Computer vision systems that work: Movies, Games, Mixed Reality

ANDREW FITZGIBBON

@awfidius

Microsoft

All Data AI

"Again, [the Difference Engine] might act on other things beside *number*." — Ada Lovelace

The ADA group studies AI models and techniques that apply to complex real-world data: big data, small data, data with complex structures like trees and graphs. In short: "All Data AI".

Linking language and vision with data from structured domains such as programming languages, bioinformatics, chemistry, or the web, we explore all aspects of machine learning and related disciplines. We research in Bayesian inference, Deep Learning, Optimization and Programming Languages, with multiple prize-winning papers and research. And we have contributed to some of the most advanced technologies that Microsoft builds, with contributions to Xbox, Office, HoloLens, Bing, Visual Studio, and many others. Our superpower is our combined focus on fundamental research and real-world engineering, to push the boundaries of AI.

All Data AI

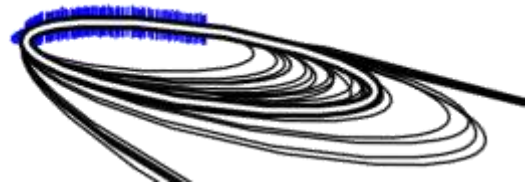
"Again, [the Difference Engine] might act on other things beside *number*." — Ada Lovelace

The ADA group studies AI models and techniques that apply to complex real-world data: big data, small data, data with complex structures like trees and graphs. In short: "All Data AI".

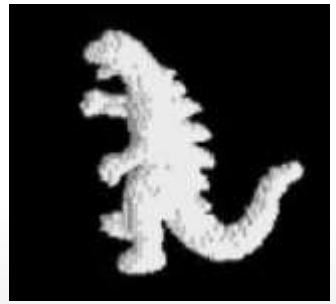
Pinch Slider
Touch Slider

- School and university in Cork, Ireland
  - First job: Water-taxi driver

- Master's and PhD in Edinburgh, Scotland
  - Topic: 2D shapes, mainly the ellipse



- Researcher in Engineering, Oxford University
  - 3D vision—things that stay still

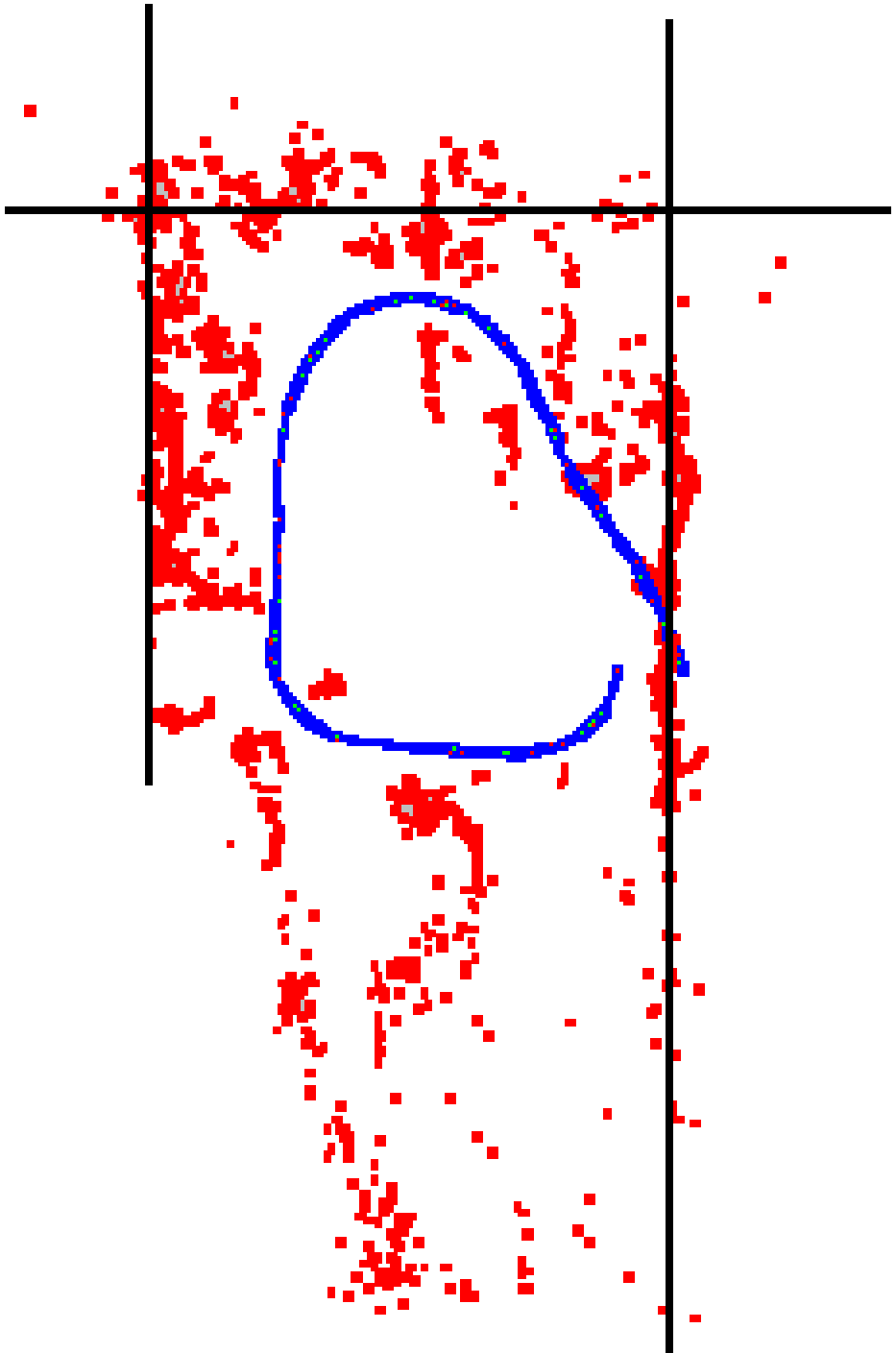


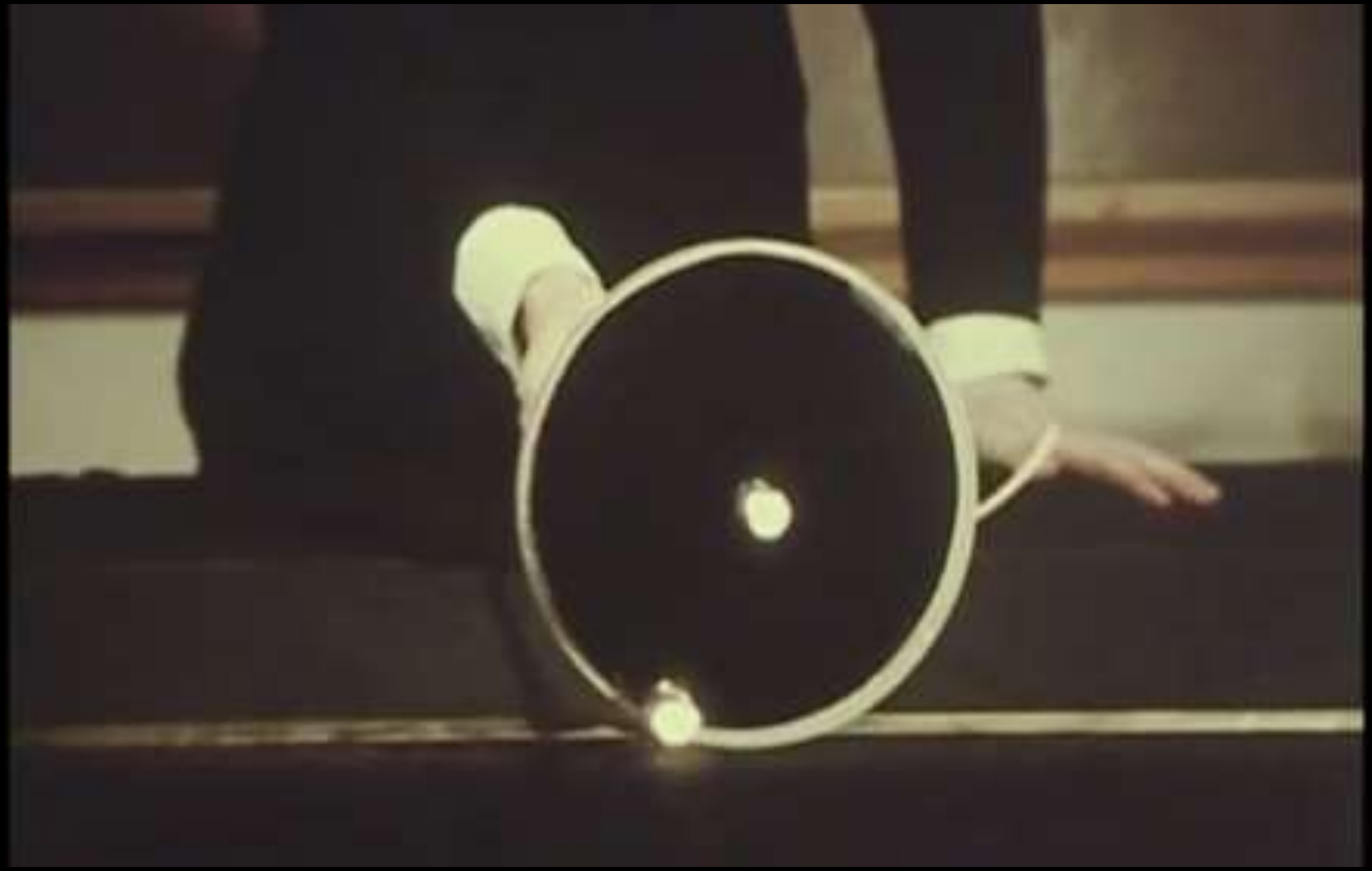
- Researcher and Scientist, Microsoft
  - 3D vision—things that move

1996

UNIVERSITY OF OXFORD
ROBOTICS RESEARCH GROUP

Microsoft

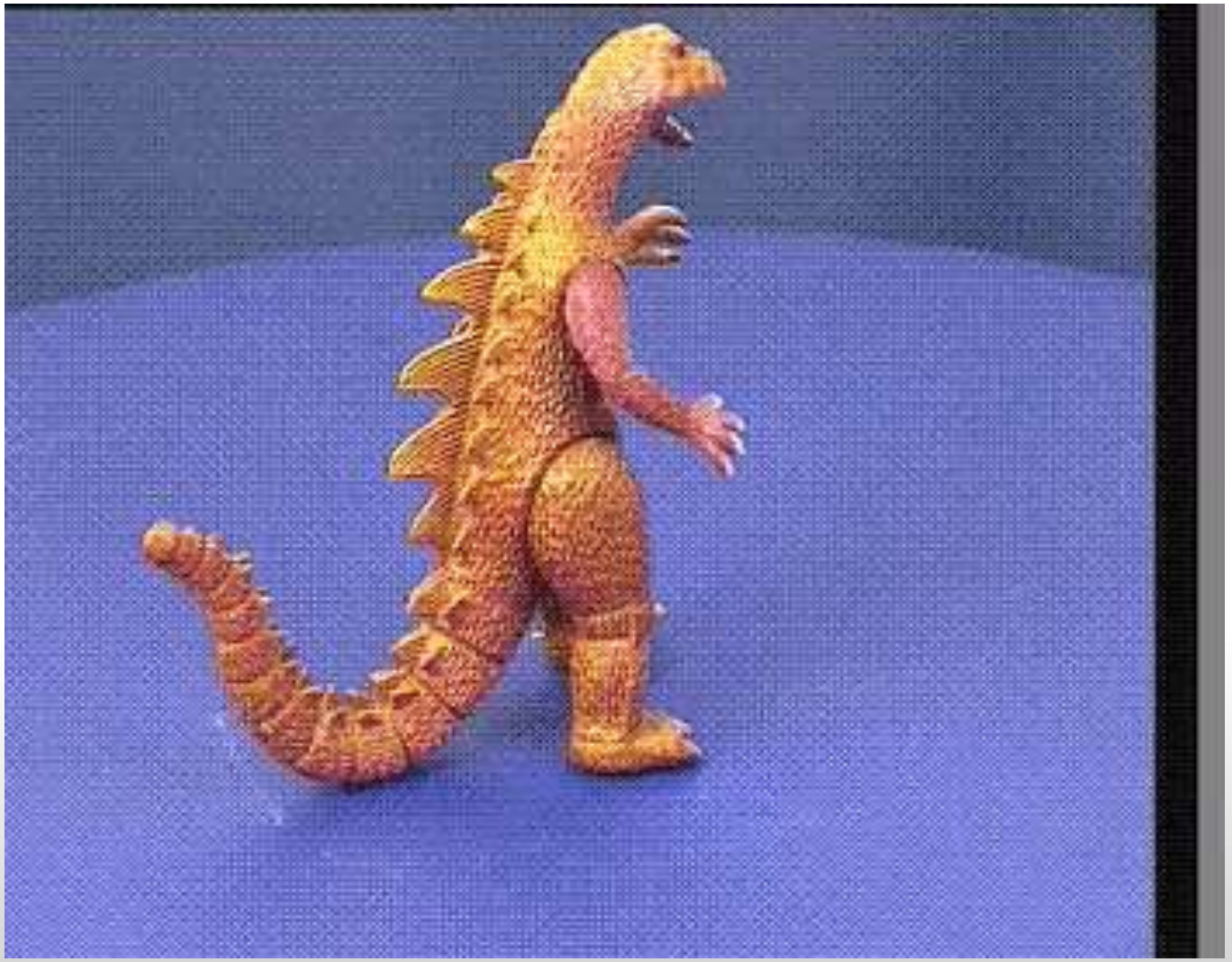




G Johansson, 1973

G Johansson, 1973

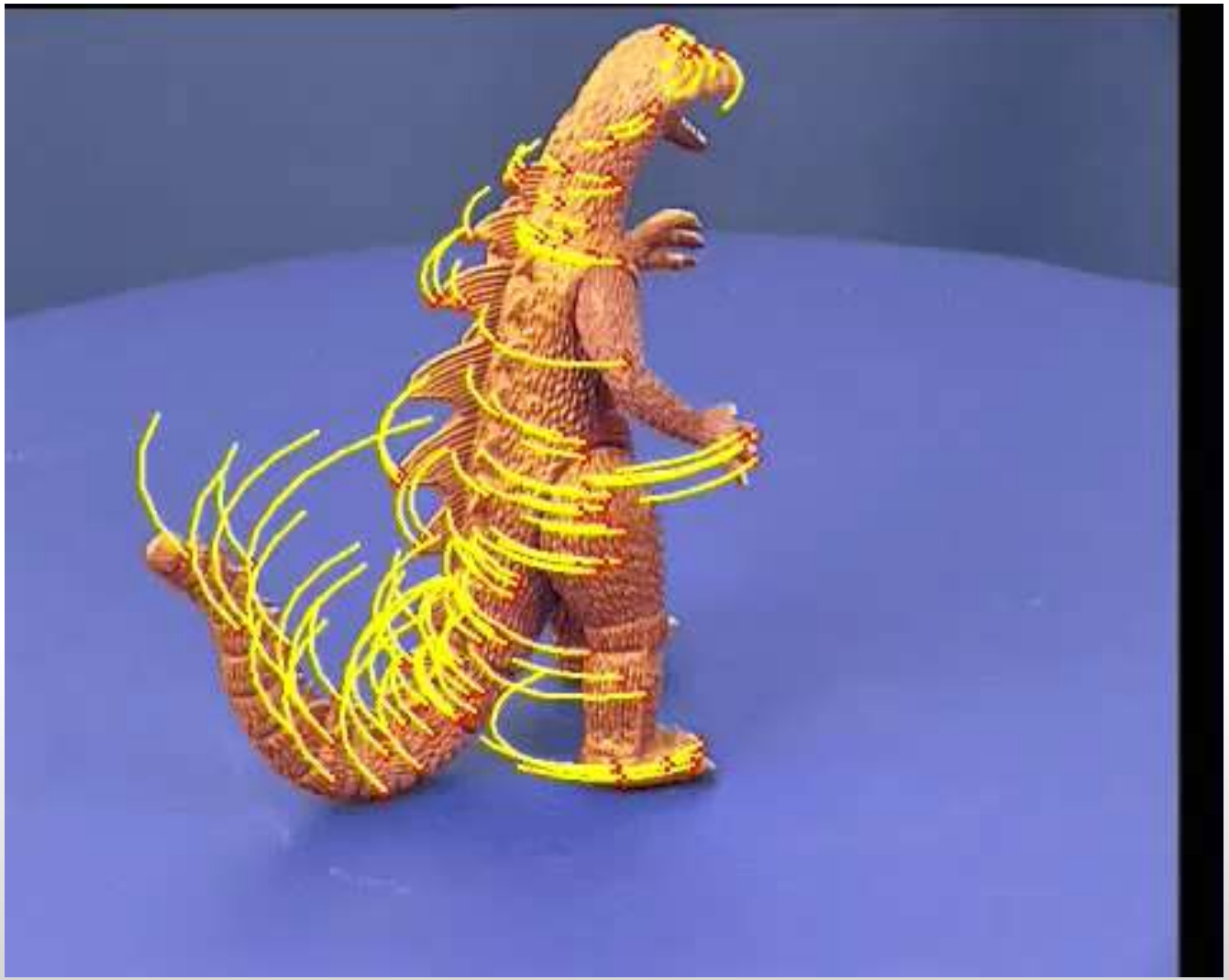
3D modelling from uncalibrated turntable sequences



