3D Goes Mainstream: A developer's perspective

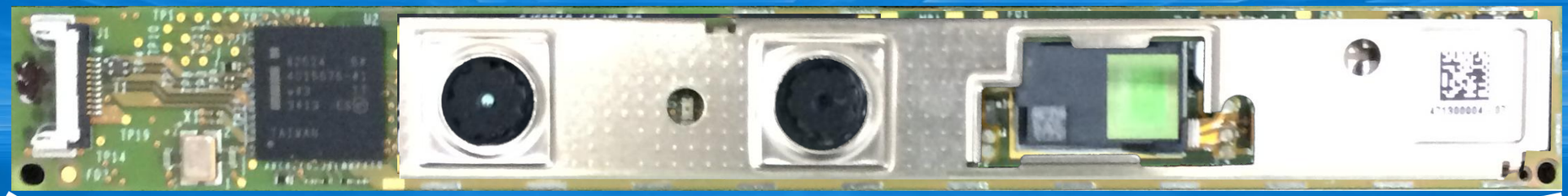
Gershom Kutliroff, Ph.D.
Perceptual Computing Software Group

Intel RealSense announcement: CES 2014

Smaller, Cheaper, Mobile …

1990 2000 2010 2013

The Embedded 3D Camera



- **Form Factor**
- **Total Cost**
- **Power (Battery Life)**

Partners

ASUS® acer NEC DELL hp FUJITSU lenovo

Experiences Will Drive Adoption

Capture and Share

Immersive Collaboration/Creation

Interact Naturally

Gaming and Play

Learning and Edutainment

RealSense is …

Applications

Perceptual Computing SDK

3D Camera

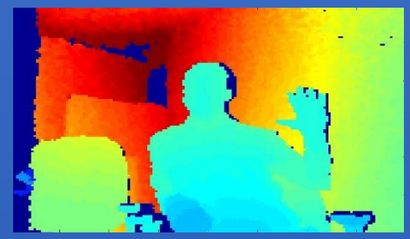
Face

Hand/Gesture

3D Scanning

Background Removal

Camera Stream (Depth + RGB)



RealSense is …

Applications

Perceptual Computing SDK

3D Camera

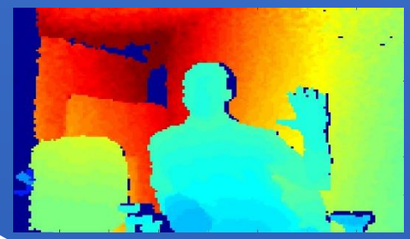
Face

Hand/Gesture

3D Scanning

Background Removal

Camera Stream (Depth + RGB)



What does it mean?

High-quality, easily accessible 3D data

New algorithms & approaches

Computing platforms

What does it mean?

