Think Like an Amateur, Do As an Expert: Lessons from a Career in Computer Vision

Dr. Takeo Kanade

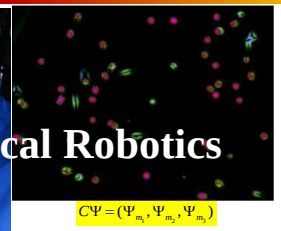
Carnegie Mellon University
The Robotics Institute

May 2018

What Kanade did or got involved

Computer Vision
face, color,
stereo, motion,
video, 3D recognition
etc. etc

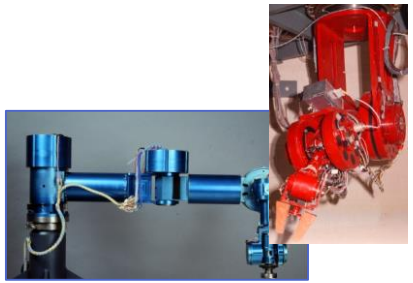
80 90 00 10



$C\Psi = (\Psi_{m_1}, \Psi_{m_2}, \Psi_{m_3})$

Medical Robotics

80 90 00 10

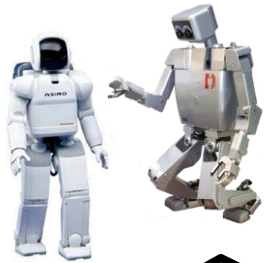


Manipulator Design and Control

80 90 00 10

Virtualized Reality

80 90 00 10



Humanoid

80 90 00 10

Autonomous Air Vehicle

80 90 00 10

80 90 00 10

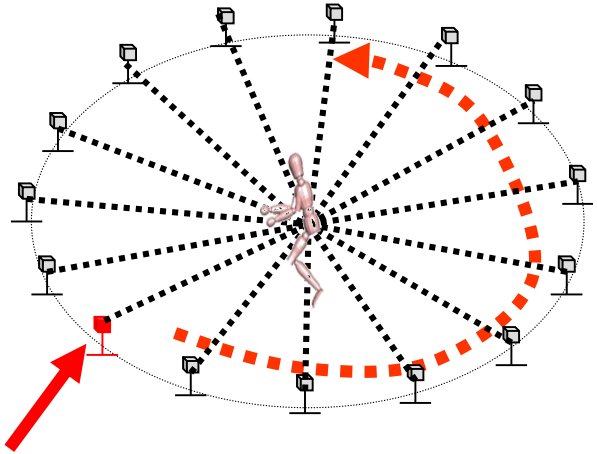
My Moment of Fame:

EyeVision

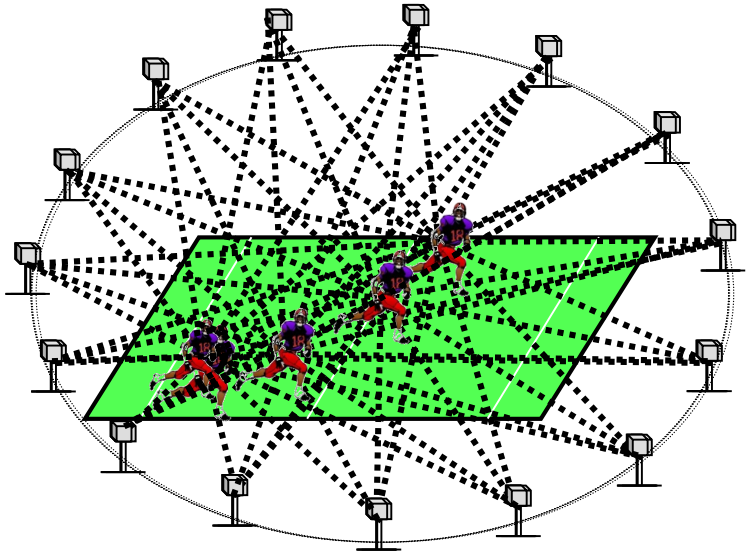
at Super Bowl XXXV

Matrix-like replay anywhere
in the American Football field

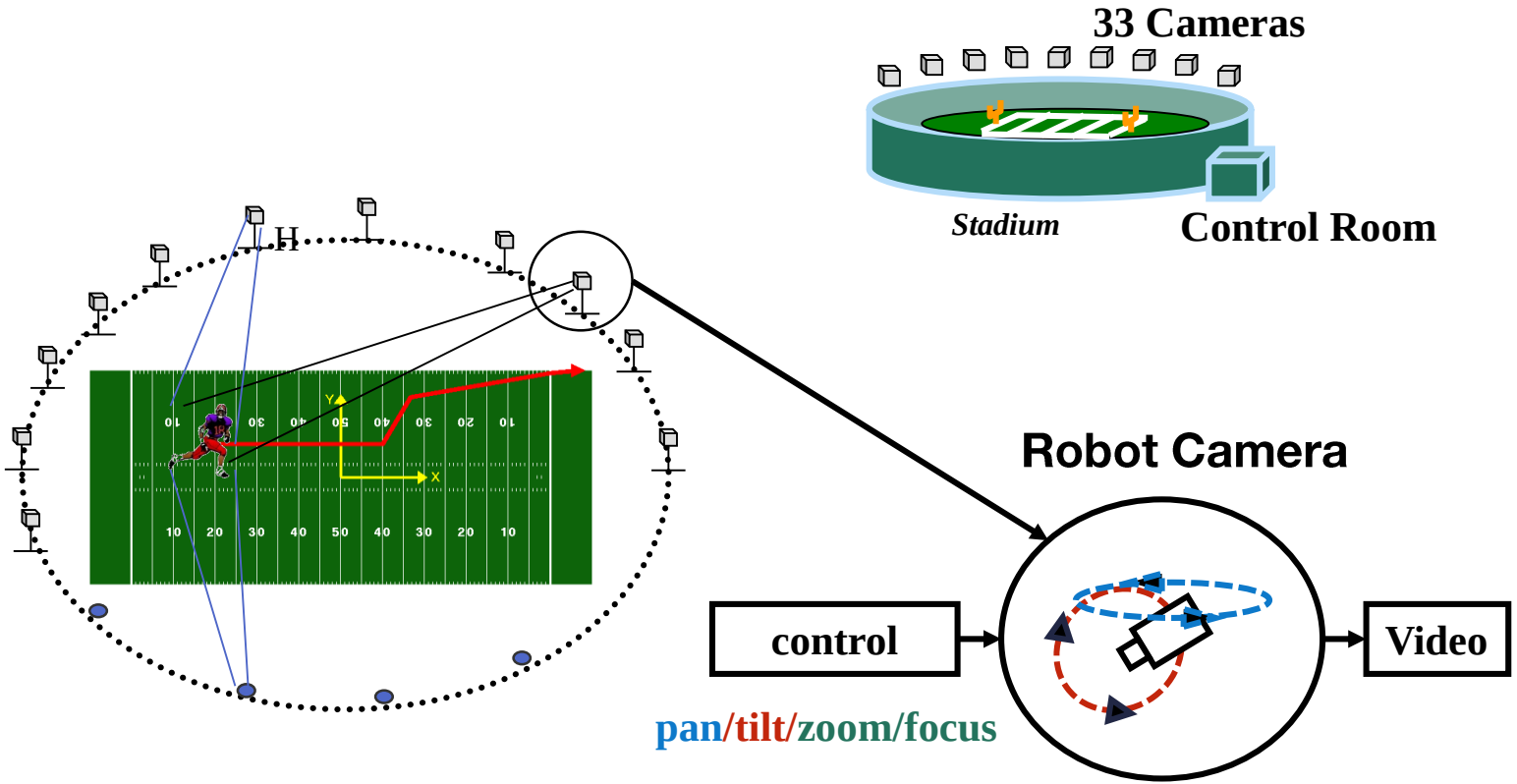
Movie Matrix



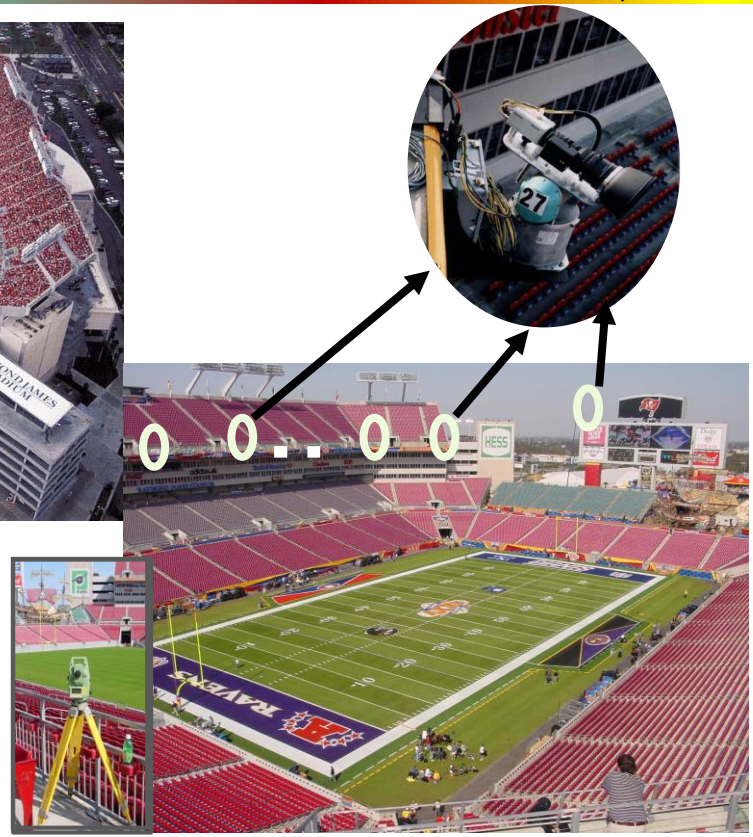
EyeVision



EyeVision



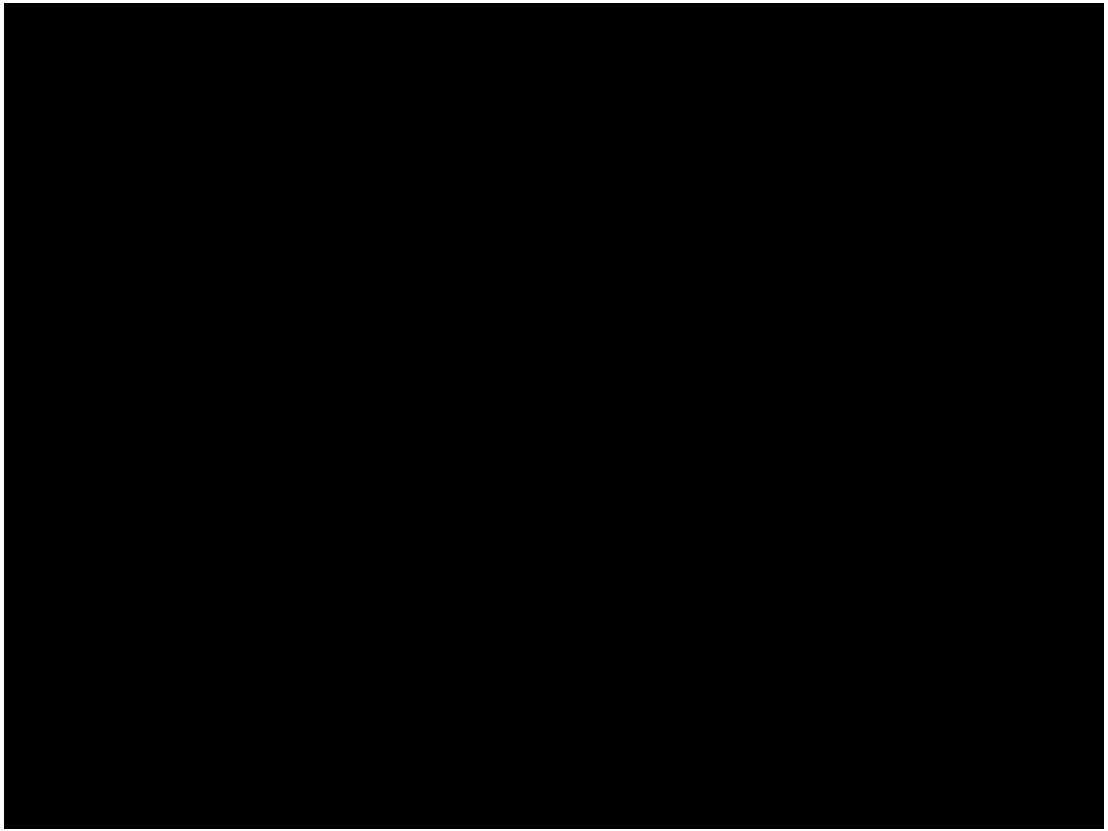
Tampa, Florida



Trailer and wires

EyeVision "Best of"

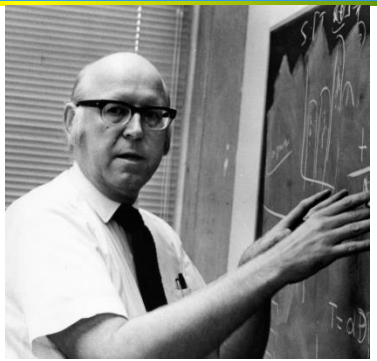
In Bruce Willis' movie "Surrogates"



What is the wish as a researcher/developer?

“Perform good research and development”

What is "good" research?



late Allen Newell

Professor, Carnegie Mellon University

"Good science responds to real phenomena or real problems."

"Good science is in the details."

"Good science makes a difference."

Useful *and Fun* Research: Make a Scenario

Picture success:

- **What can happen**
- **How and where will it be used**

Think freely with fun

The story expands, and you talk.

Other people can join

My Experience and Observation on Successful Research/Development

Successful ideas are often surprisingly simple and even naïvely trivial

Impediment to such simple thinking is the “I know” mentality (aka “Expert” knowledge)

Reliable execution requires substantive knowledge and skill

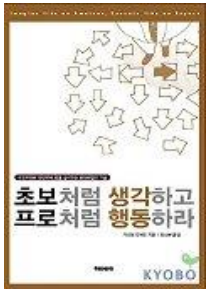
Substantive, professional skill and knowledge

Examples:

- A motion system is not stable without a control theory.
- Programs ignorant of subtlety in numerical computation behave very “strangely” and even "dangerously“
  - e.g. Most geometric theorems don’t hold inside a computer
- Representation of object
  - e.g. Line equation $y = ax + b$ vs. $ax + by + c = 0$
- Algebraic distance vs. geometric distance
- Math and semantics
- Solving $F_i(x, y, z) = 0 \ (i = 1, \ldots n)$ vs. Minimizing $\sum_{i=1}^{n} F_i(x, y, z)^2$

“Think like an Amateur, Do as an Expert”

PHP研究所出版



There are translations to Korean, Chinese, and Tiwanese, but not an English version.

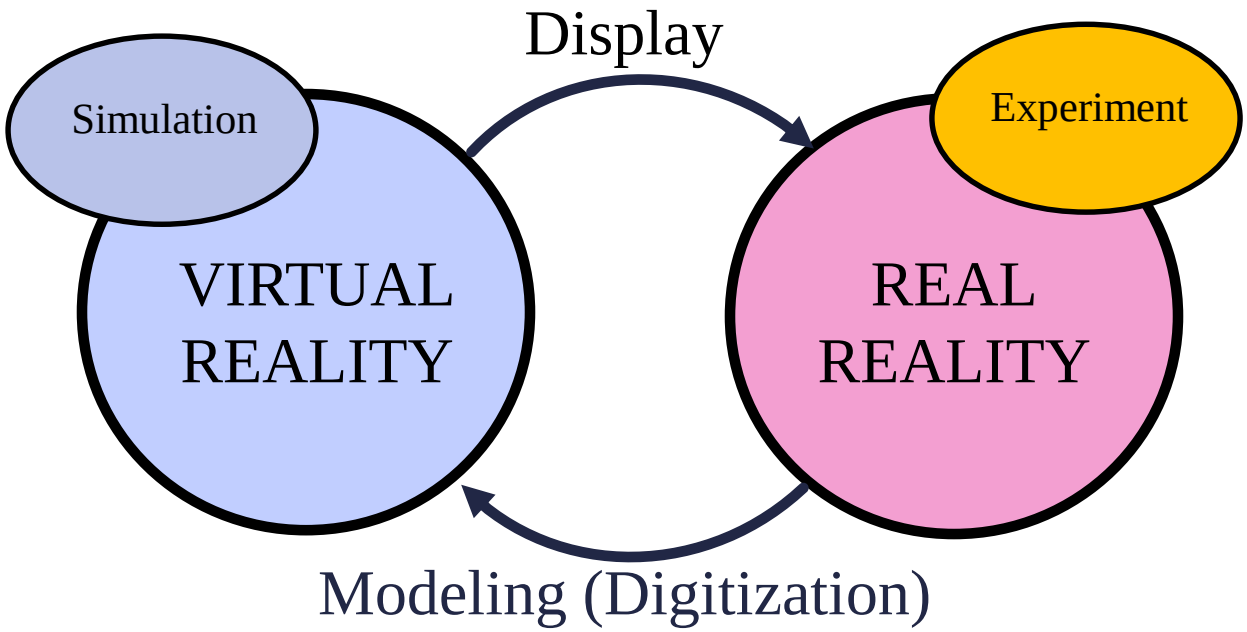
Many-Camera System

Research scenario that led to EyeVision

Can we digitize the dynamic real world into a computer?

~1990

Virtual Reality and Real Reality



The real world is where we start.

