://x.com/runwayml/status/1807822396415467686

Diffusion Probabilistic Modeling for Video Generation, Yang et al. 2022

Video Diffusion Models, Ho et al. 2022

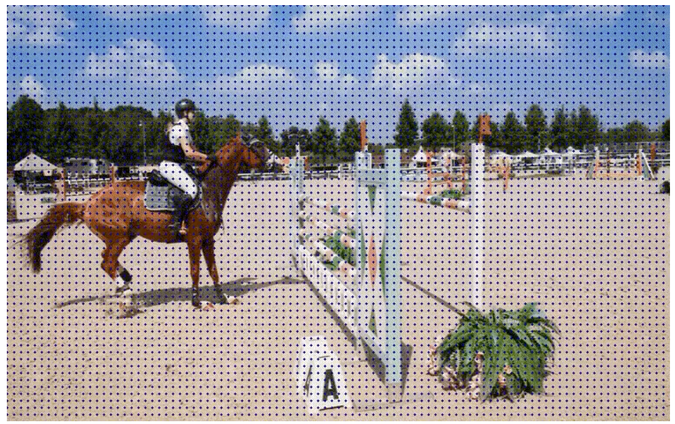
Vision = Collection of Tasks?

Image Segmentation

Video Generation

3D Reconstruction

Novel View Synthesis



Point Tracking

• • •

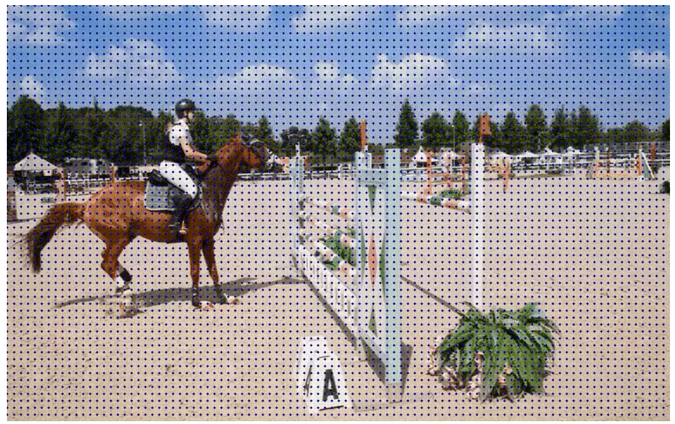
Vision = Collection of Tasks?

Image Segmentation

Video Generation

3D Reconstruction

Novel View Synthesis



Point Tracking

• • •

Adapted from Kelsey Allen, Josh Tenenbaum

Adapted from Kelsey Allen, Josh Tenenbaum

Vision = Modeling the World and Yourself Within It

Vision is the ability of an agent to build a dynamic model of its environment by perceiving it through its senses.

Vision is tightly linked to planning and intelligence.

Vision ≠ Collection of Tasks!

Vision is the ability of an agent to build a dynamic model of its environment by perceiving it through its senses.

Vision is tightly linked to planning and intelligence.

Tasks are what the scientific community identifies as ***solvable problems on the way!***

Tale: The era of geometric invariance

> Recent advances in object recognition have emphasized the integration of intensity-derived features such as affine patches with associated geometric constraints leading to impressive performance in complex scenes. Over the four previous decades, the central paradigm of recognition was based on formal geometric object descriptions with a focus on the properties of such descriptions under perspective image formation. This paper will review the key advances of the geometric era and investigate the underlying causes of the movement away from formal geometry and prior models towards the use of statistical learning methods based on appearance features.

Fig. 14. A meeting of researchers central to the geometric invariance movement at Schenectady, New York during the month of July, 1992. Top row, left to right: Andrew Zisserman, Charles Rothwell, Luc VanGool, Joseph Mundy, Stephen Maybank and Daniel Huttenlocher. Bottom row, left to right: Thomas Binford, Richard Hartley, David Forsyth and Jon Kleinberg.

"Geometric invariance" remains unsolved in vision!

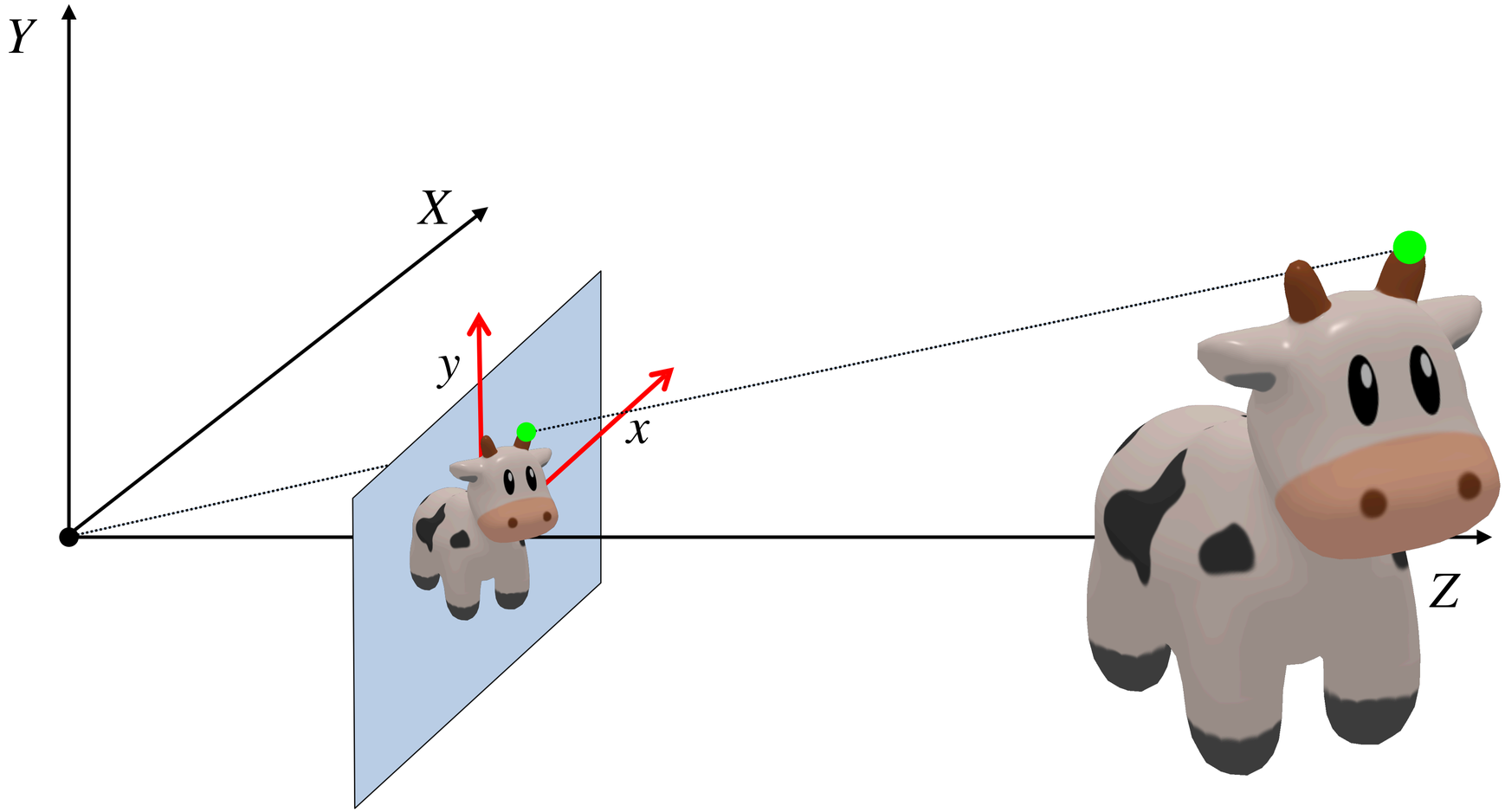
Fig. 14. A meeting of researchers central to the geometric invariance movement at Schenectady, New York during the month of July, 1992. Top row, left to right: Andrew Zisserman, Charles Rothwell, Luc VanGool, Joseph Mundy, Stephen Maybank and Daniel Huttenlocher. Bottom row, left to right: Thomas Binford, Richard Hartley, David Forsyth and Jon Kleinberg.

MODULE 1: GEOMETRY

6.8300

Prof. Vincent Sitzmann

Perspective projection



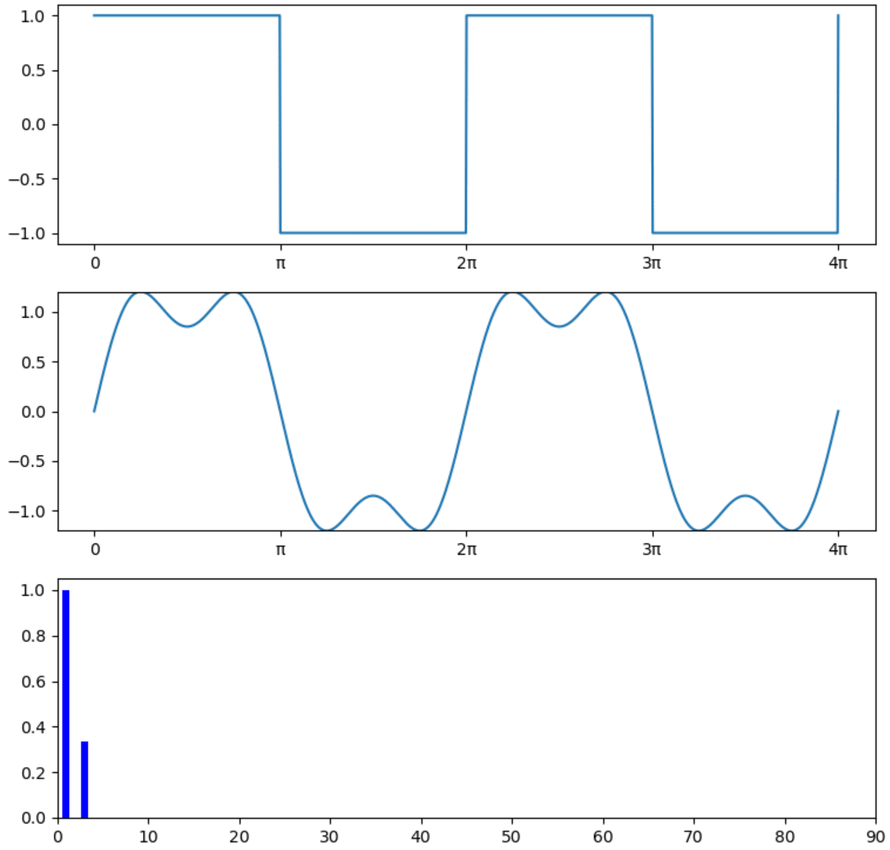
Fourier Transform

Any(**) univariate function can be expressed as a weighted sum of sinusoids of different frequencies (1807)

Jean-Baptiste Joseph Fourier (1768-1830)

Example: series for a square wave

$$\sum_{k=1,3,5,\ldots}^{\infty} \frac{1}{k}\sin(kt)$$



Slide credit: S. Lazebnik, C. Rupprecht

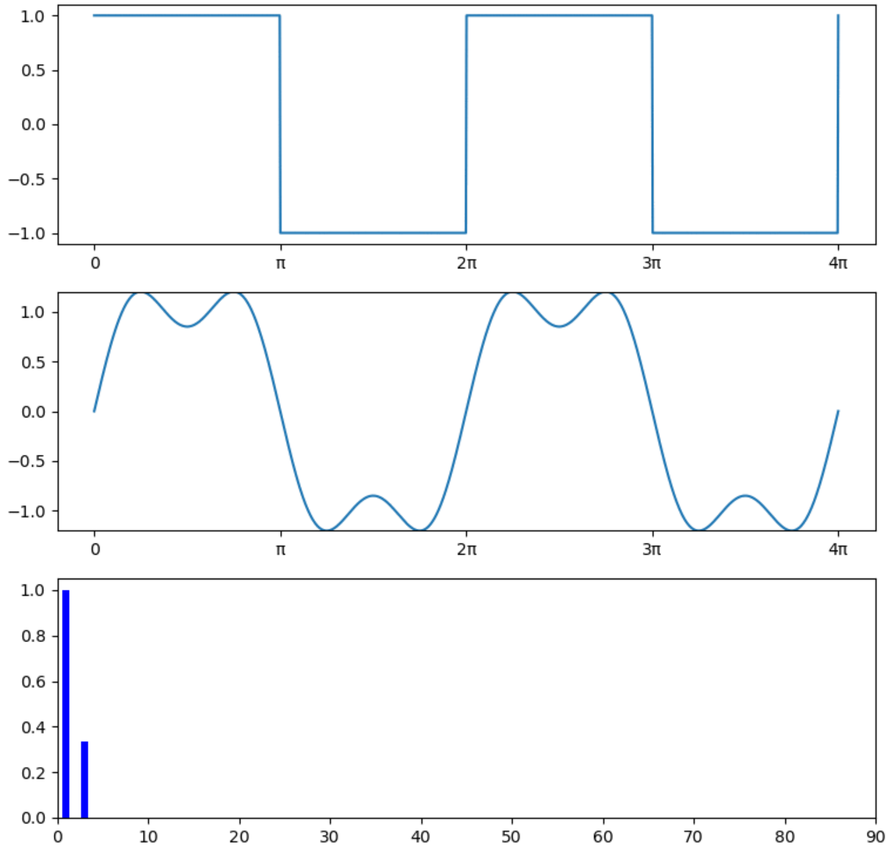
Fourier Transform

Any(**) univariate function can be expressed as a weighted sum of sinusoids of different frequencies (1807)