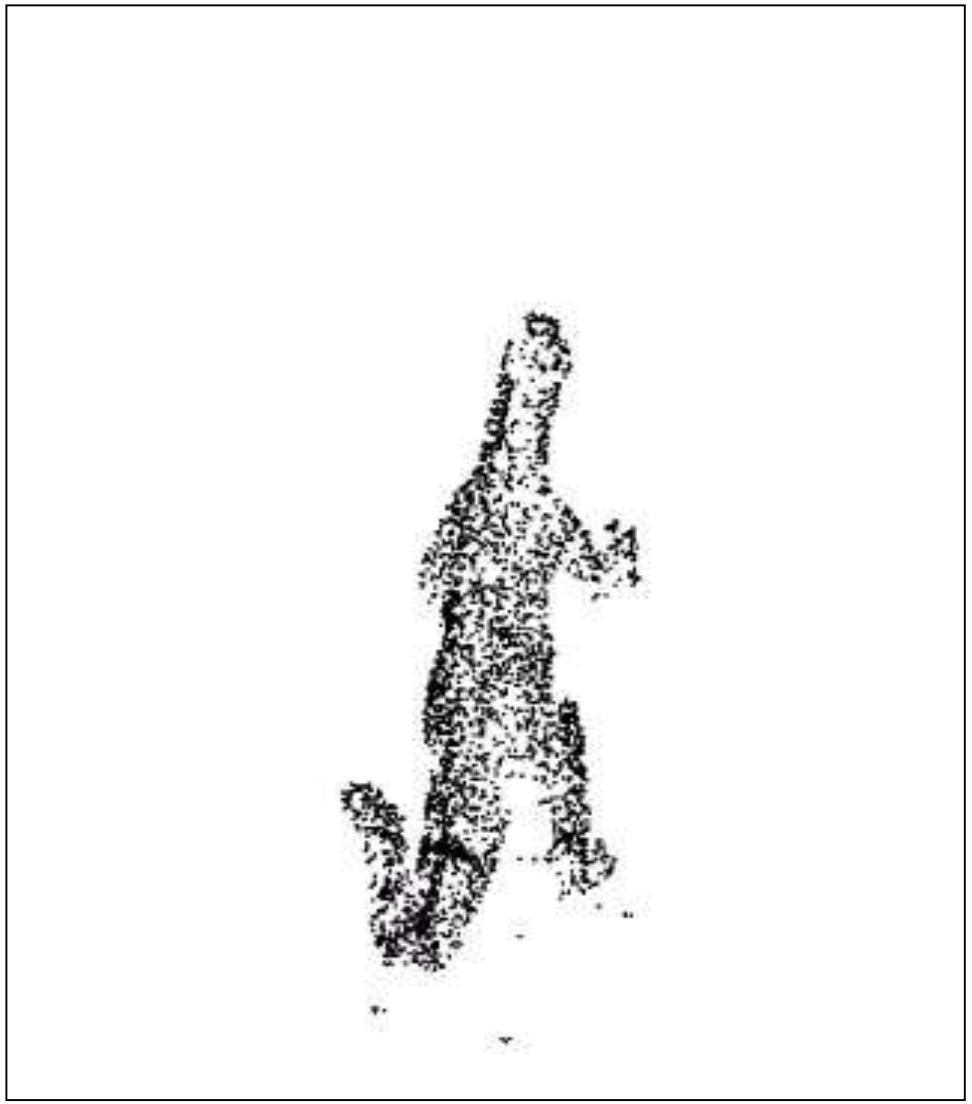
CASE STUDY (1998)

Microsoft

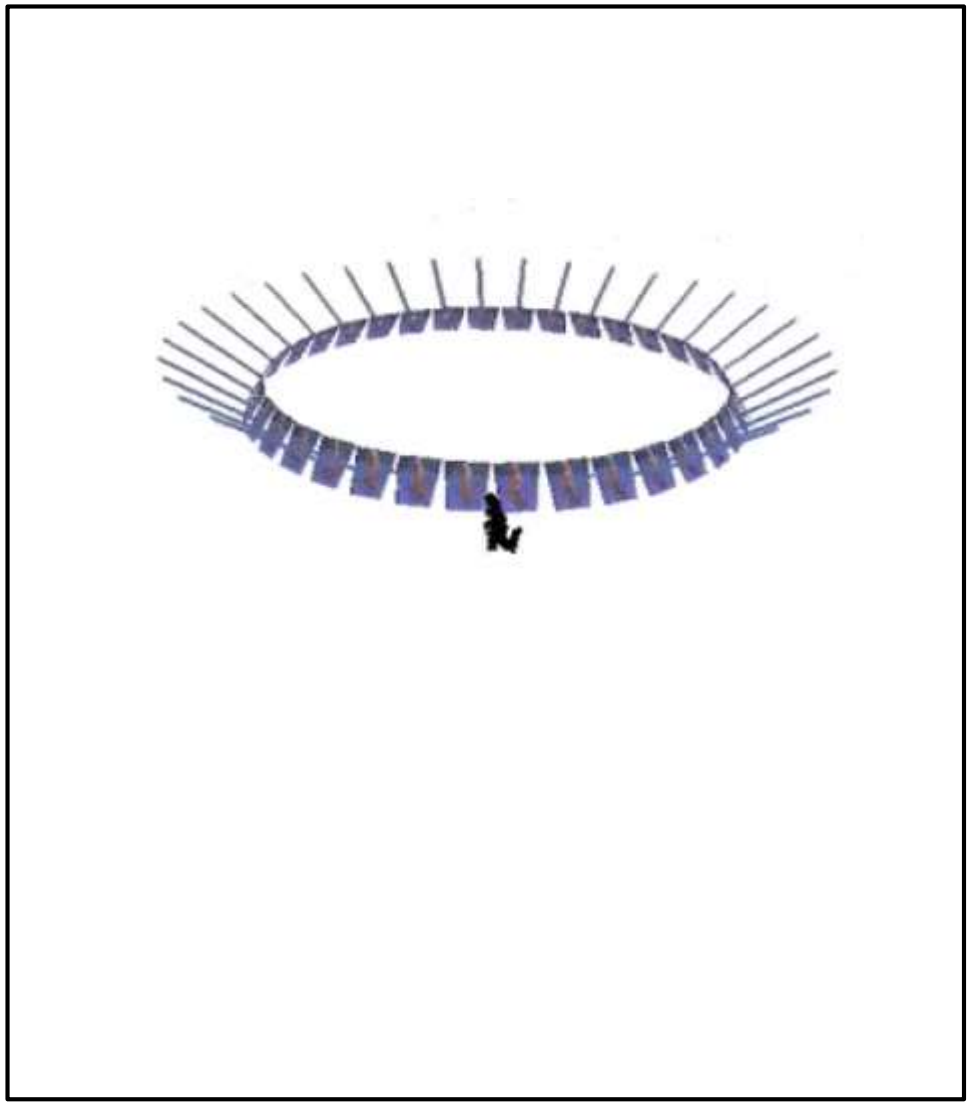
Automatic 2D point tracking

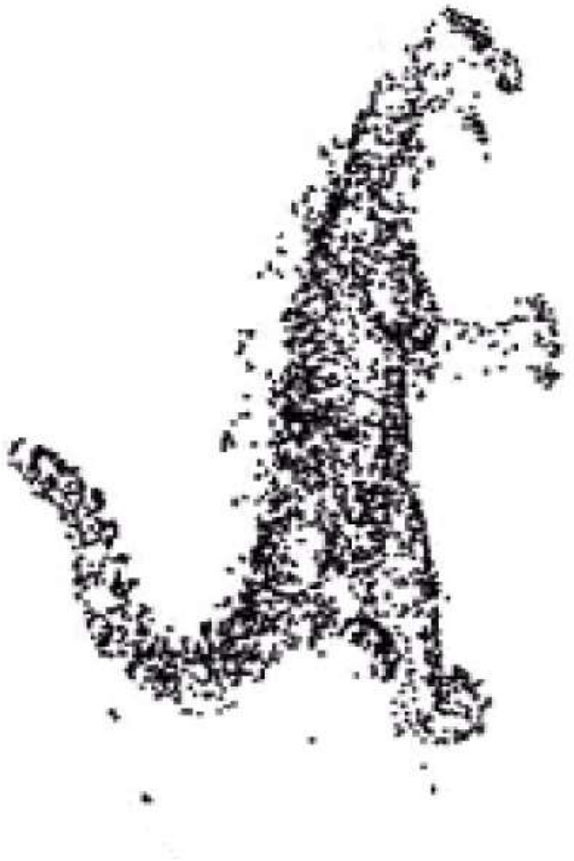


Automatically determined calibration and 3D structue

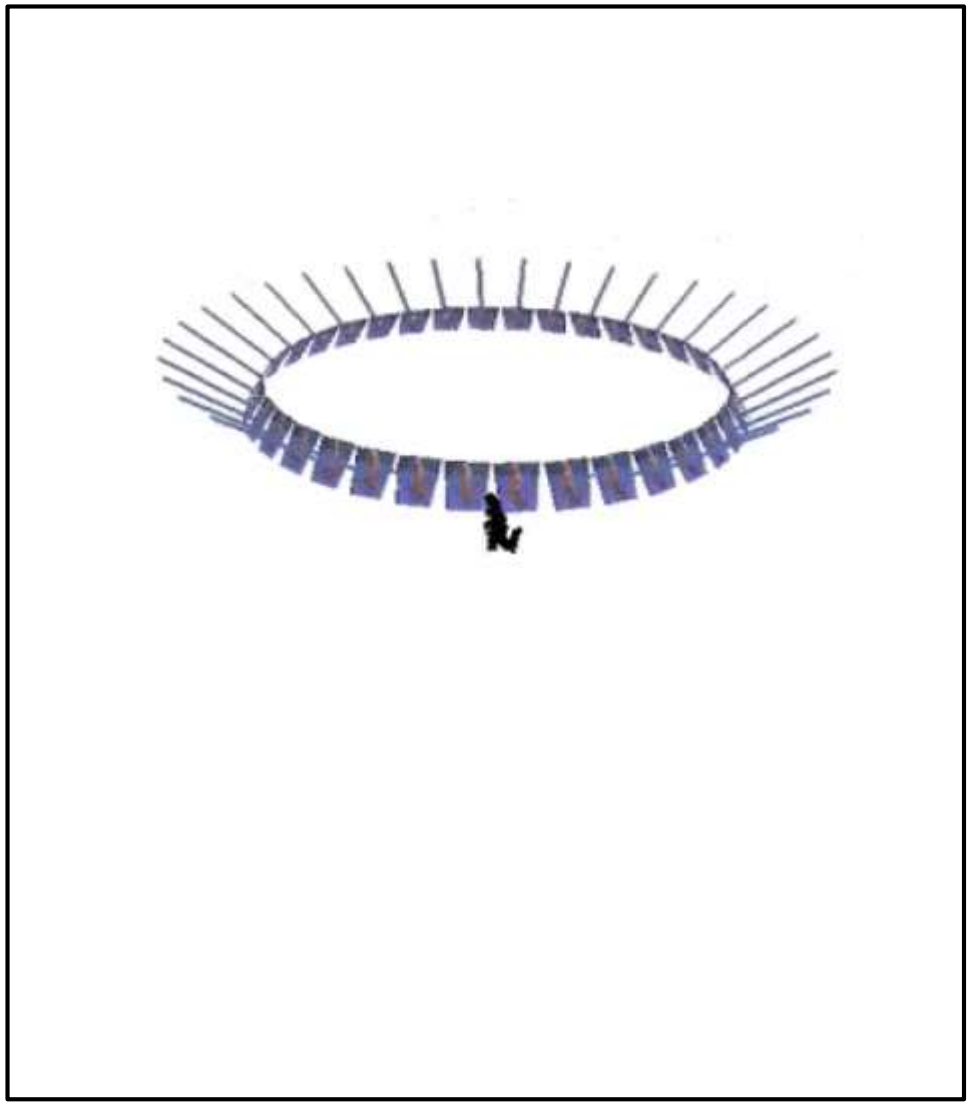


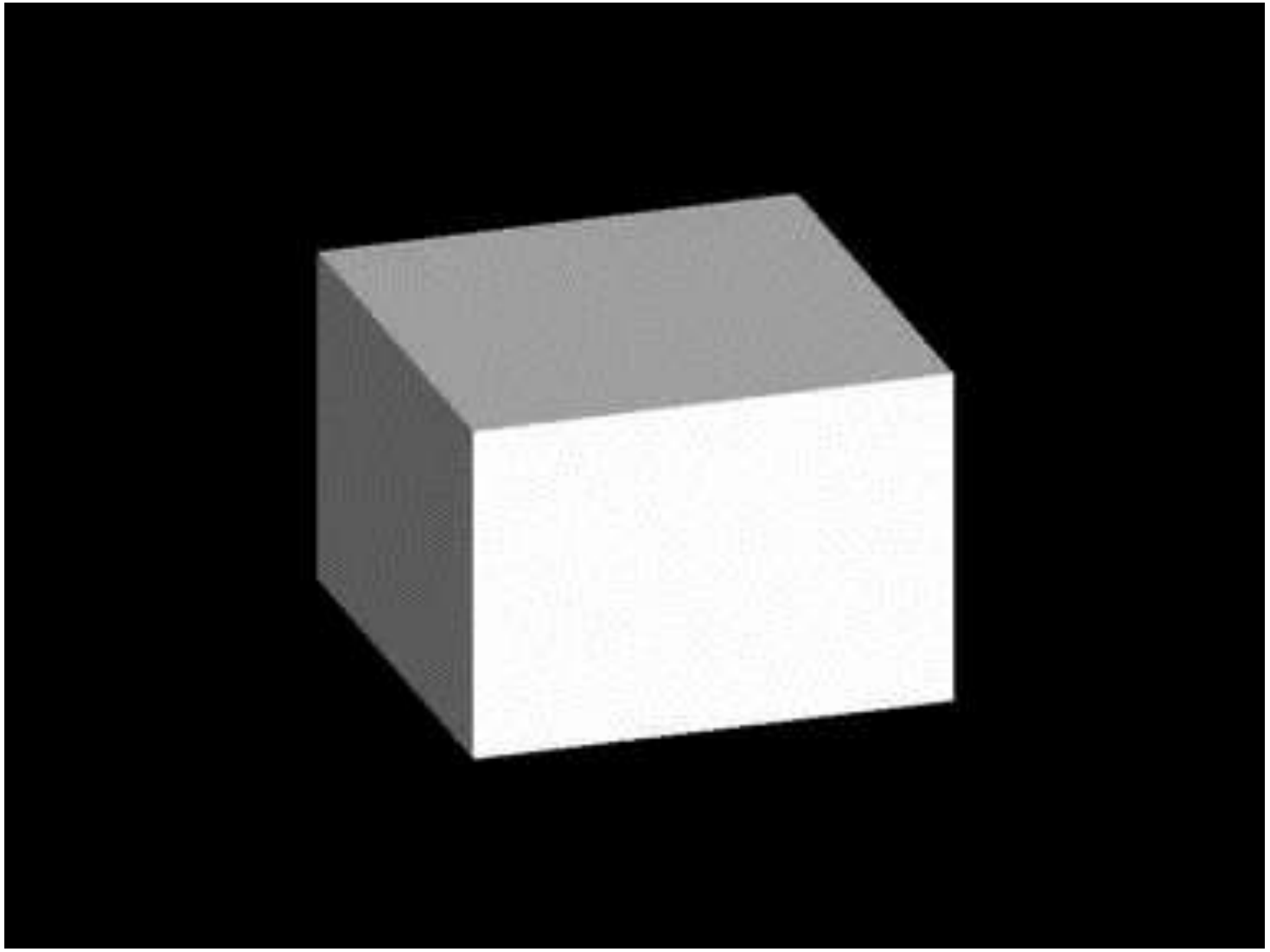
Automatically determined calibration and 3D structue