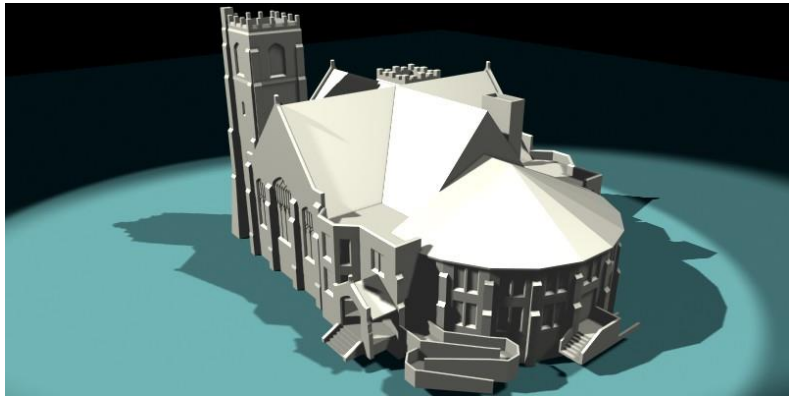
Scanning Laser Range Imager (Ladar)

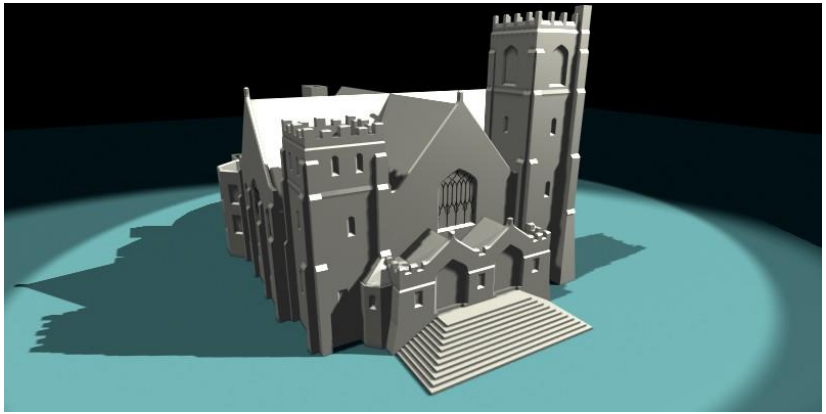
Time of Flight:	Phase Detect 2 Freq. AM
Field of View:	
Horizontal	360 ^{o}
Vertical	60 ^{o}
Speed:	125 k points/sec
Range:	60 m
Accuracy	12.8 mm
Intensity Dynamic Range:	87 dB

QuantaPoint, Inc.

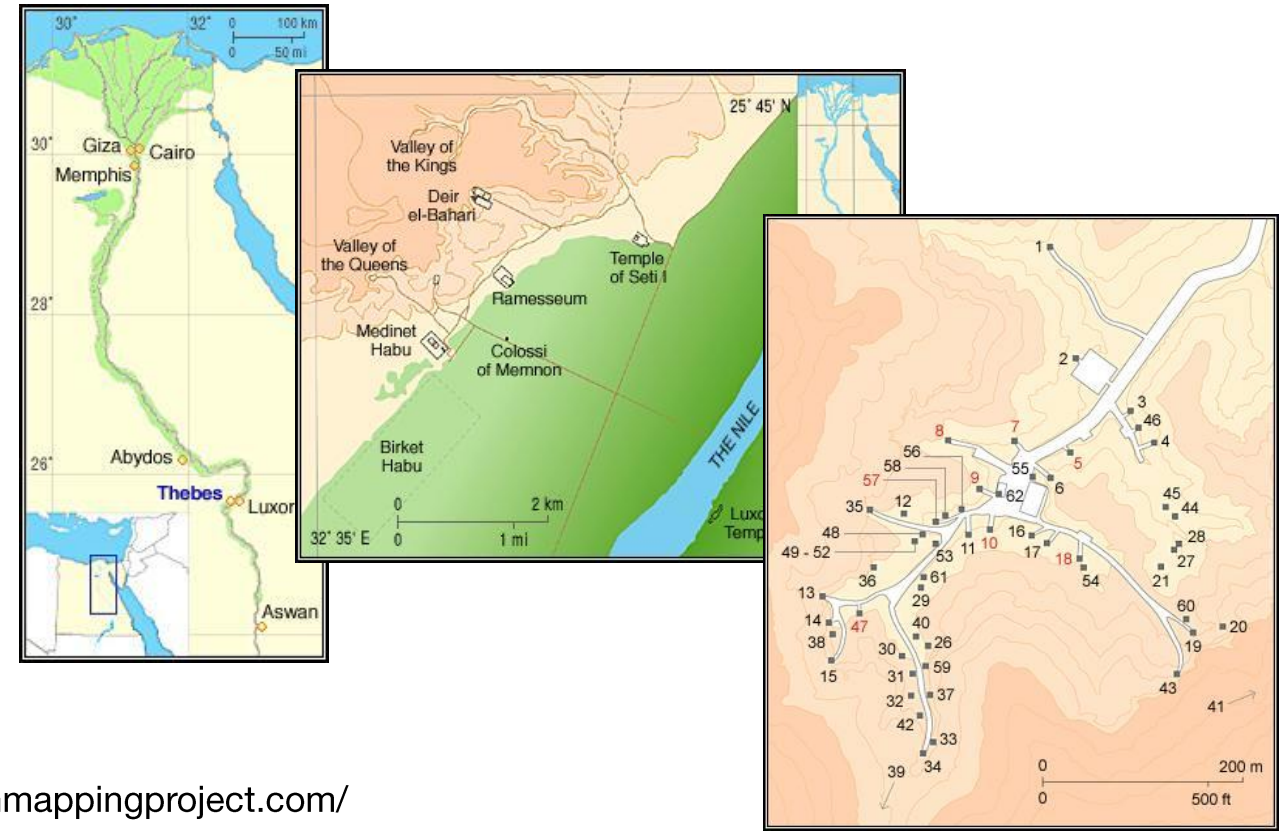
Circa 1993

Church



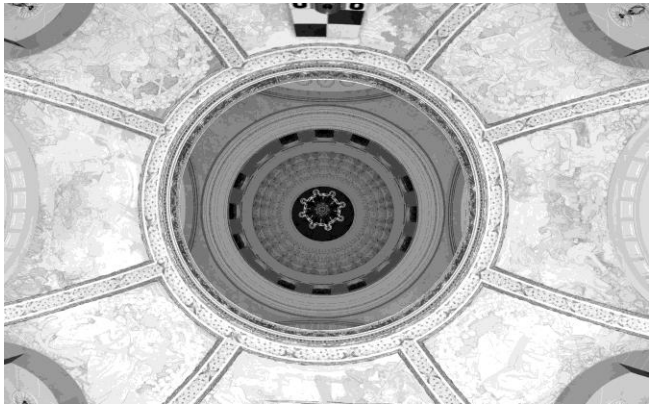


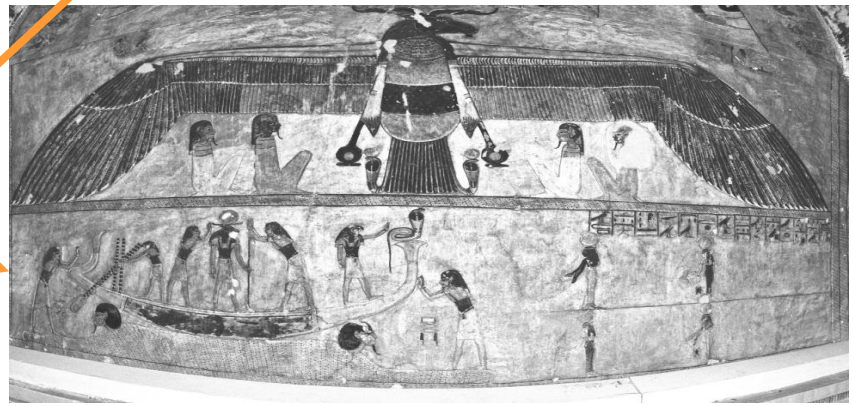
Modeling KV14 Tomb in Egypt

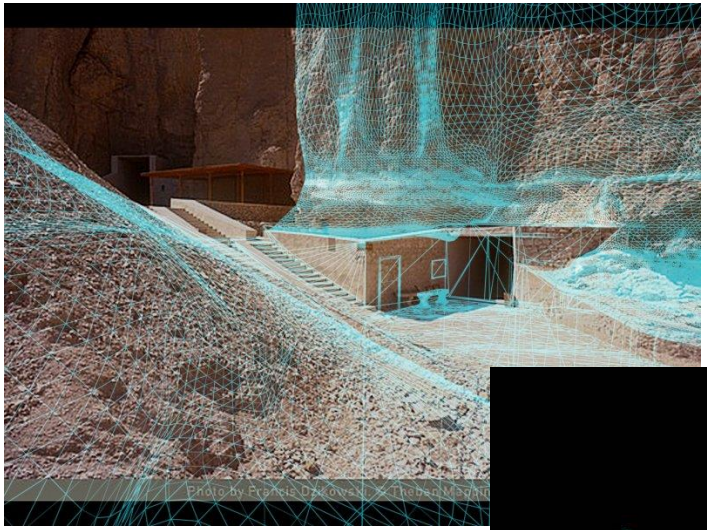


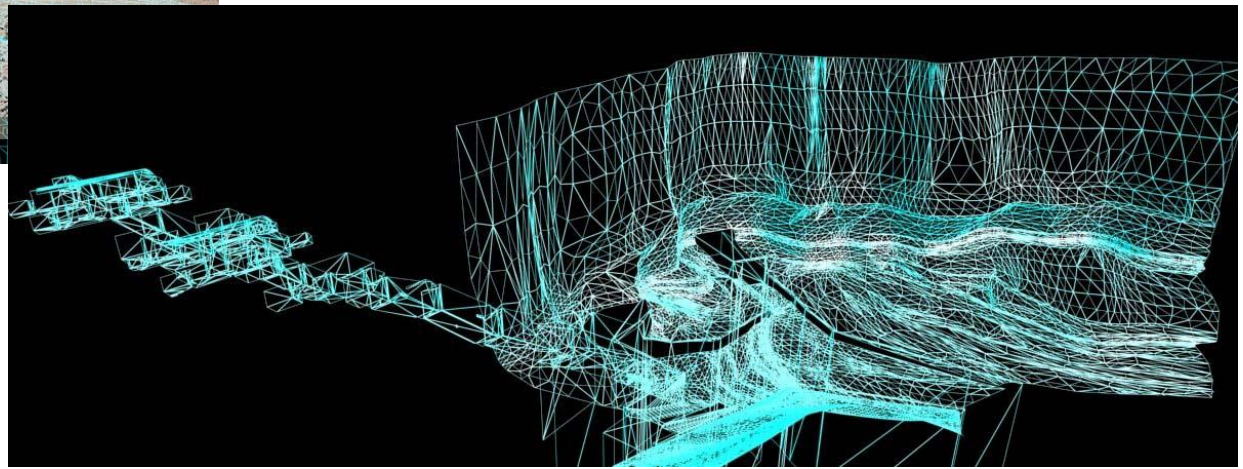
http://www.thebanmappingproject.com/

Each 1800 x 3200 3D points
230 MB/scan
1500 scans

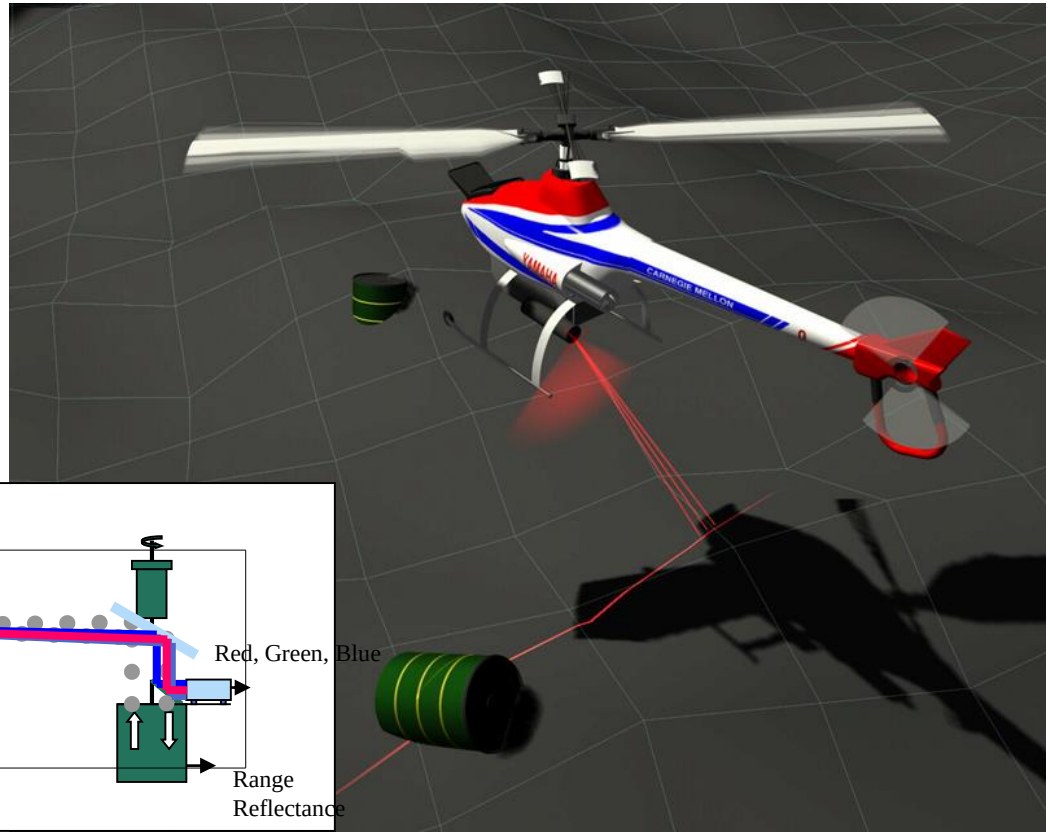






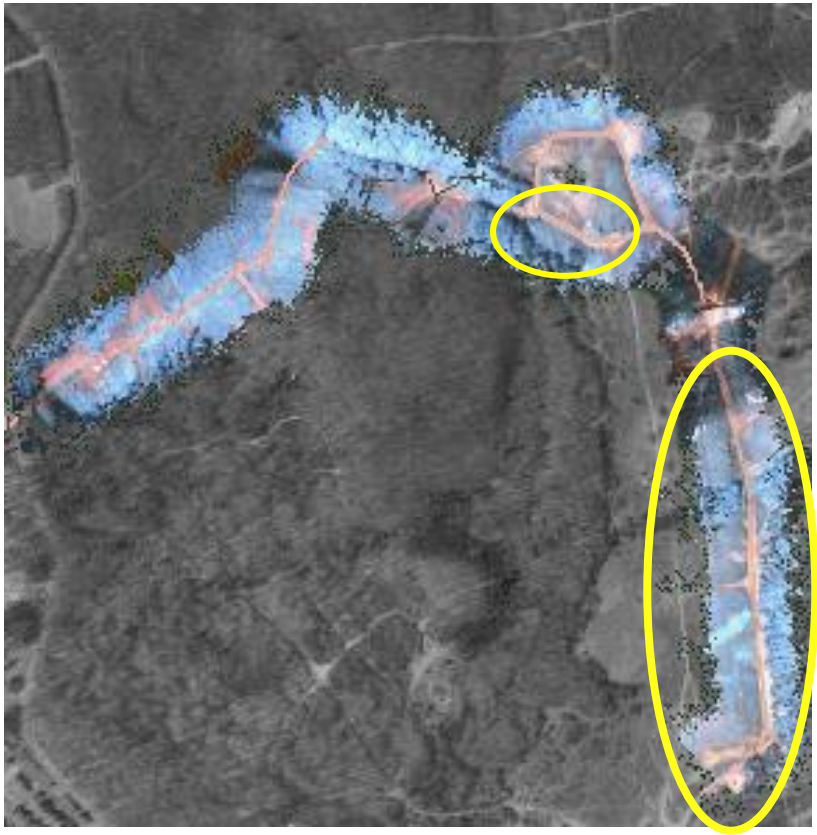


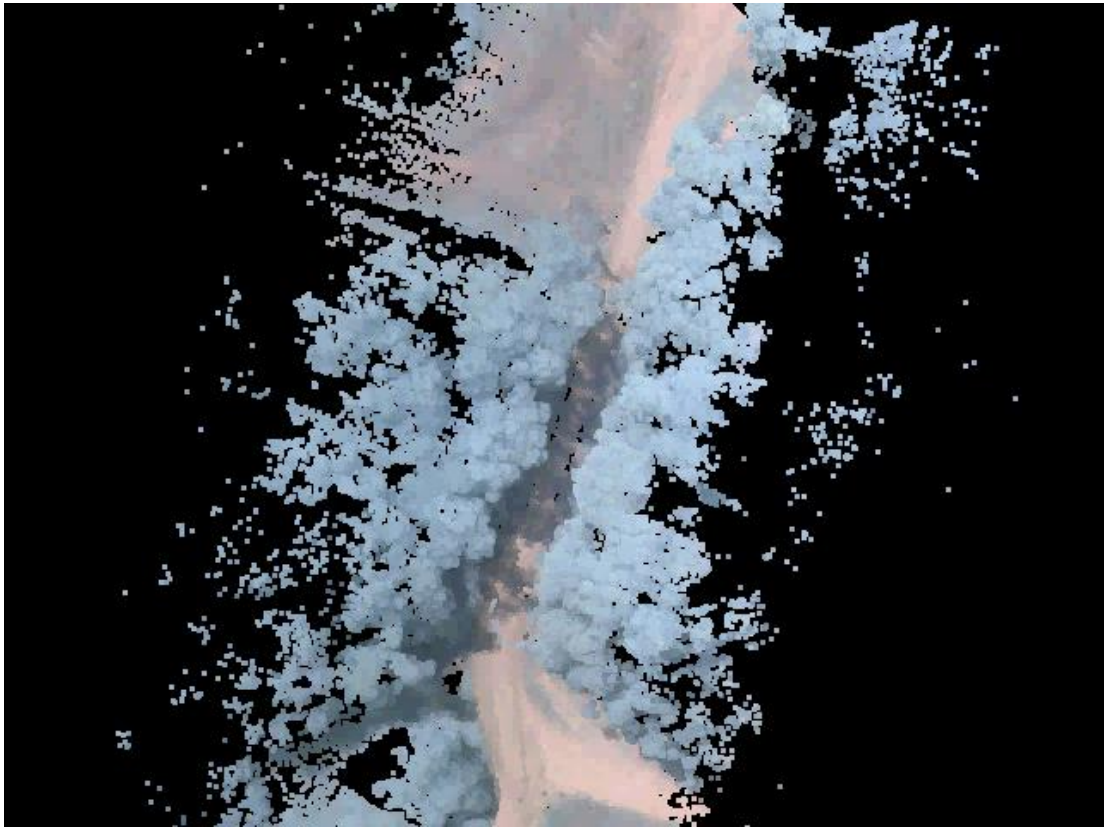
Range sensor onboard an Autonomous Helicopter