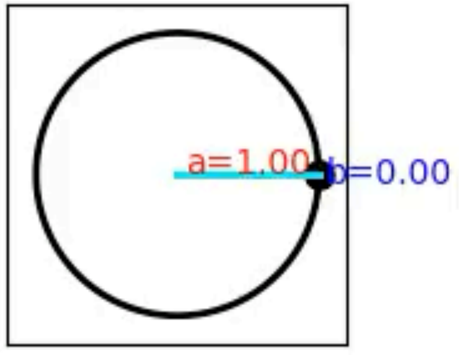
Jean-Baptiste Joseph Fourier (1768-1830)

Example: series for a square wave

$$\sum_{k=1,3,5,\ldots}^{\infty} \frac{1}{k}\sin(kt)$$



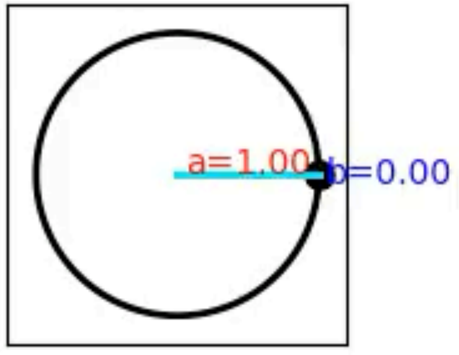
Slide credit: S. Lazebnik, C. Rupprecht



Cosine

Sine

a·cos(s) + b·sin(s)



Cosine

Sine

a·cos(s) + b·sin(s)

Geometry is an integral part of human vision

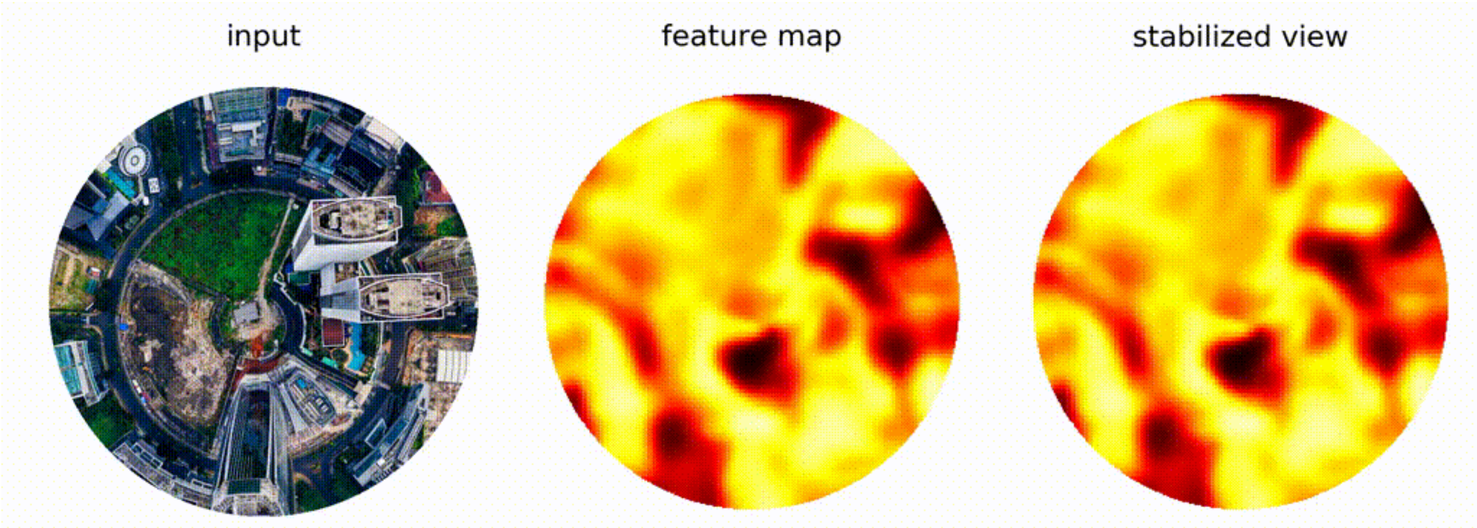
Credit: Helmet and the Norwegian University of Science and Technology's Kavli Institute for Systems Neuroscience

Geometry is an integral part of human vision

Credit: Helmet and the Norwegian University of Science and Technology's Kavli Institute for Systems Neuroscience

Geometric guarantees (equivariance)

CNN

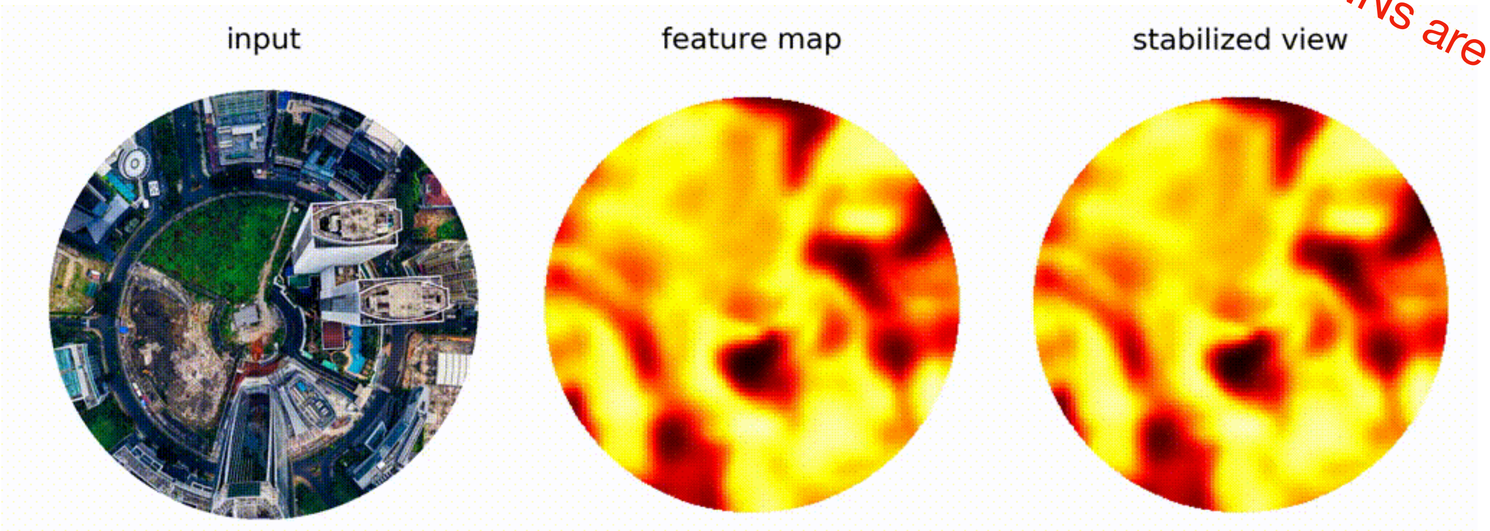


Figures source:
https://github.com/QUVA-Lab/e2cnn

Slide curtesy of Erik Bekkers from UVA Deep Learning II Course

Geometric guarantees (equivariance)

CNN



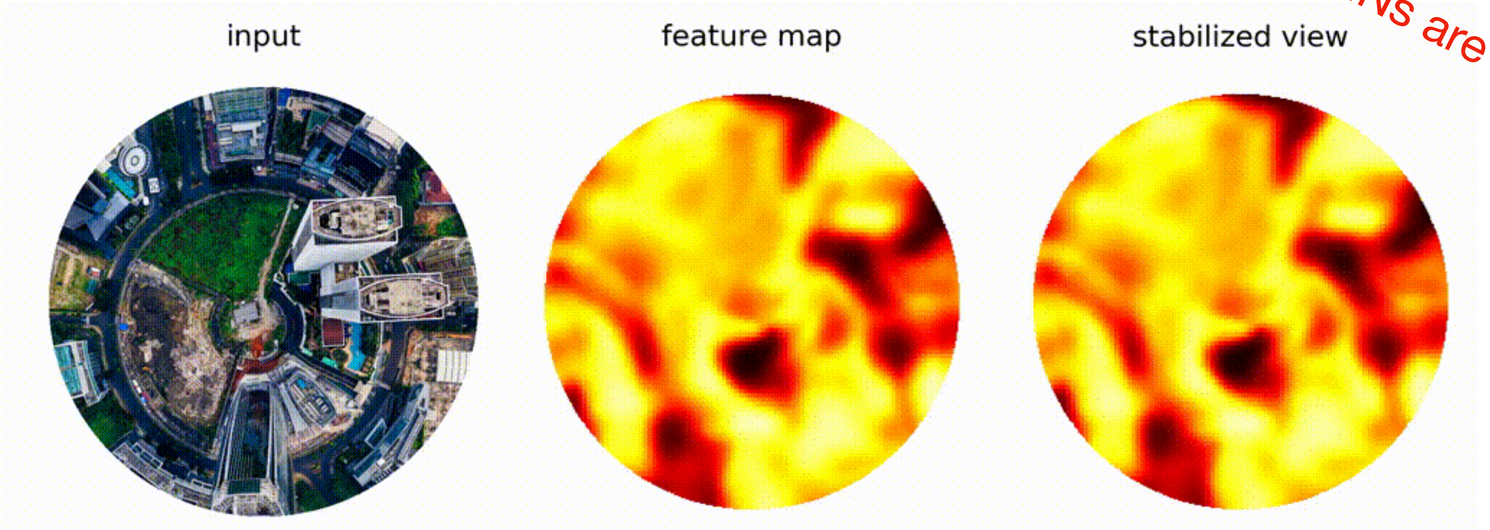
CNNs are not rotation equivariant

Figures source:
https://github.com/QUVA-Lab/e2cnn

Slide curtesy of Erik Bekkers from UVA Deep Learning II Course

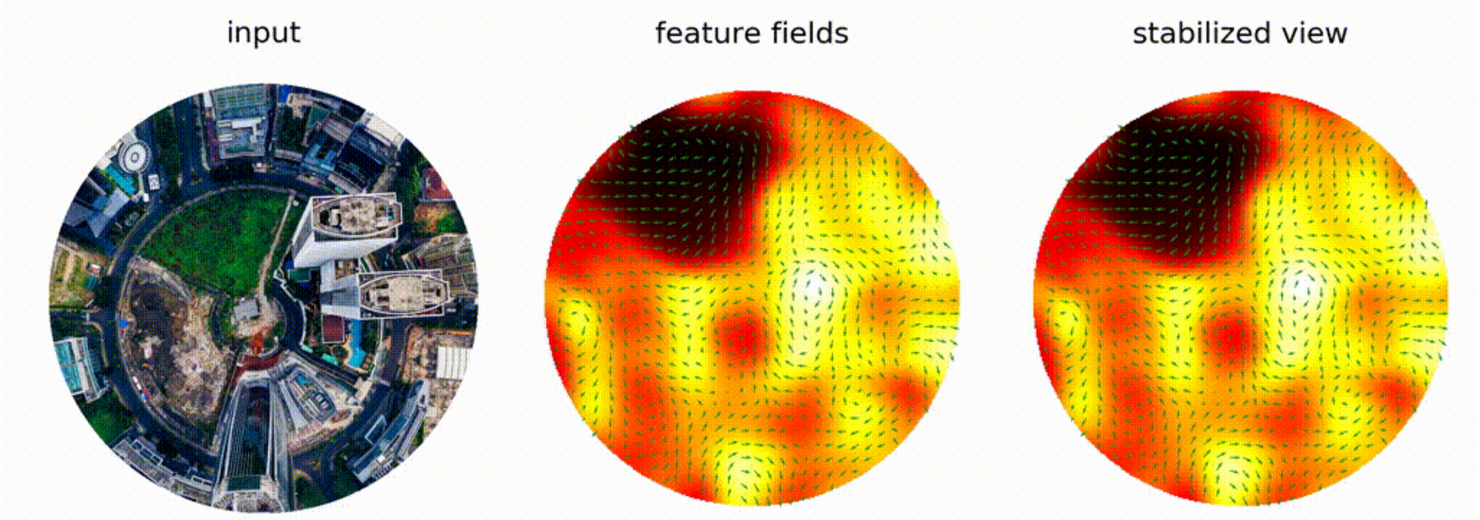
Geometric guarantees (equivariance)

