Semi-Autonomous Surgery of the Lateral Skull Base: Computer Vision, Robotics and Deep Learning

Francis "Pete" Creighton, MD

Assistant Professor, Department of Otolaryngology-Head and Neck Surgery, Division of Otology, Neurotology and Lateral Skull Base Surgery

Feb 18th 2021

Disclosures

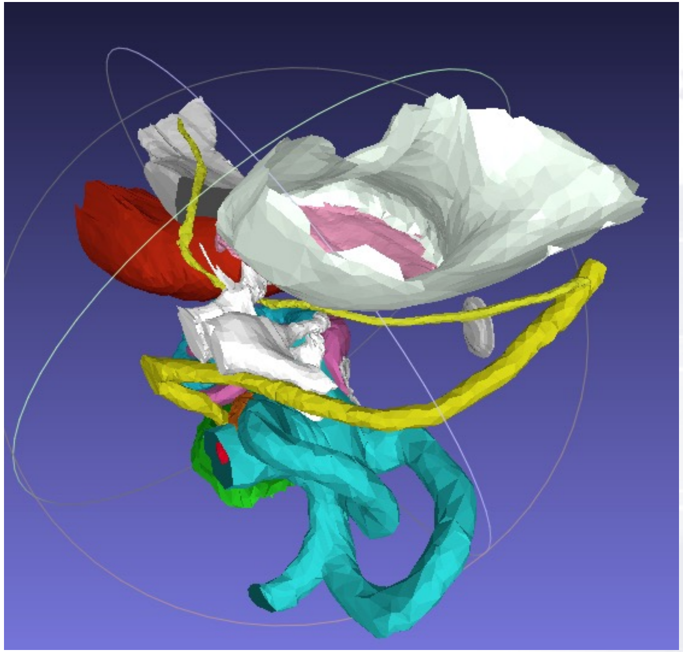
- I have no relevant financial conflicts of interest
- I will not be discussing any off-label medications

Outline

- Motivation/Background
  – Why do we need new ways of doing lateral skull base surgery
- Current Limitations
- Overview of our work in addressing these limitations
  – Robotics
  – Automated Image Segmentation
  – Stereovision, Microscope based image navigation

Temporal Bone and Skull Base Surgery

Temporal Bone and Skull Base Surgery



Temporal Bone and Skull Base Surgery

Why Do Need New Technologies in Ear and Skull Base?