Ability to solve new -- & important -- Problems

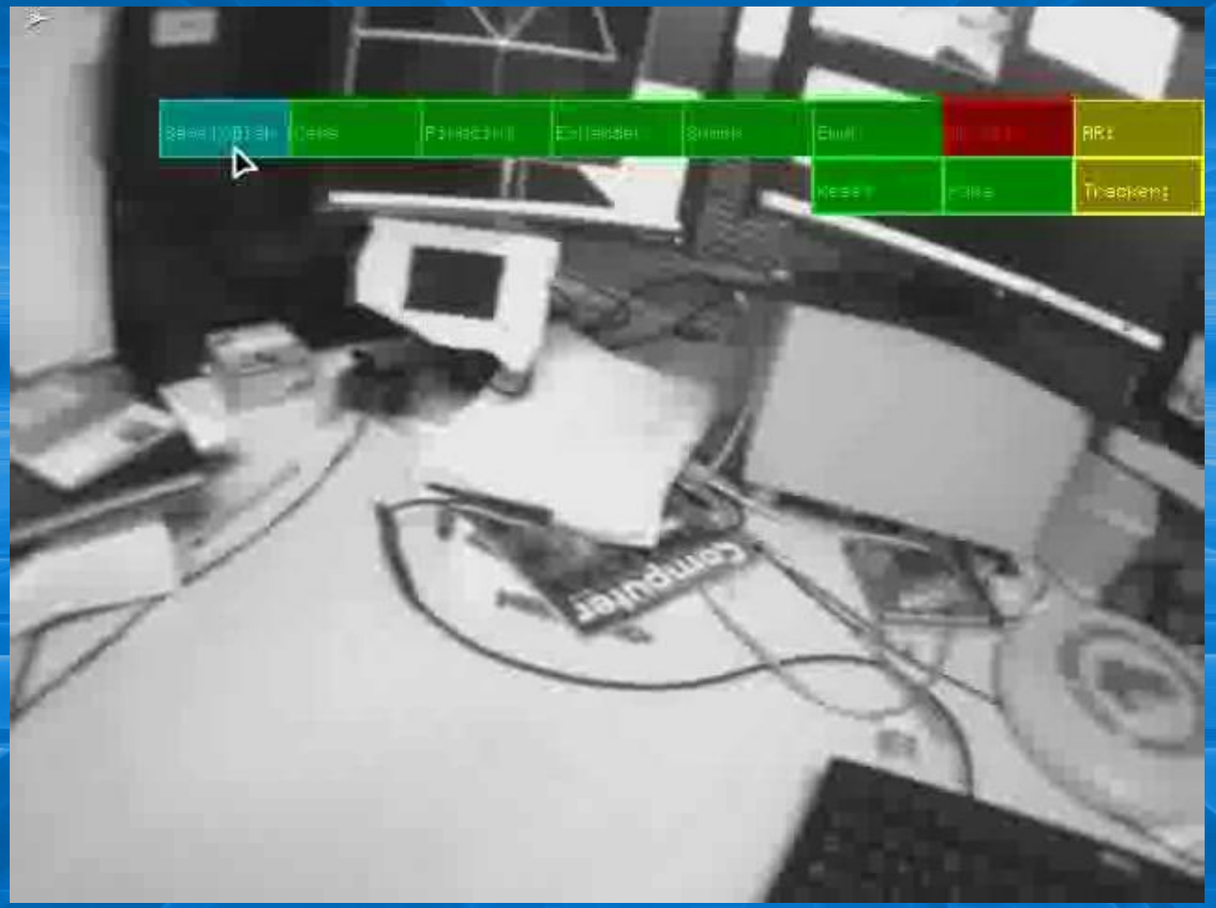
SLAM

- Simultaneous Localization and Mapping
  - Map the current environment
  - Keep track of the current location within the environment
- VSLAM (visual SLAM) relies on camera input
- Applications to
  - Robotics and autonomous vehicles
  - Augmented Reality

RGB feature-based (2D) SLAM

- Track the 6DOF camera pose at every frame
- Extract feature points from 2D images (e.g., SIFT/SURF/FAST)
- Match feature points across multiple frames to compute 3d structure
- AKA “structure from motion”

“PTAM”



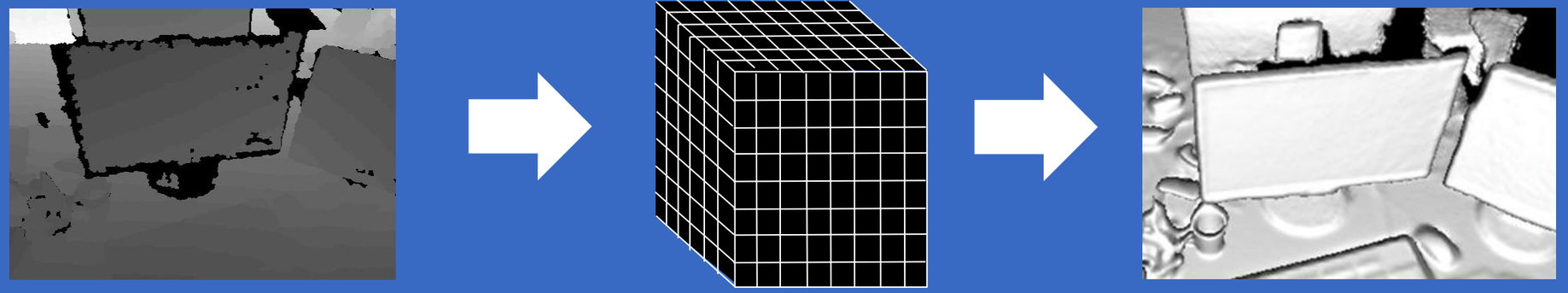
“Parallel Tracking and Mapping for Small AR Workspaces”, ISMAR 2007, Georg Klein and David Murray