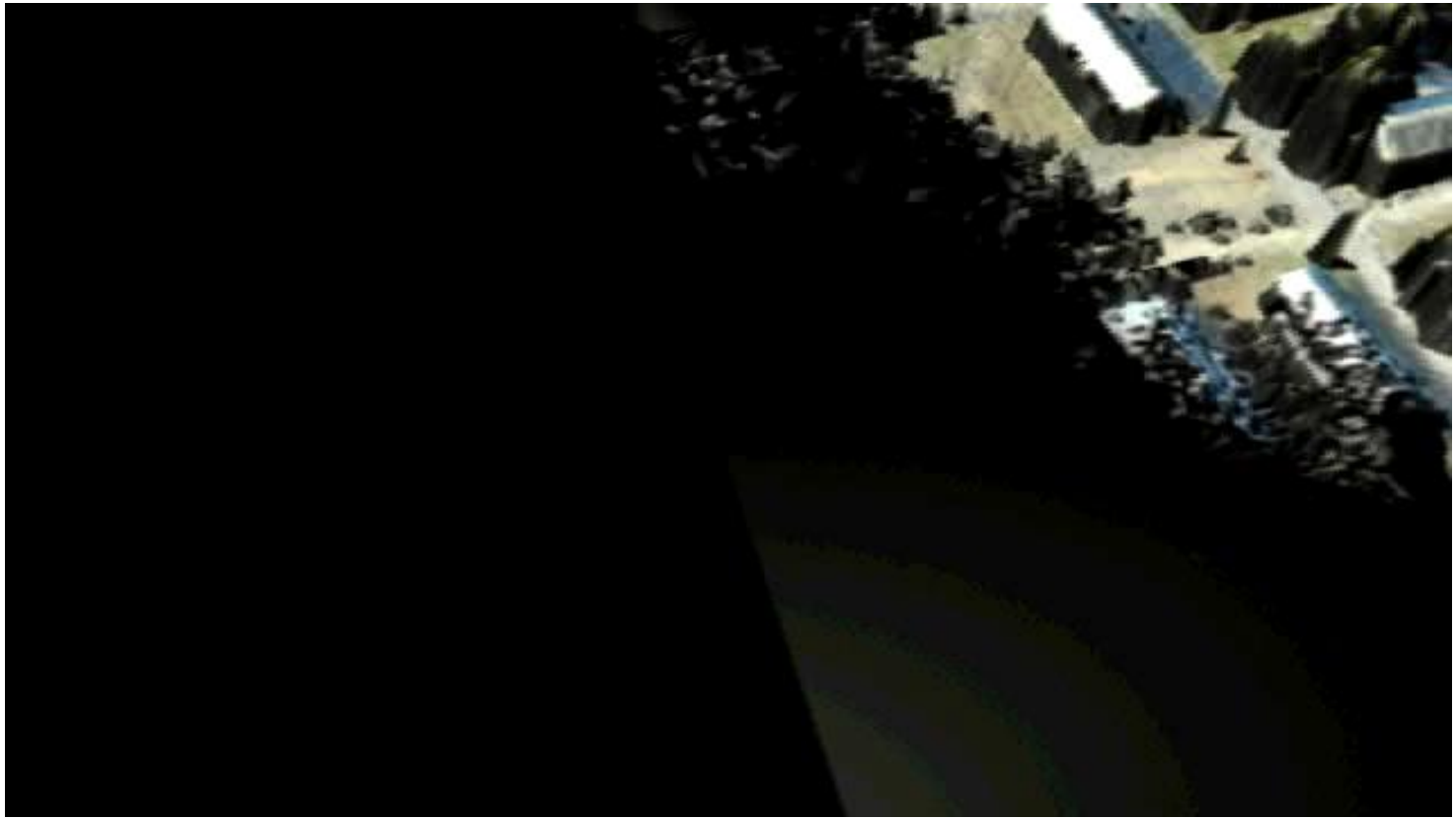
3D Mapping

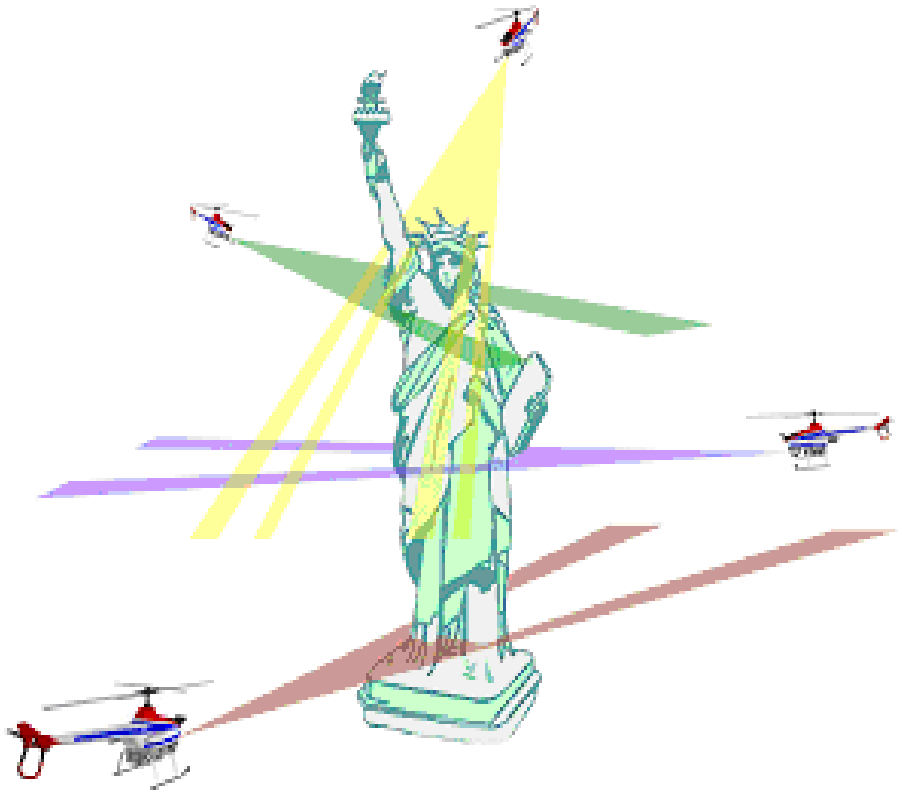
4m / pixel
1.2km
square



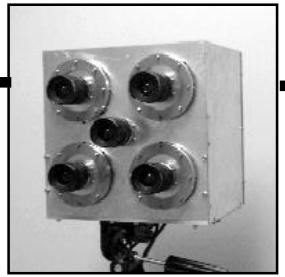


A large area mapping from air





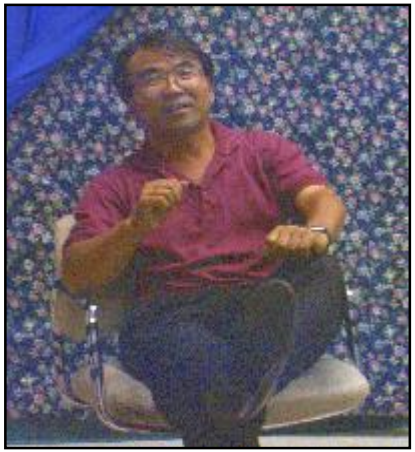
How it started: CMU Video-Rate Stereo Machine for Real-time Range Mapping (circa 1992)



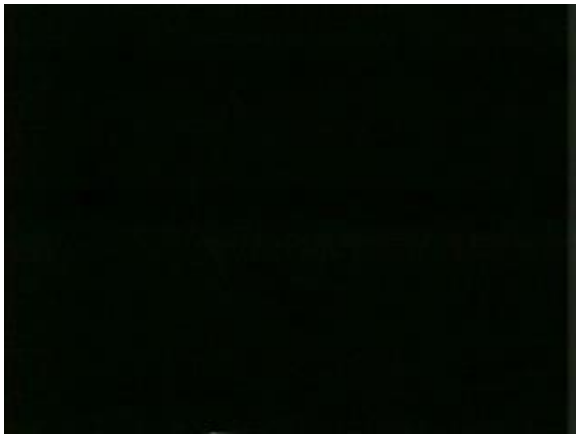
"3D Camera"

Color Image
256 x 256 x 24 bits

Depth Image
200 x 200 x 8 bits
at 30 fps

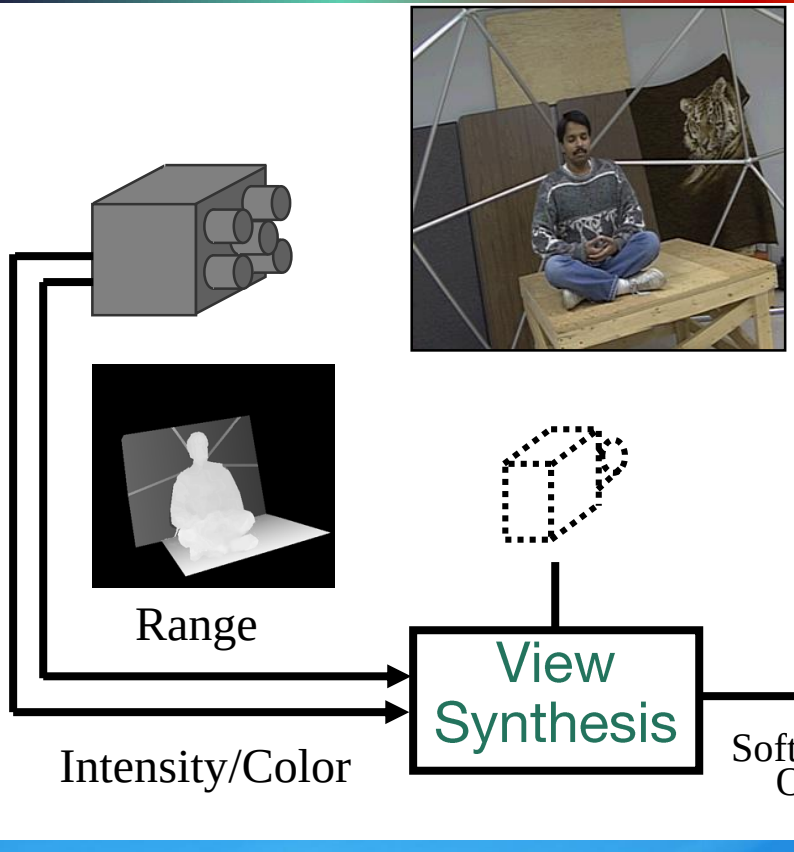


aligned
pixel by
pixel

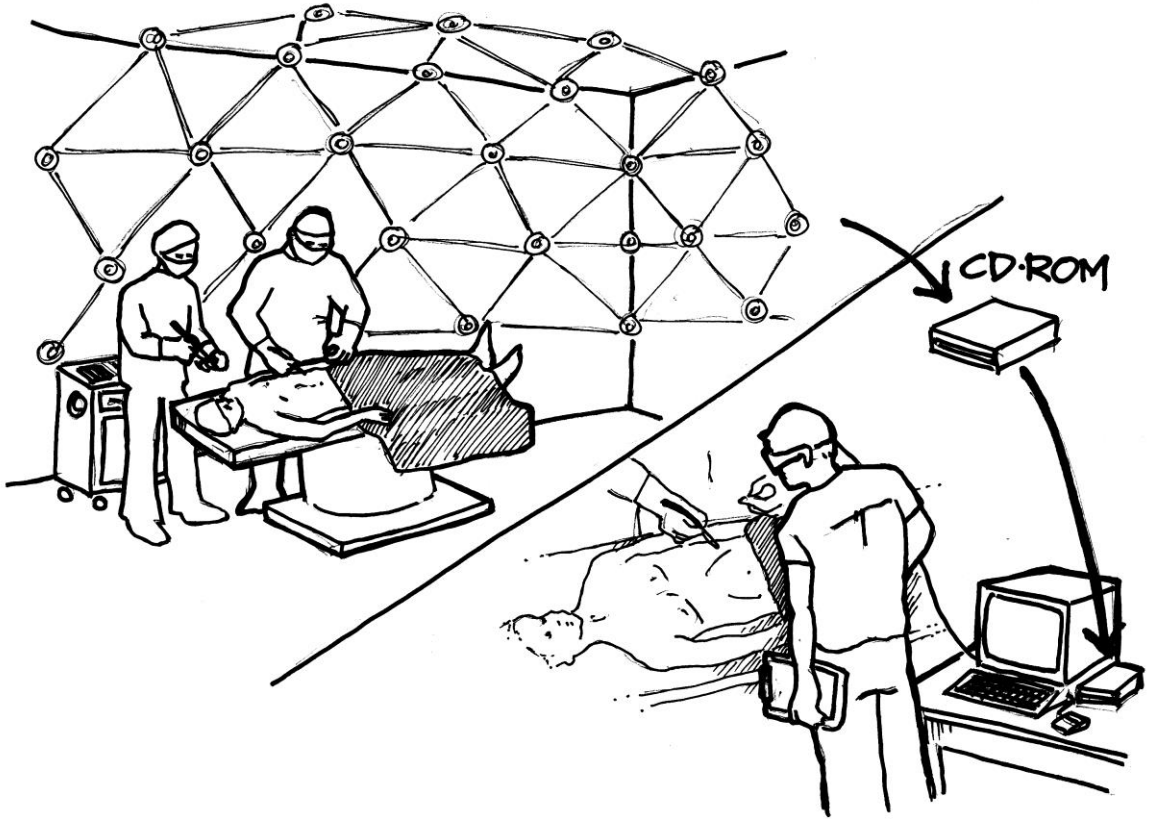


[with Kimura, Kano, Tanaka, Oda]

Soft Camera

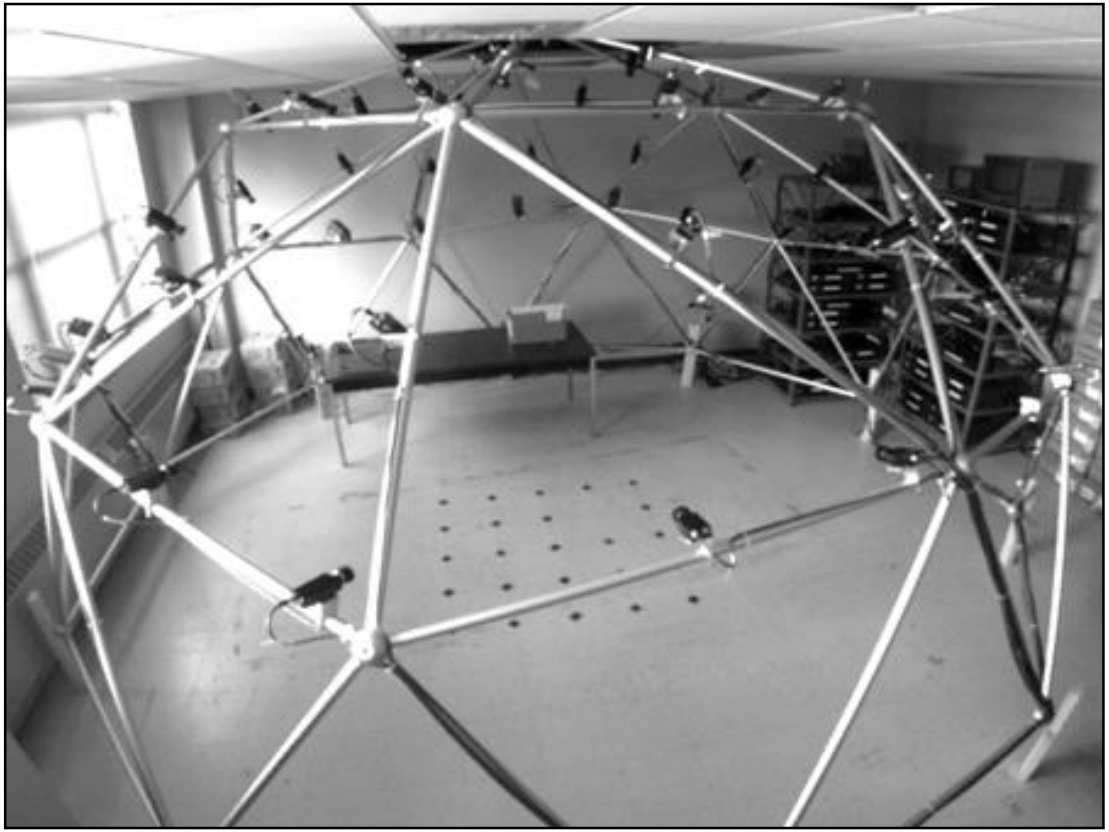


Whole-scene Modeling



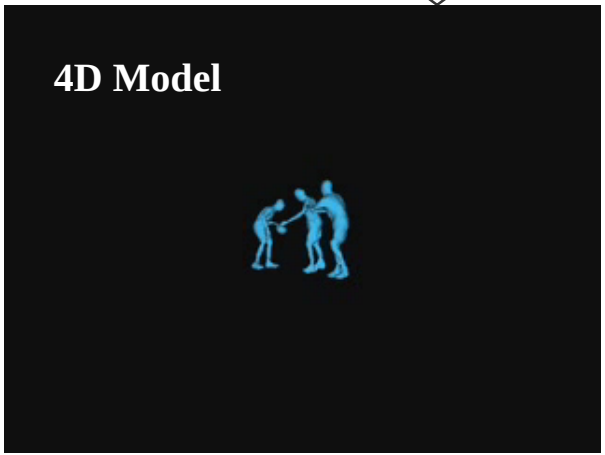
3D Dome (with 51 Cameras) – Analog system, circa 1995

51 analog cameras:
Synchronized and time-stamped
Recorded on VHS video tapes
Digitized later one by one

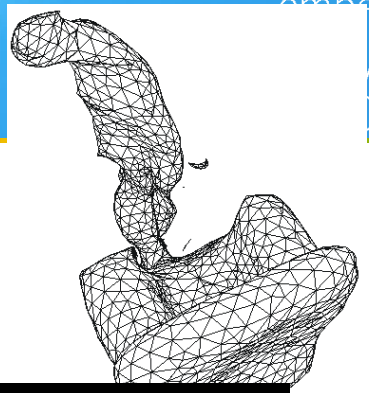


4D Digitization (1995)

Example:
3-Man Basketball



4D Digitization (~2000): Man-Sofa-Ball

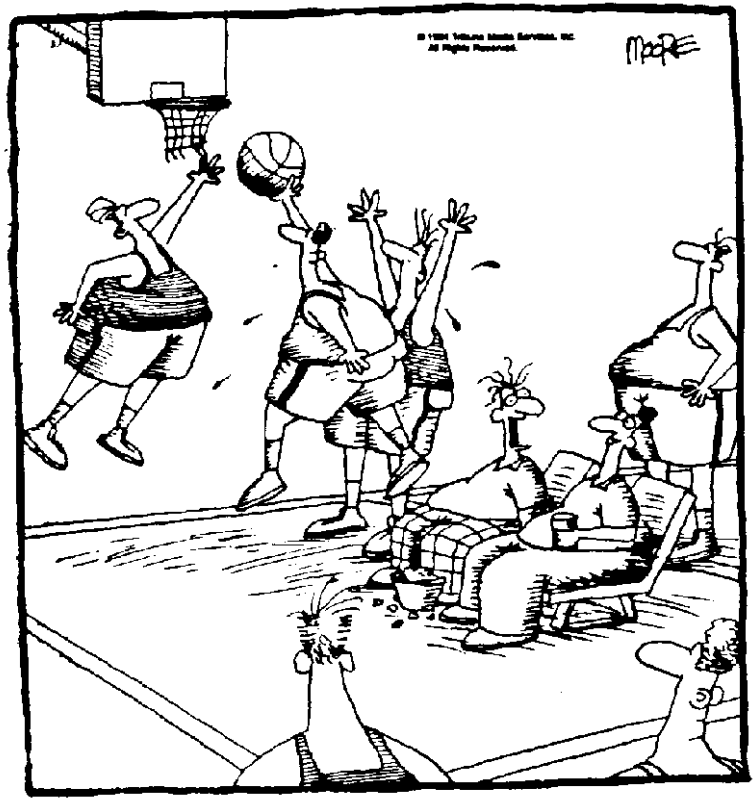


Digital 3D Room: 39 High Quality Cameras

(These are movies.)