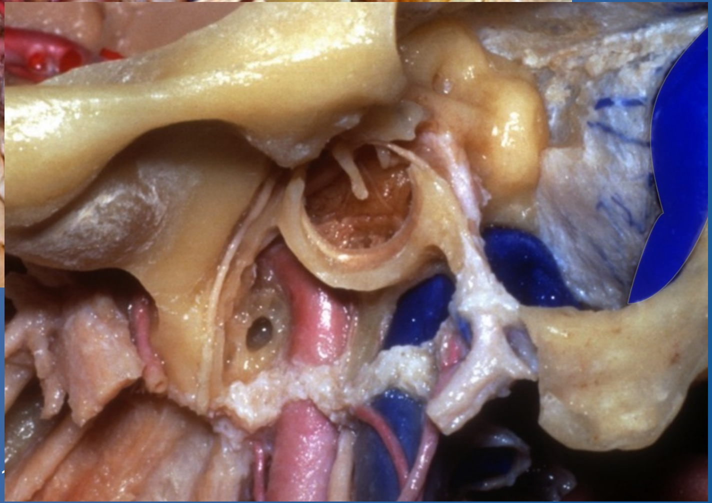
JOHNS HOPKINS MEDICINE

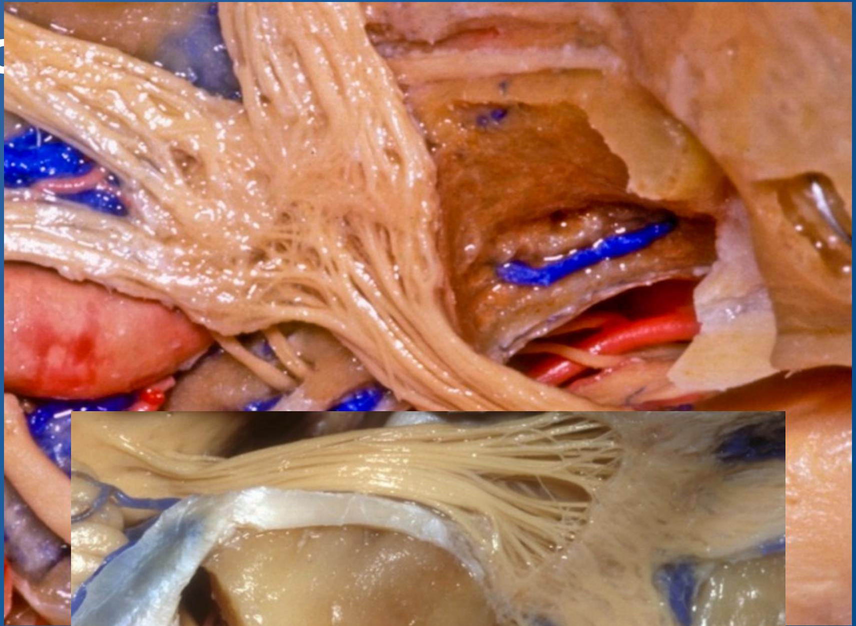
- High Degree of Technical Difficulty to Access
  - Millimeter differences between success and failure
  - Limits surgical options for skull base tumors

- Extensive training required
- Long Operative Times

- Injuries in the skull base result in significant impact to patient's quality of life