Depth-based (3D) SLAM

- Relies on depth data only
- In addition to 6DOF camera tracking, also construct dense 3D surface reconstruction
- Reconstruction itself is used to improve accuracy of tracking

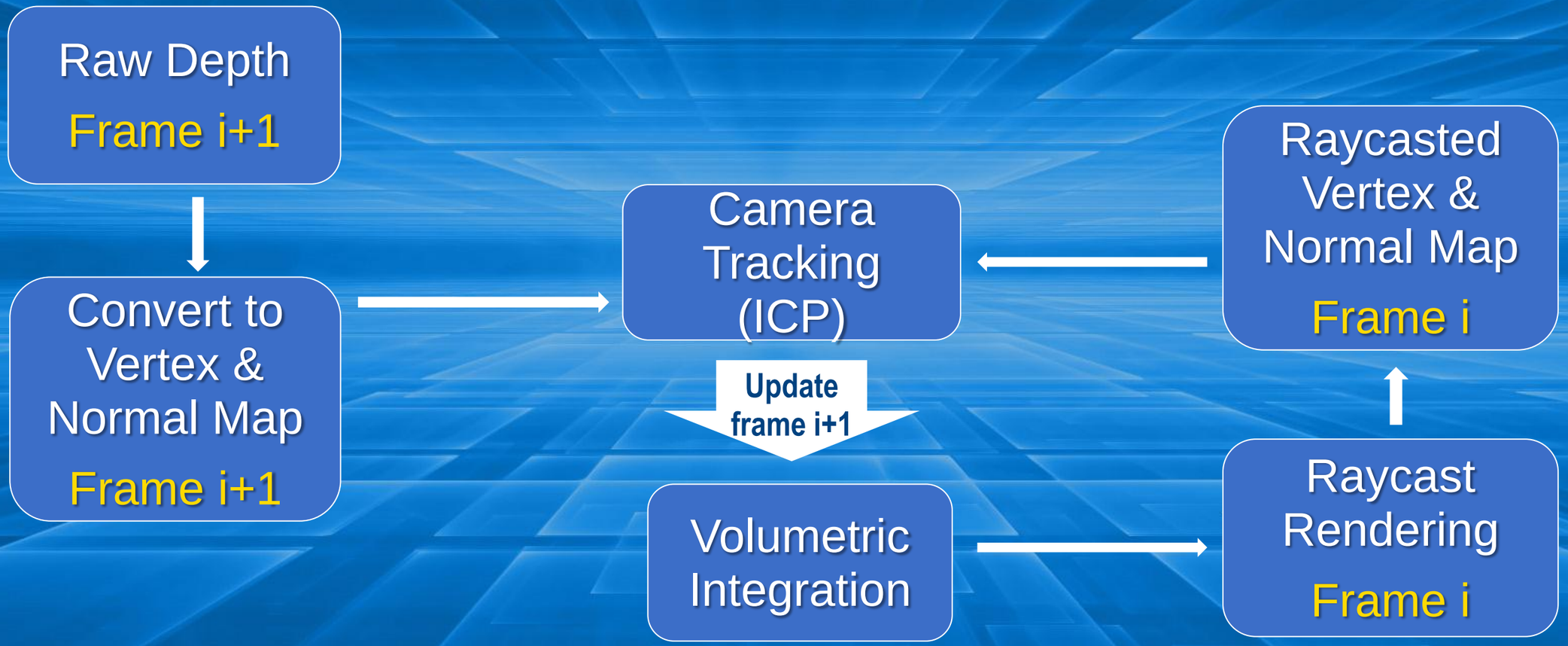
Volumetric Integration

- Map 3D data points to a 3-dimensional volume (voxel grid)



Curless and Levoy, “A Volumetric method for building complex models from range images”, 1996