CNN



CNNs are not rotation equivariant

Equivariant NN



Figures source:
https://github.com/QUVA-Lab/e2cnn

$$\mathbf{F} = \mathbf{K}_2^{-T}(\mathbf{R}[\mathbf{t}]_\times)\mathbf{K}_1^{-1}$$

$$\mathbf{F}\tilde{\mathbf{x}}_1 = \mathbf{l}_2$$

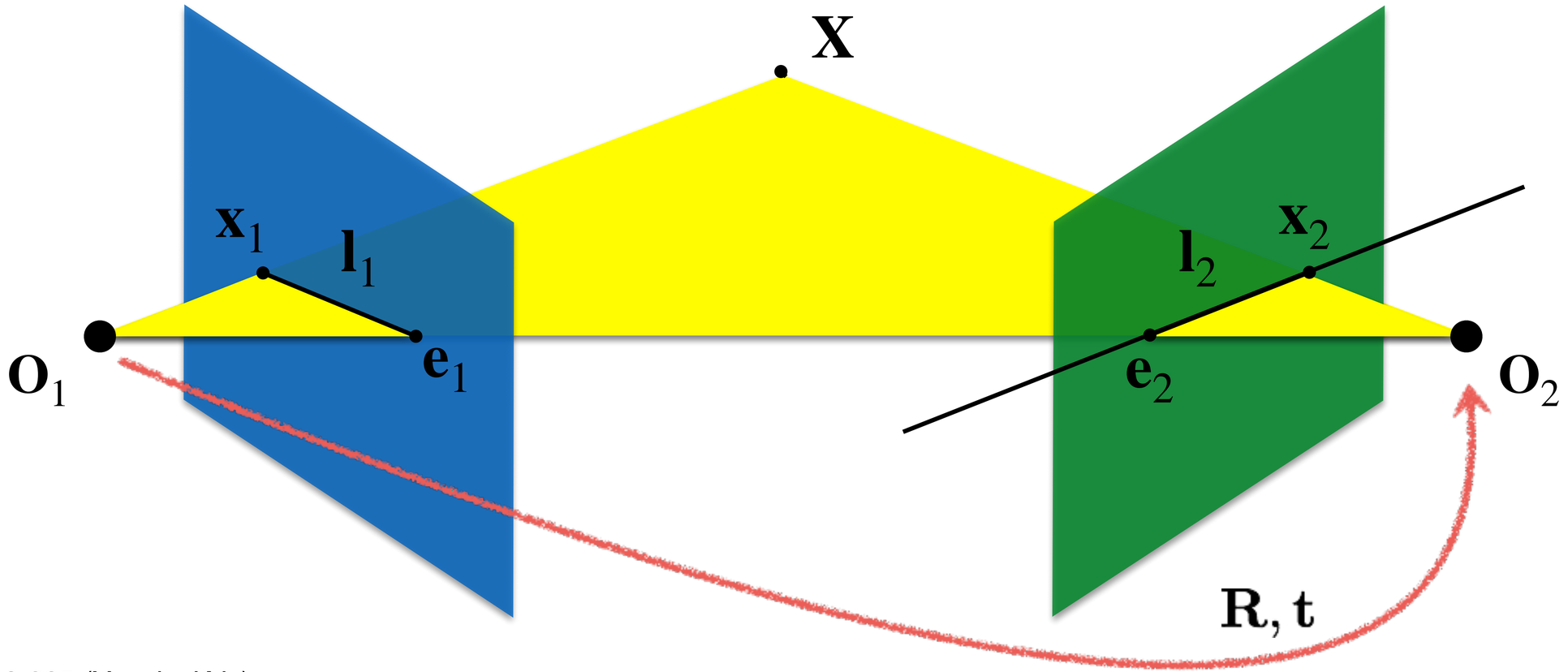
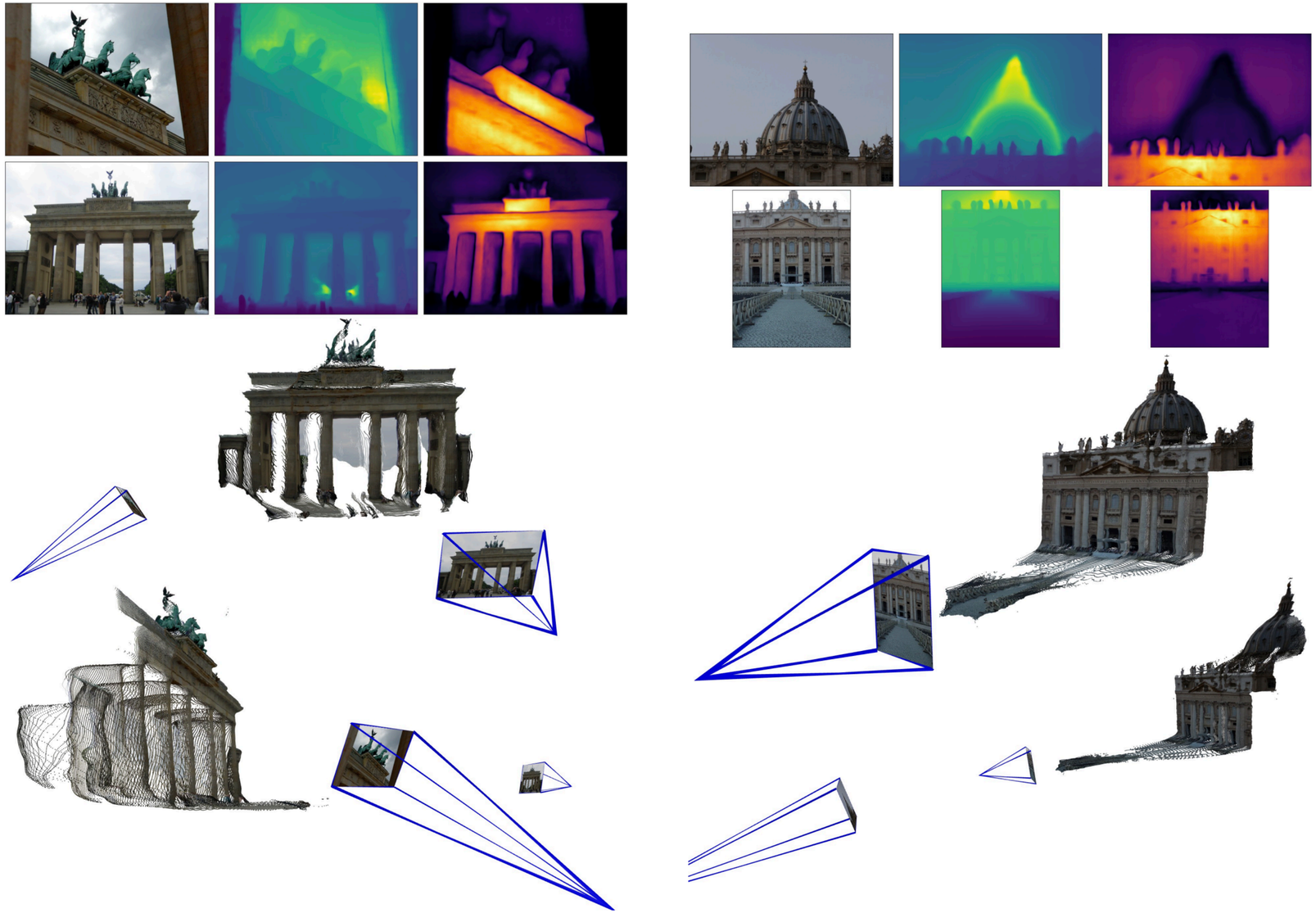


Figure credit: CMU 16-385 (Yannis, Kris)