Virtualized Reality

"Let's Watch the NBA on the Court"



"These are great seats!"

Multi-camera technologies are used in many places:

- Entertainment – Eye-vision like sports and stage video
- Large-area 3D modeling
- Multi-Aperture Camera (e.g. Lytro, Nikon, Hitachi)
  - Everywhere in-focus; Obstacle See-through
- Security – Surveillance
- Microscopy

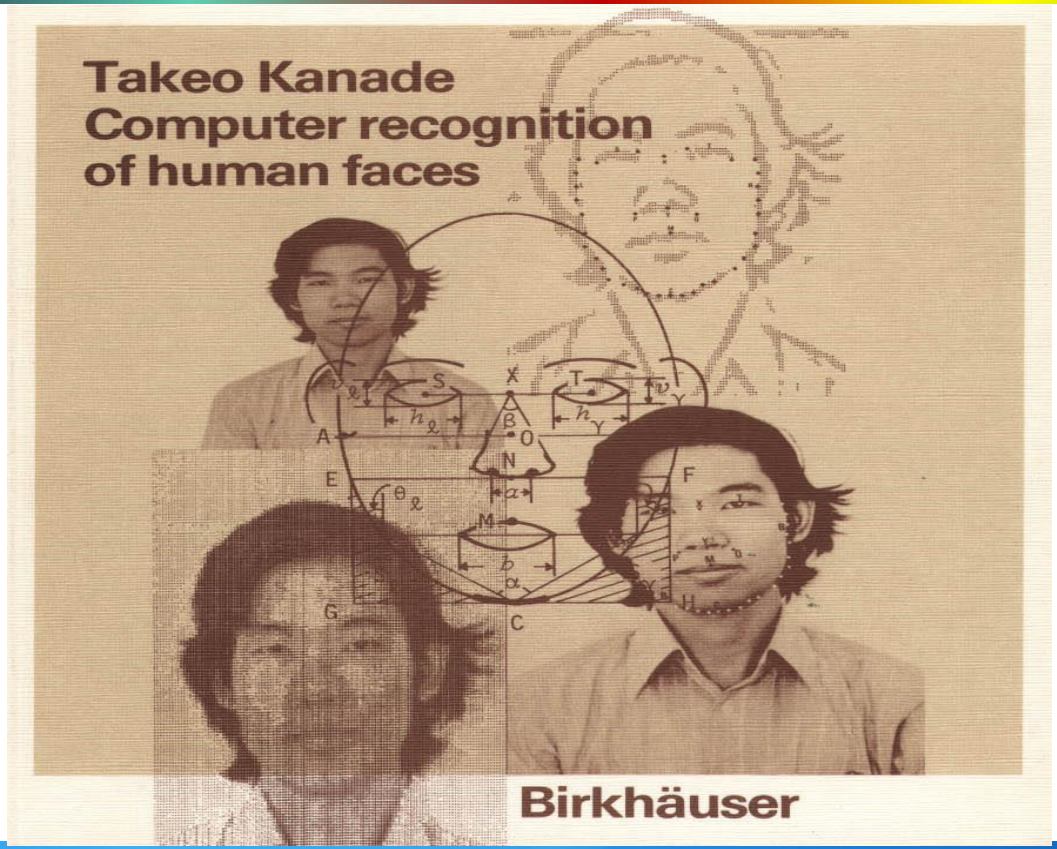
The very first paper on multi-camera stereo was rejected!

"… but, devices that use this many unnecessary cameras are too expensive to be useful." (one reviewer's comment)

Panoptic Studio: The 1000-Camera Virtualizing Engine Room (Currently 480 cameras)

Pictures of human – my favorite topic

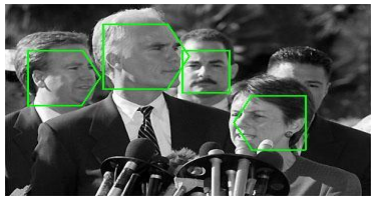
- My Ph.D Thesis

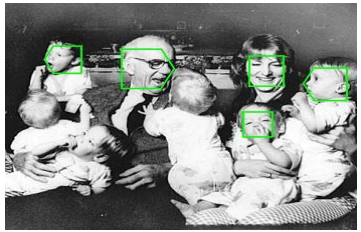


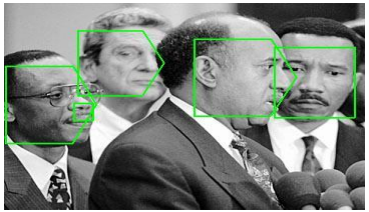
[Kanade73]

Processed 1000 images collected at World Expo 70

Face Detection



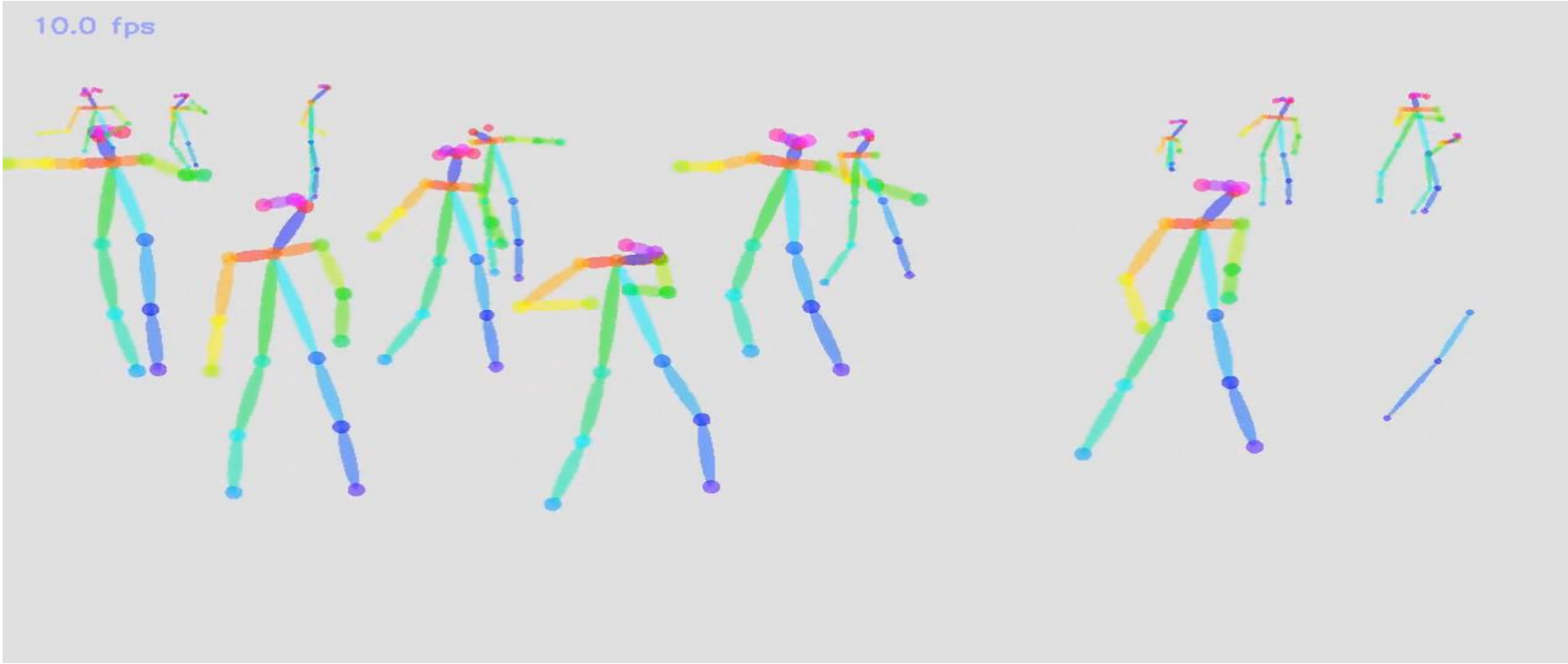




Rowley 1996, Schneiderman 1998

Recognizing Human Poses

[Hanbyul Joo, Yaser Sheikh, 2017]



Complete Visual Pose Recognition

[Hanbyul Joo, Yaser Sheikh, 2017]

Even, fingers

Further, if you use stereo, 3D body shape is available

[Hanbyul Joo, etal 2018]

Autonomous Driving: NAVLAB Project (~ 1984)

Terregator: ~1985
Campus pathway
<100m/hr, 100m total
Camera, Off-board computing

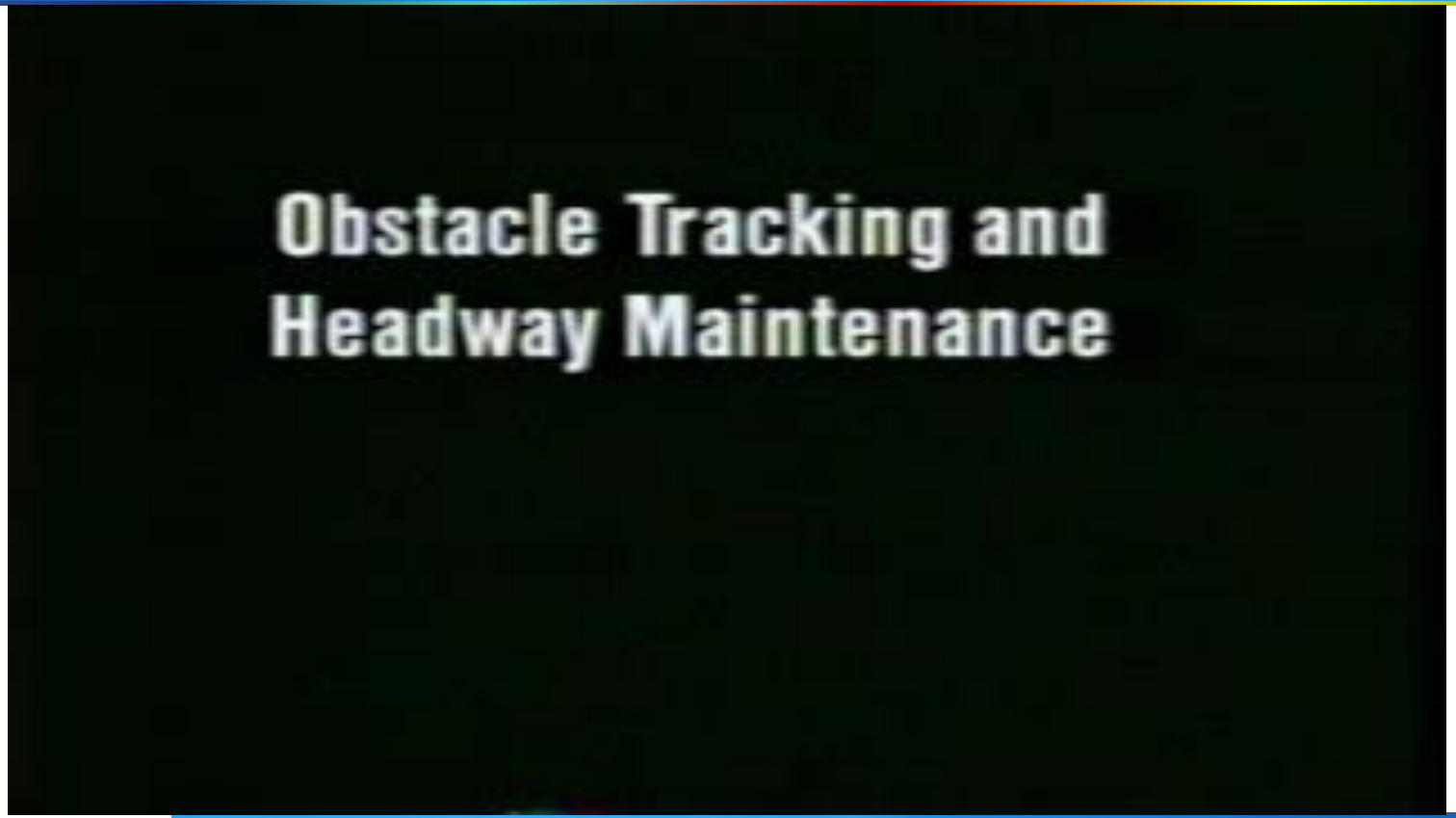
Navlab 1: 1986-89
Park pathway
~20 km/hr, 10km
Cameras, Laser range finder
Sun 3s and 4s

Navlab II: 1990-92
Natural Terrain, 20km/hr
Freeway, 70km/hr, 150km
Cameras, Laser range finder, Sparc 20s

Navlab Progress (1985~1990)

Nablab Progress (1990~1995)

Navlab Progress (1990~1995)



Navlab 1 - 10 lineups

Navlab 6-10: 1997

Integration of cameras, radars, GPS, motion sensors, PCs
AHS Demo 1997 in San Diego
1500 riders, 125,000 miles

No Hands Across America: 1995

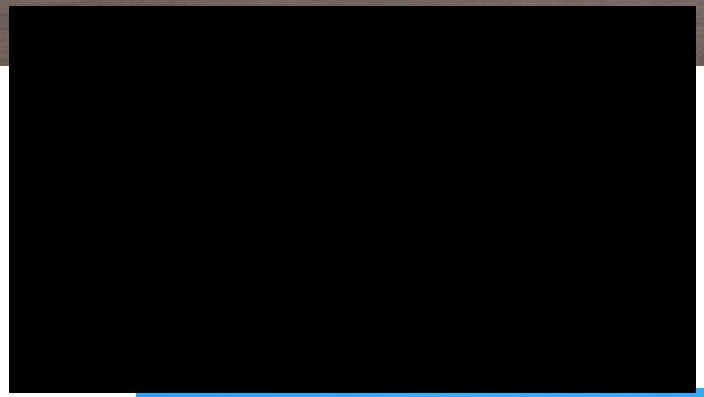
Using computer vision, the **Navlab 5** steered 98.2% of the distance from Washington, D.C. to San Diego CA. (2797 miles out of 2849)

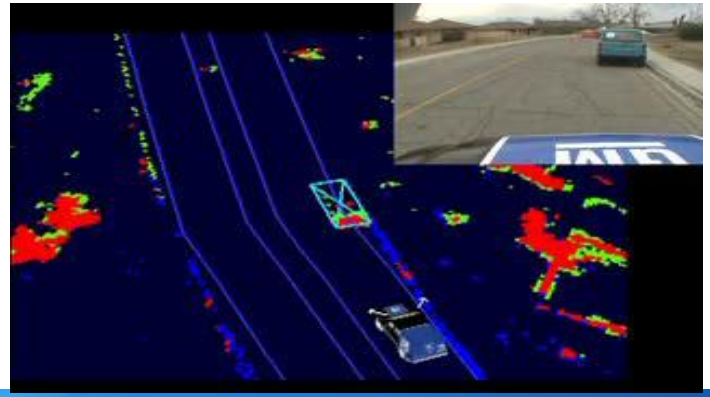
Navlab 5 mingling with the stars!