Kinect Fusion

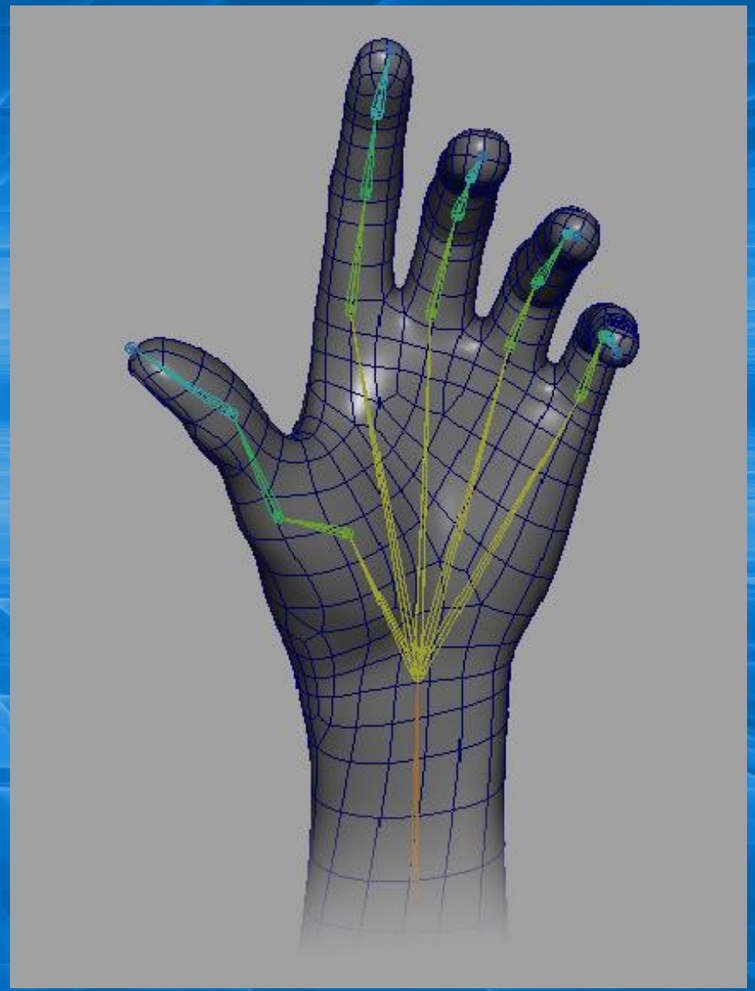


Newcombe, et al., “KinectFusion: Real-time dense surface mapping and tracking”, 2011

Depth-based fusion

The 3D model

- Control vertices $\{v_i\}_{i=1,\ldots,n}$ determine the 3D positions of all points on the model.
- For any hand pose p, a transformation $M_{i,p}$ is associated with each control vertex v_i.



Synthesis-Analysis

I

Analysis

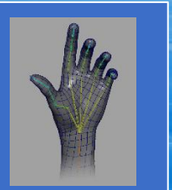
Good approximation?

Final model parameters

Yes

No

Adjust the 3D model (intelligently)



$\hat{I}$

Synthesis

Estimated parameters

Non-linear search optimization

- The synthetic image depends on the articulation:

$$\hat{I} = f\left(M_{1,p}, M_{2,p}, \ldots, M_{n,p}\right).$$

- So, the objective is to find p, such that the energy function is minimized:

$$E = \sum_{x,y} \left\| I(x,y) - \hat{I}(x,y) \right\|^2.$$

- In general, this space is nonlinear, and high-dimensional (i.e., "nasty").
  - If E is expressed as a differentiable function of $M_{i,p}$, can search