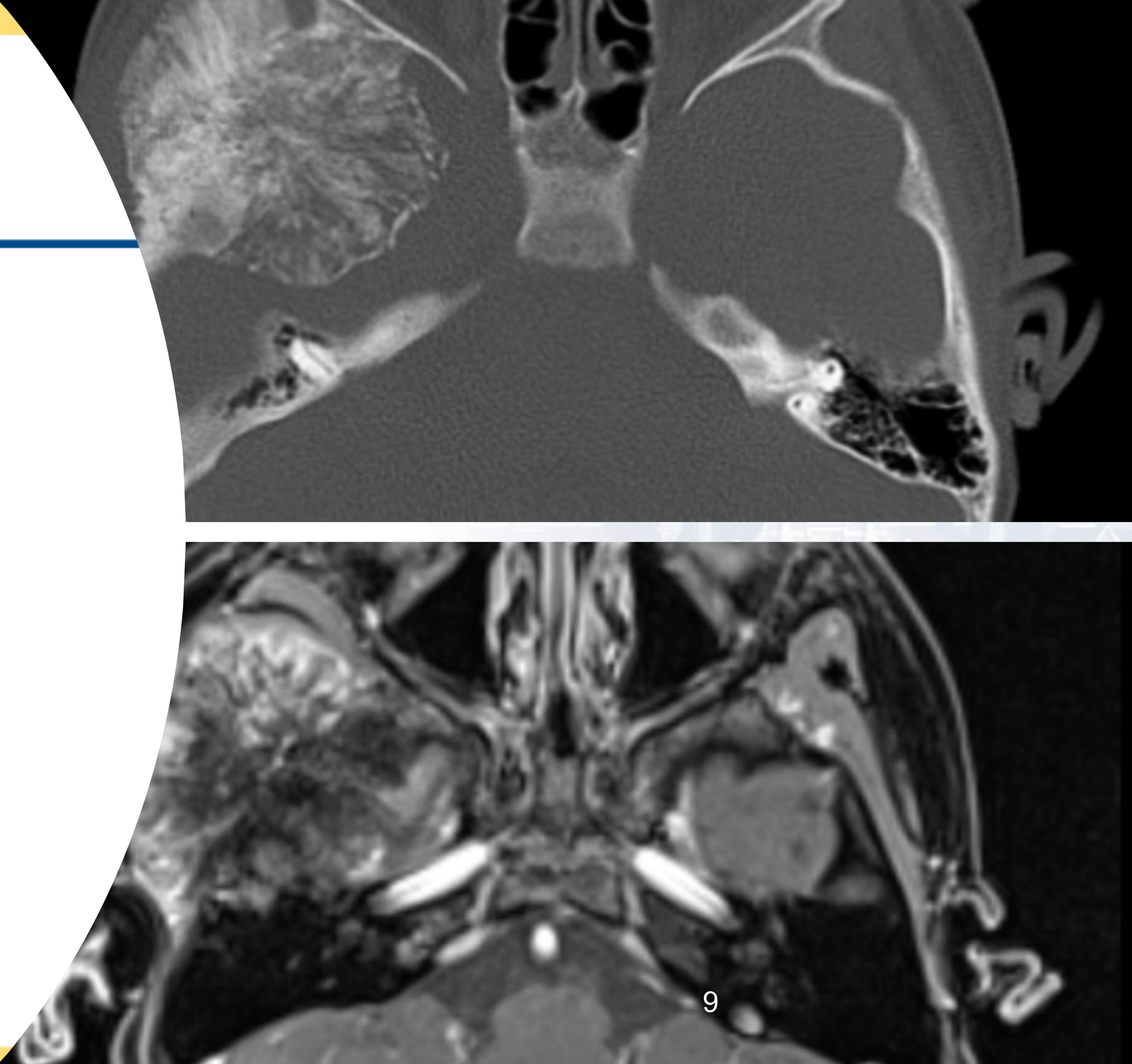
Technolo

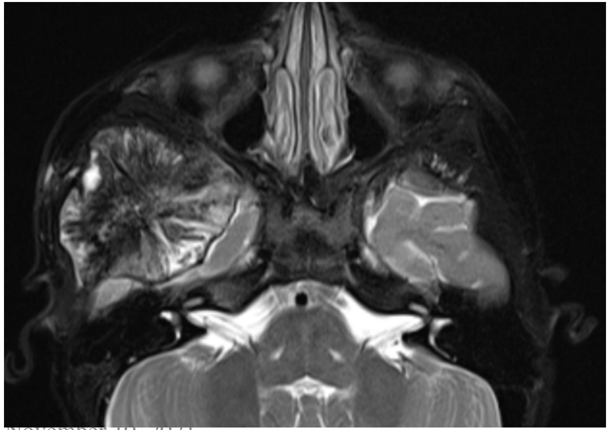




Why Do Need New Technologies in Ear and Skull Base?

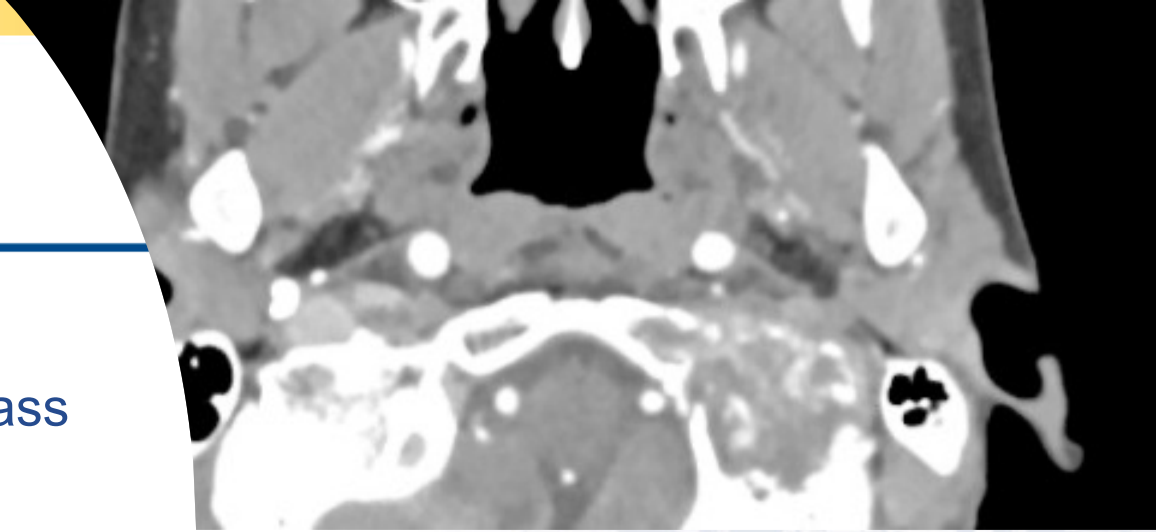
- 3.5 yr. F presents with osteosarcoma of temporal bone
- Failed chemotherapy
- Inferior edge of tumor abuts petrous carotid