DUST3R: Pose SfM as supervised learning problem



Anisotropic Volumetric 3D Gaussians

Final Rendering

3D Gaussian Visualization

Anisotropic Volumetric 3D Gaussians

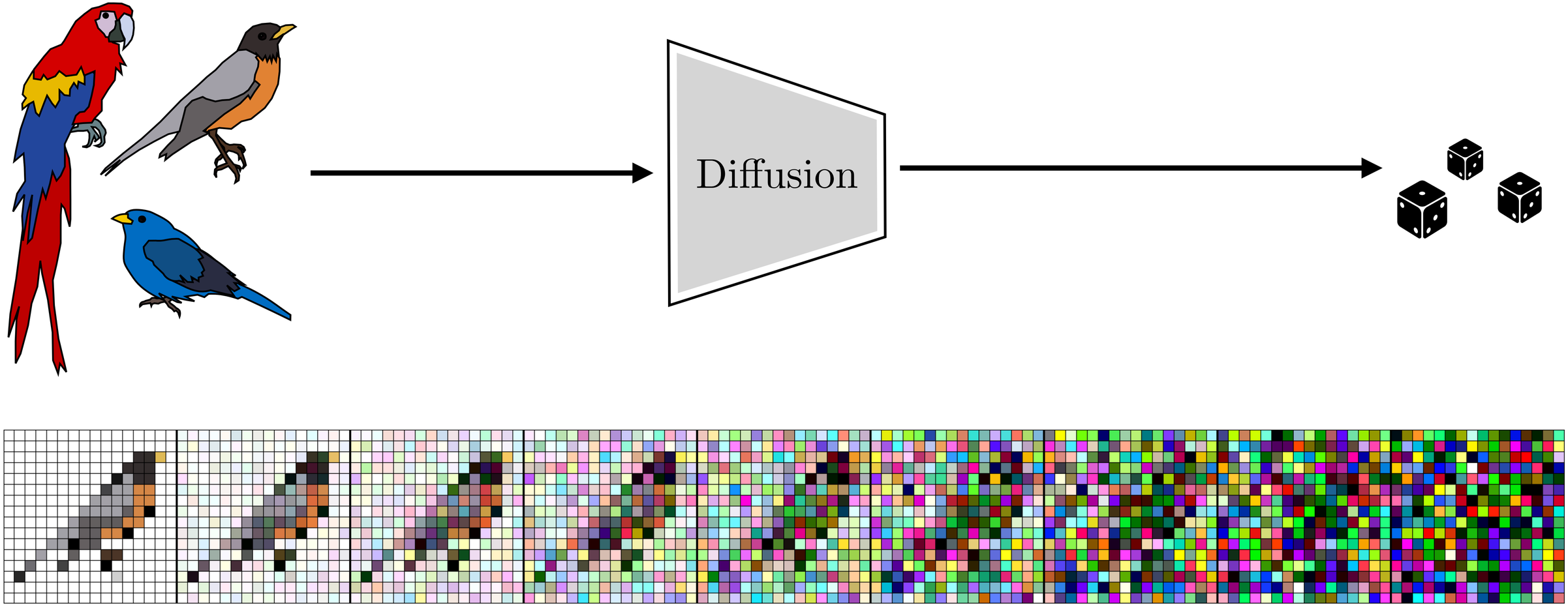
Final Rendering

3D Gaussian Visualization

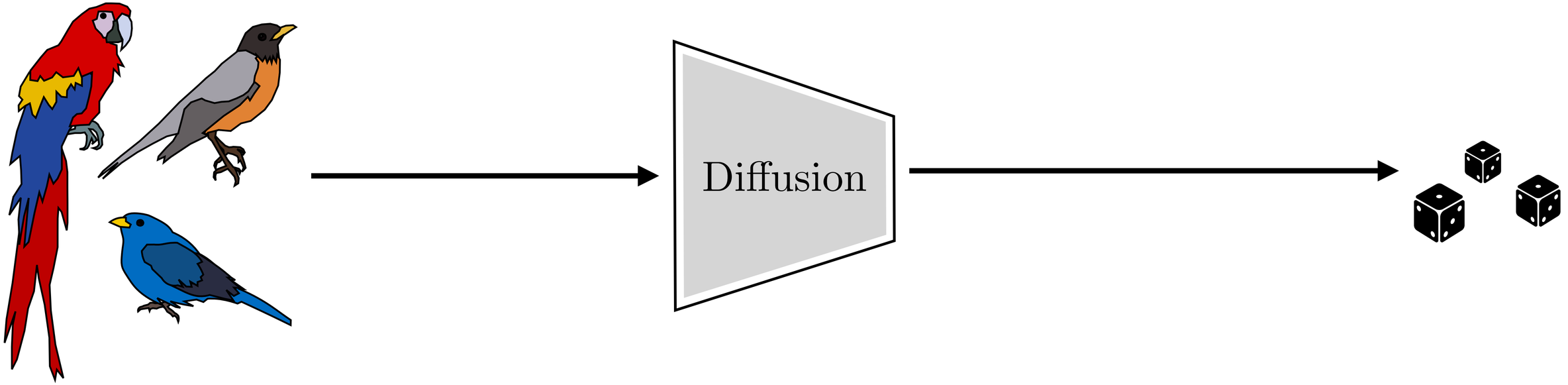
MODULE 2: GENERATIVE MODELING & REPRESENTATION LEARNING

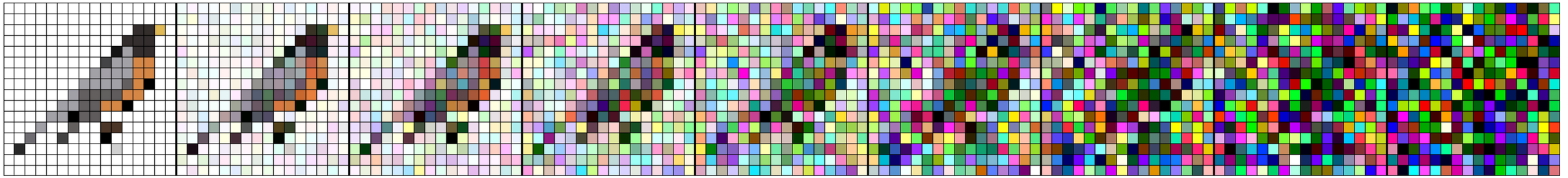
6.8300 Prof. Vincent Sitzmann

Diffusion Models



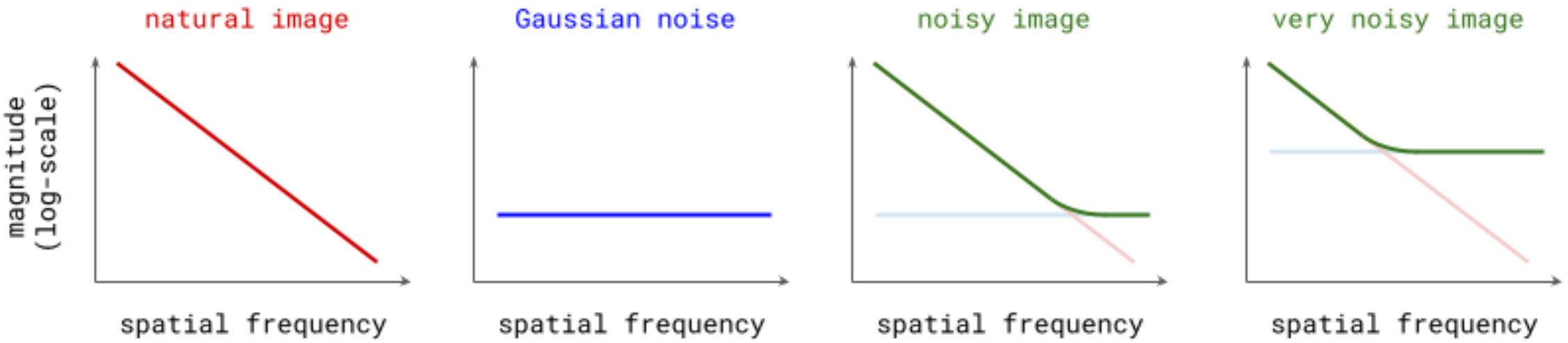
Diffusion Models

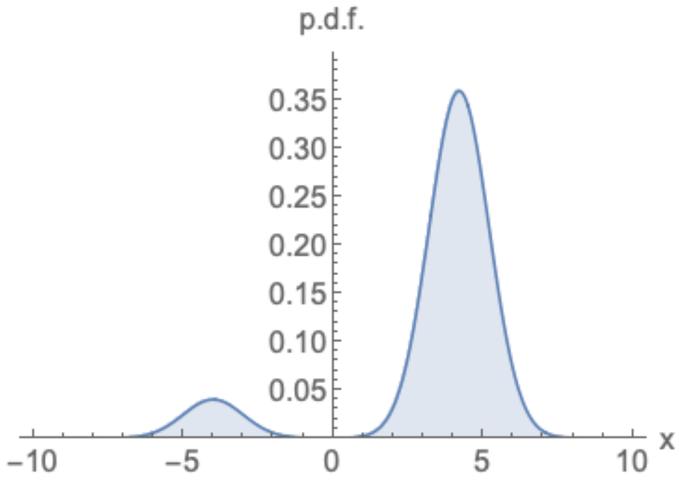


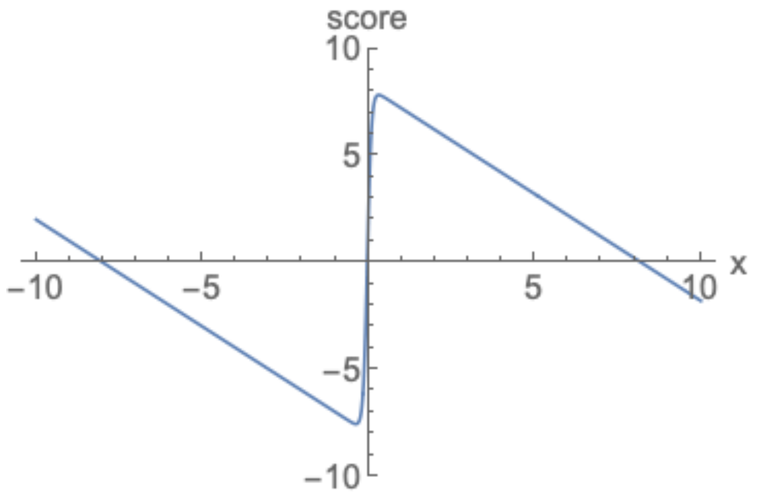


Diffusion: Just add noise

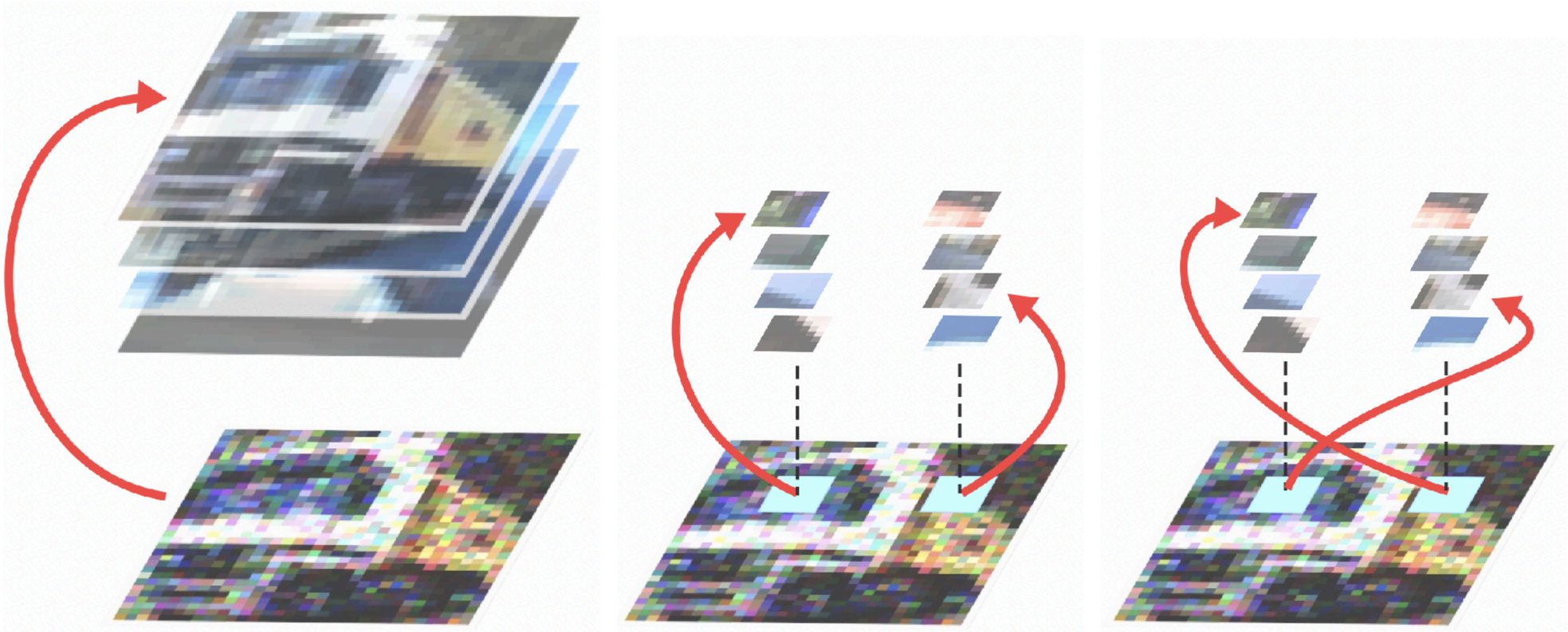
Perspectives on Diffusion





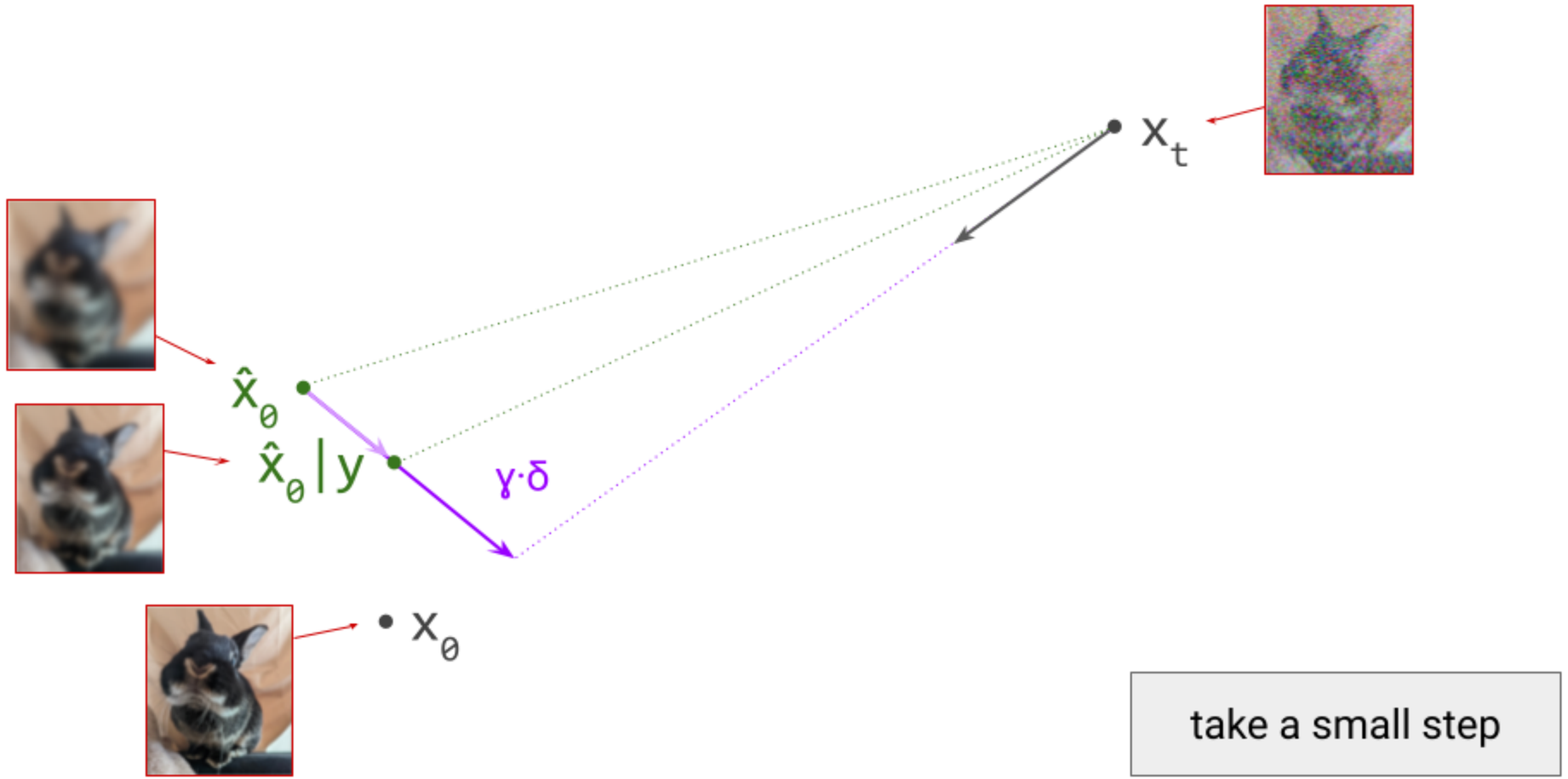


How / Why Diffusion Models Generalize



(a) IS Machine (b) LS Machine (c) ELS Machine

What is guidance and why does it work?