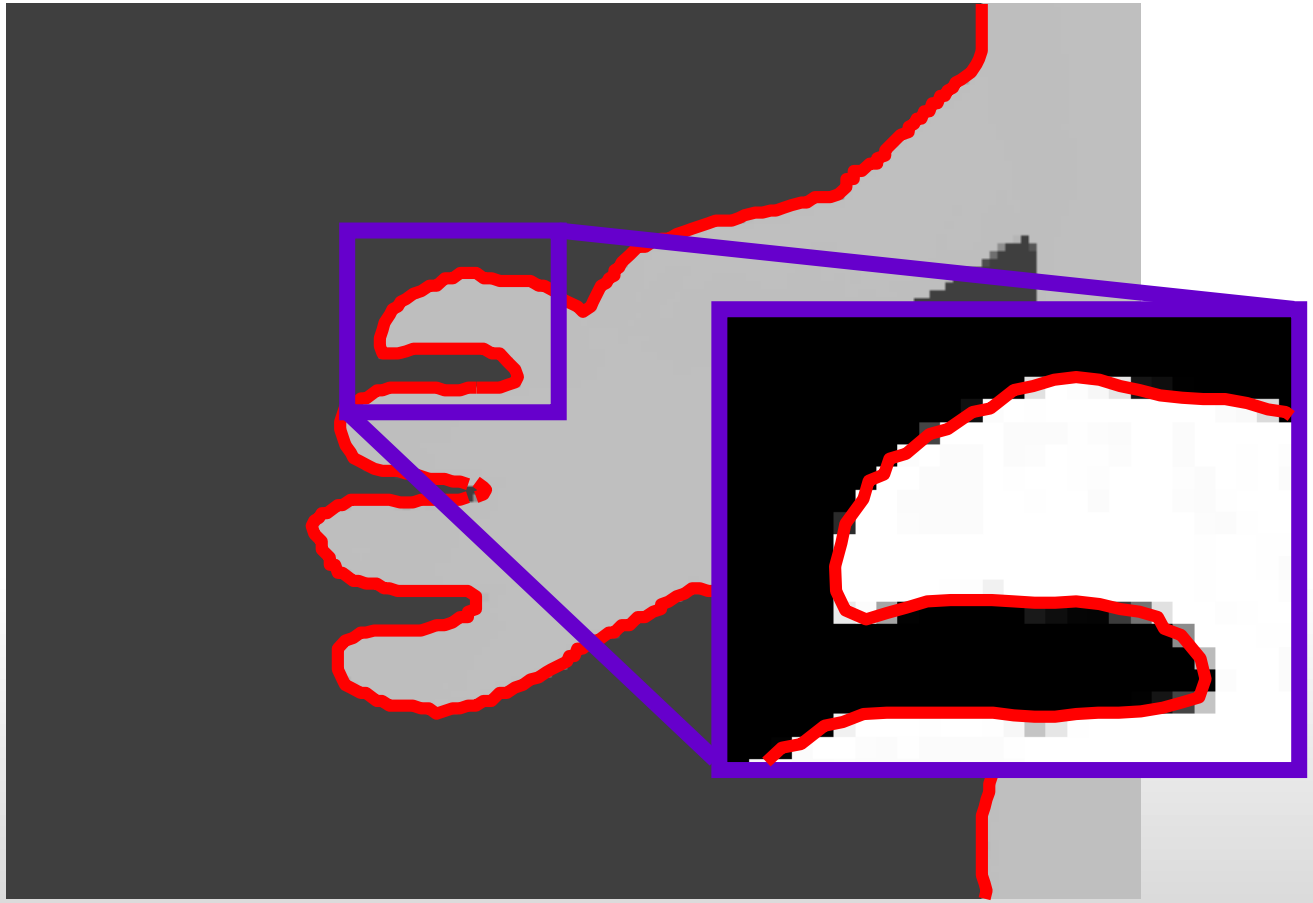
Automatically determined calibration and 3D structue

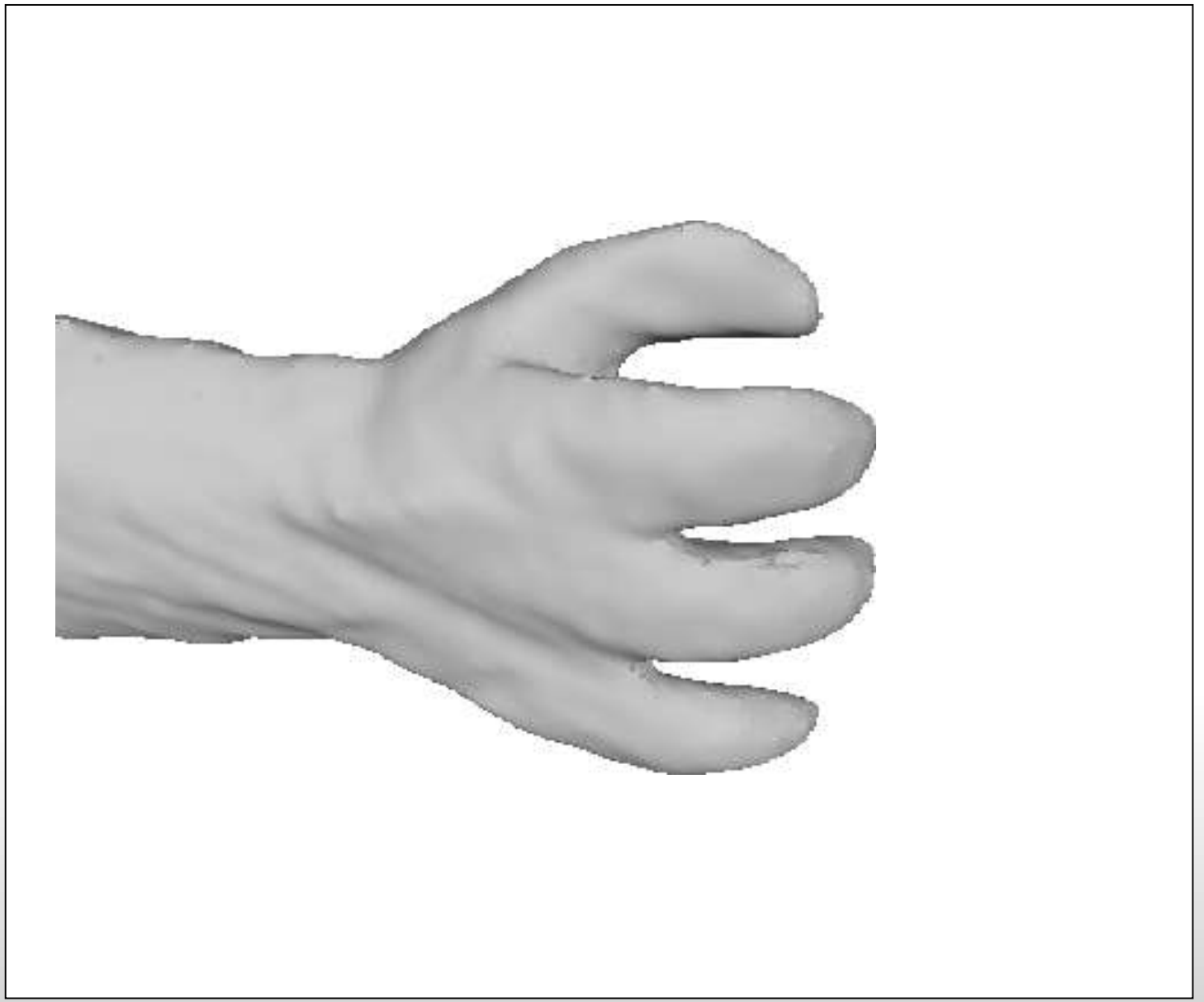




Automatically recovered model: polygon mesh

Subpixel silhouettes for subvoxel marching cubes





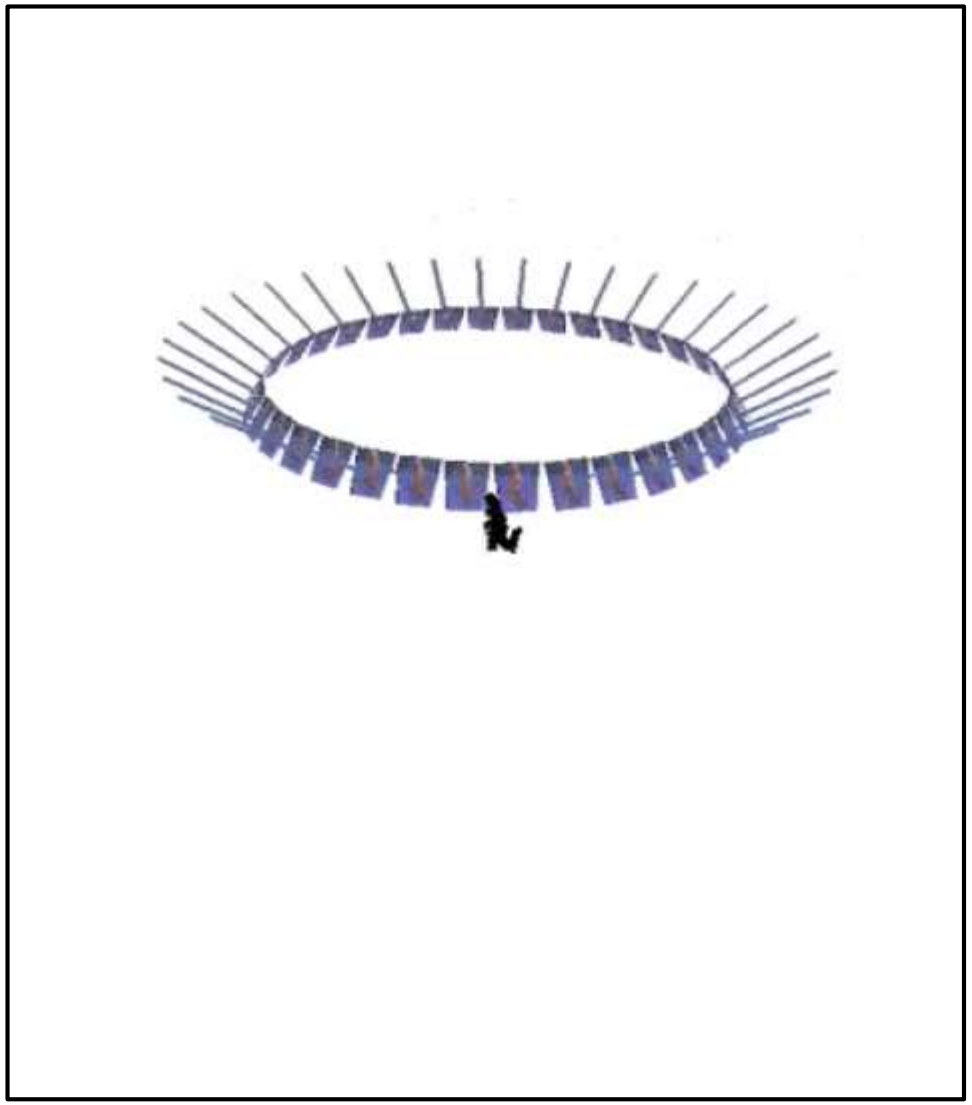
CASE STUDY

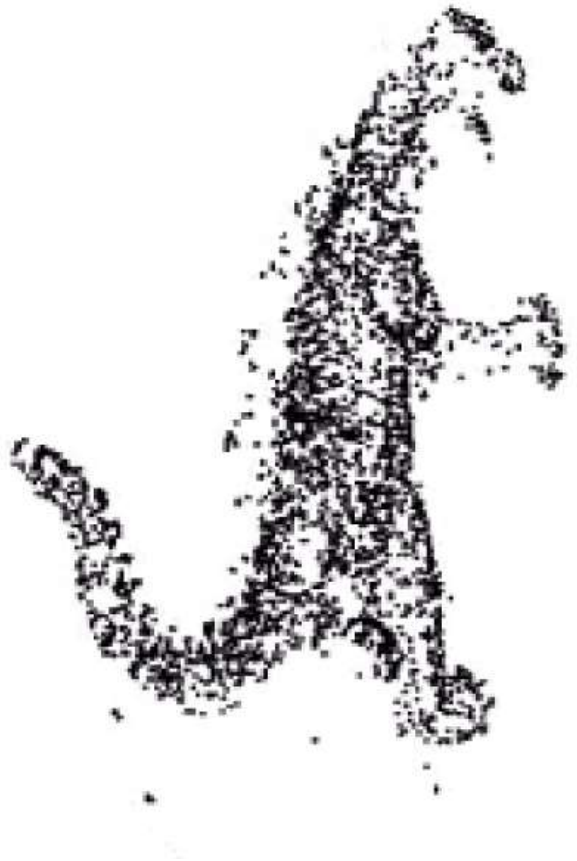
Microsoft

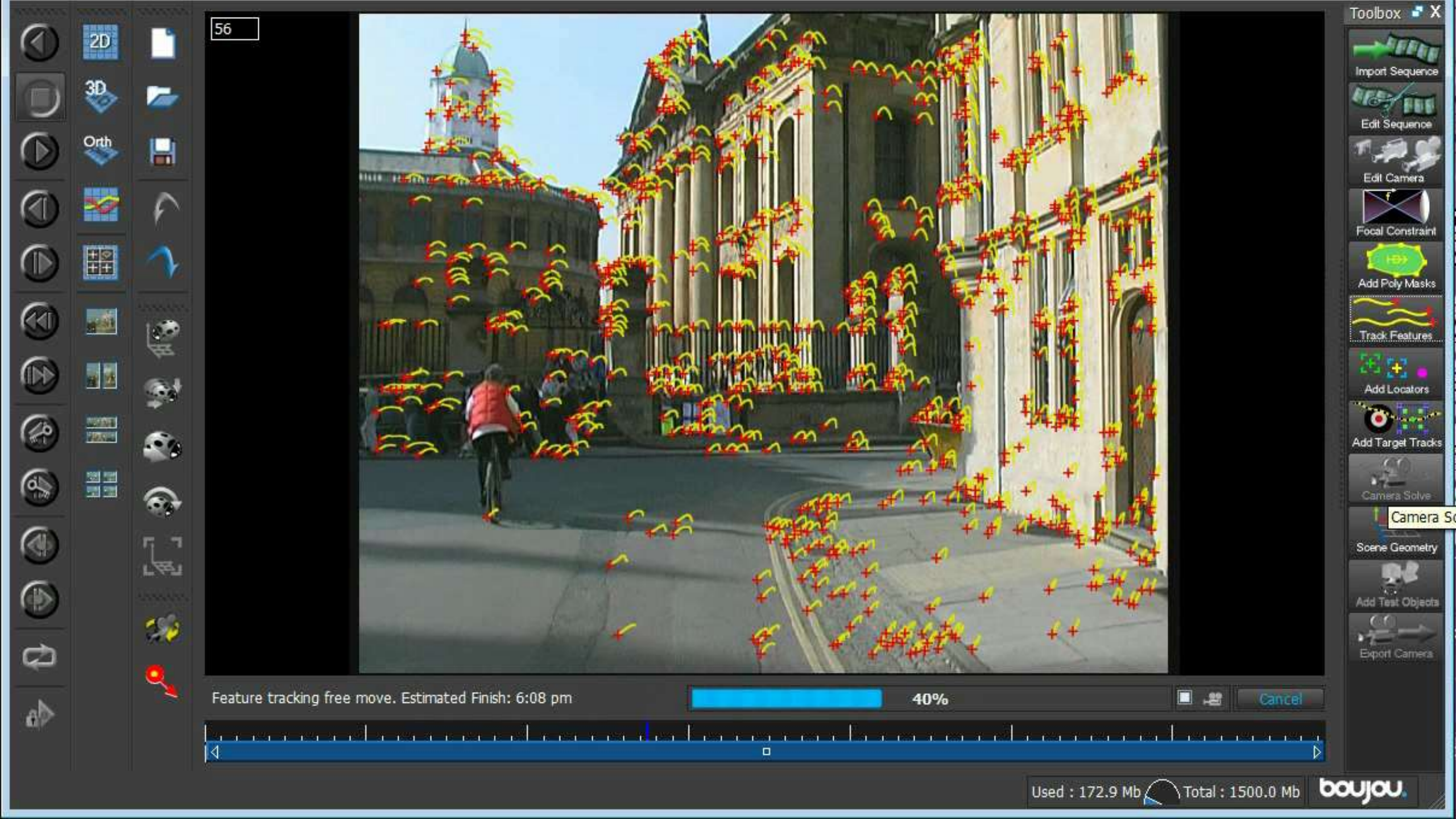
- "Let's make a computer vision startup!"
  - Somewhat unusual in 1999