Urban Grand Challenge: 2007

The “CMU Boss”: Red Whittaker and his team

Complete autonomy

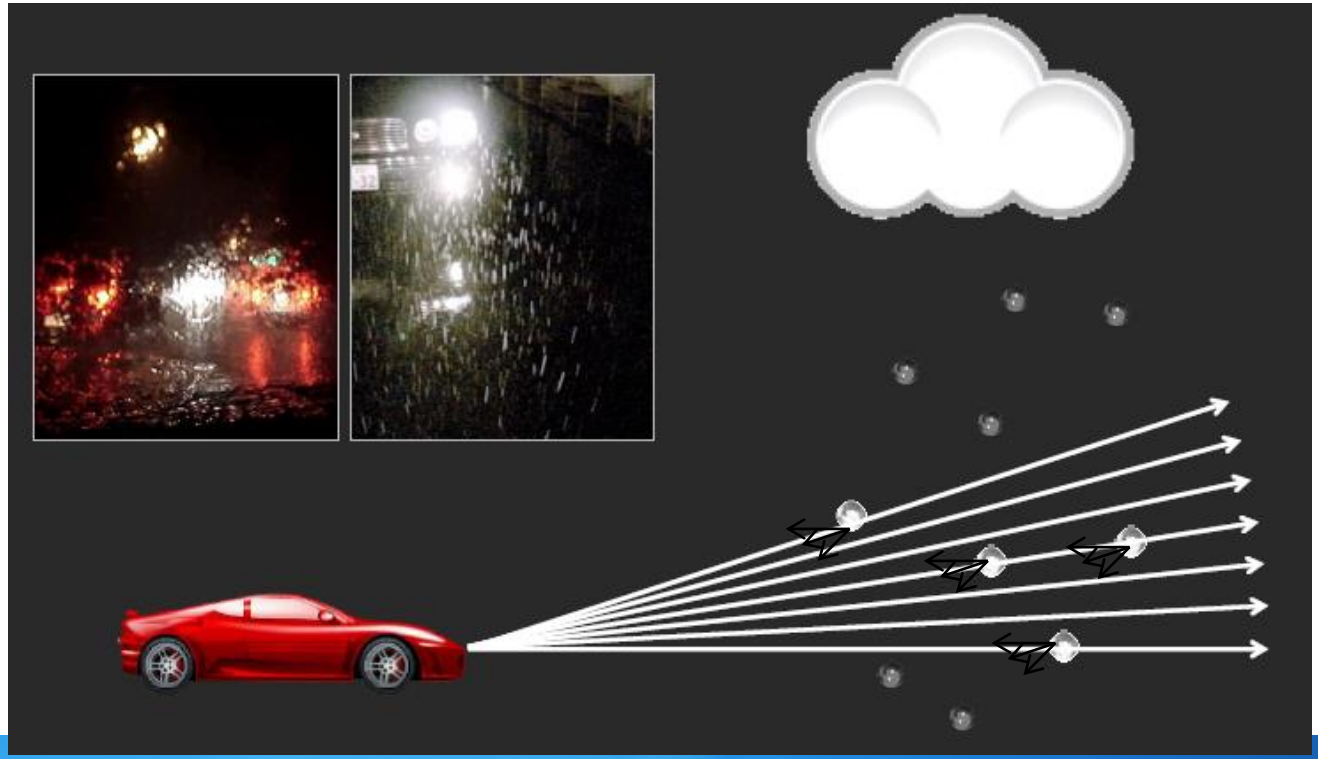




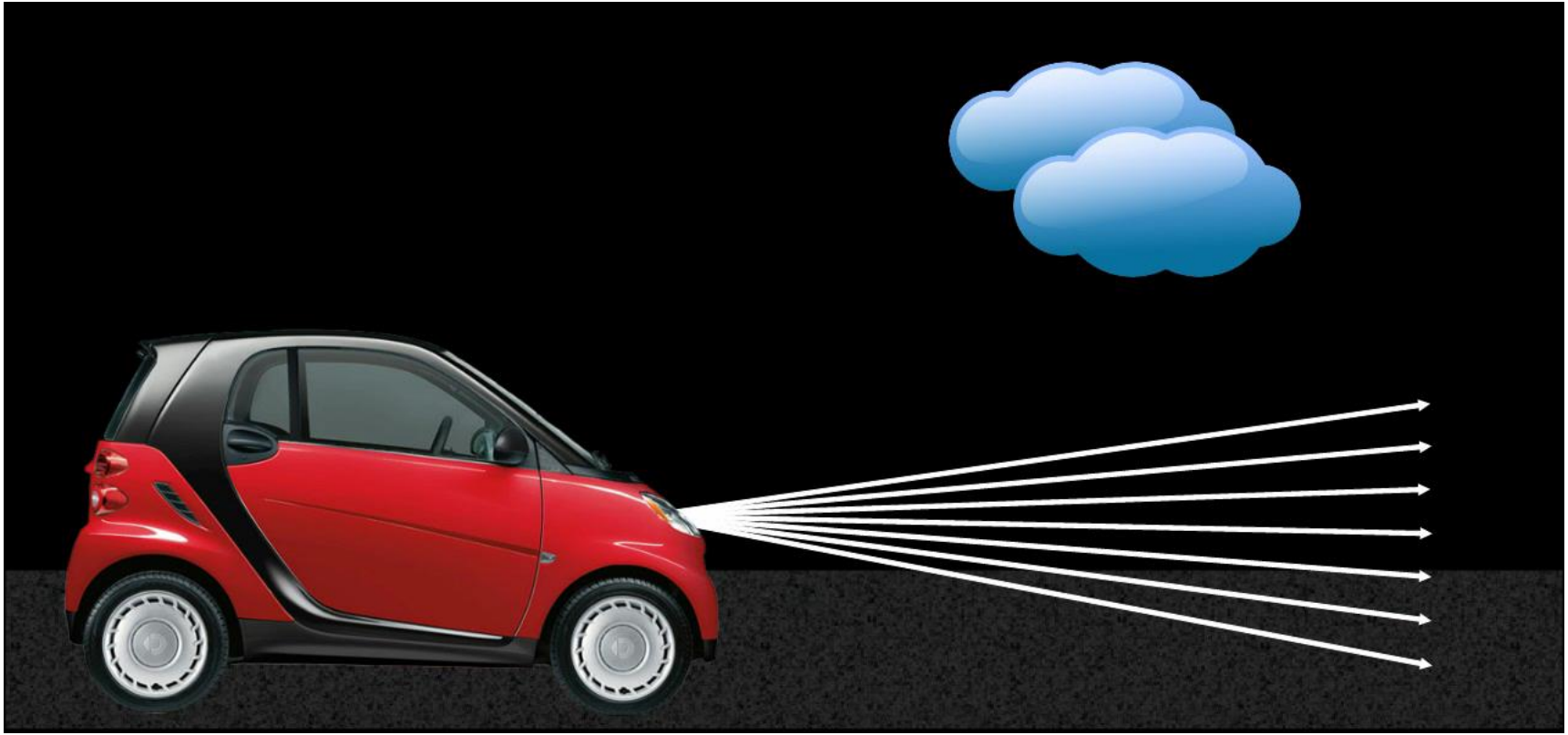
Smart Headlight

Driving in Rain or Snow at Night

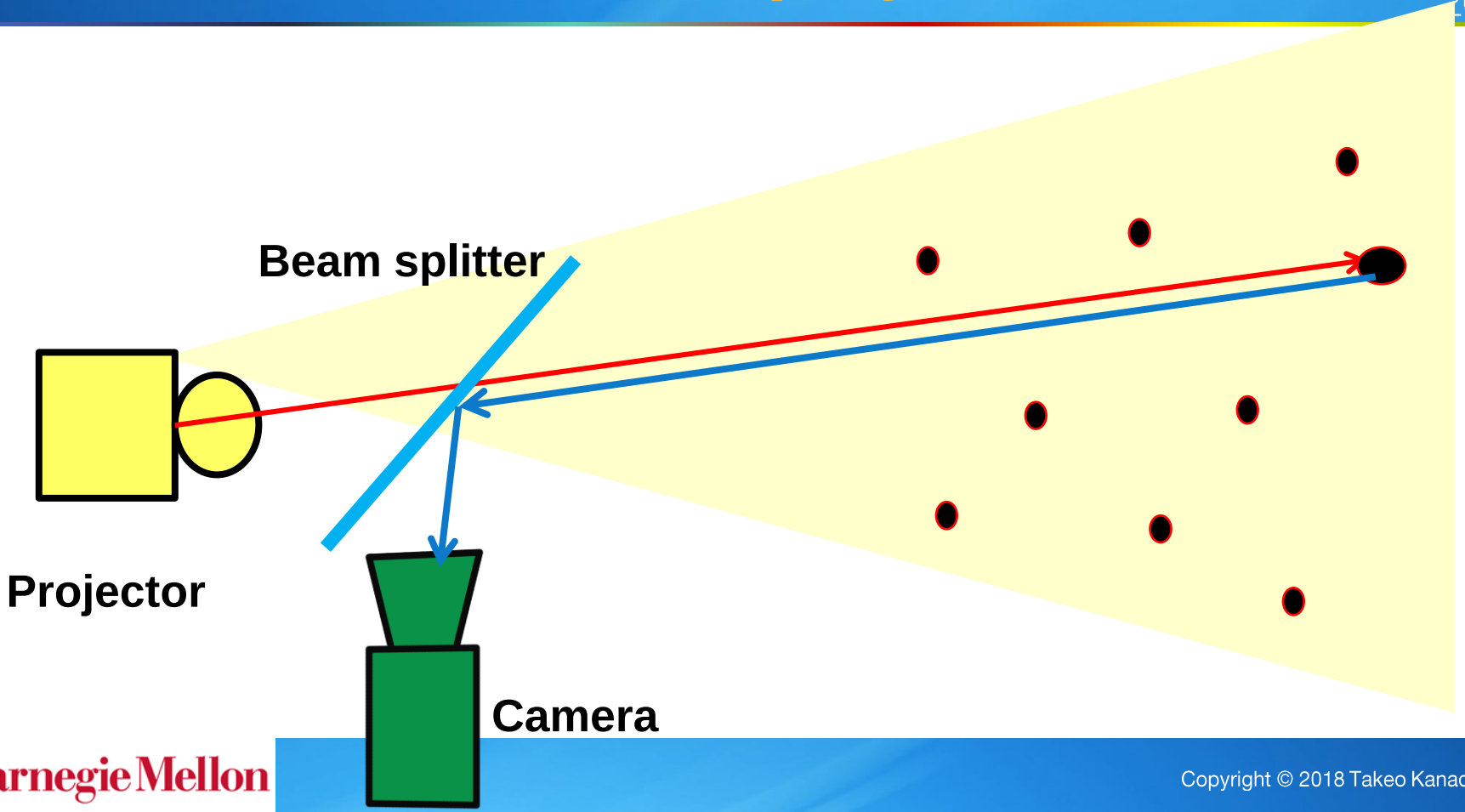
Rain drops and snow flakes are highly reflective



Control individual rays of headlight



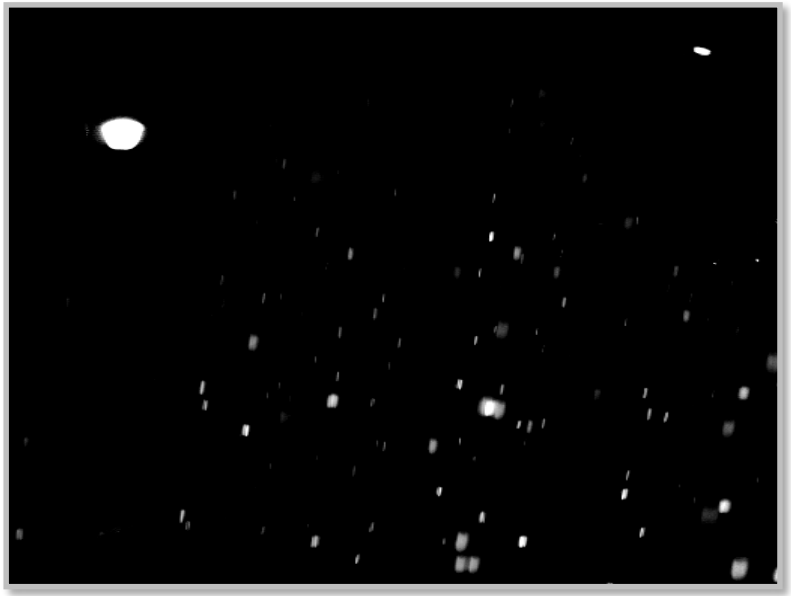
Confocal camera and projector



Rain Streaks vs. Rain Drops

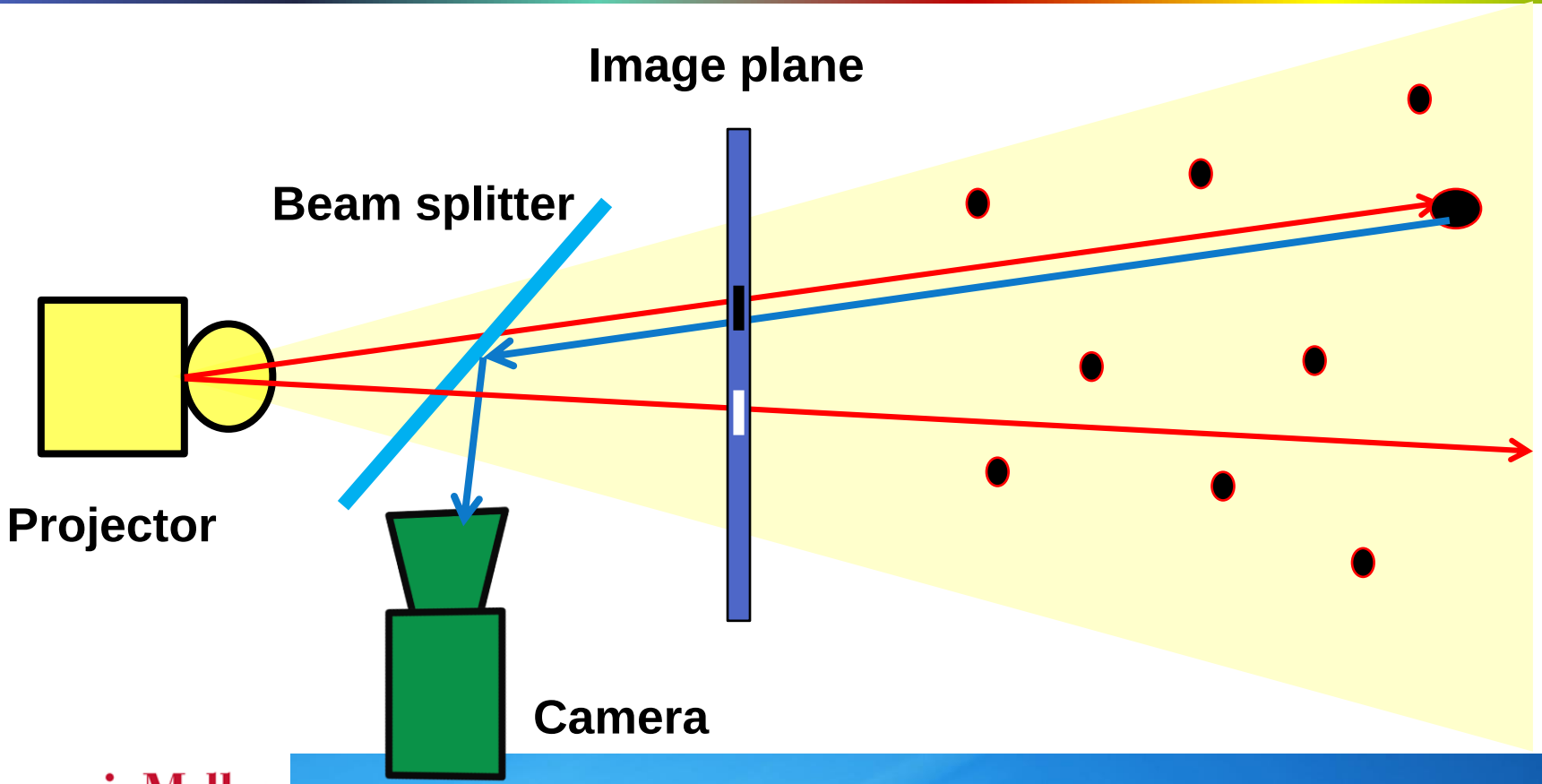
Rain streaks appear dense but drops are sparse

Long Exposure Time (12 ms)



Short Exposure Time (1 ms)

Confocal camera and projector

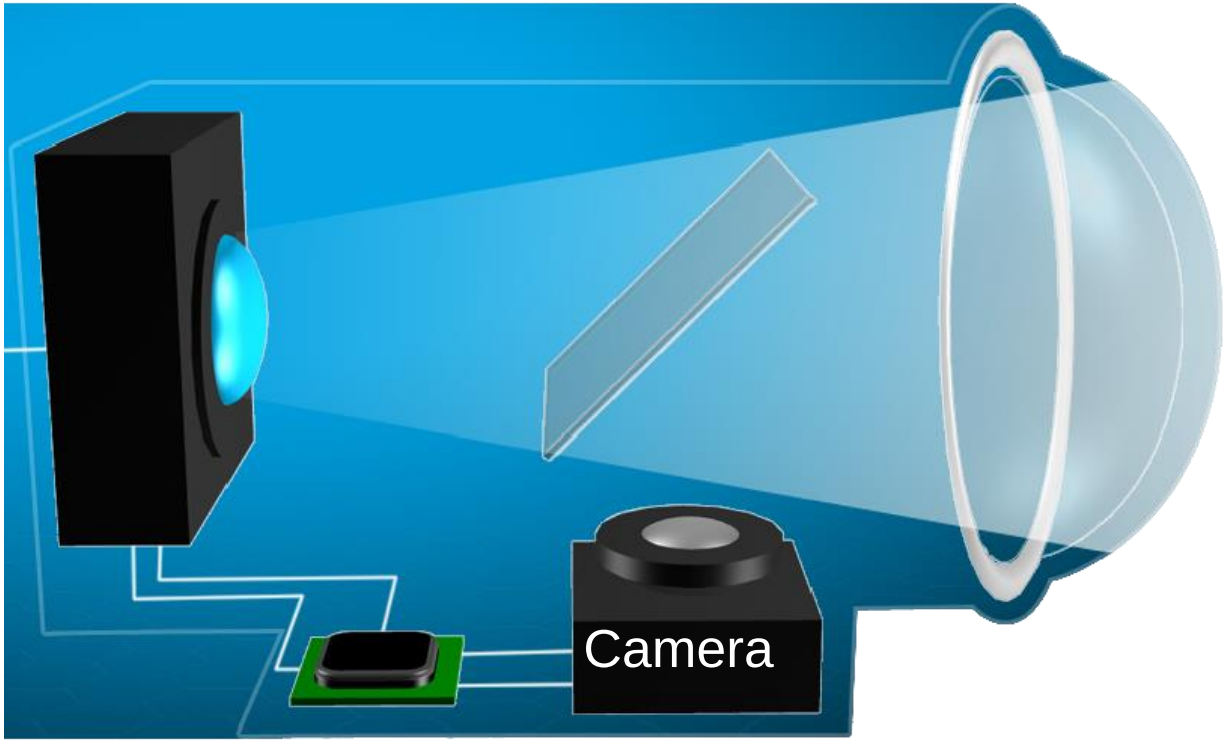


“Light Throughput” matters:

What percentage of light get through without hitting rain/snow drops.