Sequence Generative Models

Full Sequence

Diffusion Forcing

Ground Truth

Sequence Generative Models

Full Sequence

Diffusion Forcing

Ground Truth

3D Generative Models

Input: Single Image

Deterministic Reconstruction

3D Generative Models

Input: Single Image

Deterministic Reconstruction

3D Generative Models

Input: Single Image

Deterministic Reconstruction

Ours

3D Generative Models

Input: Single Image

Deterministic Reconstruction

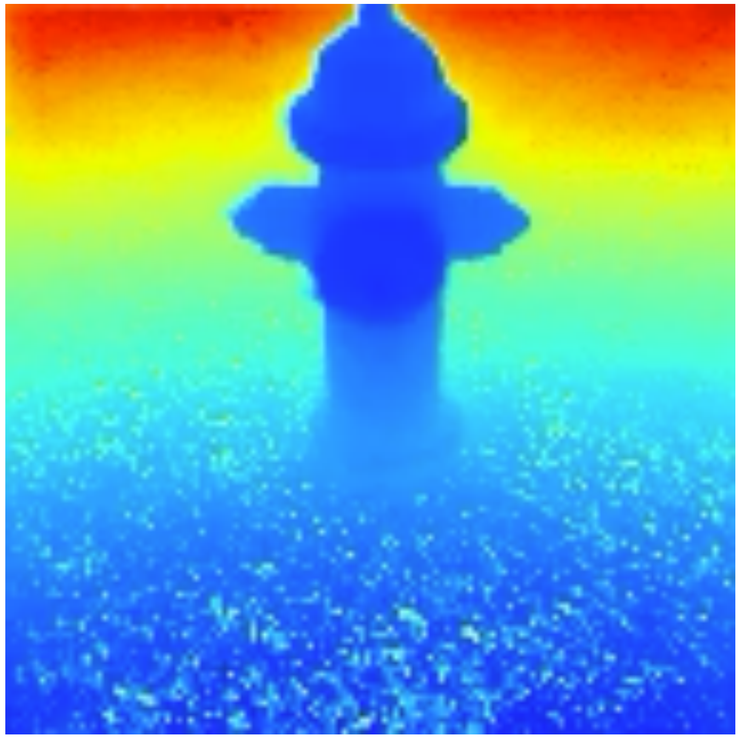
Ours

This is a **conditional generative model** that learns to **directly sample 3D scenes, trained only on images.**

3D Generative Models

Input: Single Image

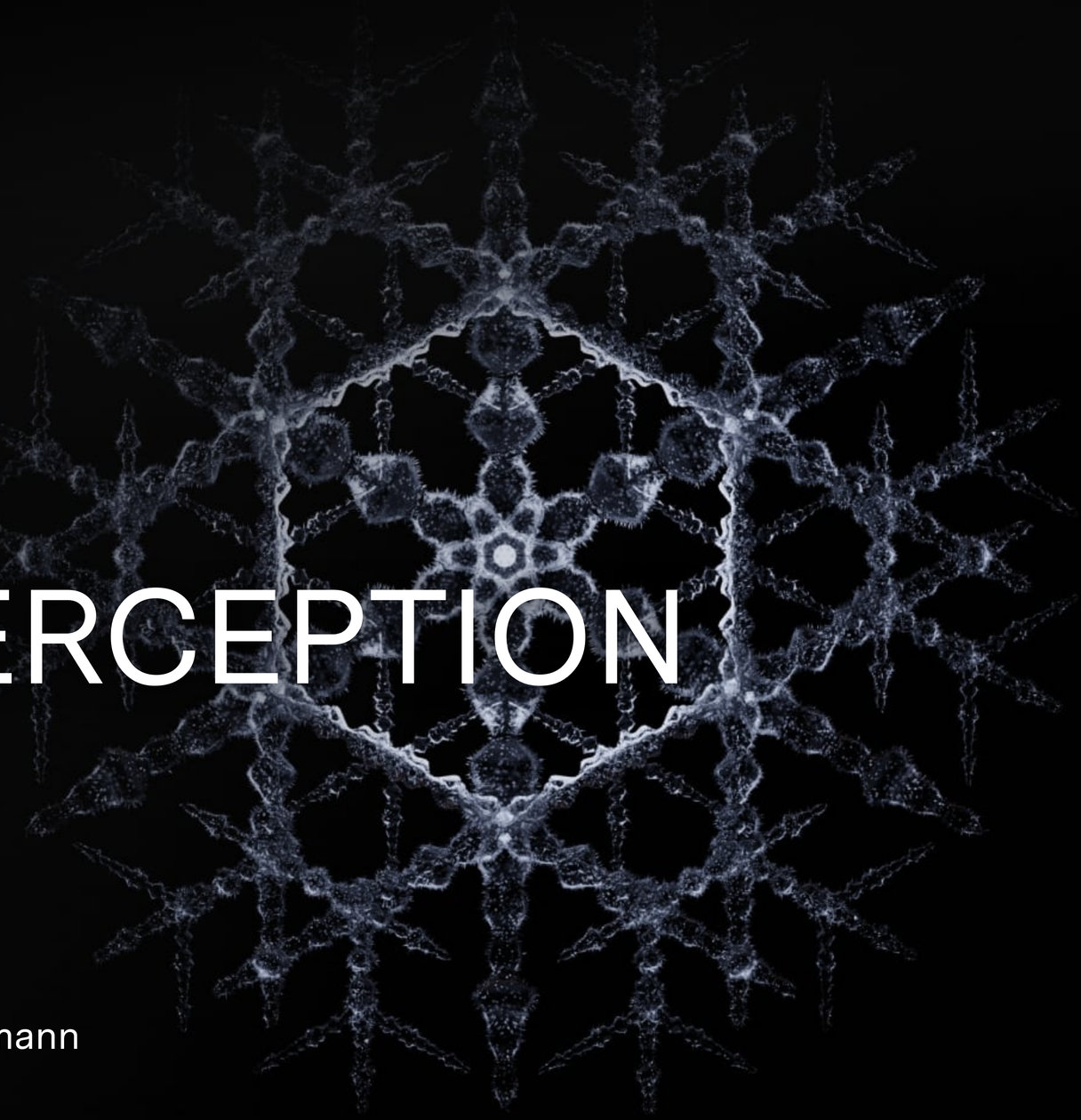
Depth



Ours

This is a **conditional generative model** that learns to **directly sample 3D scenes, trained only on images.**