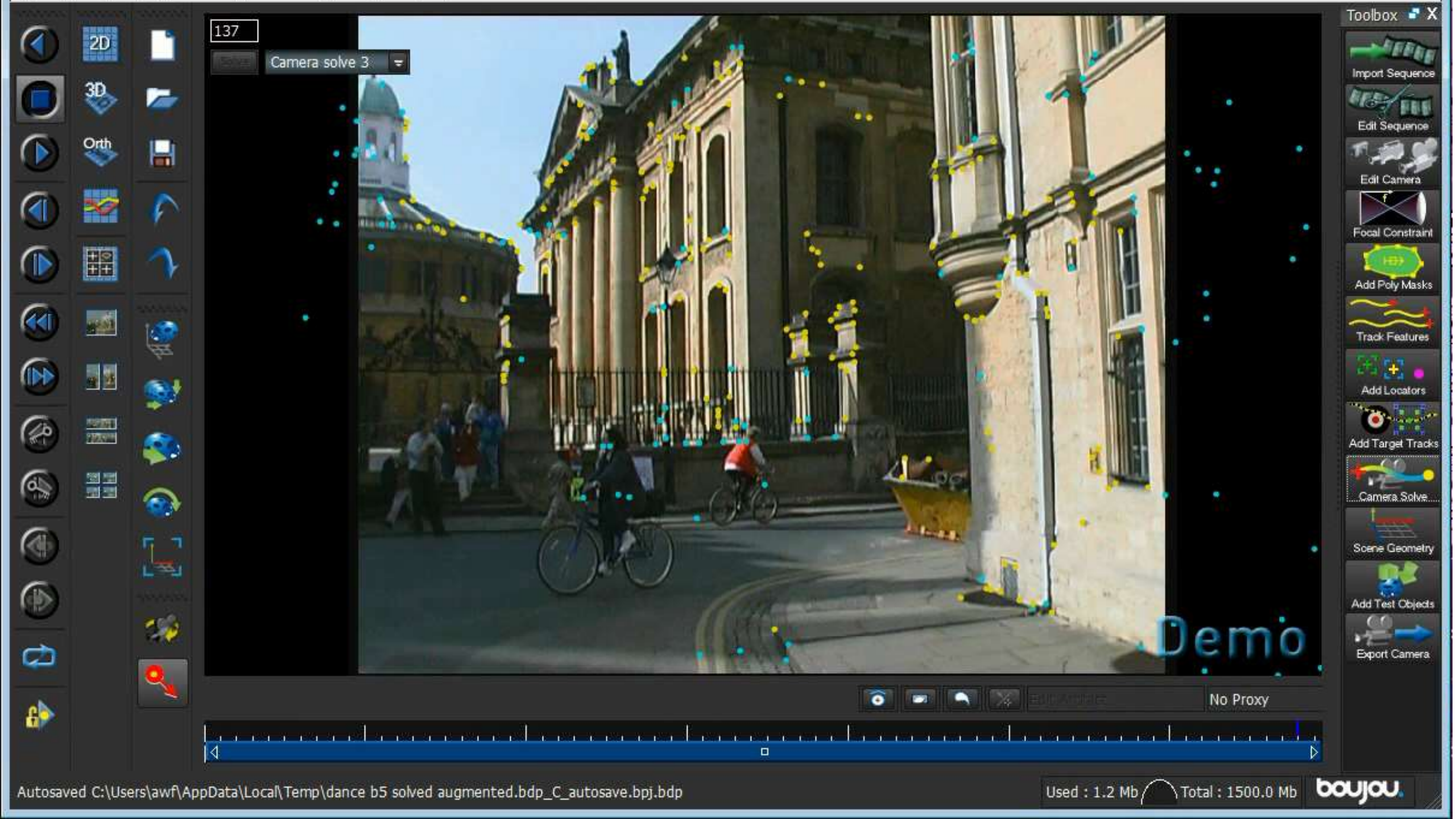
- And nearly a very bad idea…
  - Industries that need 3D models are quite particular

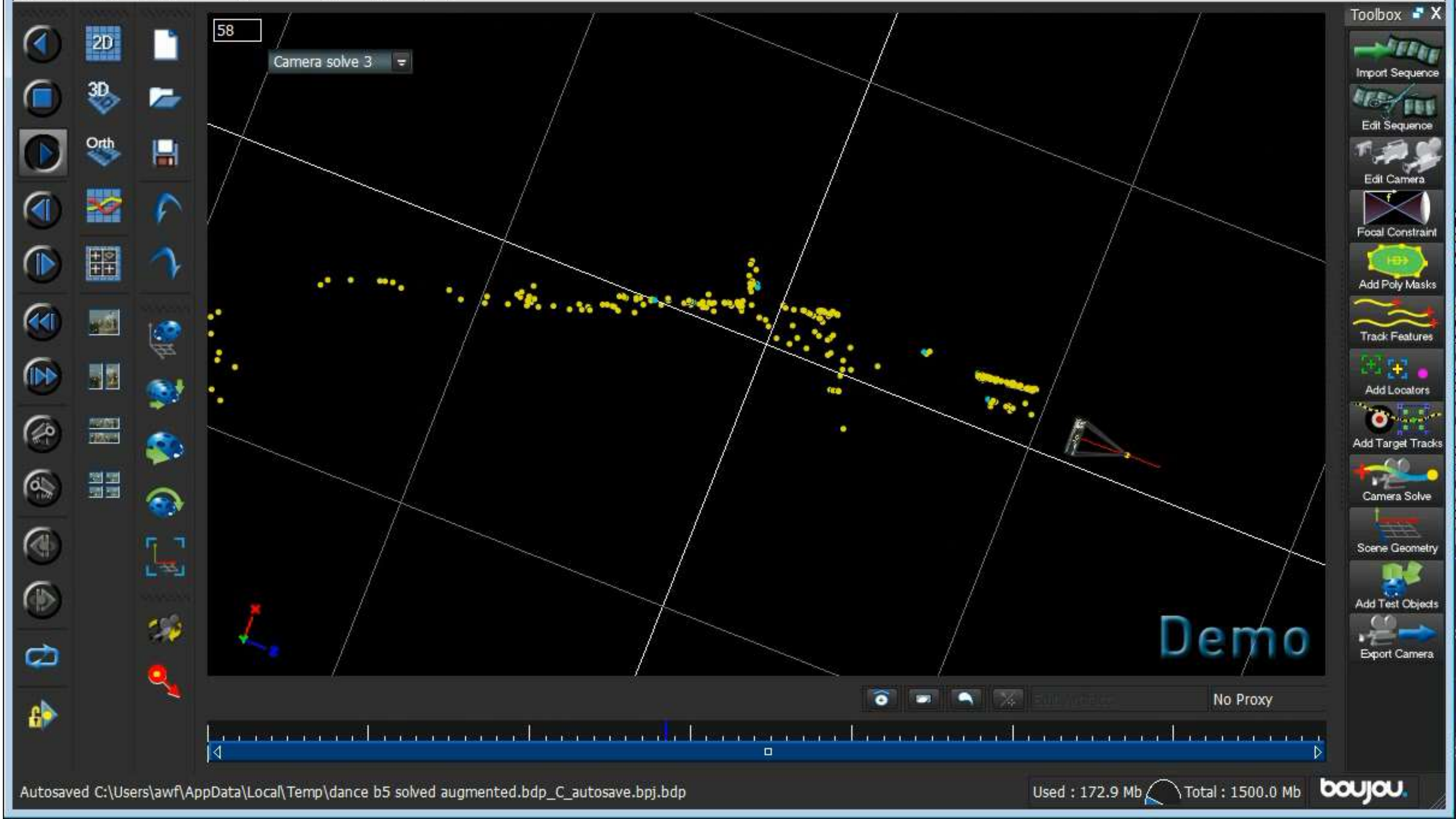
Automatically determined calibration and 3D structure

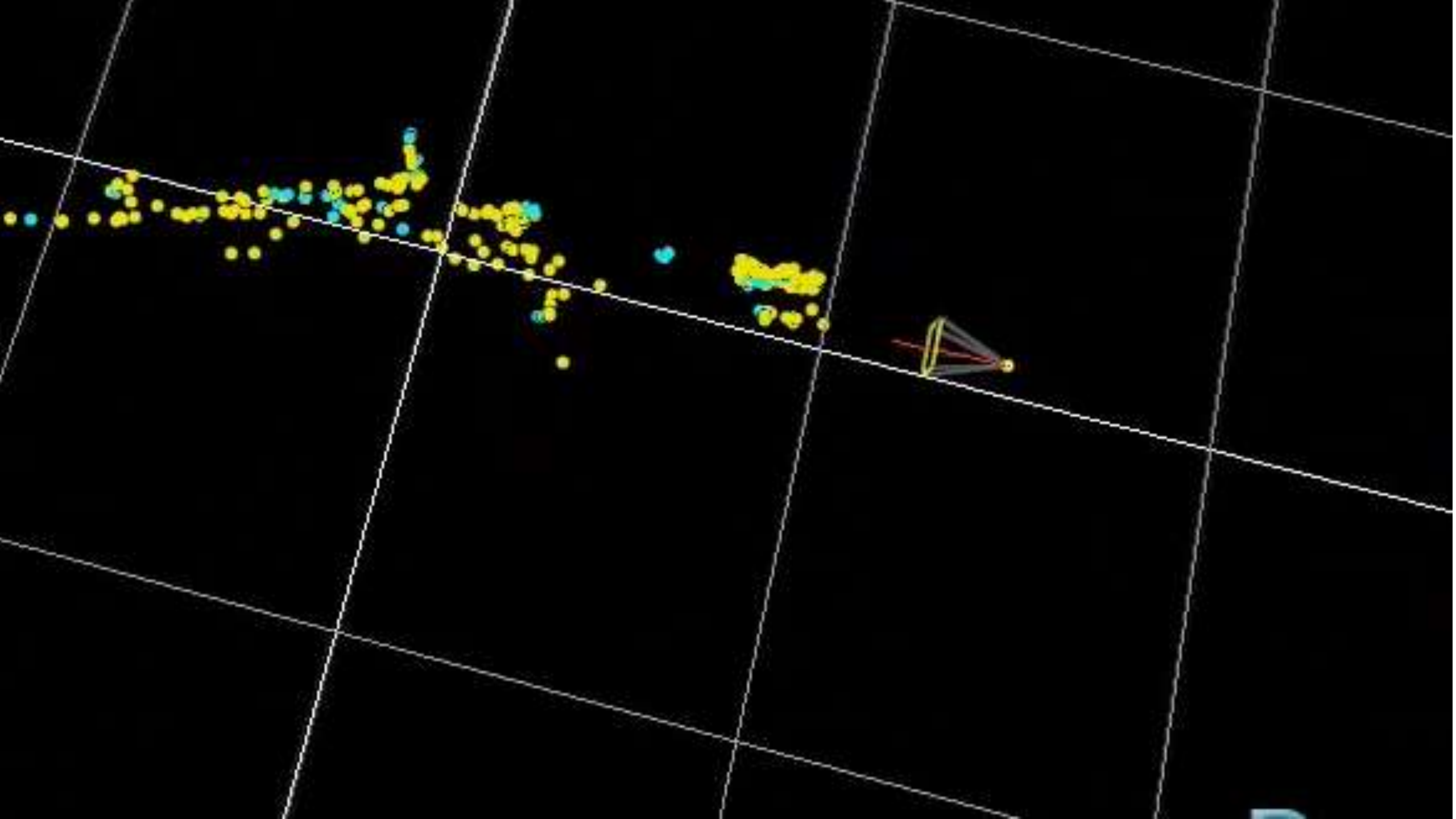


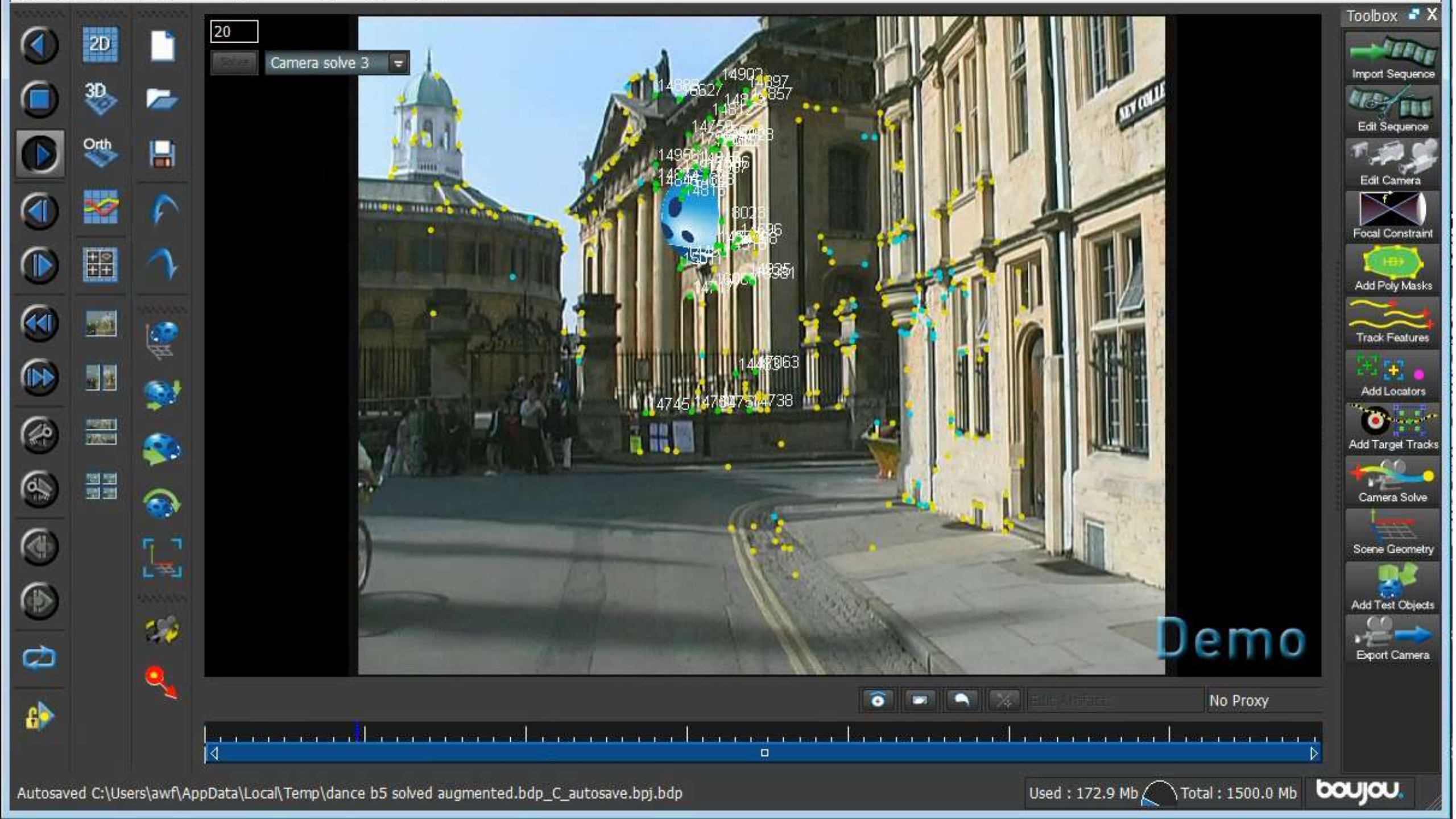



56
2D
3D
Orth
Toolbox
Import Sequence
Edit Sequence
Edit Camera
Focal Constraint
Add Poly Masks
Track Features
Add Locators
Add Target Tracks
Camera Solve
Scene Geometry
Add Test Objects
Export Camera
Feature tracking free move. Estimated Finish: 6:08 pm
40%
Cancel
Used : 172.9 Mb
Total : 1500.0 Mb
boujou.


137
Camera solve 3
Toolbox
Import Sequence
Edit Sequence
Edit Camera
Focal Constraint
Add Poly Masks
Track Features
Add Locators
Add Target Tracks
Camera Solve
Scene Geometry
Add Test Objects
Export Camera
Demo
No Proxy
Autosaved C:\Users\awf\AppData\Local\Temp\dance b5 solved augmented.bdp_C_autosave.bpj.bdp
Used : 1.2 Mb
Total : 1500.0 Mb
boujou.


58
Camera solve 3
Toolbox
Import Sequence
Edit Sequence
Edit Camera
Focal Constraint
Add Poly Masks
Track Features
Add Locators
Add Target Tracks
Camera Solve
Scene Geometry
Add Test Objects
Export Camera
Demo
No Proxy
Autosaved C:\Users\awf\AppData\Local\Temp\dance b5 solved augmented.bdp_C_autosave.bpj.bdp
Used : 172.9 Mb
Total : 1500.0 Mb
boujou.




20
Camera solve 3
2D
3D
Orth
Toolbox
Import Sequence
Edit Sequence
Edit Camera
Focal Constraint
Add Poly Masks
Track Features
Add Locators
Add Target Tracks
Camera Solve
Scene Geometry
Add Test Objects
Export Camera
Demo
No Proxy
Autosaved C:\Users\awf\AppData\Local\Temp\dance b5 solved augmented.bdp_C_autosave.bpj.bdp
Used : 172.9 Mb
Total : 1500.0 Mb
boujou

How?