Programmable and Reactive Headlight

Spatial light modulator



DIY Smart Headlight: Lamp

Custom-built DLP projector turned into a Smart Headlight

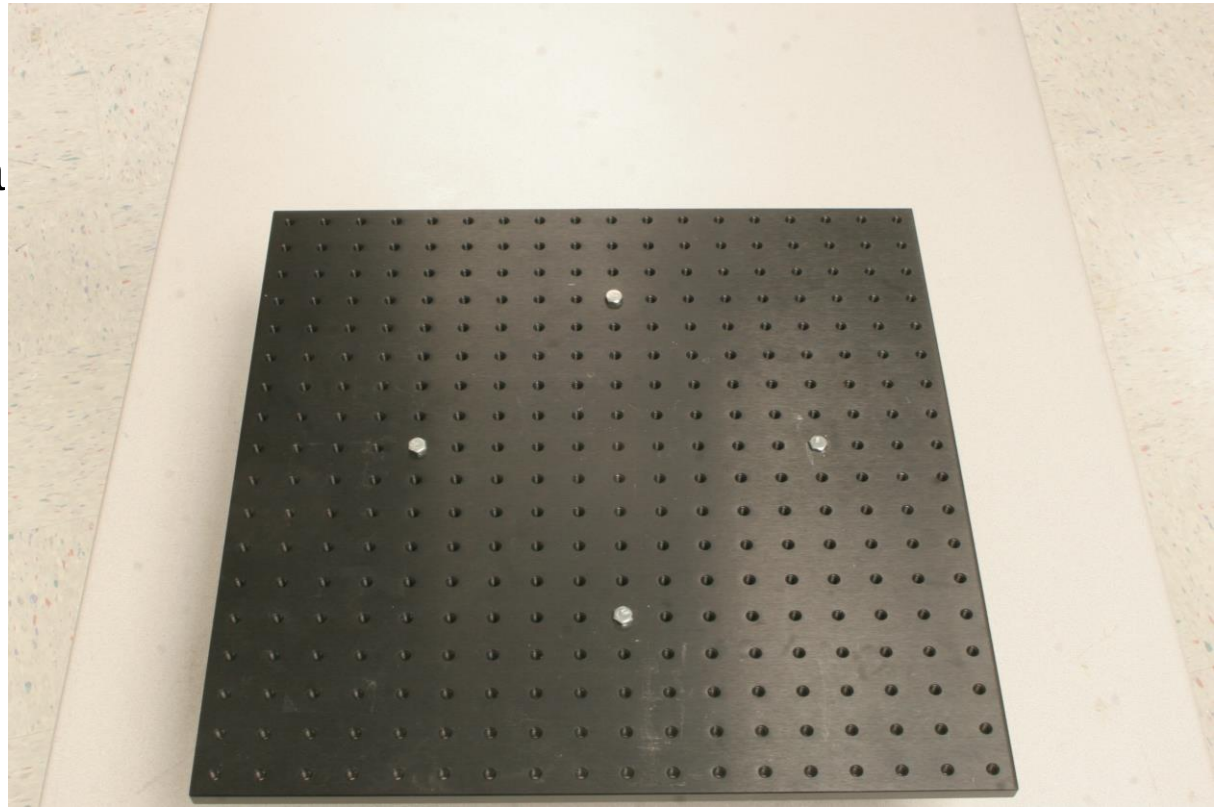
4000 Lumens running at 1700 Hz

revised

DIY Smart Headlight: Lamp, Camera, Processor combined

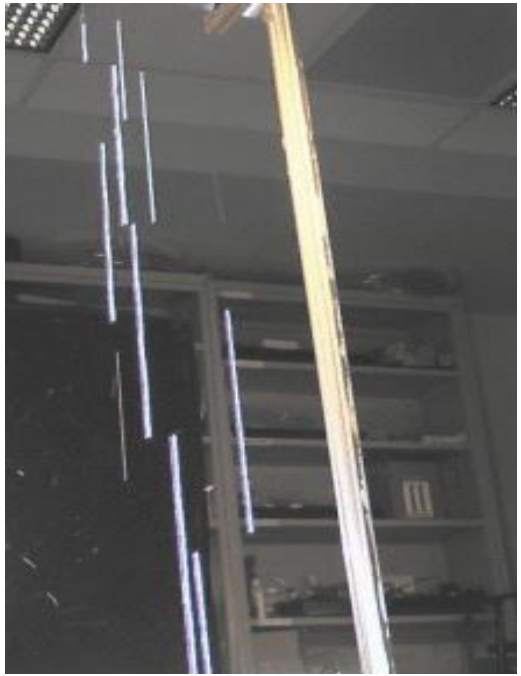
Co-axial setting of Projector and Camera

Reaction time 1.5 milli-seconds: 500 times faster than the average driver reaction time




VIRGINIA
WSS-7424

Making Rain Disappear



Standard Headlight

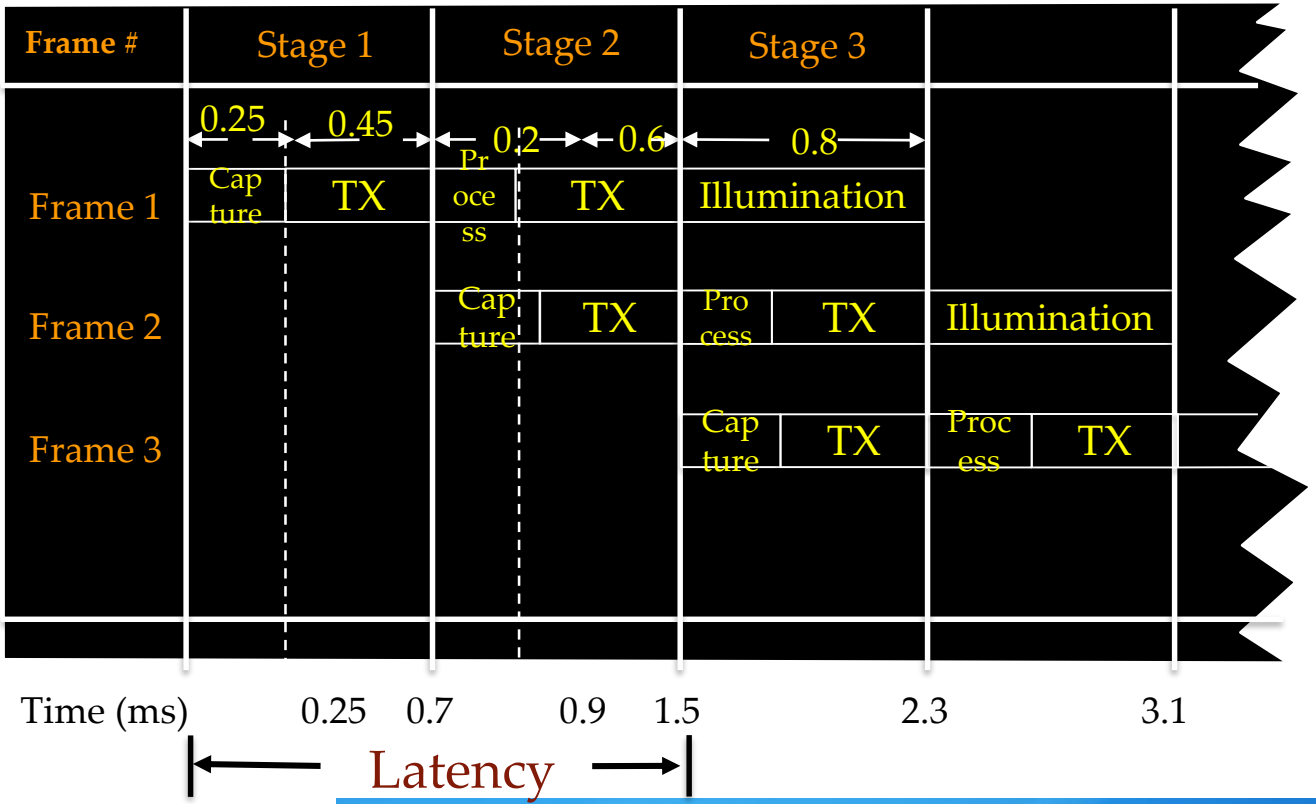
Smart Headlight

70% Visibility Improvement (5-10% Light Loss)

Dis-Illuminating Snow

60% Visibility Improvement (10-15% Light Loss)

The Key Technical Issue -Latency



Frame rate ~1000 fps
Latency: 1.5 msec

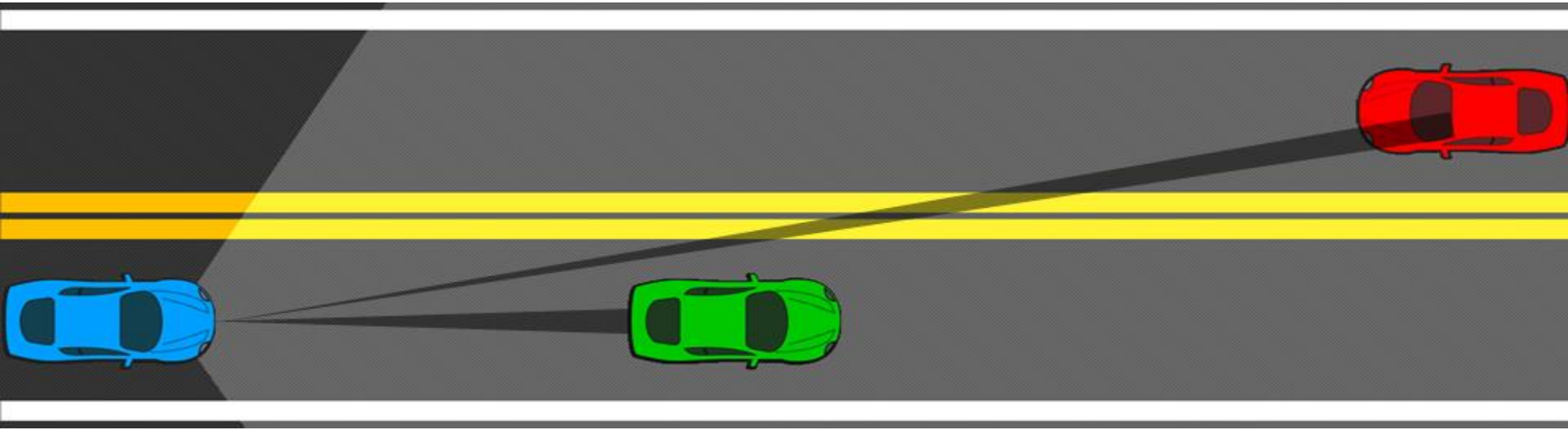
Limitation of the current approach: Separation of sensor, processor, and projector

Many problems become easier as you get faster!

Headlight High-Beam Glare

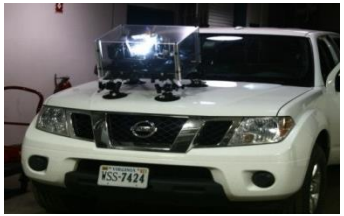
“A 55-year-old takes eight times longer to recover from glare than a 16-year-old” – AAA Safety Research
450 Fatalities per year – NHTSA 2007

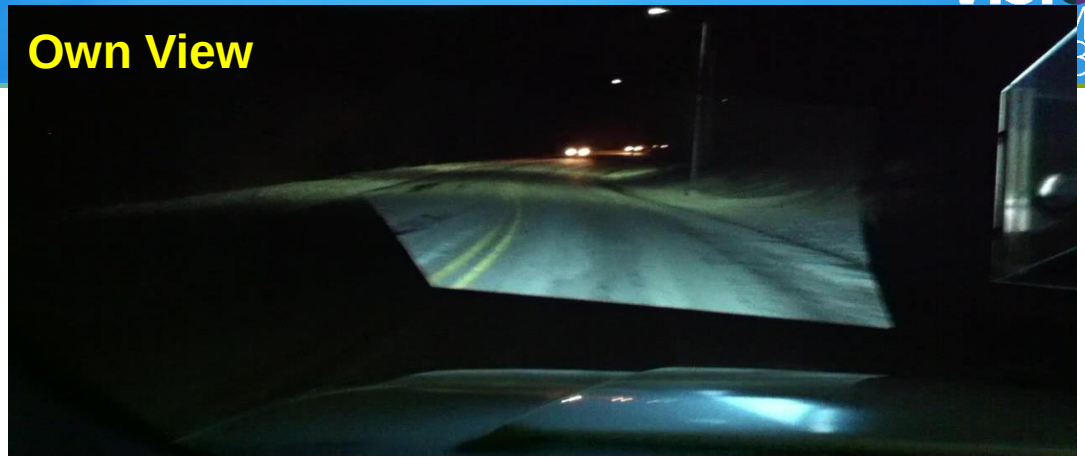
Anti-Glare Persistent High-Beams



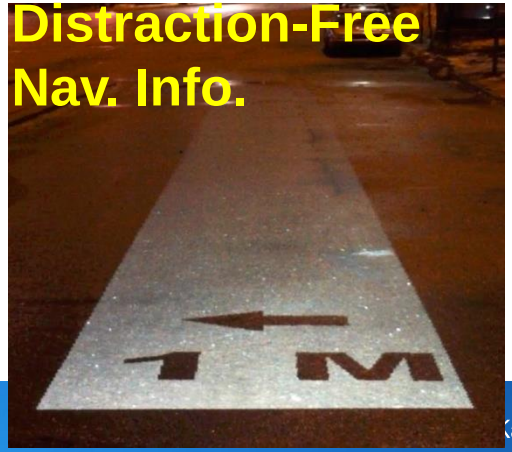
Turn off rays that go into the oncoming driver's eye
Little perceivable difference to smart-headlight driver.

Car with a Smart Headlight





Better Visibility and Inf. Display



“Genuine” Augmented Reality !?

Alter how the reality appears to human, not the display of the reality

On a Pickup Track

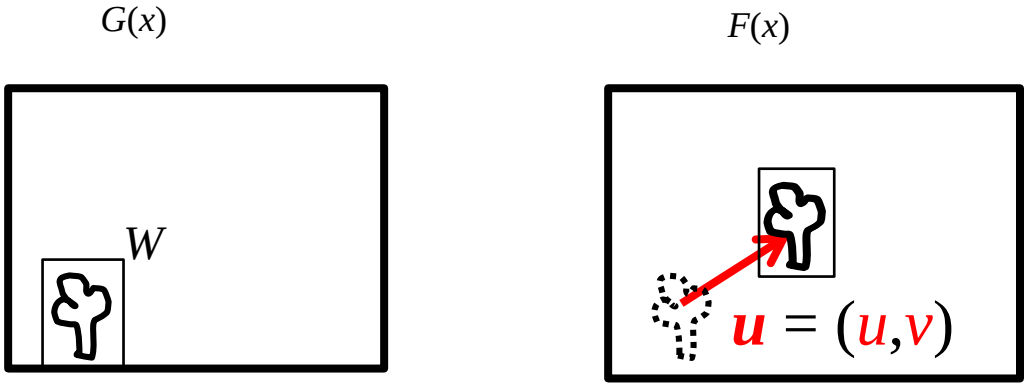
Value of Research

Research has to be new?

“Newness itself is not virtue, usefulness is.”
(Takeo Kanade)

In general, however, it is hard to predict the value of research.

We are working on a tracking problem, circa 1980

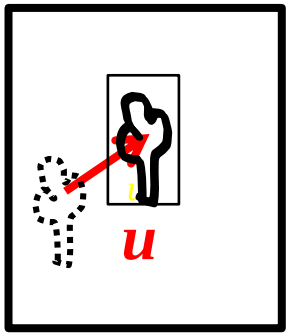


$$E(\boldsymbol{u}) = \sum_{x \in W} (G(\boldsymbol{x}) - F(\boldsymbol{x} + \boldsymbol{u}))^2$$

Search for $\boldsymbol{u}$ that minimizes this SSD error

Bruce Lucas

Lucas-Kanade Optical Flow [Lucas, Kanade 81]



minimize $$E(\boldsymbol{u}) = \sum_{x \in W} (G(\boldsymbol{x}) - F(\boldsymbol{x}+\boldsymbol{u}))^2$$

Small $\boldsymbol{u}$ → Taylor-expand $F(\boldsymbol{x}+\boldsymbol{u})$

$$F(\boldsymbol{x}+\boldsymbol{u}) \cong F(\boldsymbol{x}) + \frac{\partial F}{\partial \boldsymbol{x}}^T \boldsymbol{u}$$

→ $E(\boldsymbol{u})$ becomes quadratic of $\boldsymbol{u}$

→ The answer $\boldsymbol{u}$ can be obtained explicitly as

$$\boldsymbol{u} = \sum (G(\boldsymbol{x}) - F(\boldsymbol{x})) \frac{\partial F}{\partial \boldsymbol{x}} [\sum \frac{\partial F}{\partial \boldsymbol{x}} \frac{\partial F}{\partial \boldsymbol{x}}^T]^{-1}$$

Make it work: THE best way to convince

“If your audience ask why your method works so well, they are not yet convinced. If they were, they would ask ”How much is it?” (Takeo Kanade)

Do fast.