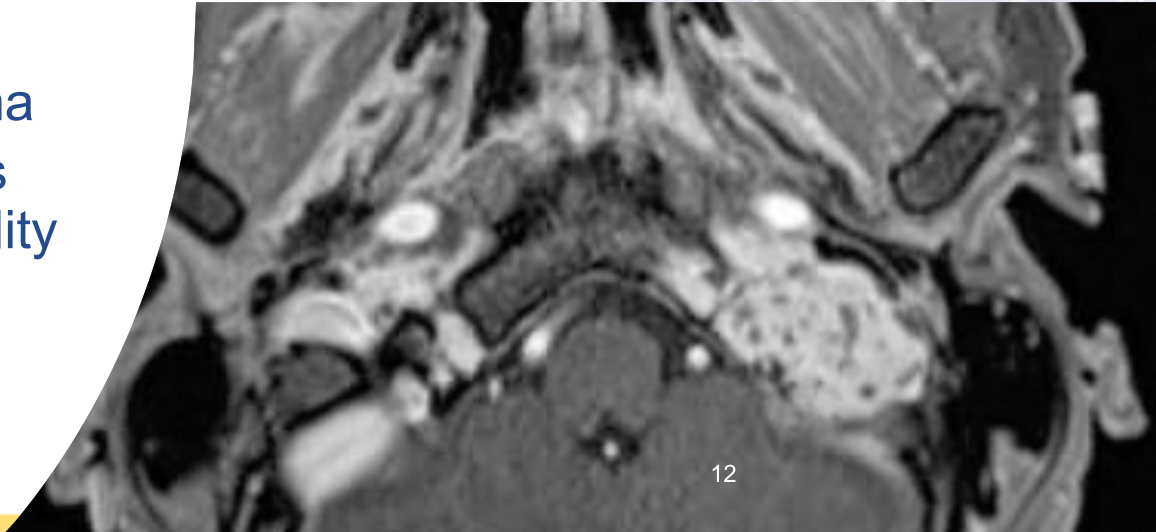




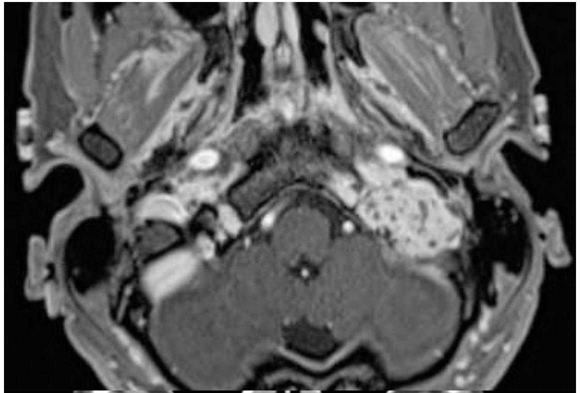
Why Do Need New Technologies in Ear and Skull Base?

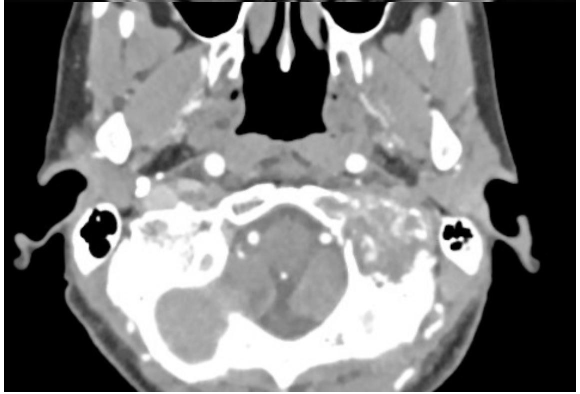
- 42y F presents with rapidly progressive mass in temporal bone, condyle and clivus
- H/O epithelioid hemangioendothelioma
- Developing numbness due to cervical instability
- Vocal weakness and tongue weakness