1. Gather 100 test videos
2. Build metrics
3. Implement all known algorithms
4. Fix the broken metrics
5. Add 100 more videos

Troy (2004)

Warner Bros, MPC

Troy (2004)

Warner Bros, MPC

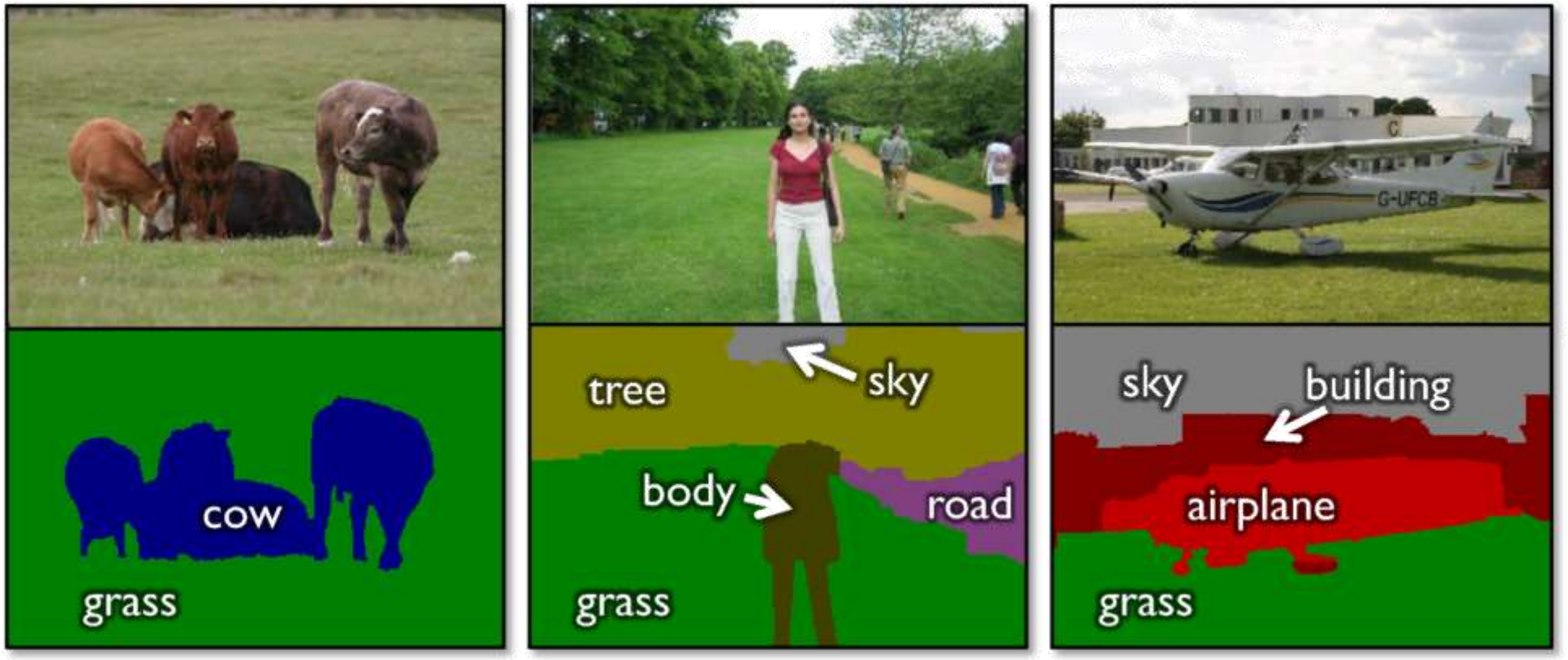
- A computer program which makes inserting 3D objects easy
- Developed at Oxford University and at company "2d3"
- Now used in almost every movie
  - *Lord of the Rings* series
  - *Harry Potter* series
  - *District 9*

and, of course…. *Bridget Jones's Diary*

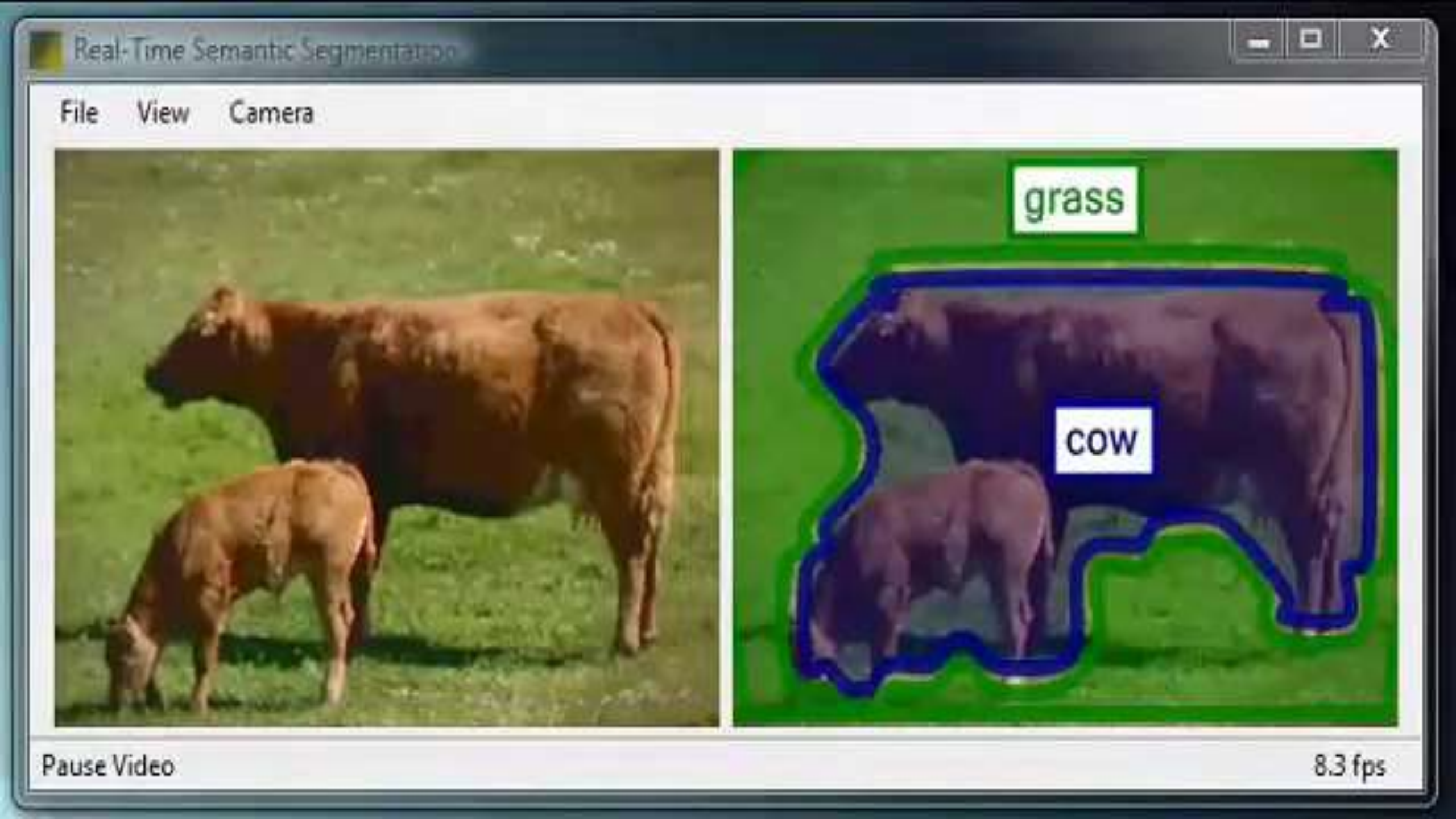
4 wide-angle tracking cameras

2004

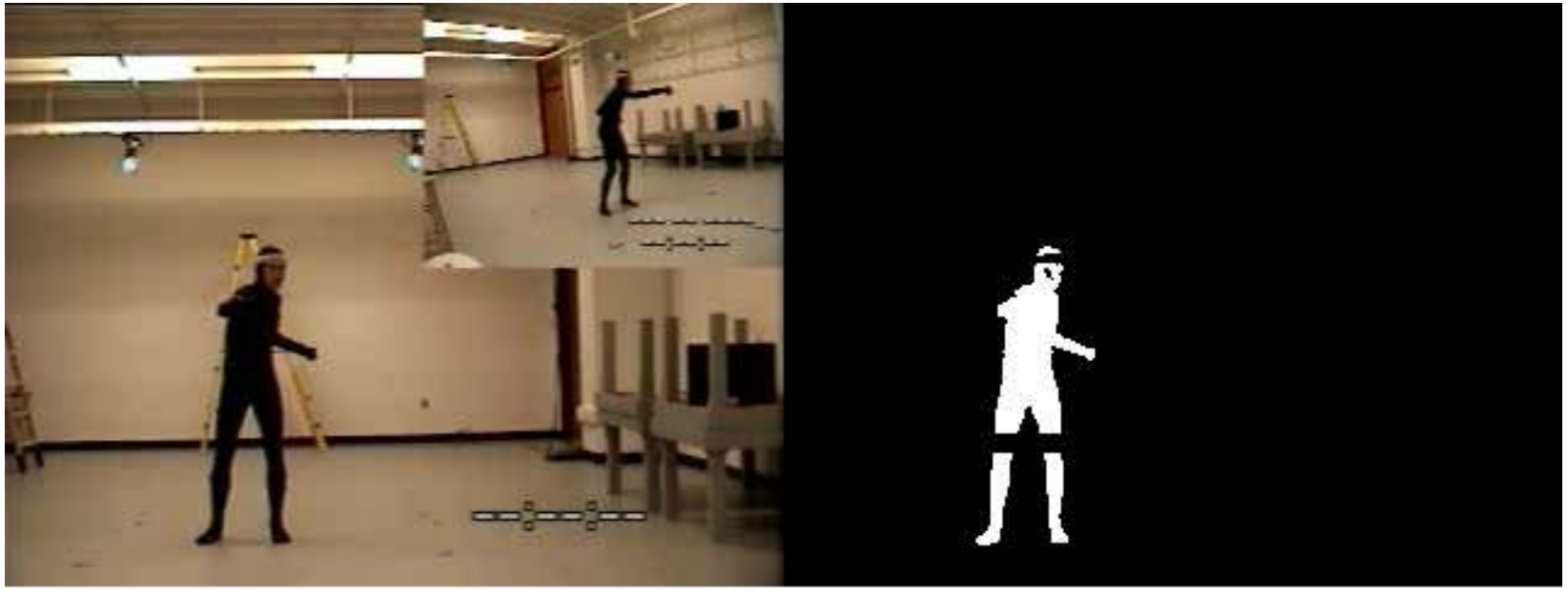
THINGS THAT MOVE



J. Shotton, J. Winn, C. Rother, A. Criminisi, ***TextonBoost*: Joint Appearance, Shape and Context Modeling for Multi-Class Object Recognition and Segmentation.** European Conference on Computer Vision, 2006


Real-Time Semantic Segmentation
File
View
Camera
grass
cow
Pause Video
8.3 fps

- ☑ Wide range of motion
  - ☒ But limited agility
  - ☒ And not realtime



R Navaratnam, A Fitzgibbon, R Cipolla
The Joint Manifold Model for Semi-supervised Multi-valued Regression
IEEE Intl Conf on Computer Vision, 2007

“We need a body tracker with

- ☑ All motions…
- ☑ All agilities…
- ☑ 10x Realtime…
- ☑ For multiple players…

“We need a body tracker with

☑ All motions…

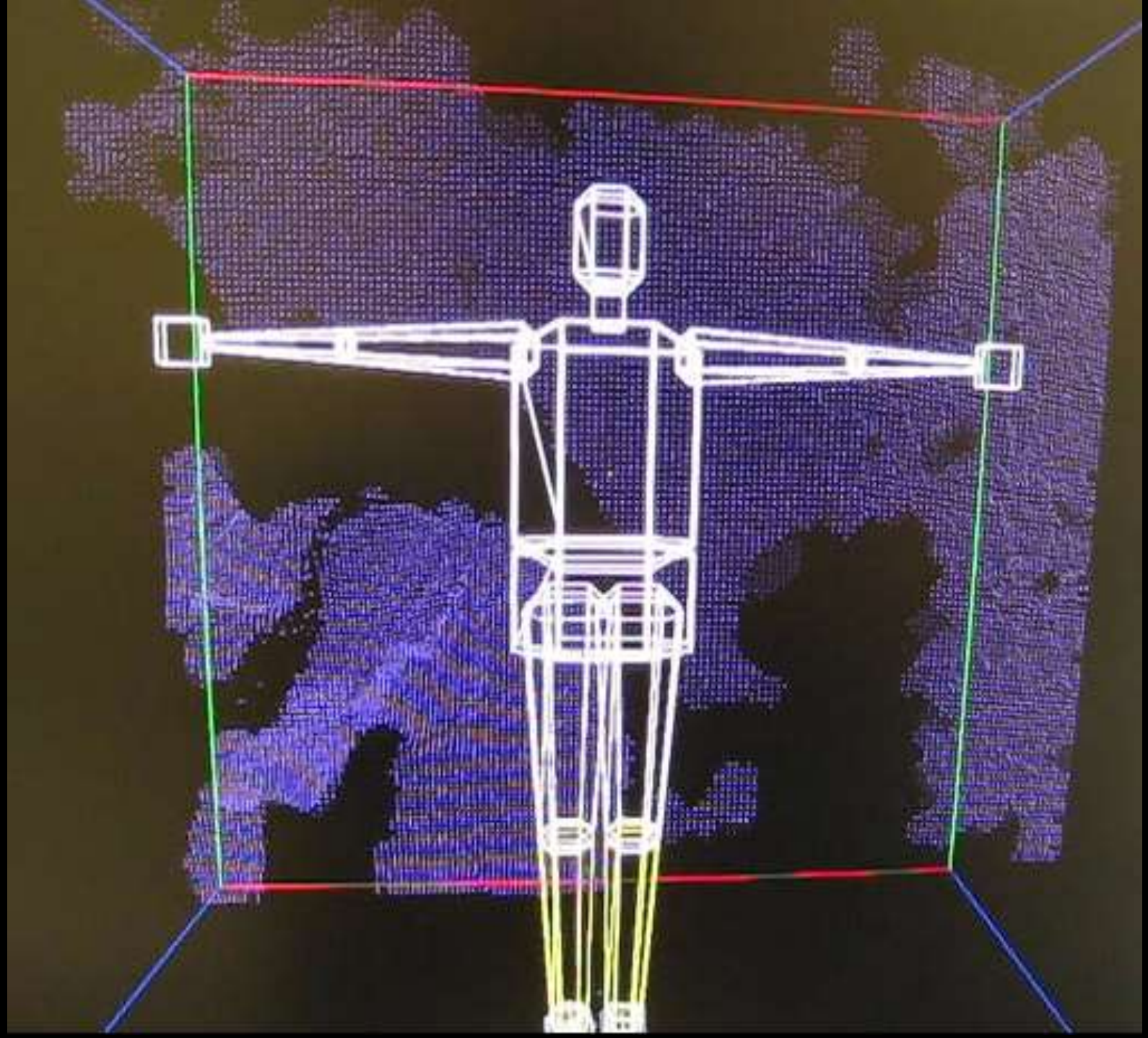
☑ All agilities…

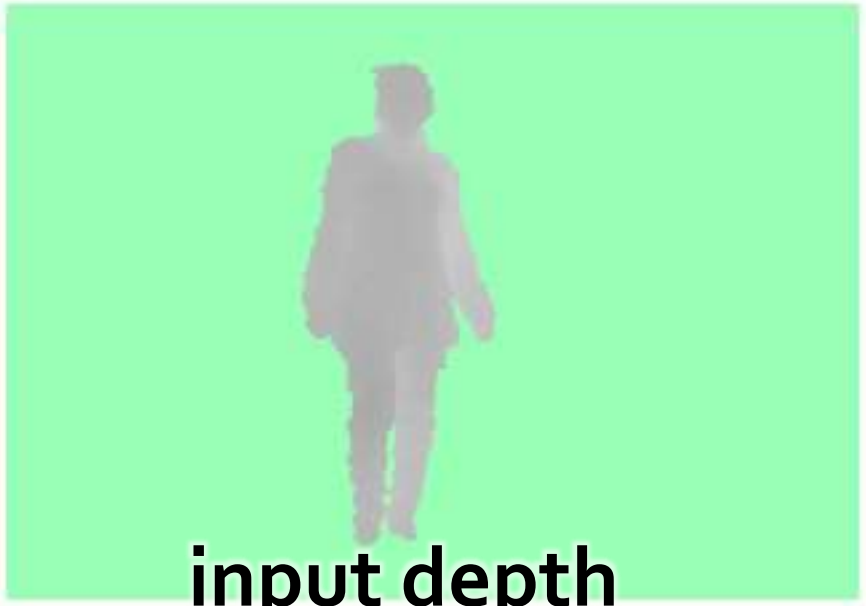
☑ 10x Realtime…

☑ For multiple players…

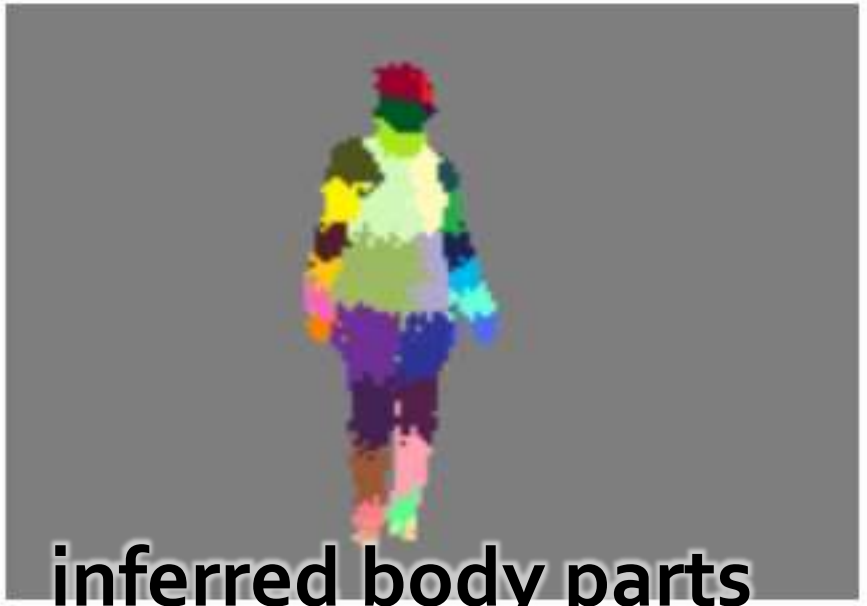
… but you have got 3D ☺”

"… but you have got 3D ☺"