Synthesize an RGB image

Render the 3D Model

Add Texture

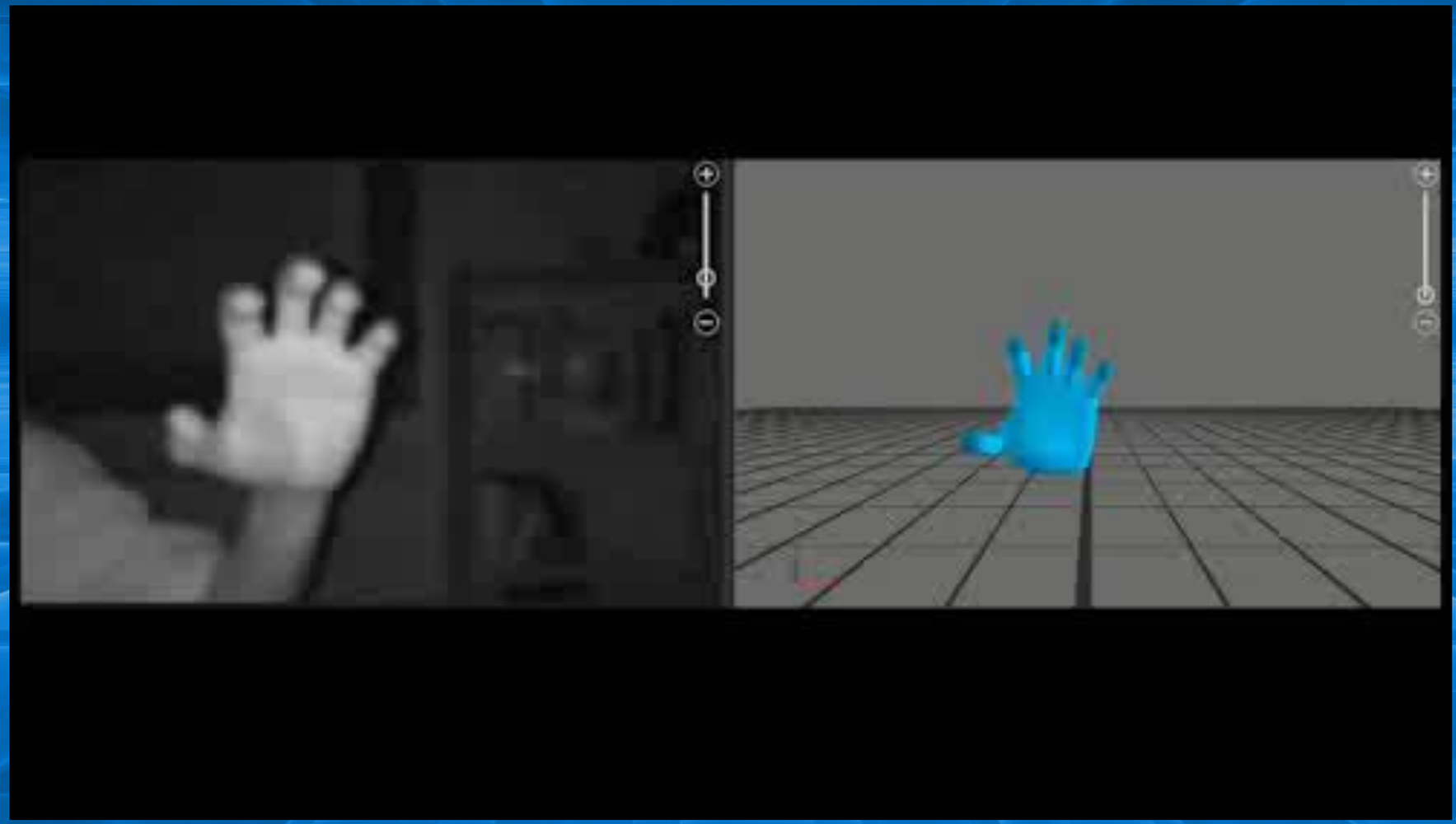
Add Shading

Add Special Treatment for Nails

Add Lighting

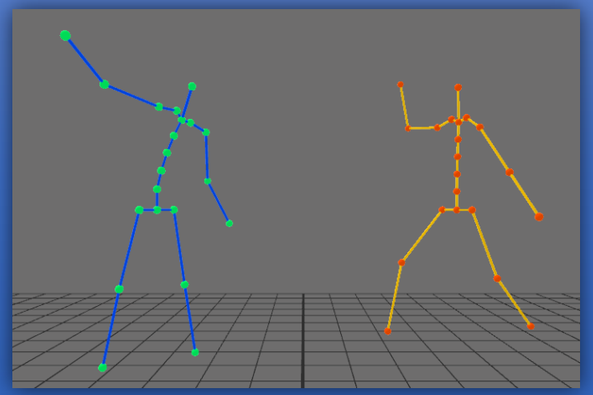
Synthesize a Depth Image

Real-time Hand Skeleton Tracking



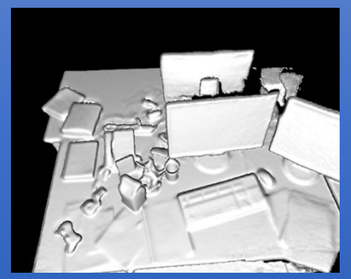
What Next?