“If you come up with a good idea, there are at least two more people who think the same.” (Takeo Kanade)

“Who said it first is not important, who gets there first is.” (Takeo Kanade)

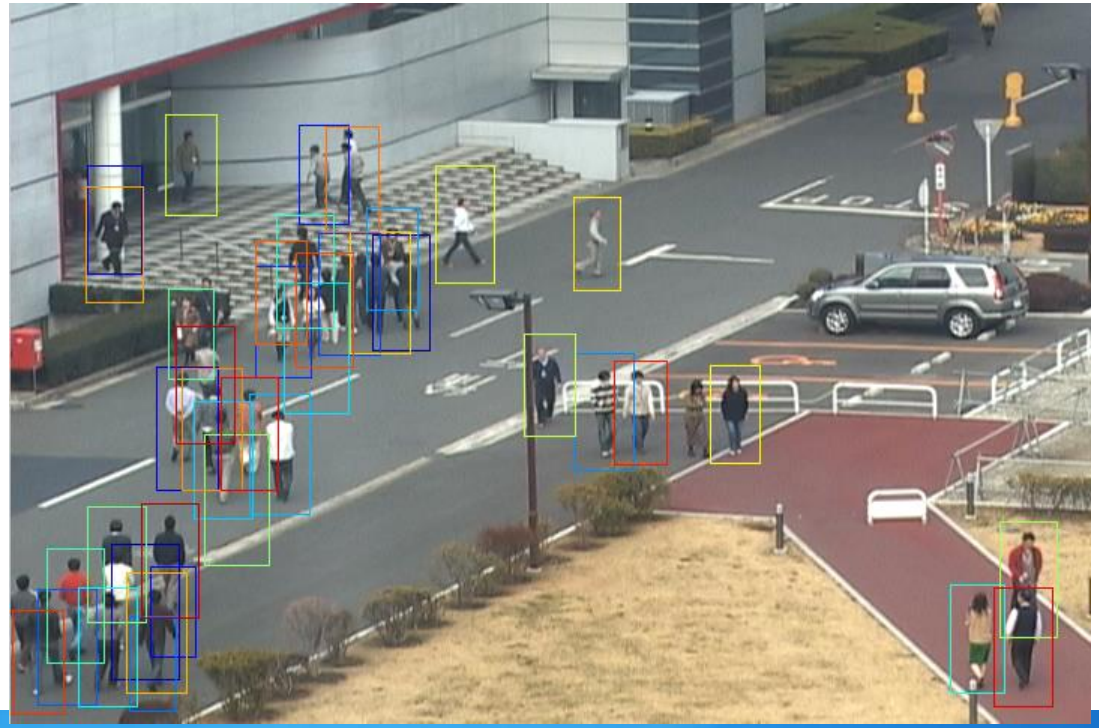
Computer Vision is in the state of "Perfect Storm"

- New, ubiquitous, embedded, sometimes super-human vision sensors
- Powerful vision and learning algorithms, readily available on the net
- Low-power high-performance vision processors, high-speed connected to cloud

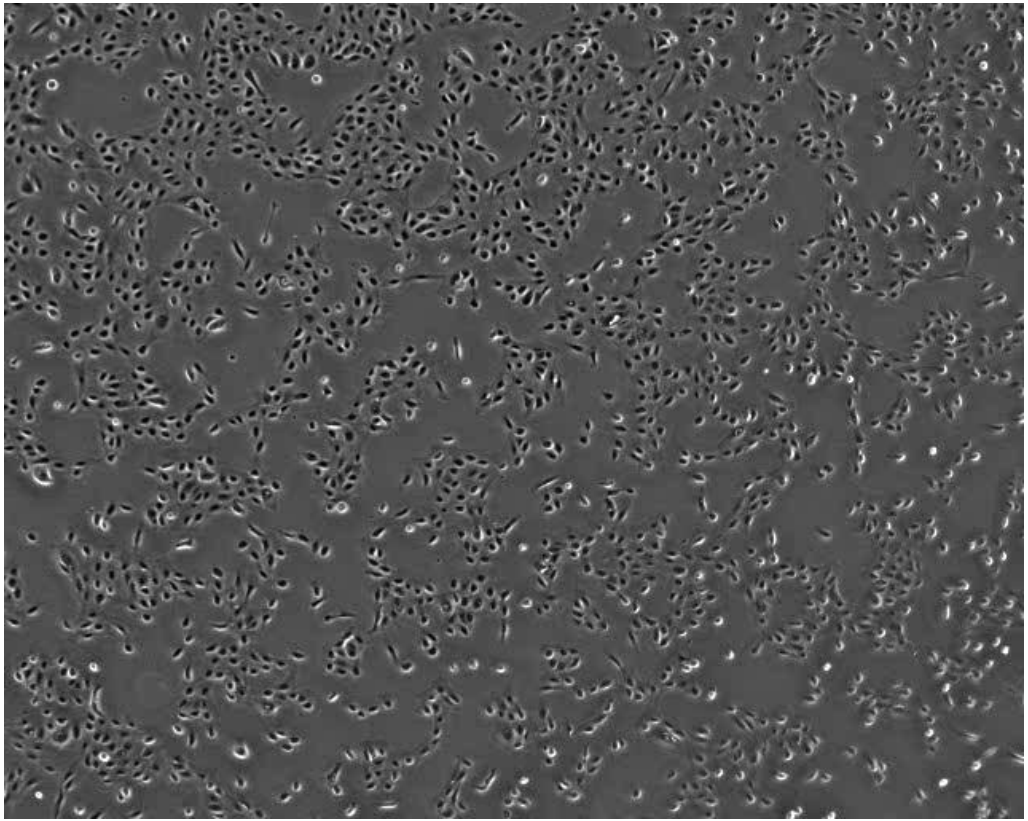
Advice! "In CV, when images are taken, it is too late." (Takeo Kanade)

“Problems are waiting for you to solve.”

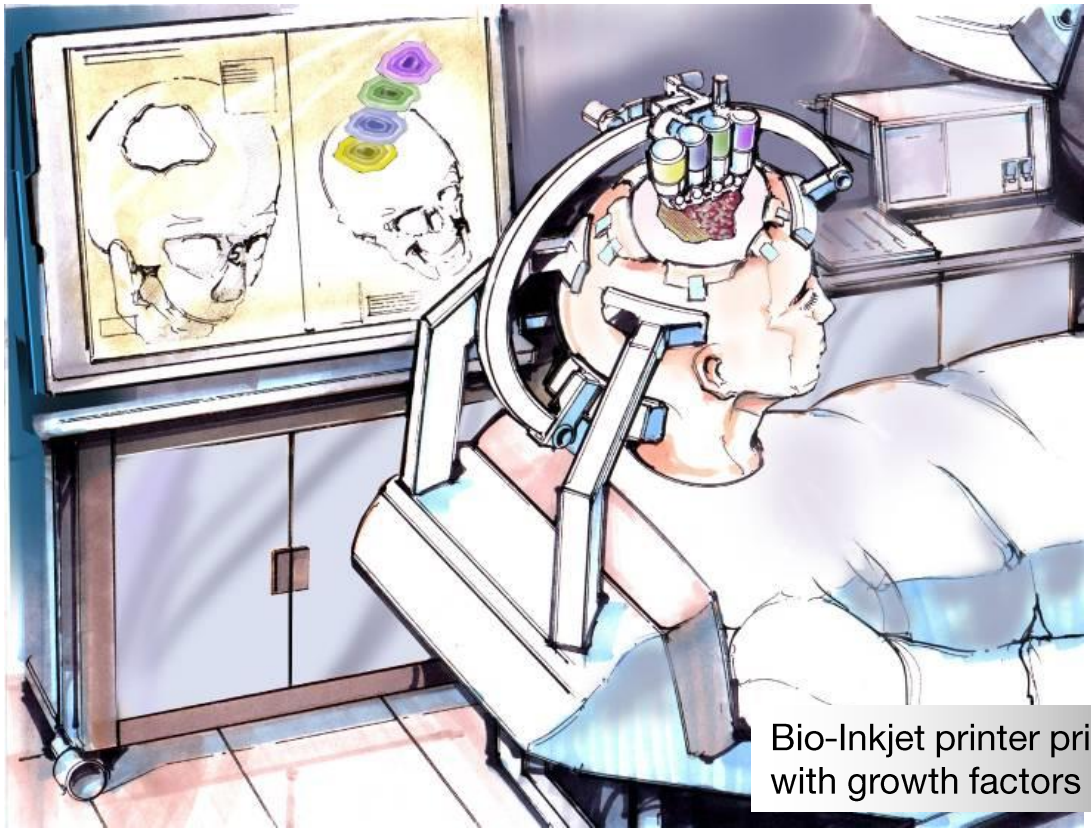
A Typical Tracking Problem in Vision



A bigger-scale tracking problem?

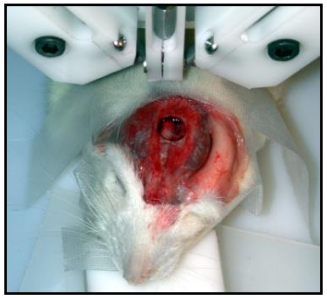


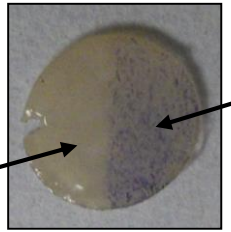
Tissue Repair by Bio-Inkjet Printing



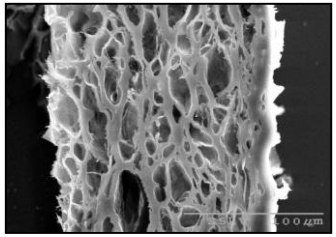
Bio-Inkjet printer printing scaffolds with growth factors in situ

Spatial Control of Bone Regeneration *In Vivo*





Printed directly onto collagen-based scaffold



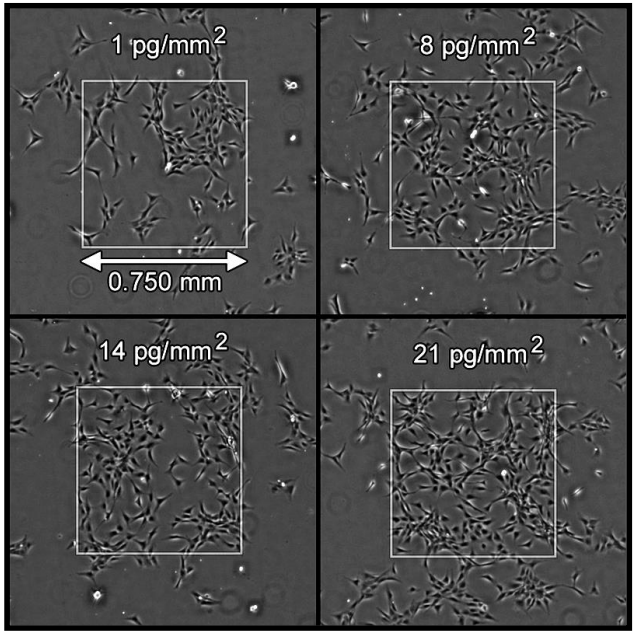
Cross-section through scaffold

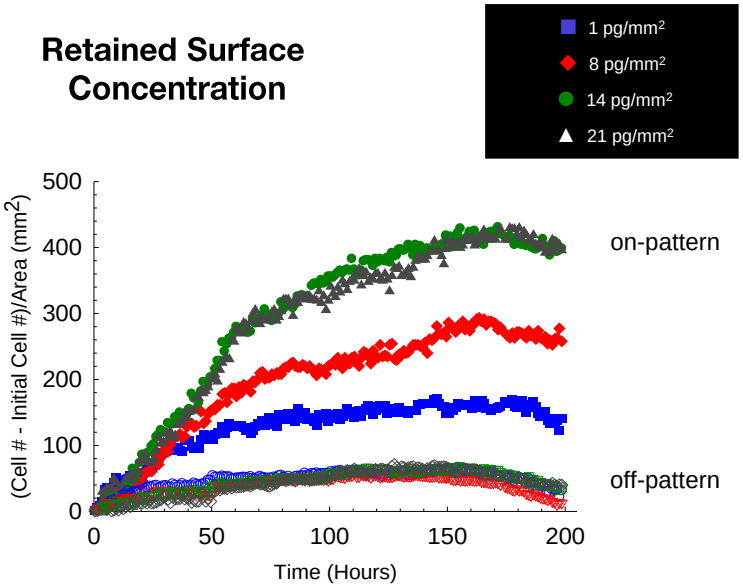
[Weiss, Campbell, et al,]

Dose-Dependent Cell Response

Manual or Semi-manual post-data acquisition is a bottleneck to progress

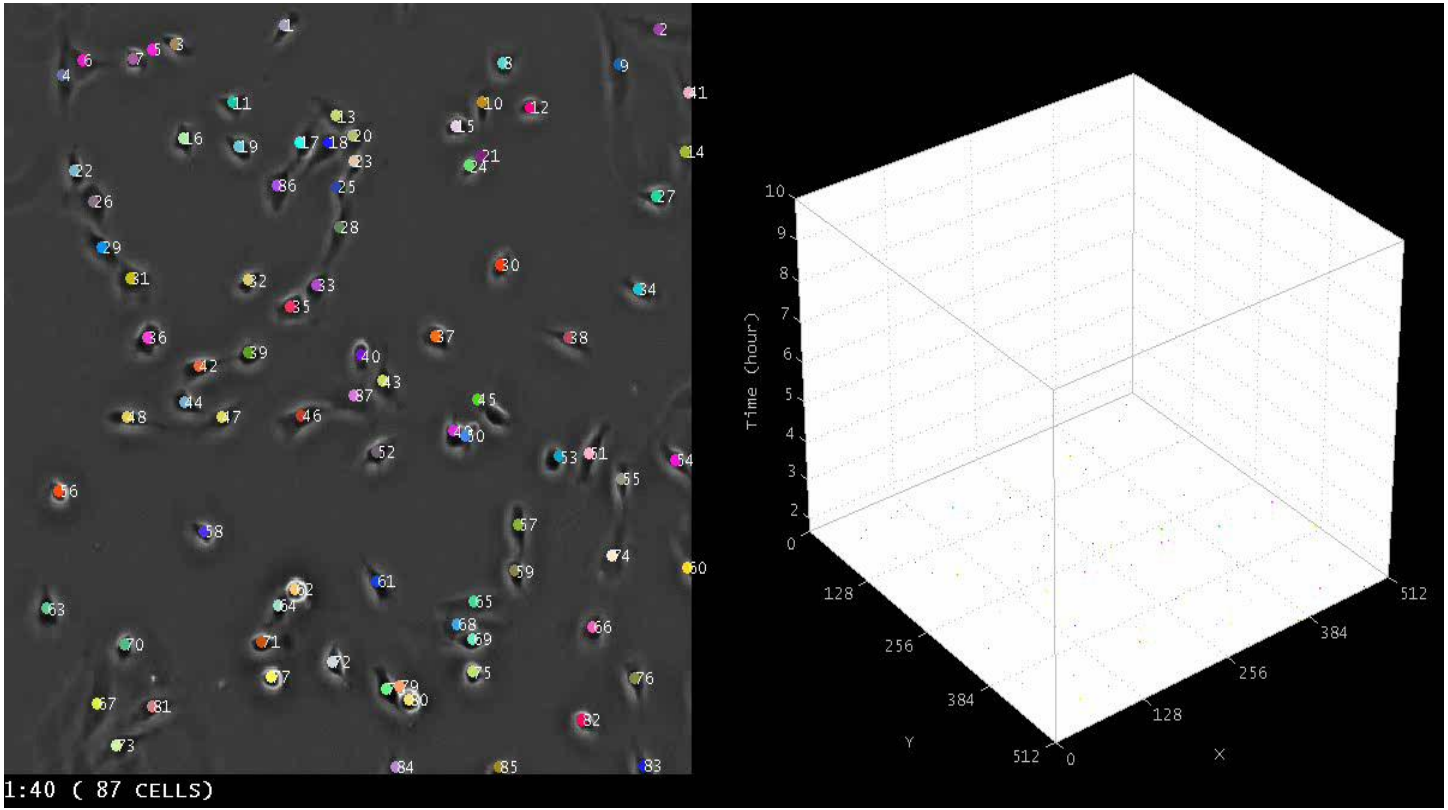
More information can be extracted from these image stacks