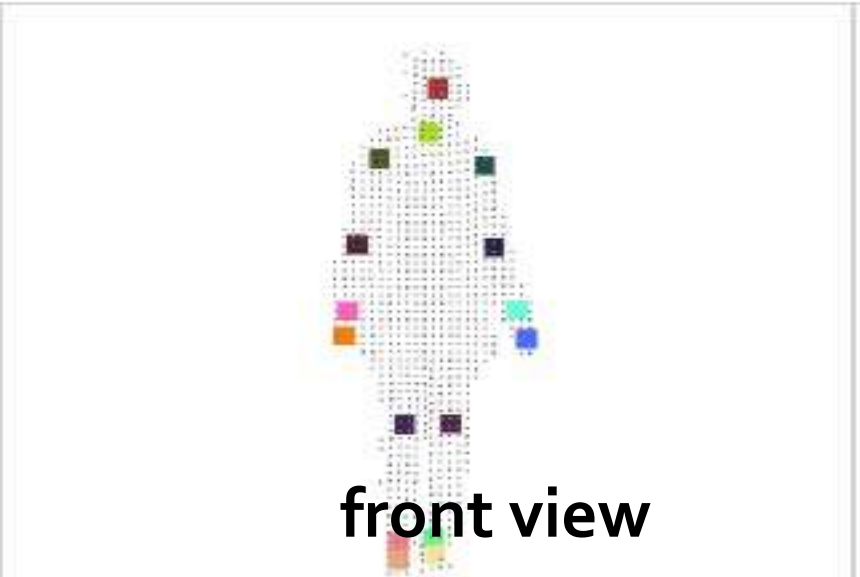




input depth

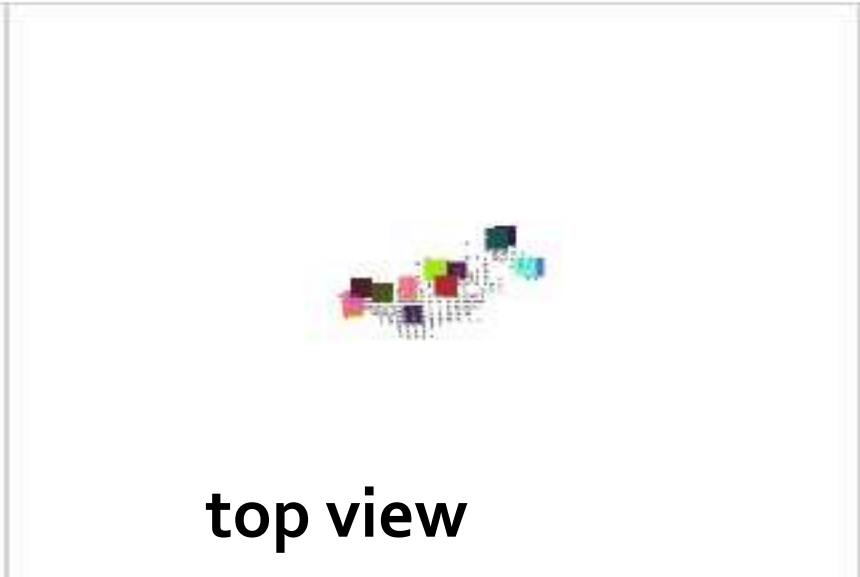


inferred body parts



front view

side view



top view

inferred joint positions: no tracking or smoothing

TRAINING DATA

Microsoft

10 March 2011 Last updated at 11:09 GMT

Microsoft Kinect 'fastest-selling device on record'

Microsoft has sold more than 10 million Kinect sensor systems since launch on 4 November, and - according to Guinness World Records - is the fastest-selling consumer electronics device on record.

The sales figures outstrip those of both Apple's iPhone and iPad when launched, Guinness

Azure Kinect DK

Developer kit with advanced AI sensors for building computer vision and speech models

Health and life sciences

Enhance physical therapy, improve and monitor

Retail

Design a shopping experience that your

Logistics and manufacturing

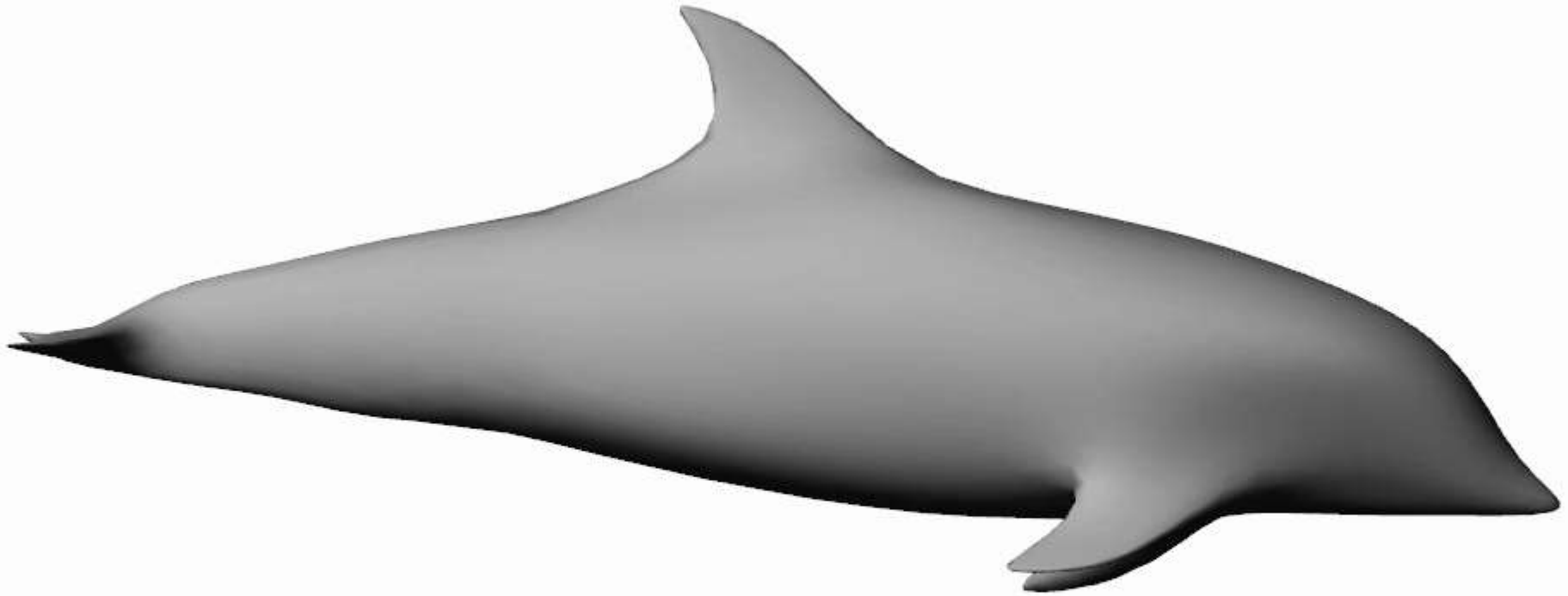
Ship and receive items

Robotics

Gain new environmental understanding with depth,

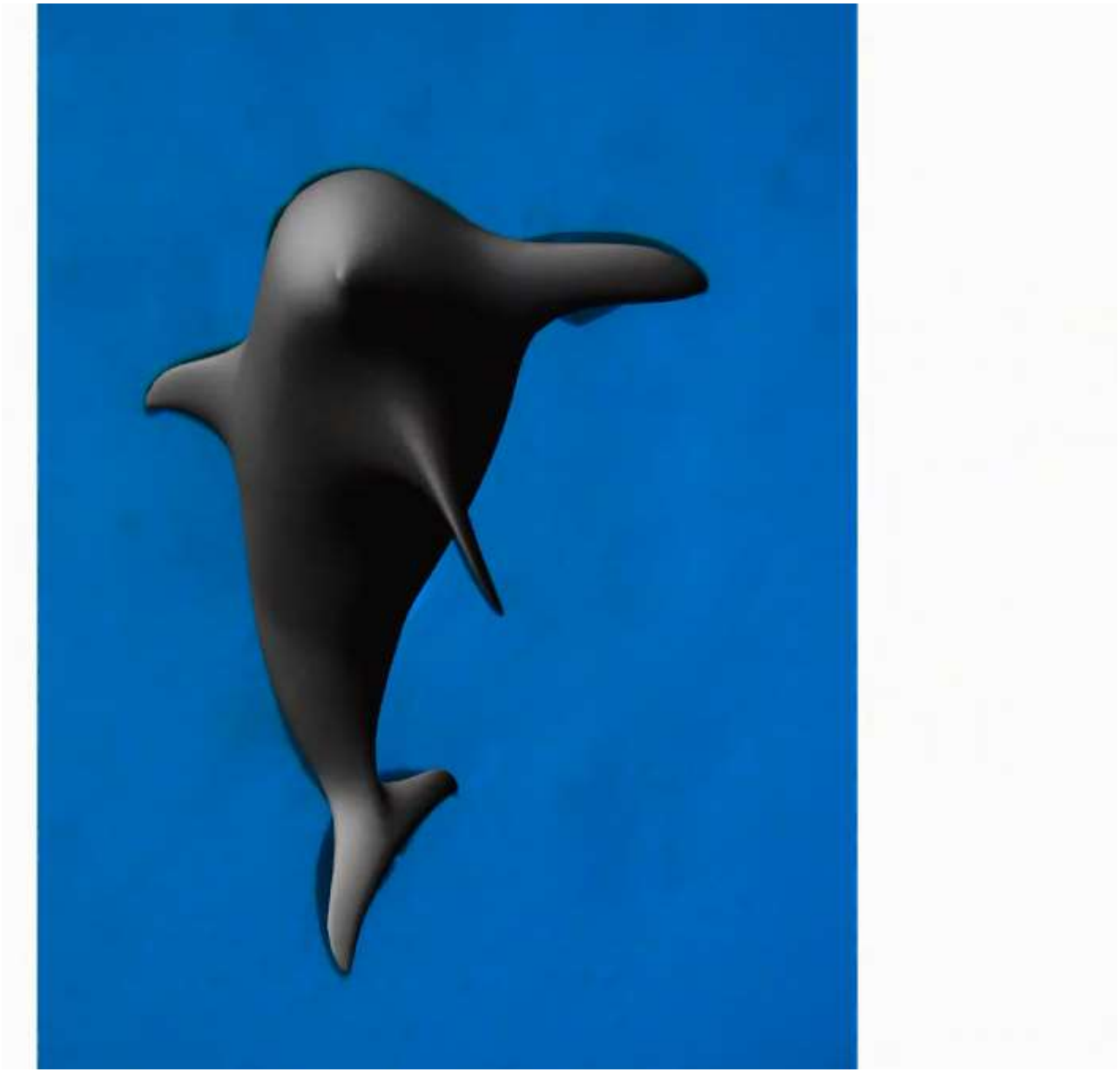
- I was the one working on body tracking – yet my algorithms played no part