Markerless Motion Capture



Hand Skeleton Tracking

3D Reconstruction



Object Recognition?

Action Recognition?

3D Video Compression?

3D Video Search?

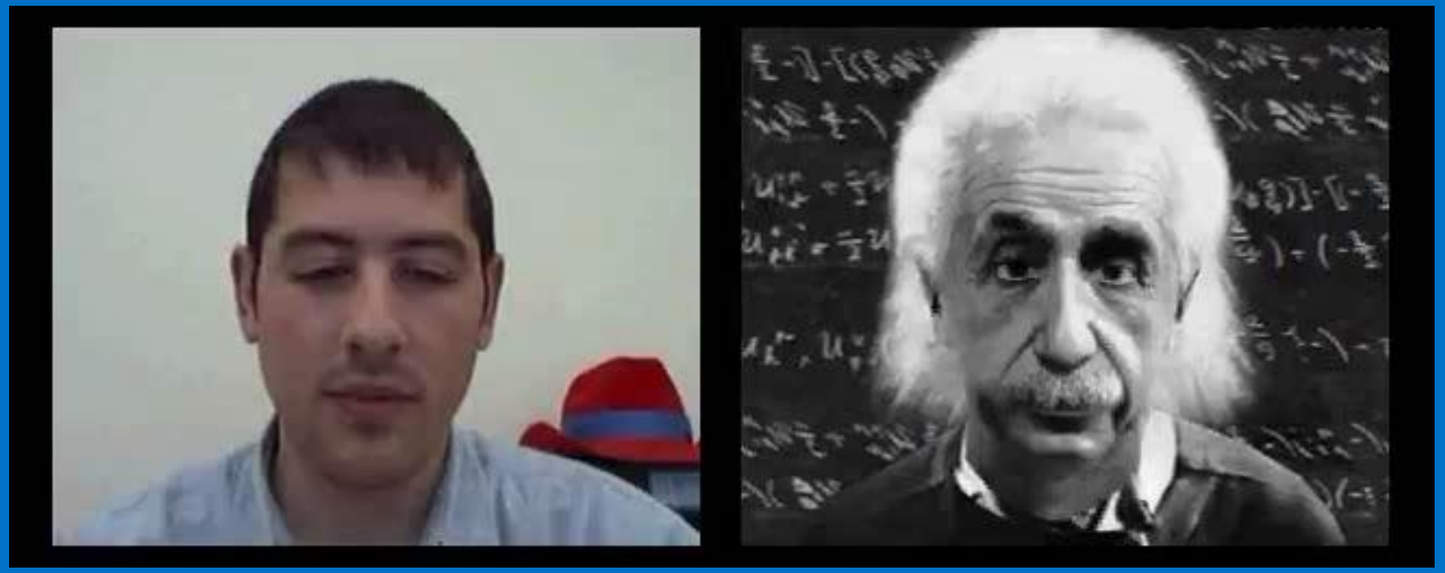
2014


intel®
Look Inside.™

Changing Computing Paradigms

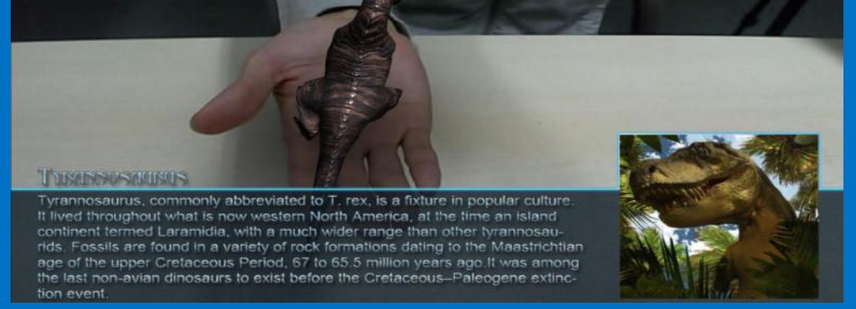
- GPU
  - Rendering: Algorithm components that replicate real-world structure
  - GPGPU: Large amounts of data, highly parallelized
- API functionality (e.g., OpenCL)
- Acceleration libraries (OpenCV, PCL)