MODULE 1:
Geometry

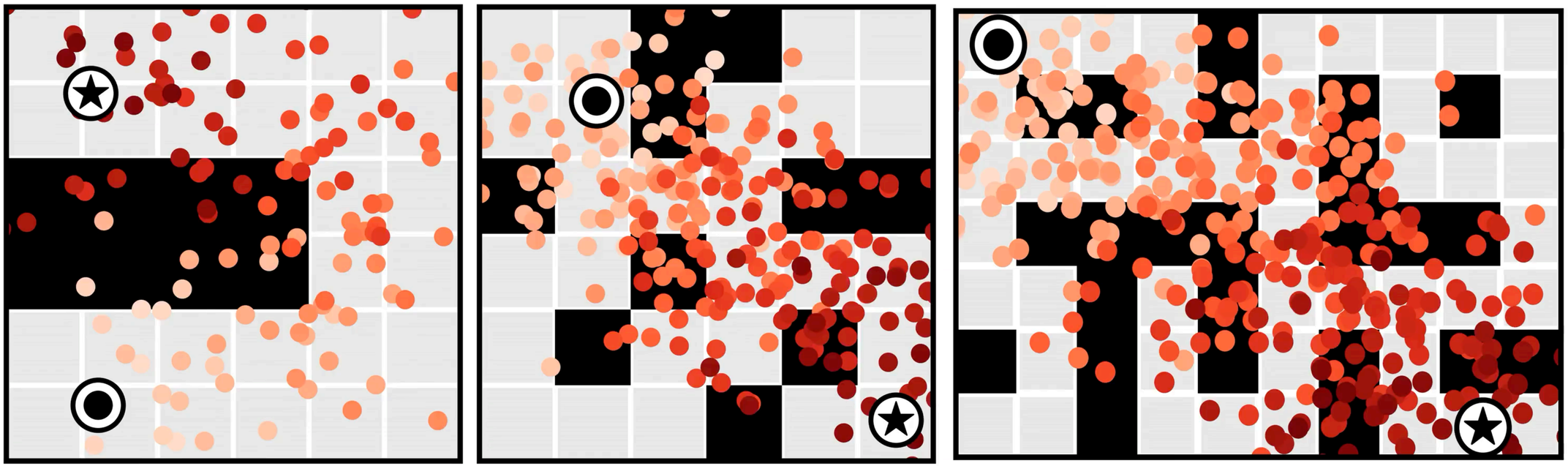


MODULE 3: ROBOTIC PERCEPTION

6.8300

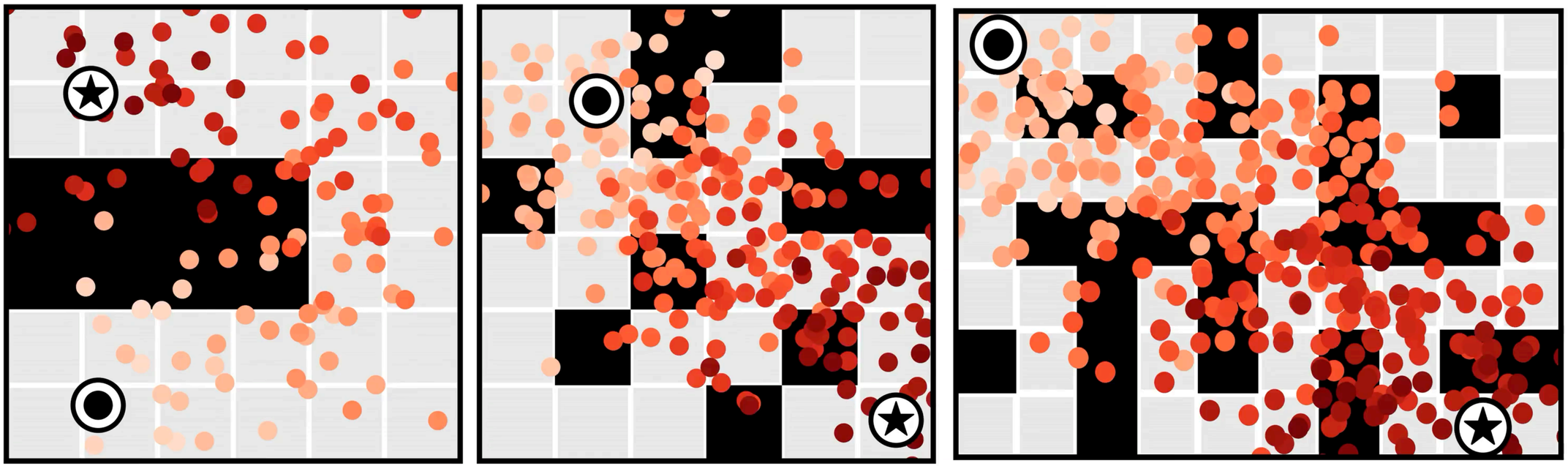
Prof. Vincent Sitzmann

Diffusion Planning



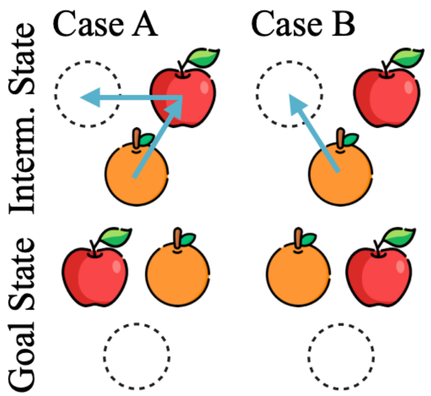
Diffusion Forcing, Chen et al. 2024

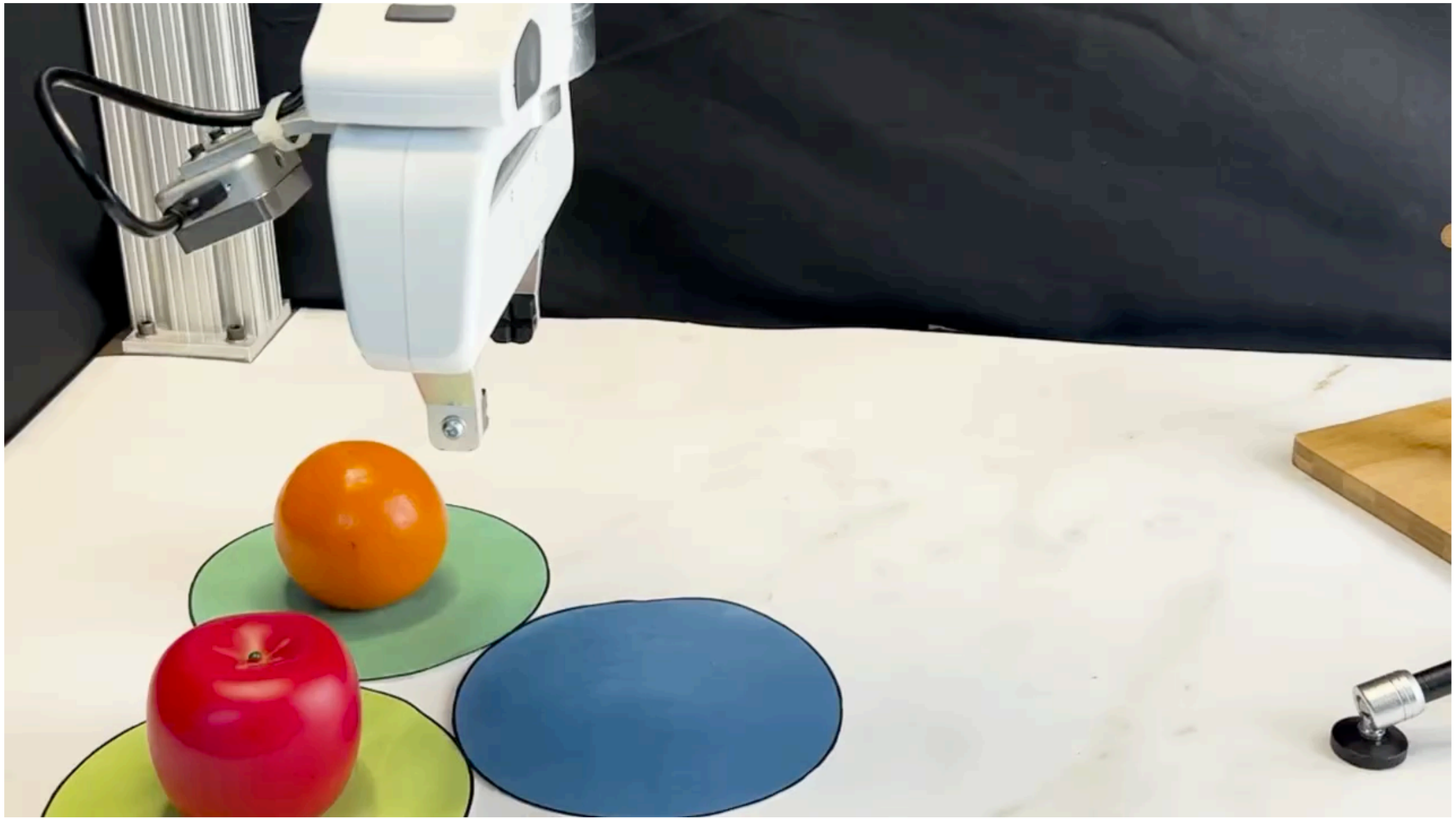
Diffusion Planning



Diffusion Forcing, Chen et al. 2024

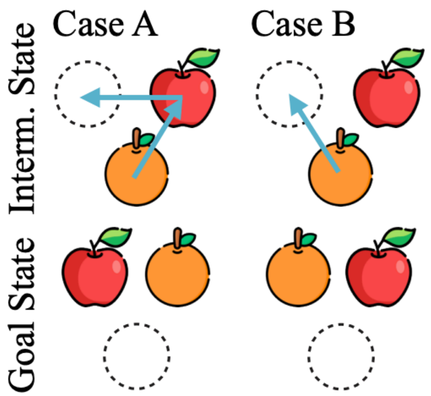
Long-Horizon, Stateful Behavior Cloning

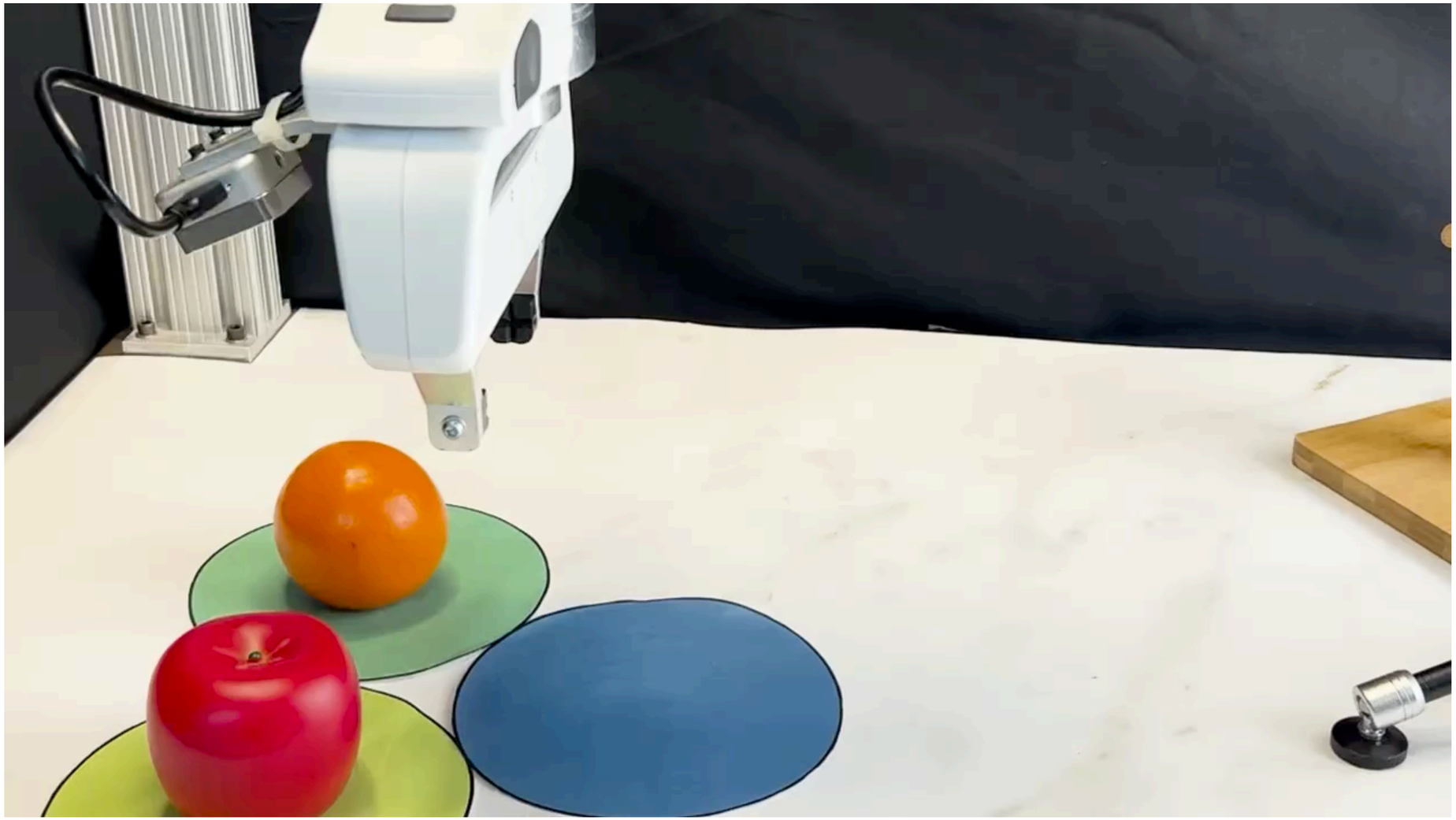




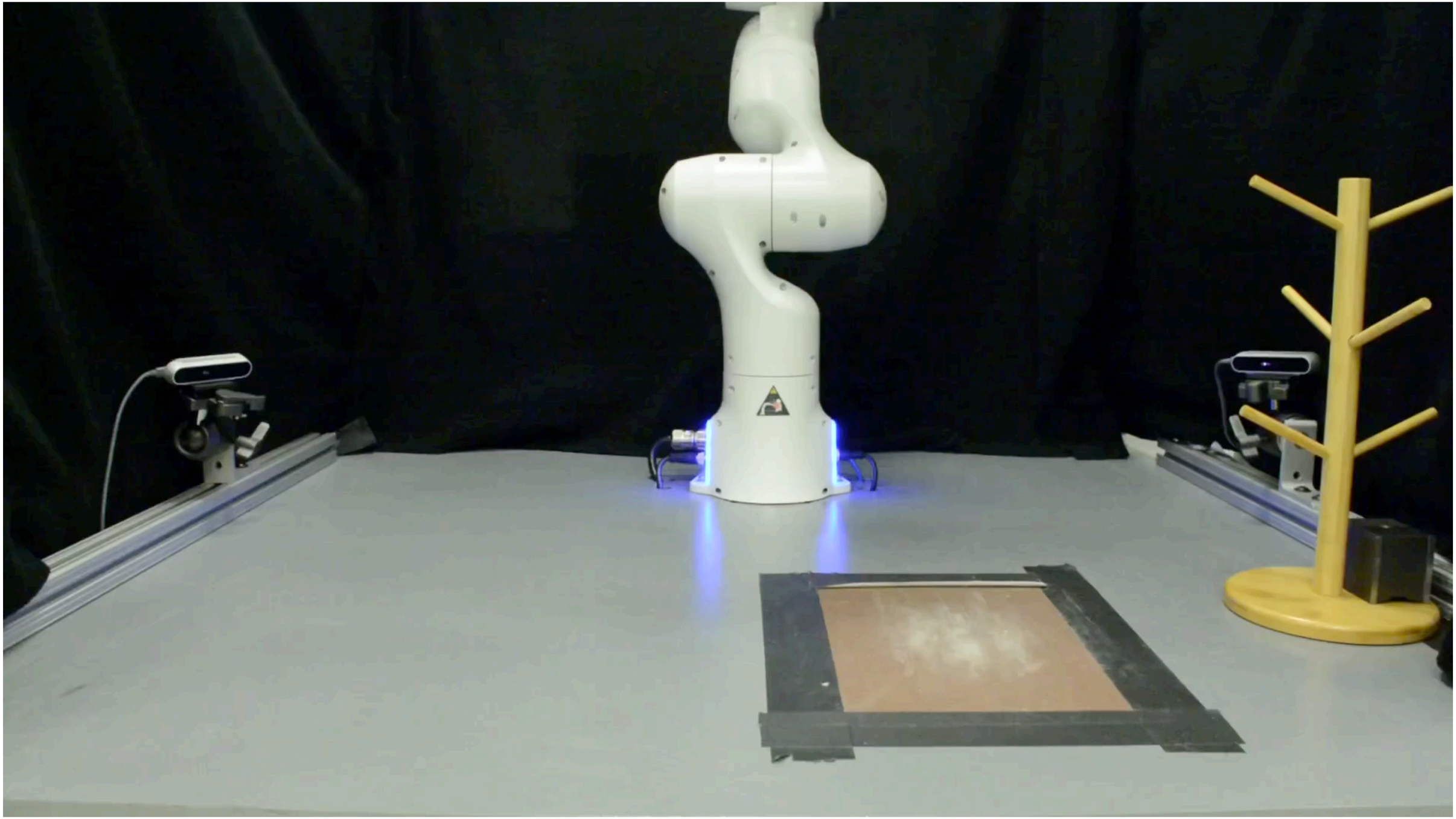
Diffusion Forcing, Chen et al. 2024

Long-Horizon, Stateful Behavior Cloning

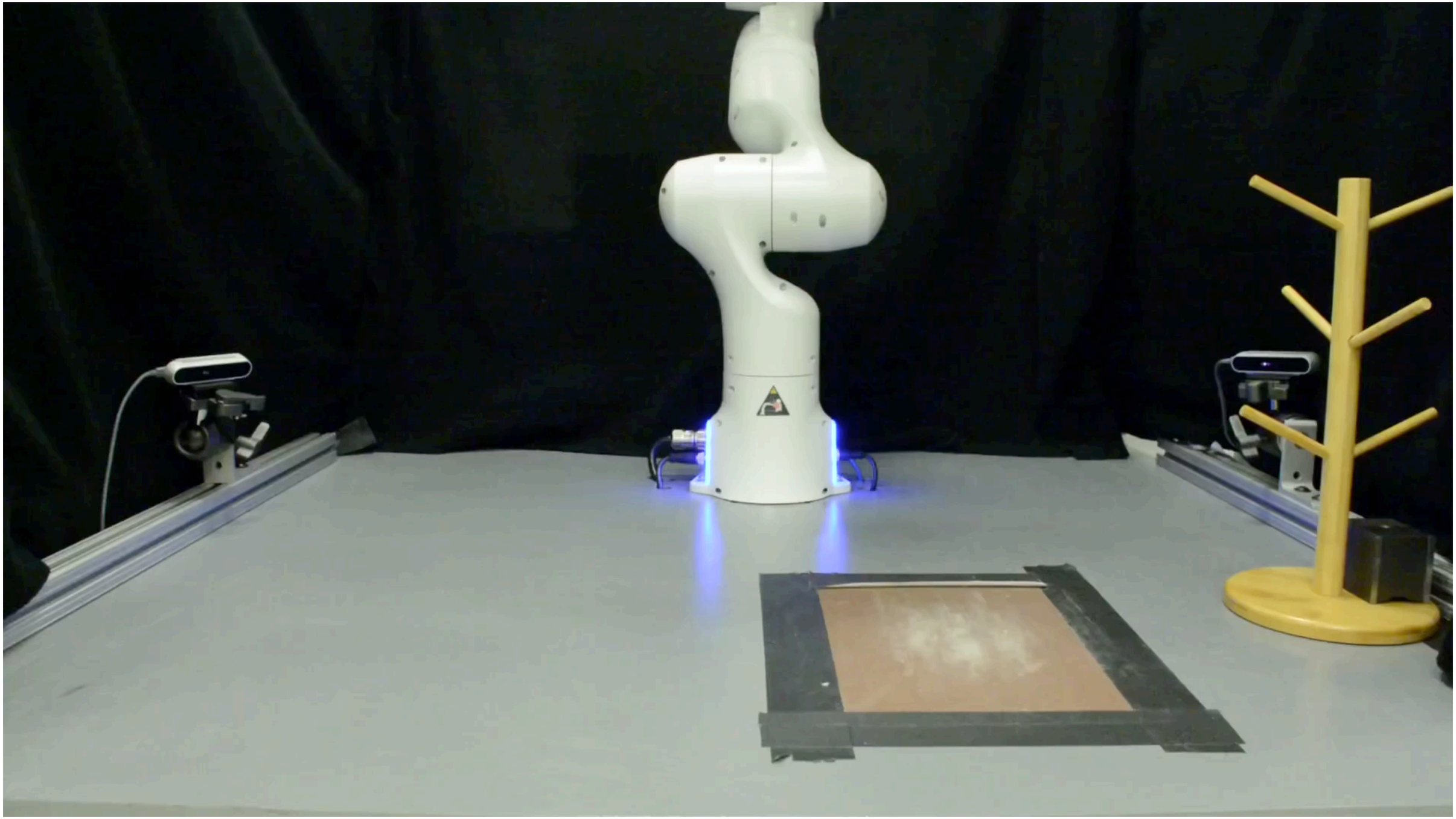




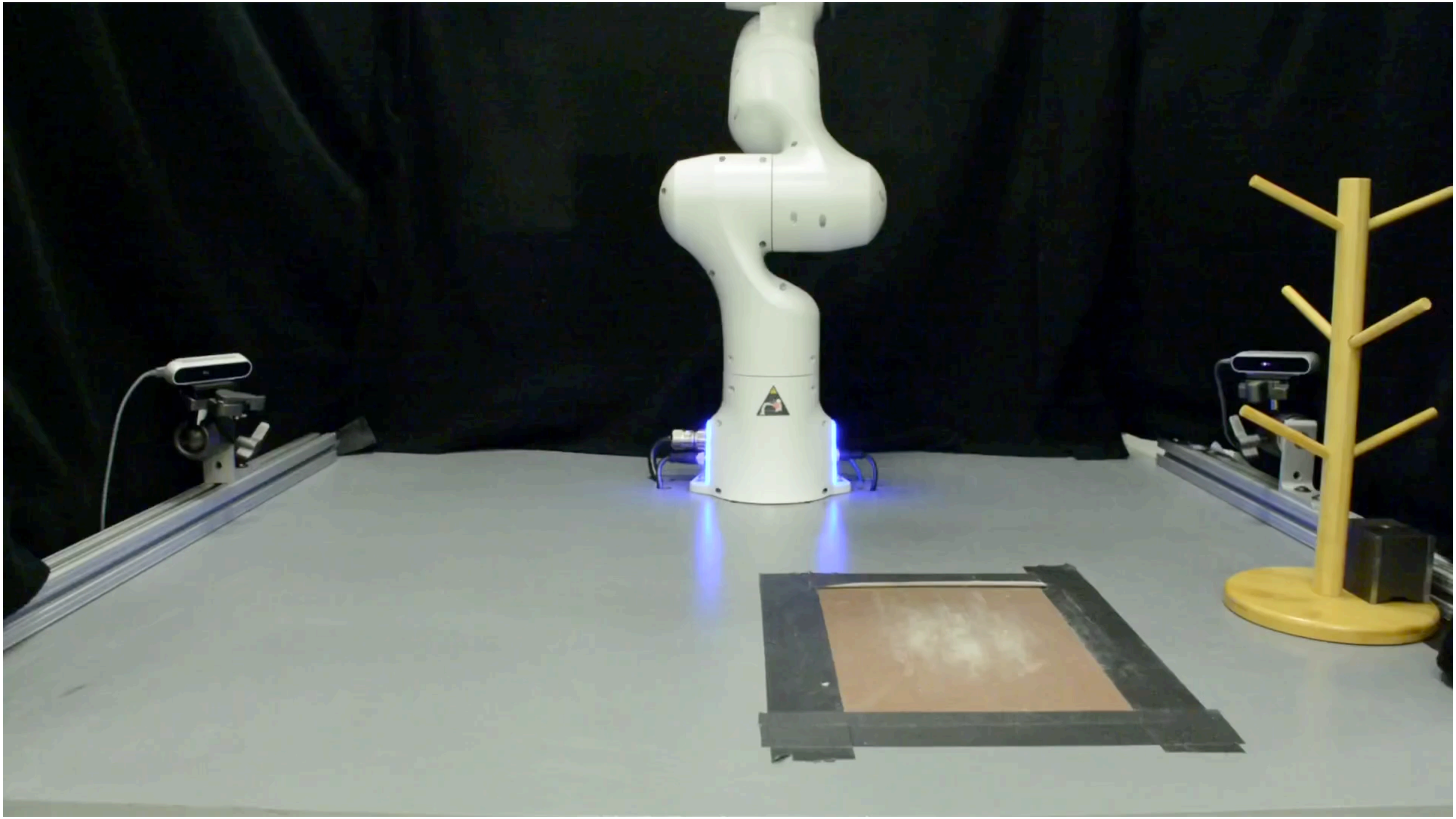
Diffusion Forcing, Chen et al. 2024