- Jamie was working on farm animals – turned out to be the key



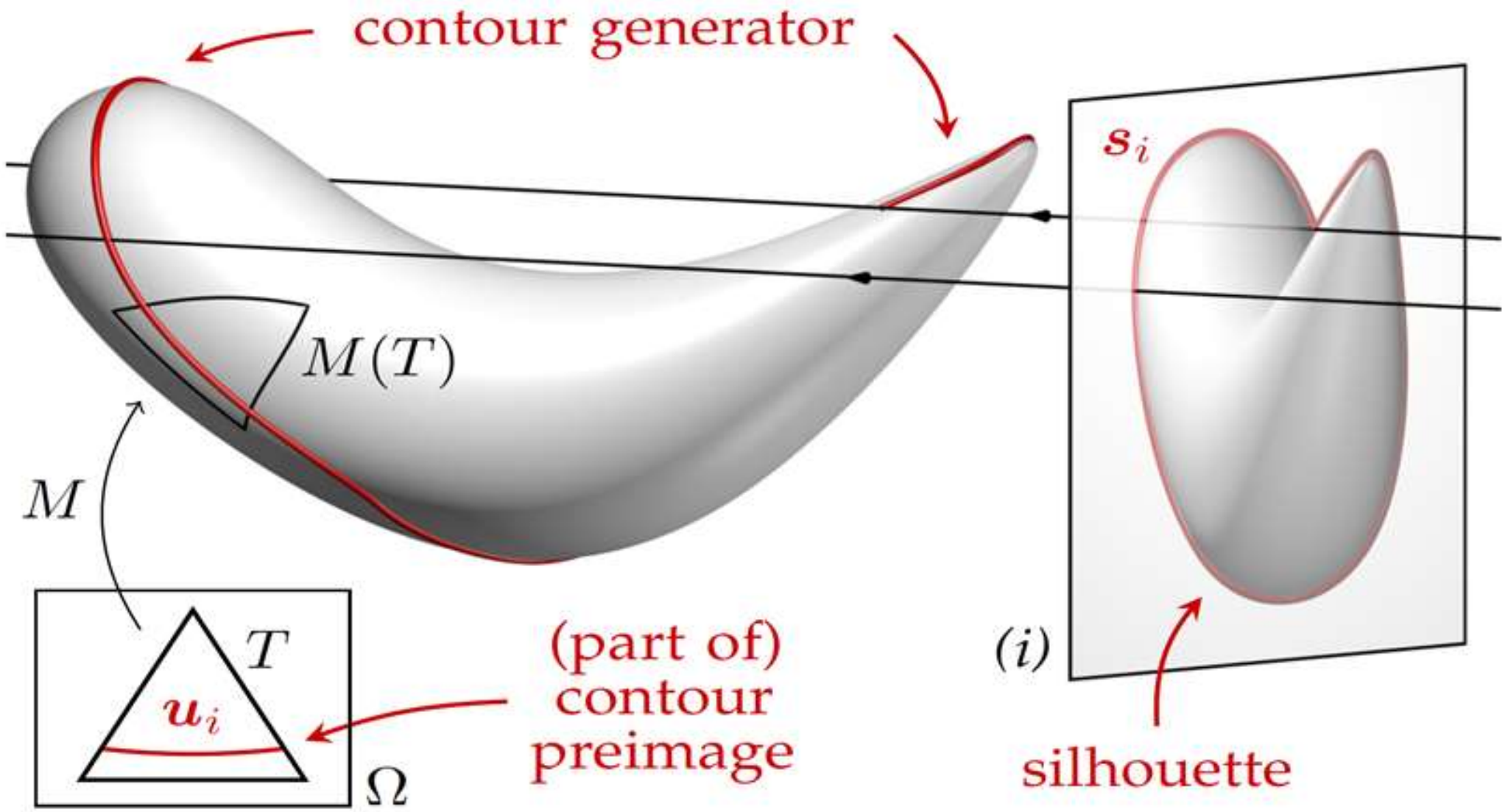
Fitting subdivision surfaces to 2D data

Microsoft

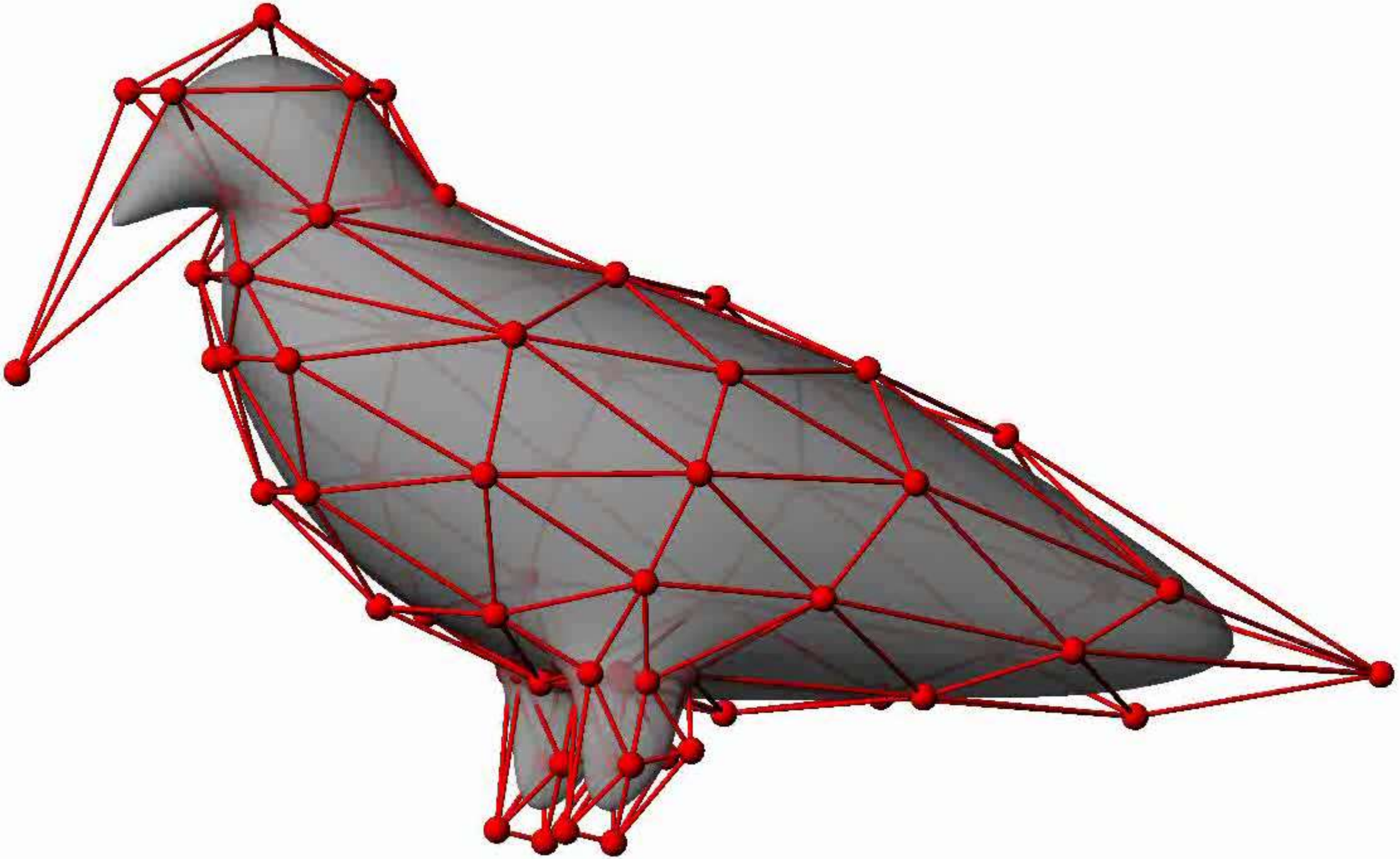


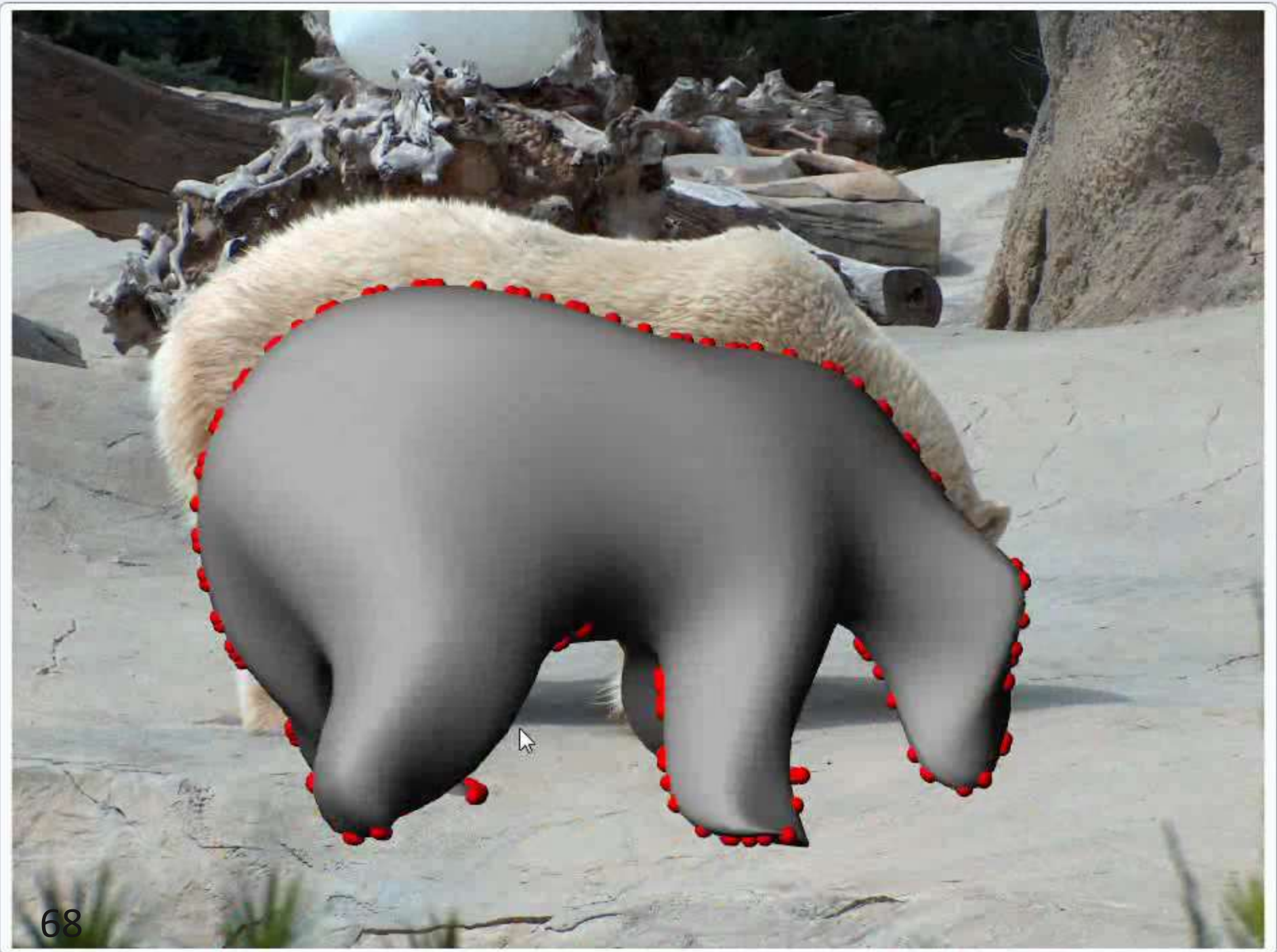
Fitting subdivision surfaces to 2D data

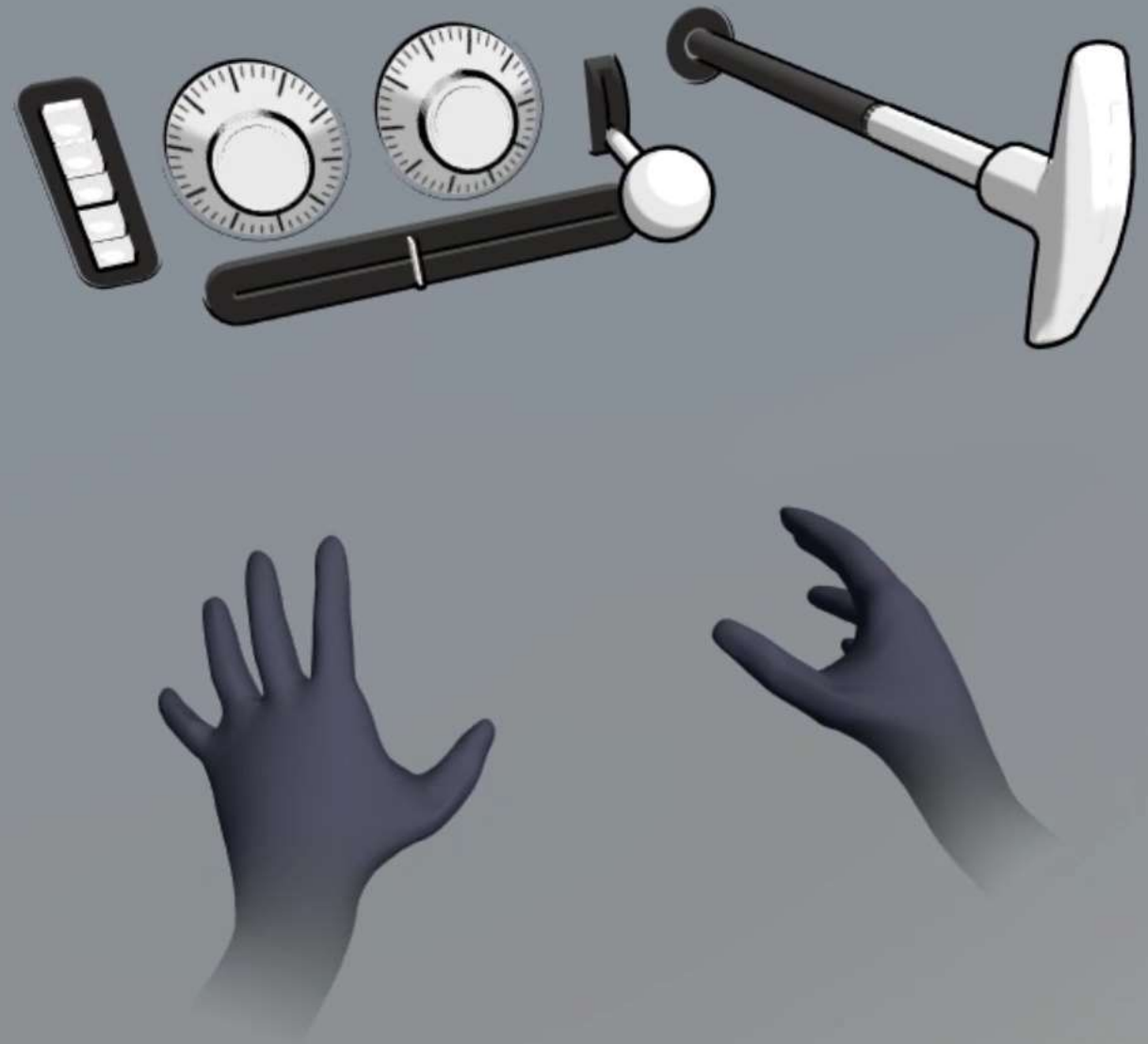
Microsoft



Silhouette formation: 3D to 2D







Efficient and precise interactive hand tracking through joint, continuous optimization of pose and correspondences
Taylor et al., *ACM Transactions on Graphics 35(4), pp. #143, 1–12, Proc. SIGGRAPH 2016*

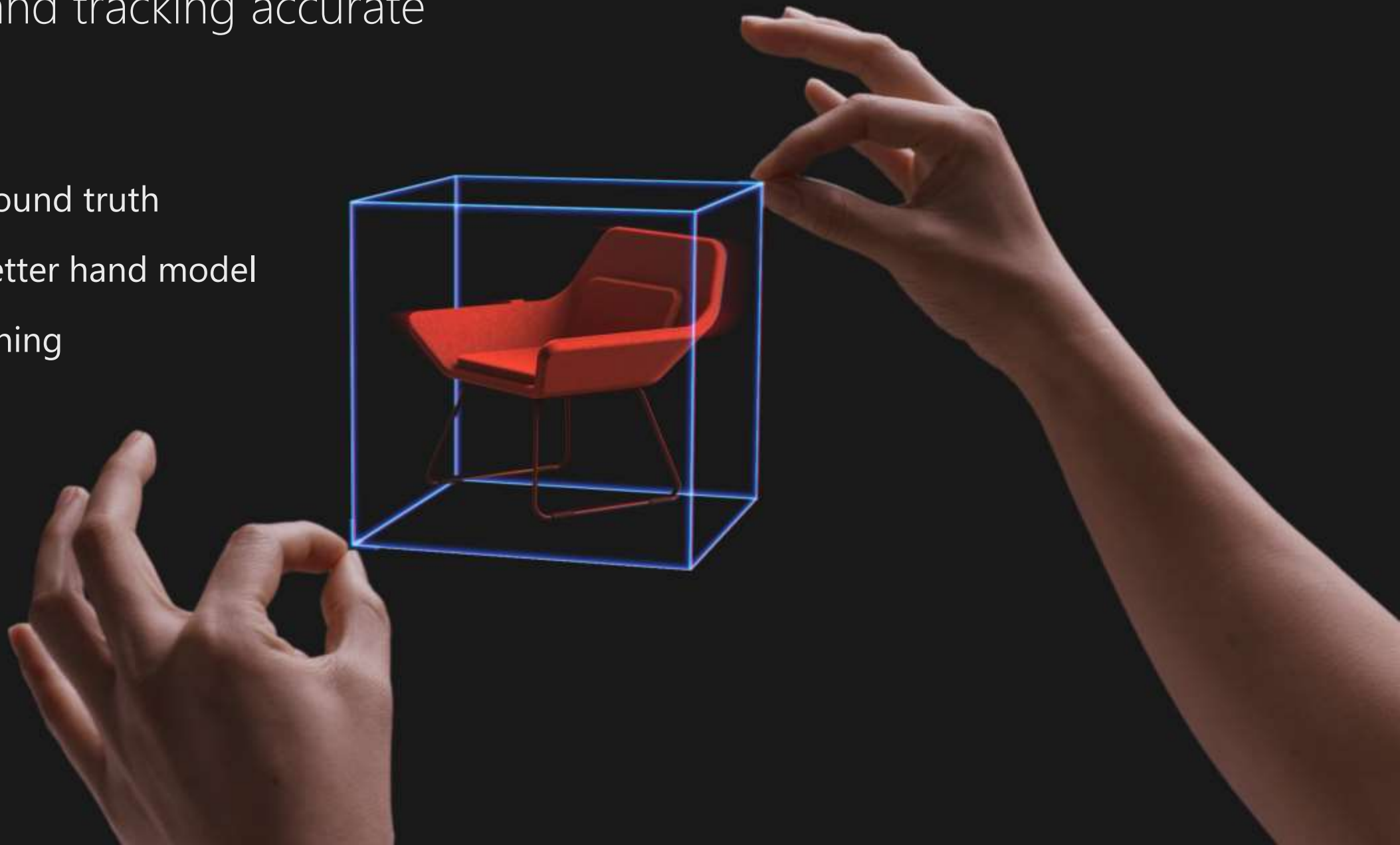
HoloLens 2

Making hand tracking accurate

Collecting ground truth

Learning a better hand model

Machine learning



4D Ground Truth

148,000 hand poses

97 hand shapes

Synthetic training data

Synthetic training data are labelled images made using computer graphics.

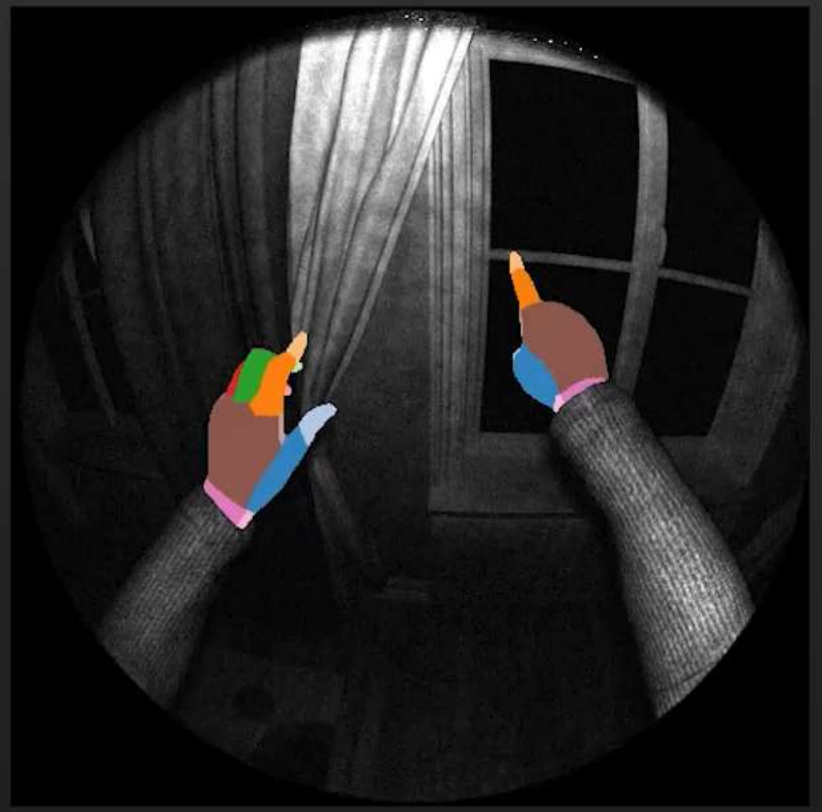
Why use synthetic data?

- Clean labels without annotation noise or error
- Can make GT that is impossible to label by hand
- Easy to control variation in dataset

Synthetic Training Data for Hand Tracking

Visible Light - RGB

Depth Camera



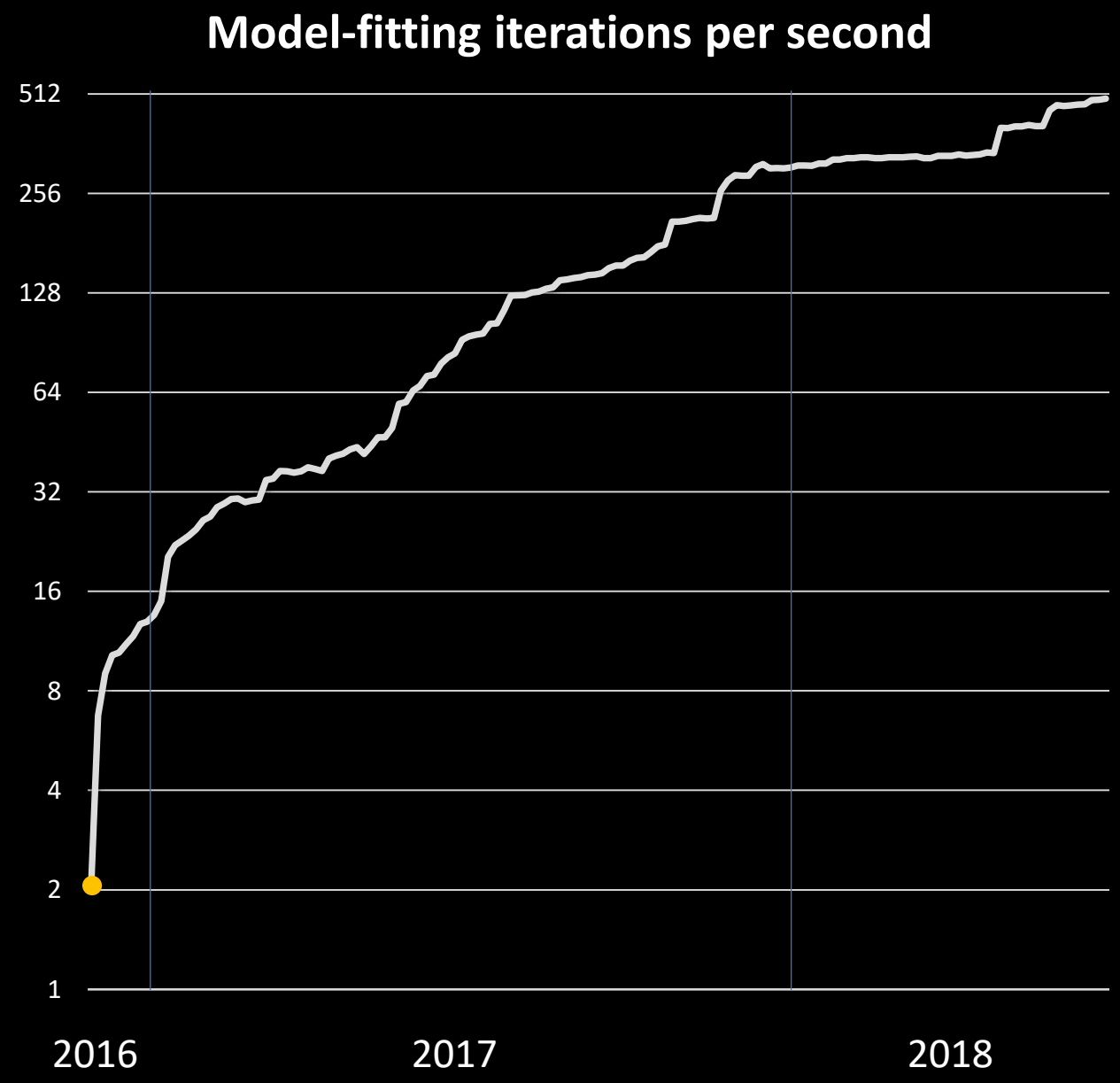
Ground Truth for Machine Learning

Making it fast on a (pair of) 500MHz machines with 128K RAM

- Optimize code
  Great, we went to Floptimal

- Whiteboard malloc

- Use simpler models?
  Great, but lose accuracy.