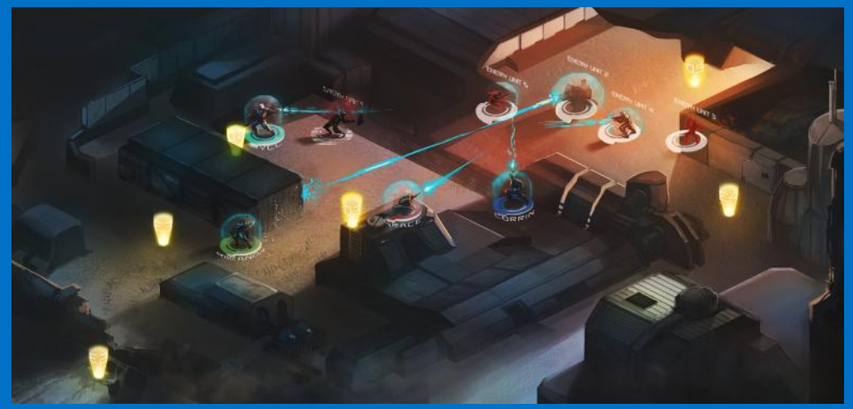
A 2D instantiation

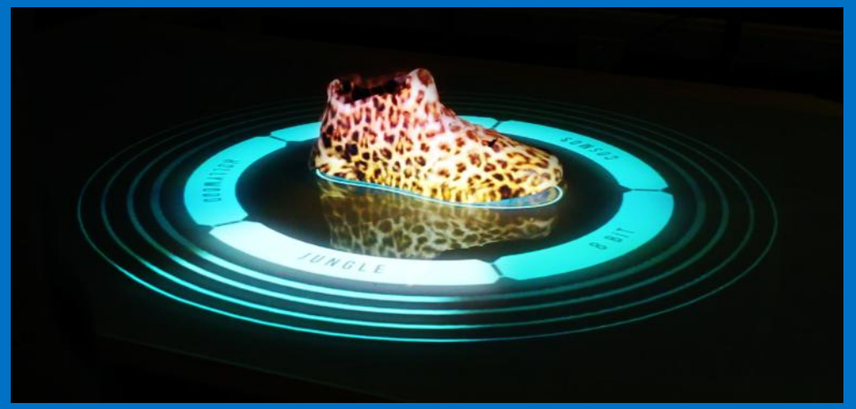


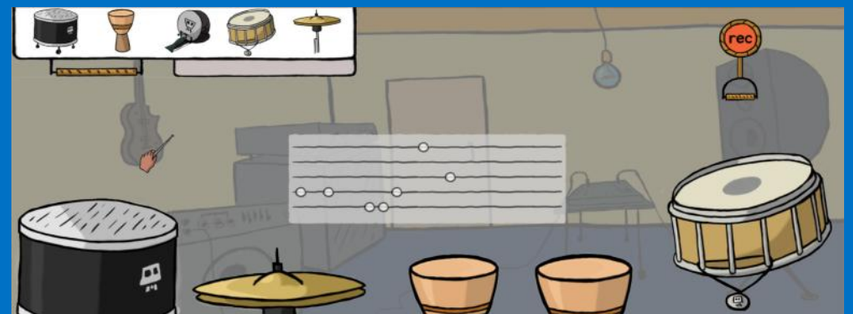
- Add an illumination model
- And shaders
- Account for the eyes
- And the hair
- And then ….

Enabling a wide developer ecosystem…









Synthesis-Analysis

1. Create a 3D parameterized model of an object.
2. Render (“synthesize”) the 3D model to a 2D image
3. Compare (“analyze”) the rendered image and the data from the camera.
4. Adjust the parameters for the 3D model (intelligently).
5. Repeat.

RealSense is …

Applications

Perceptual Computing SDK

3D Camera

Interact Naturally

Capture and Share

Gaming and Play

Immersive Collaboration/Creation

Learning and Edutainment