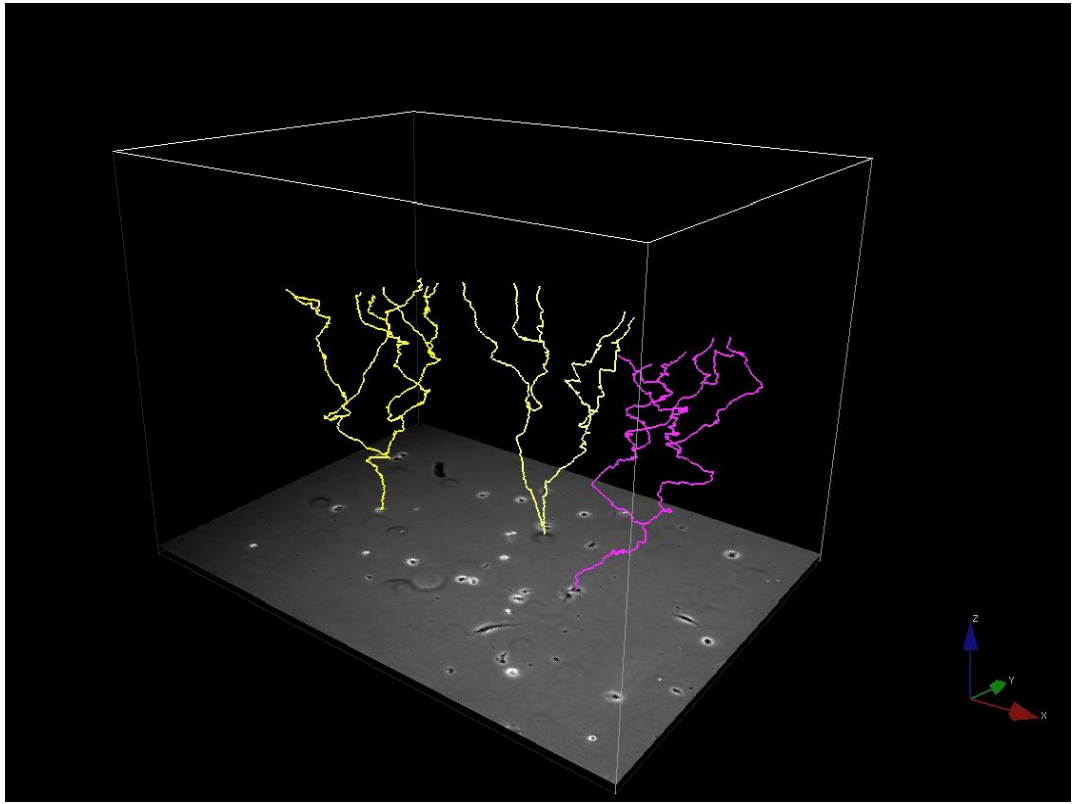




Miller, E.D., et al. **Biomaterials** 2006;27:2213-2221

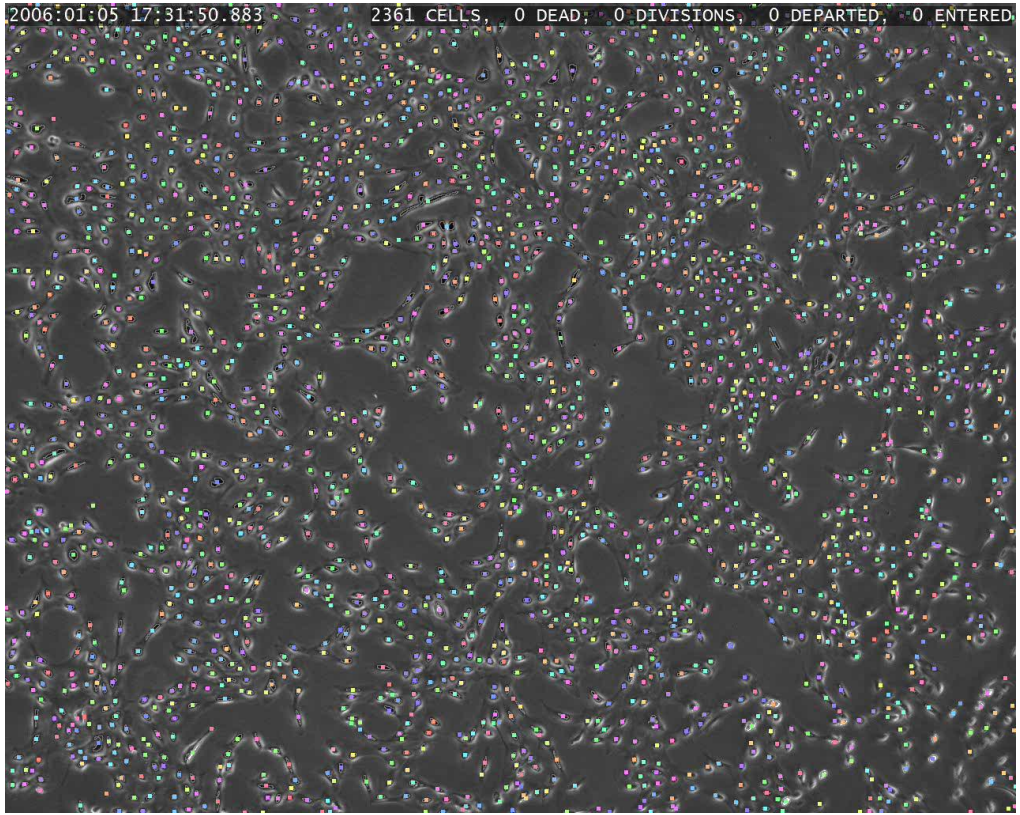
Tracking Result: Spatial-Temporal View





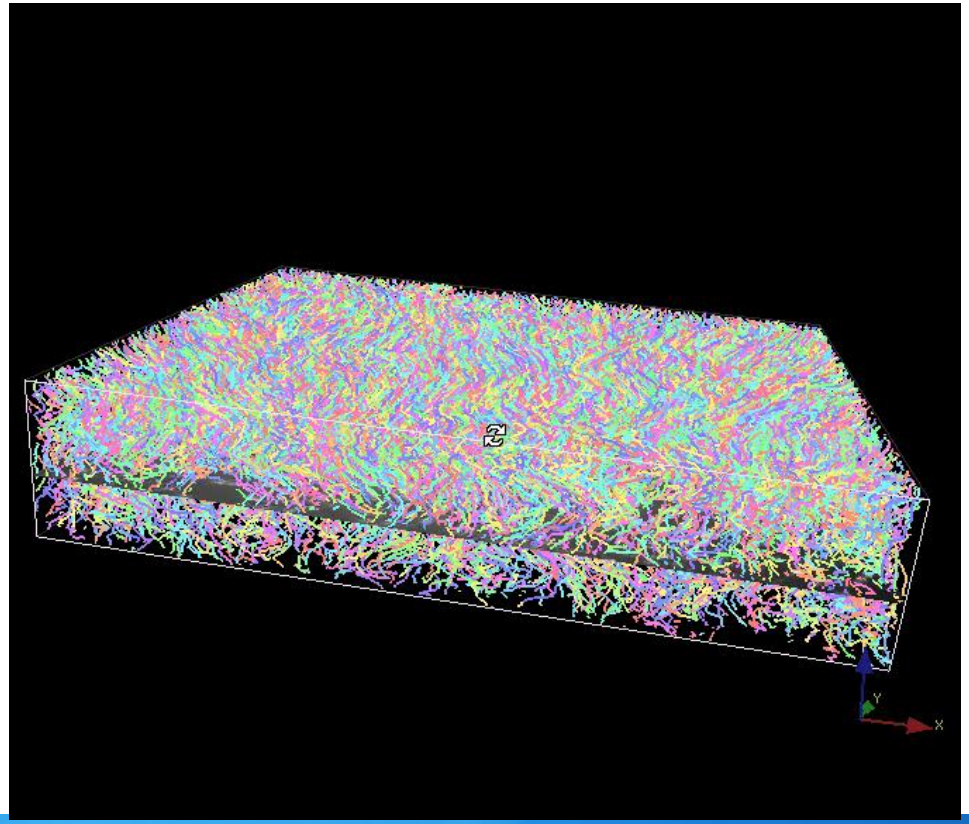
Tracking Thousands of Stem Cells

256 Frames, 10-minute frame interval, 1280x1024 pixels/frame30 Seconds/Frame Tracking Time Intel Xeon 2.66 GHz CPU 4Gb RAM

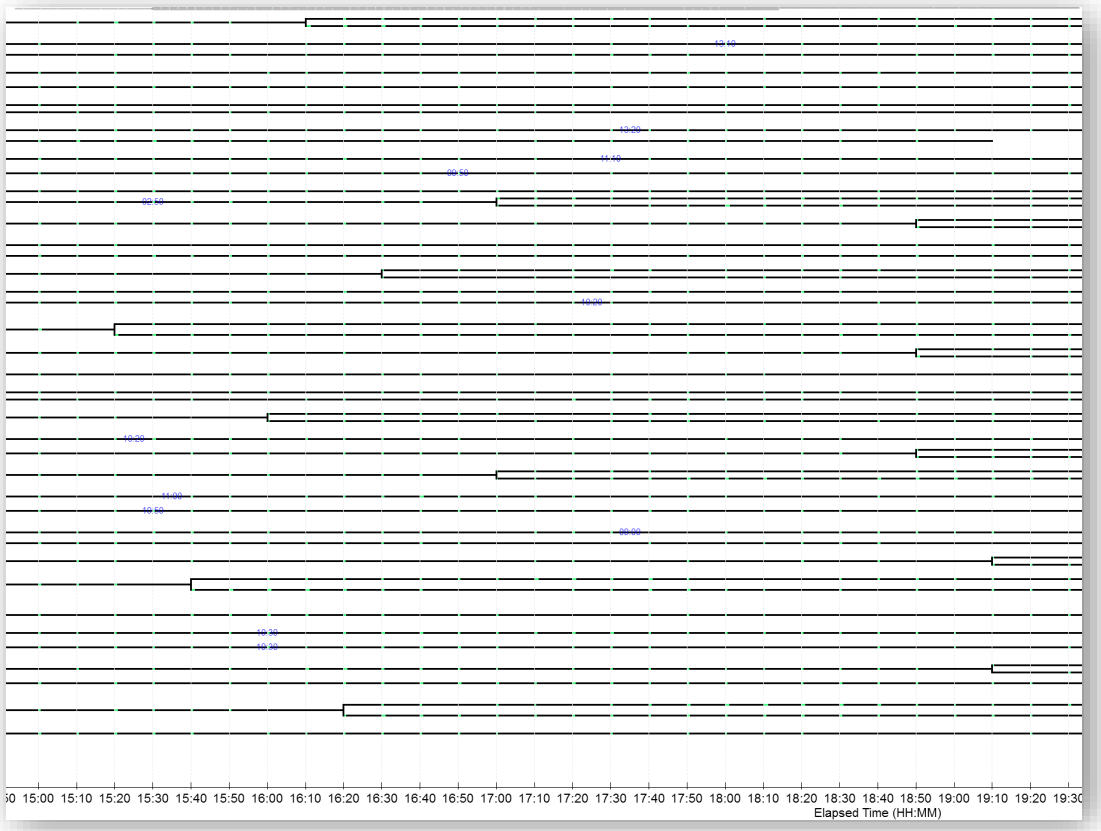


Noodle video

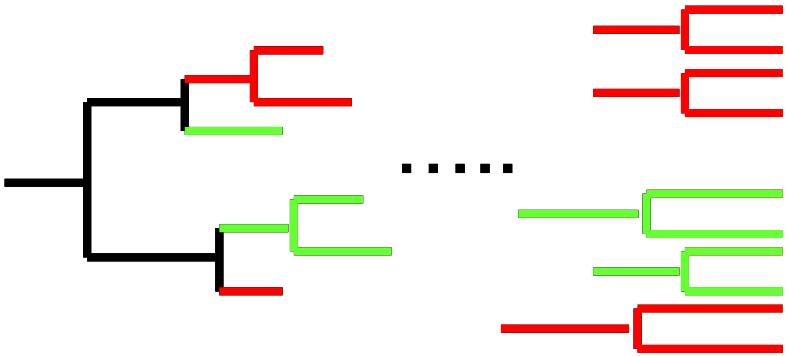
30 Seconds/Frame Tracking Time Intel Xeon 2.66 GHz CPU 4Gb RAM



Cell Population Lineage

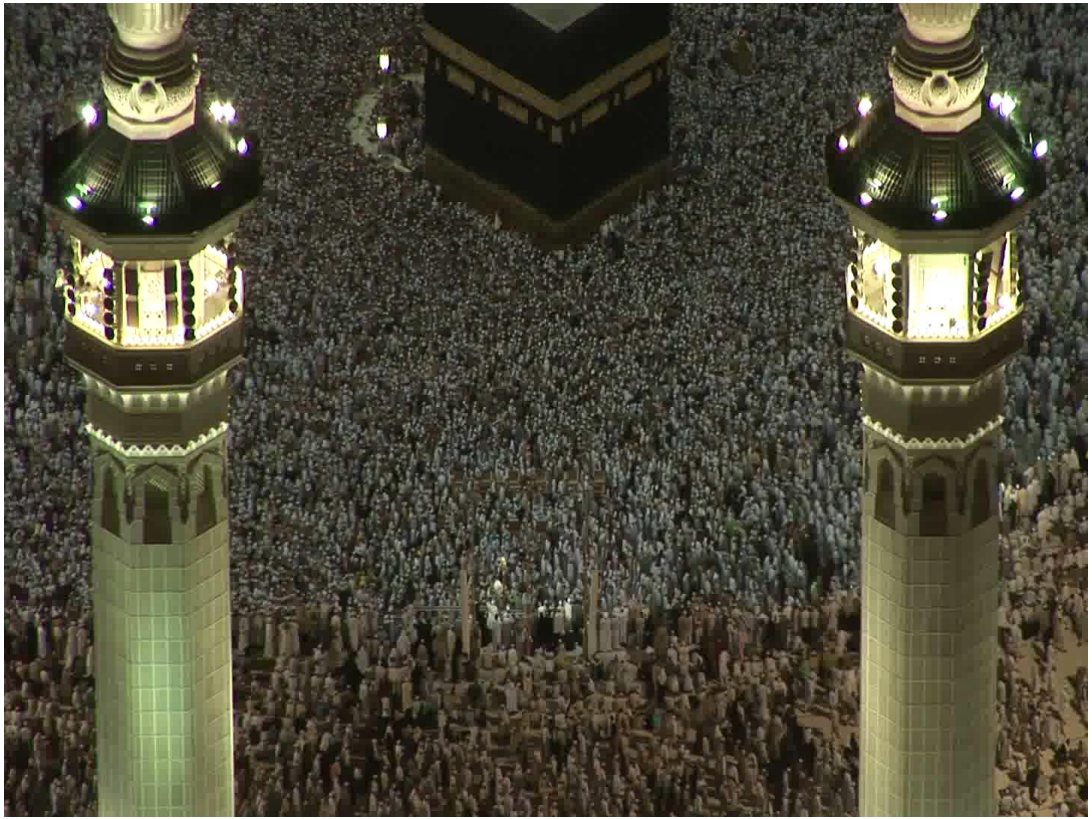


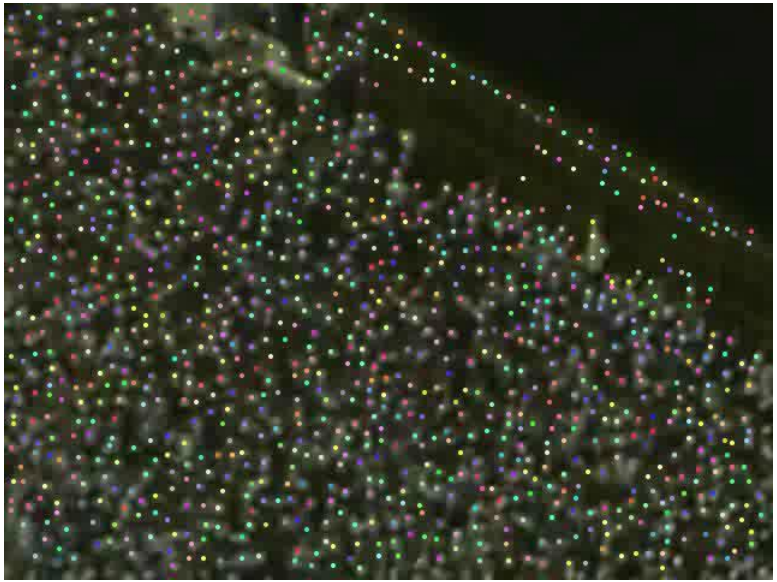
Many interesting questions

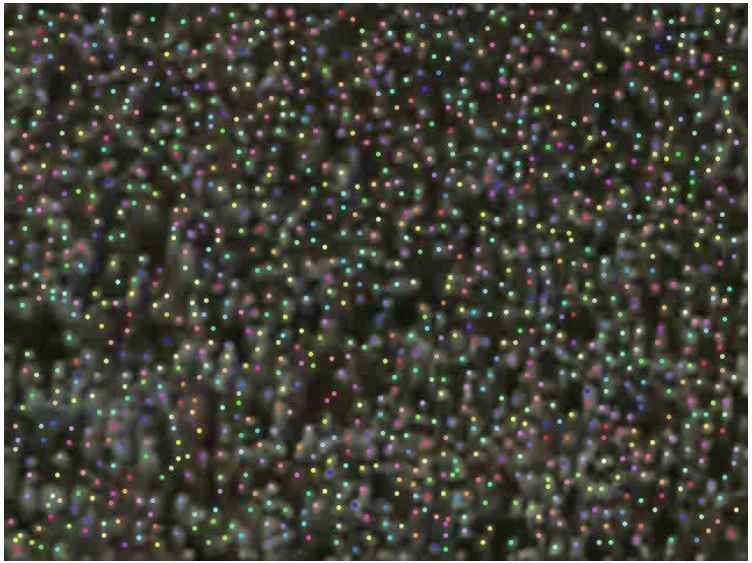


Synchronization - extent
Division rate – what
Differentiation – when and how
…. And what controls these ?

A really large-scale tracking problem?







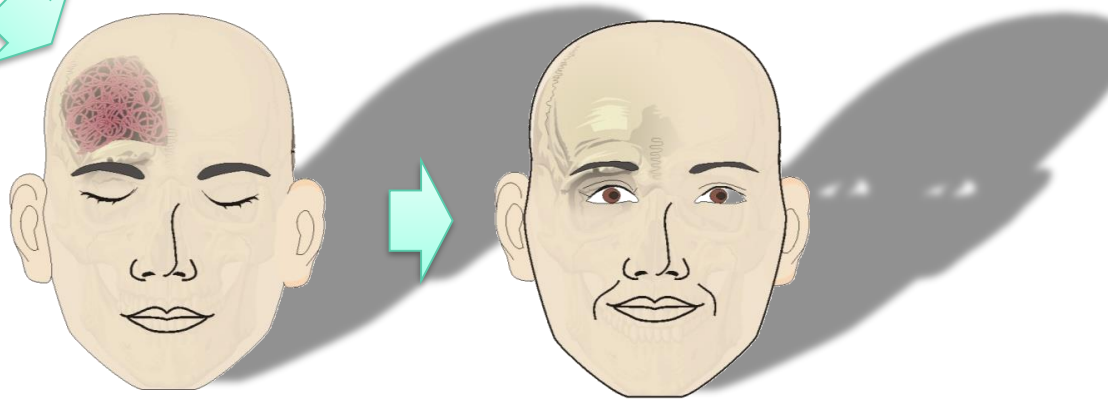
再生工学

Courtesy of Weiss and Campbell

Growth Factor

“Scaffold”
(Biomimetic Extracellular Matrix)

Critical-Size Defect



Implantation

Regeneration

Computer progress

- **First, large computer**:

  "Enshrined" in a computer room,
  Special stabilized power source,
  Used for Scientific computation

- **Then, mini-computer:**

  Plug-in to wall outlet, Toy-like magnetic tape device,
  System control – Embedded computing (in today's term)

- **Then, personal computer:**

  "Wasteful" use of computer - one for one person,
  Everyday-life information processing

- **Current data-oriented computing**

  No "theory or model"? Data contains the reality!