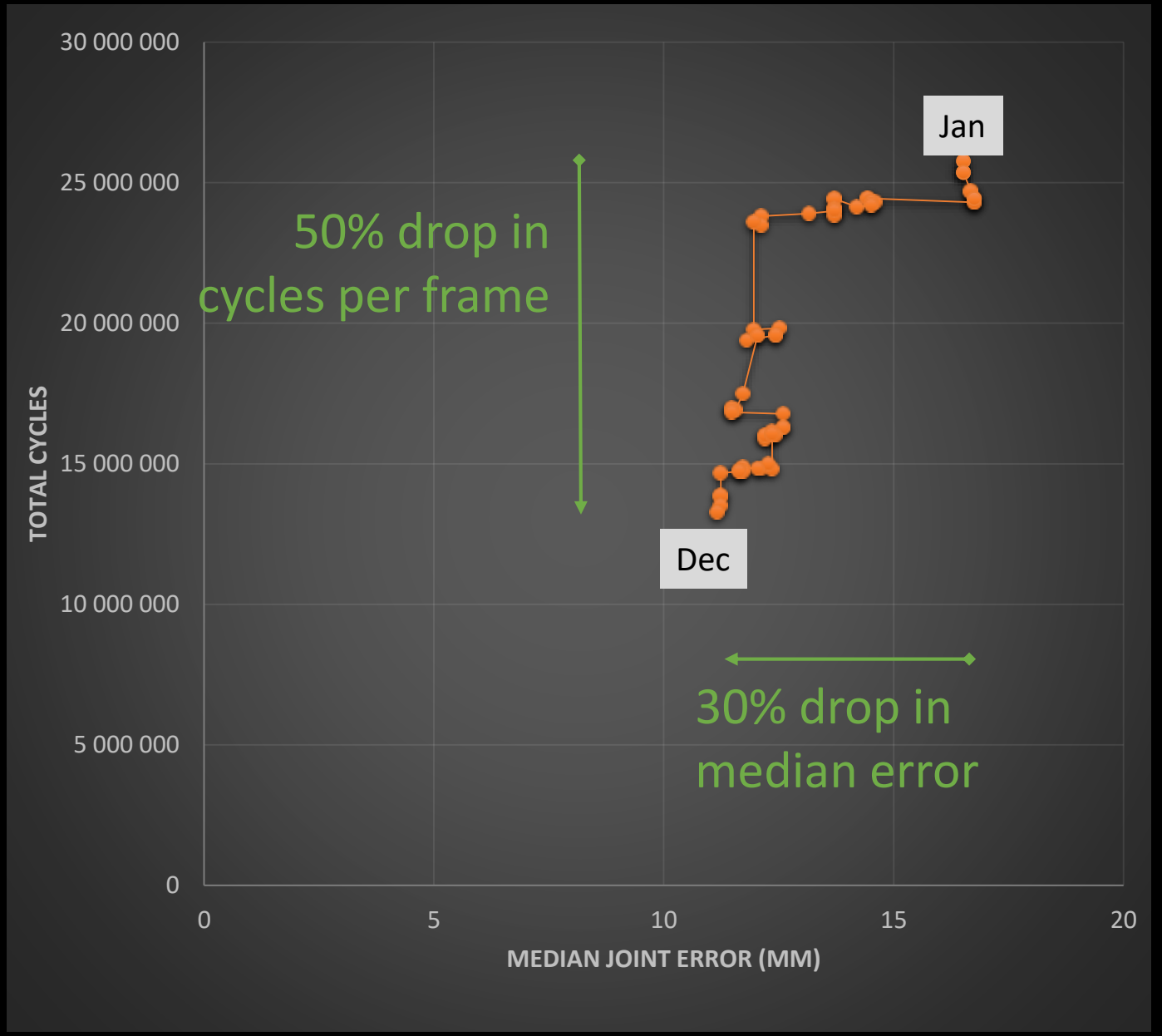


Making it fast on a (pair of) 500MHz machines with 128K RAM

- Optimize code
  Great, we went to Floptimal

- Whiteboard malloc

- Use simpler models?
  Great, but lose accuracy.

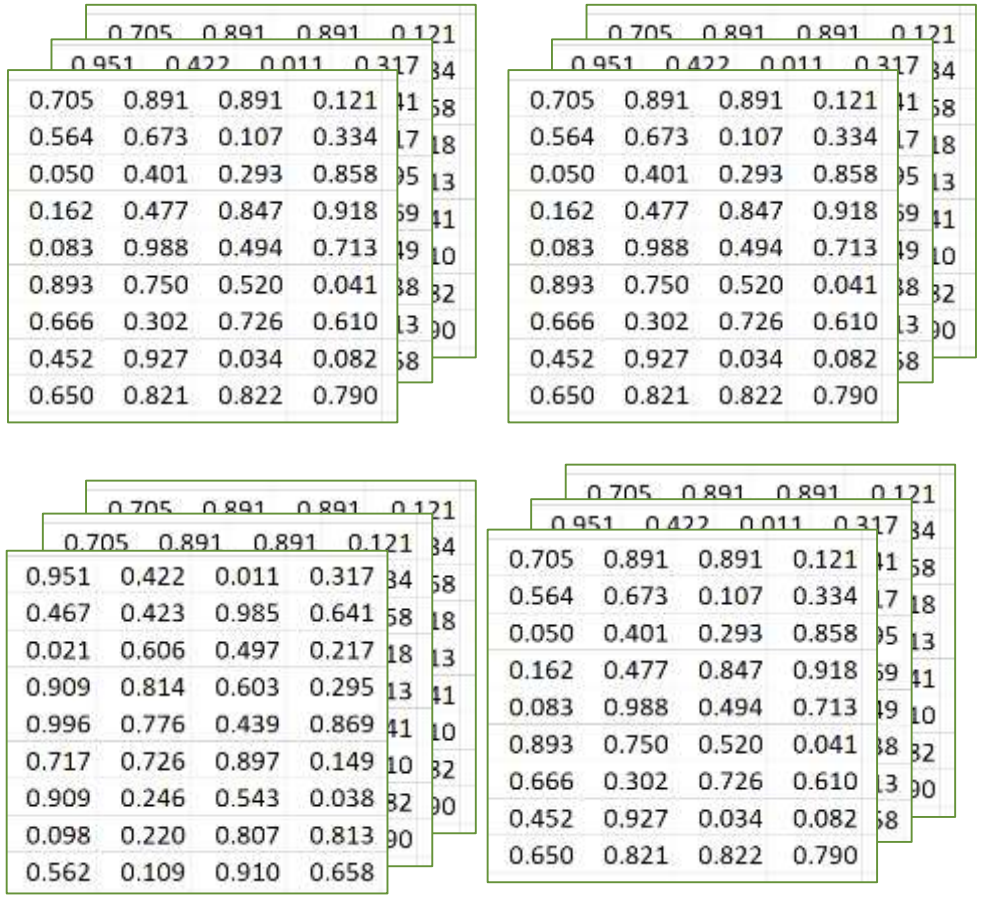


All Data AI

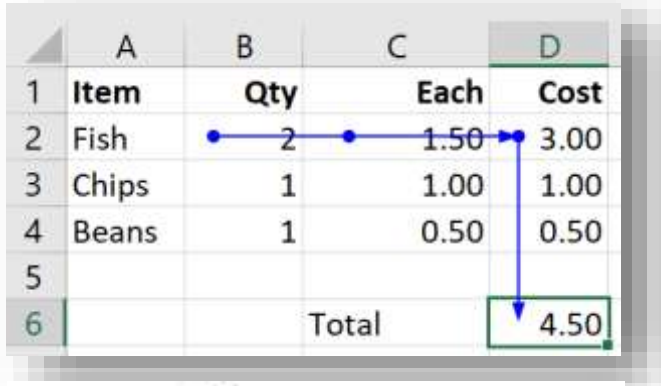
"Again, [the Difference Engine] might act on other things beside *number*." — Ada Lovelace

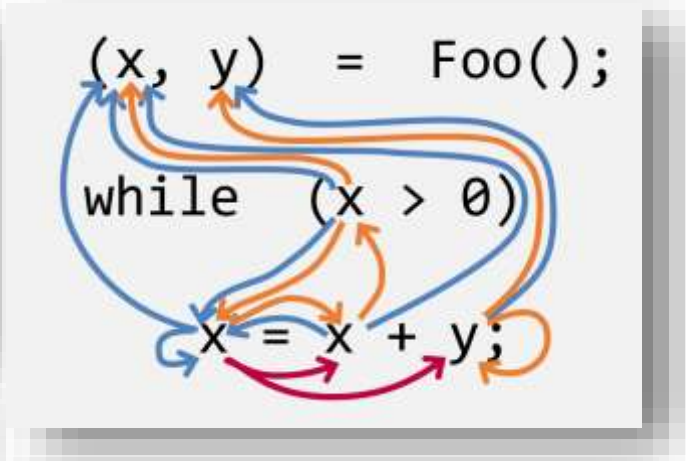
The ADA group studies AI models and techniques that apply to complex real-world data: big data, small data, data with complex structures like trees and graphs. In short: "All Data AI".

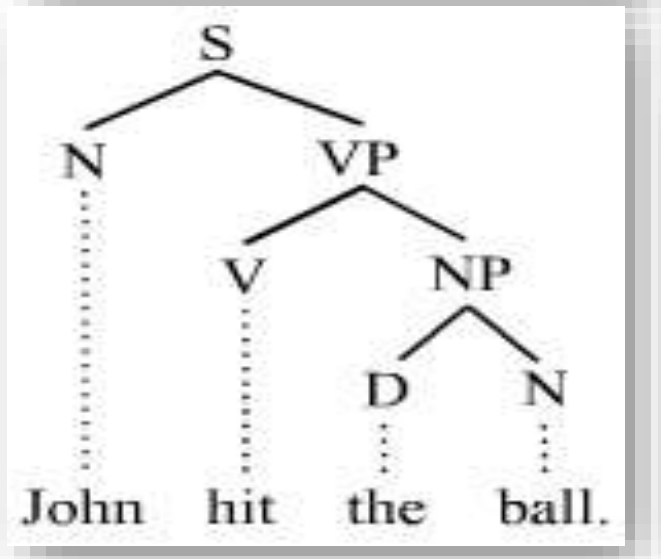
Linking language and vision with data from structured domains such as programming languages, bioinformatics, chemistry, or the web, we explore all aspects of machine learning and related disciplines. We research in Bayesian inference, Deep Learning, Optimization and Programming Languages, with multiple prize-winning papers and research. And we have contributed to some of the most advanced technologies that Microsoft builds, with contributions to Xbox, Office, HoloLens, Bing, Visual Studio, and many others. Our superpower is our combined focus on fundamental research and real-world engineering, to push the boundaries of AI.

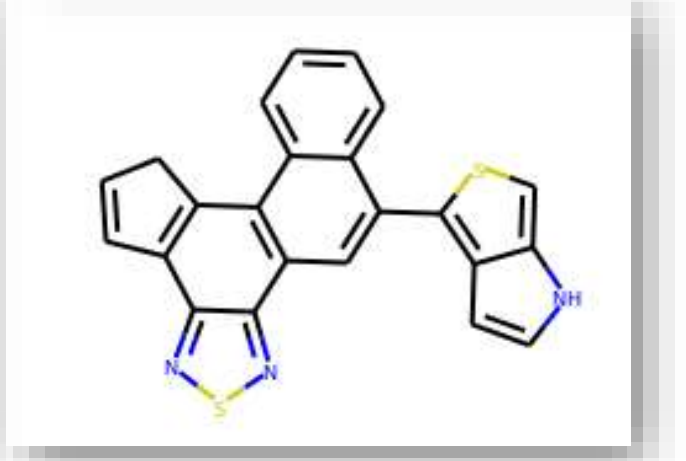


Gen 2: Everything's a Tensor
And frameworks constrain us to think this way

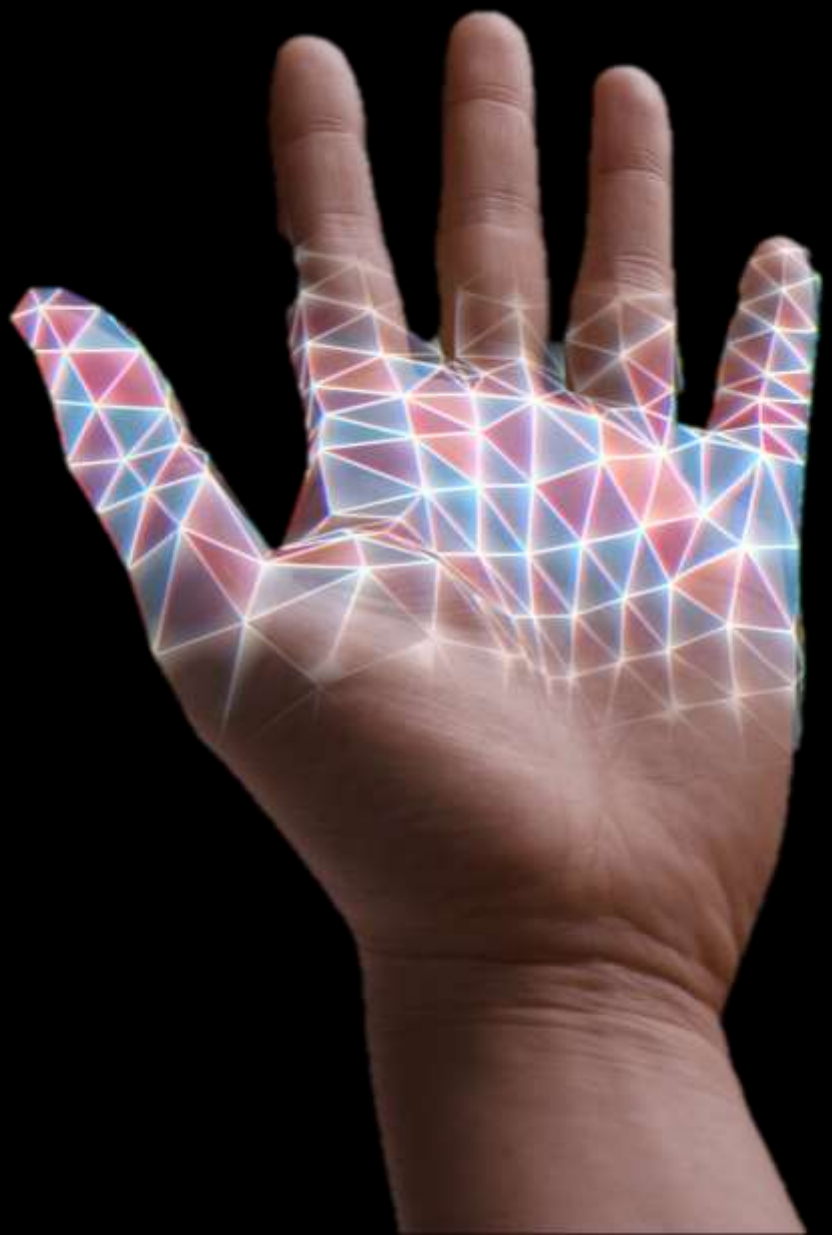








Gen3. Everything has its natural structure
But we need frameworks to work with these structures



Three (well, four) computer vision systems that work

1. Pure geometry
2. Machine learning with huge training data
3. Geometry + deep learning, again with big data

The future:

- More hybrids
- More data
- More models
- More hardware