Neural Descriptor Fields, Simeonov, Du et al. 2021

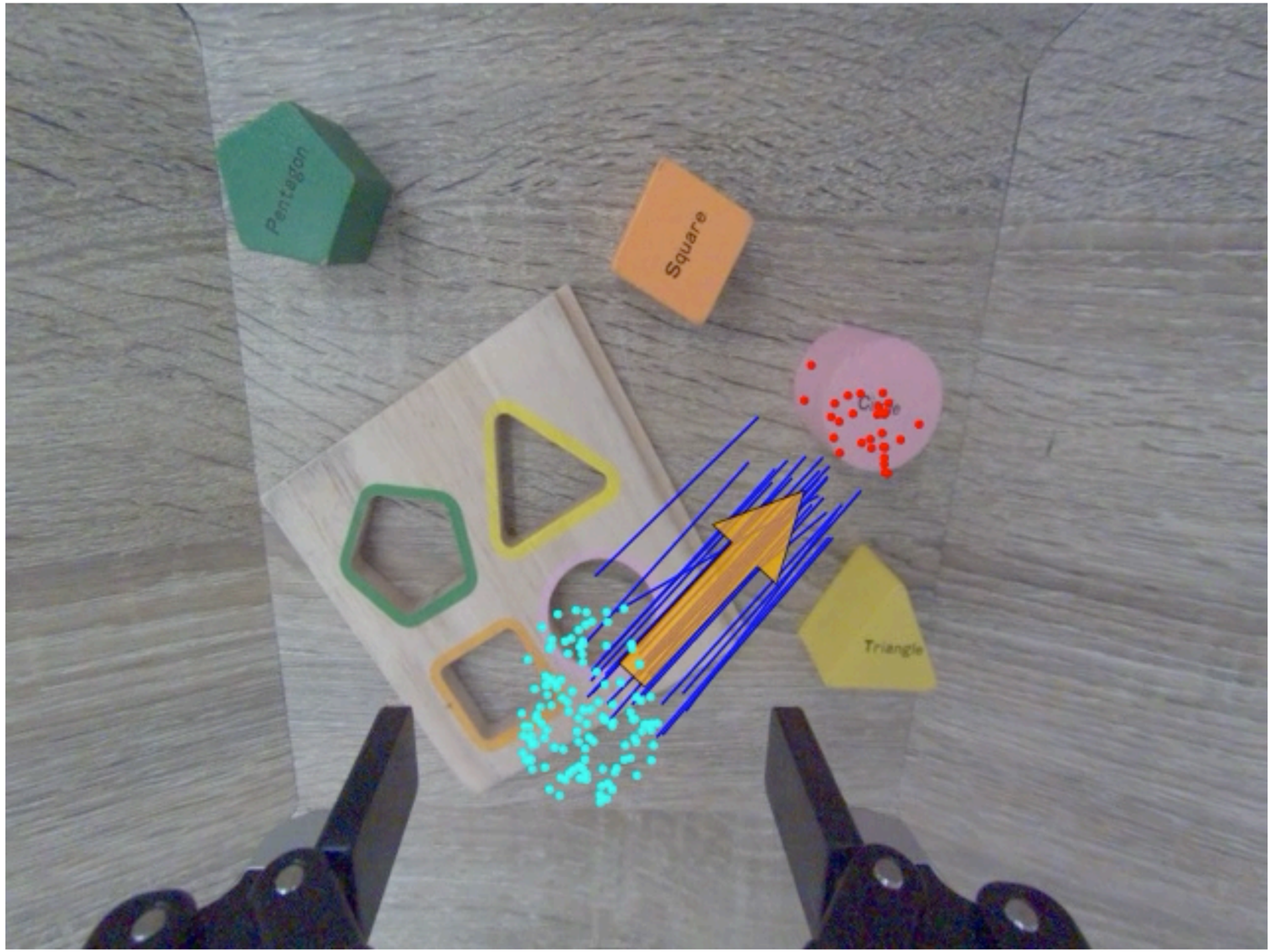


Neural Descriptor Fields, Simeonov, Du et al. 2021



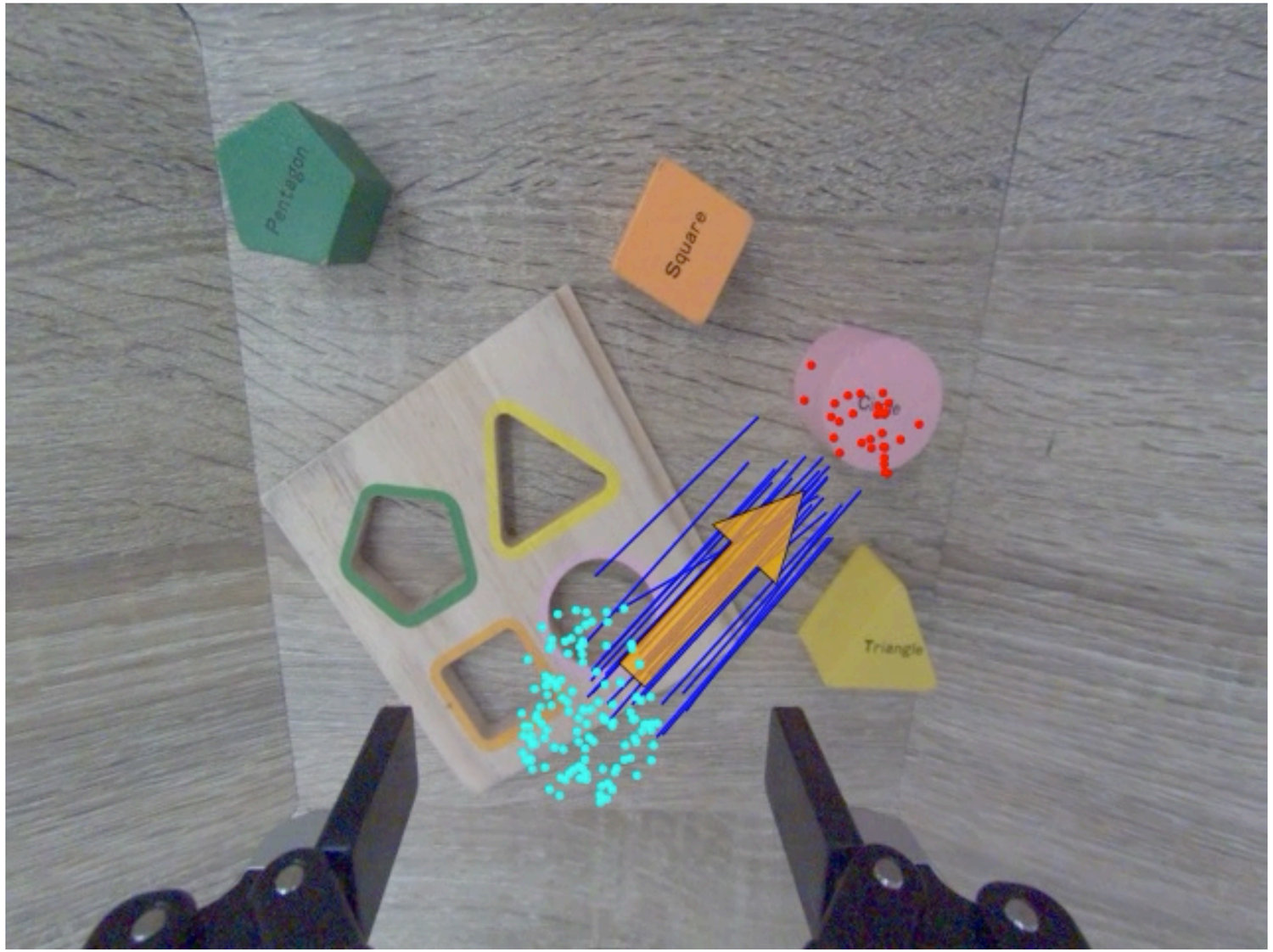
Neural Descriptor Fields, Simeonov, Du et al. 2021

Visual Imitation Learning



Robotap: Tracking Arbitrary Points for Few-Shot Visual Imitation, Vecerik et al.

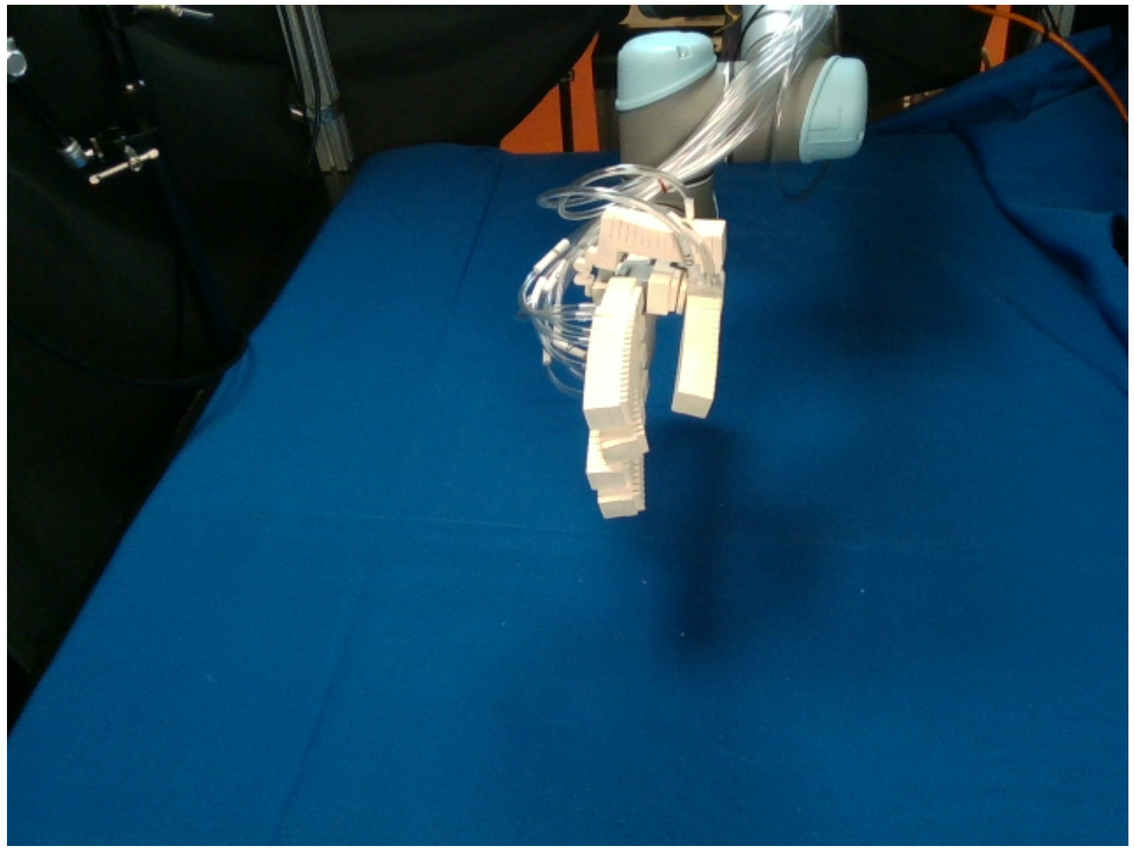
Visual Imitation Learning



Robotap: Tracking Arbitrary Points for Few-Shot Visual Imitation, Vecerik et al.

Proposal: Learn to Reconstruct a Jacobian Field, i.e., a function that maps every 3D point to its Body Jacobian!

Input: Single Image

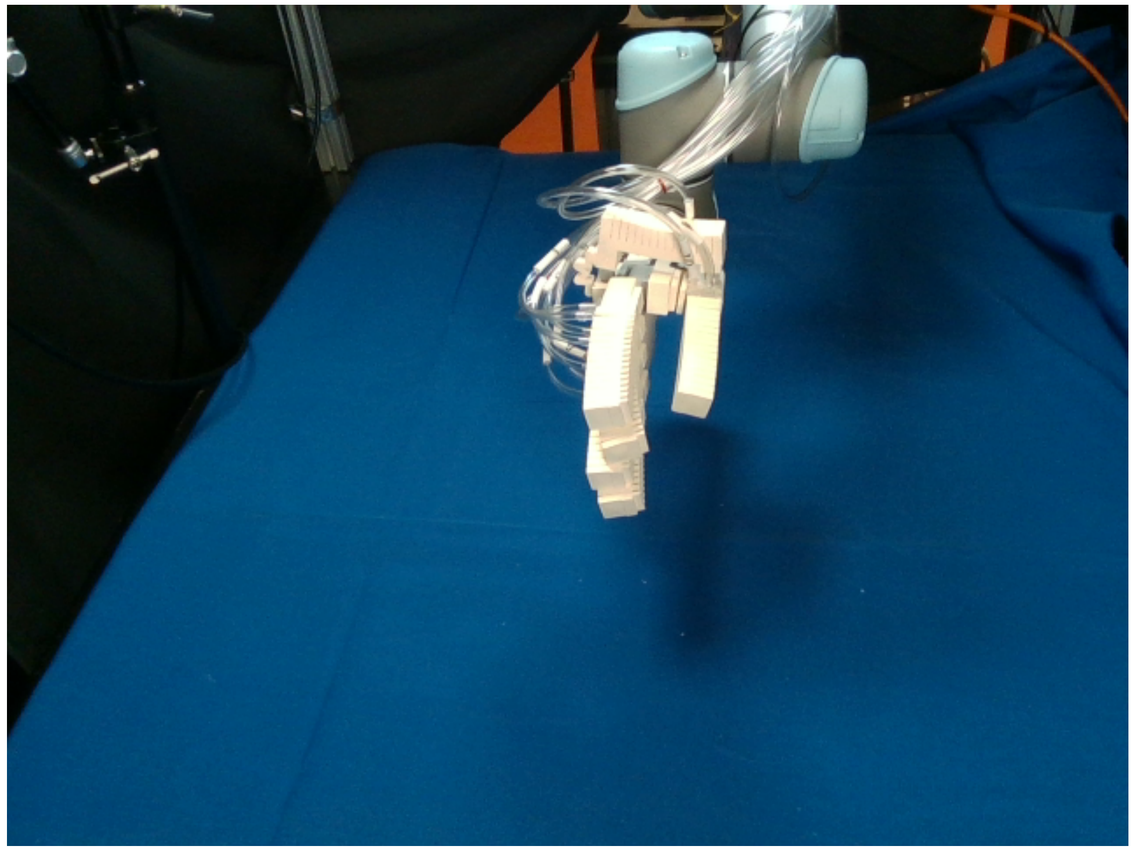


Neural Jacobian Field $\mathbf{J}(\mathbf{x}, \mathbf{I})$

Unifying 3D Representation and Control of Diverse Robots with a Single Camera, Li et al., 2024

Proposal: Learn to Reconstruct a Jacobian Field, i.e., a function that maps every 3D point to its Body Jacobian!

Input: Single Image



Neural Jacobian Field $\mathbf{J}(\mathbf{x}, \mathbf{I})$

Unifying 3D Representation and Control of Diverse Robots with a Single Camera, Li et al., 2024

Recap

- We went over prerequisites & other logistics.
- We discussed what problem scientists have studied historically in computer vision.
- We discussed *survivorship bias* in the tasks that concern the vision community at any point in time.
- We discussed a broader scope of vision.
- We discussed the broad modules of this class.