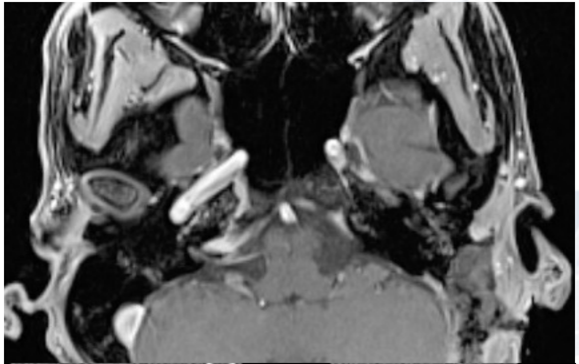




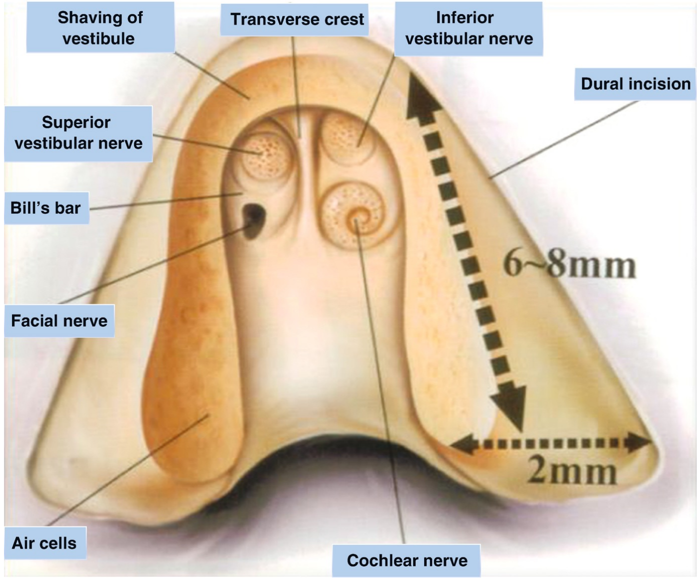
Why Do Need New Technologies in Ear and Skull Base?

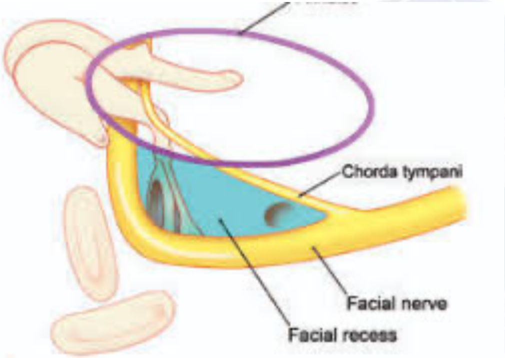


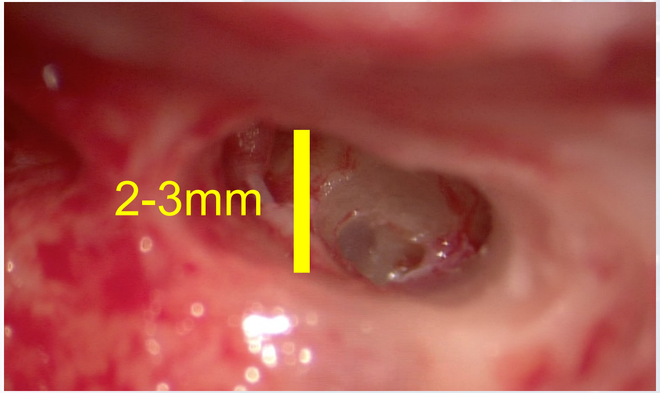




Why Do Need New Technologies in Ear and Skull Base?







Research Goals

Computer Vision & Machine Learning

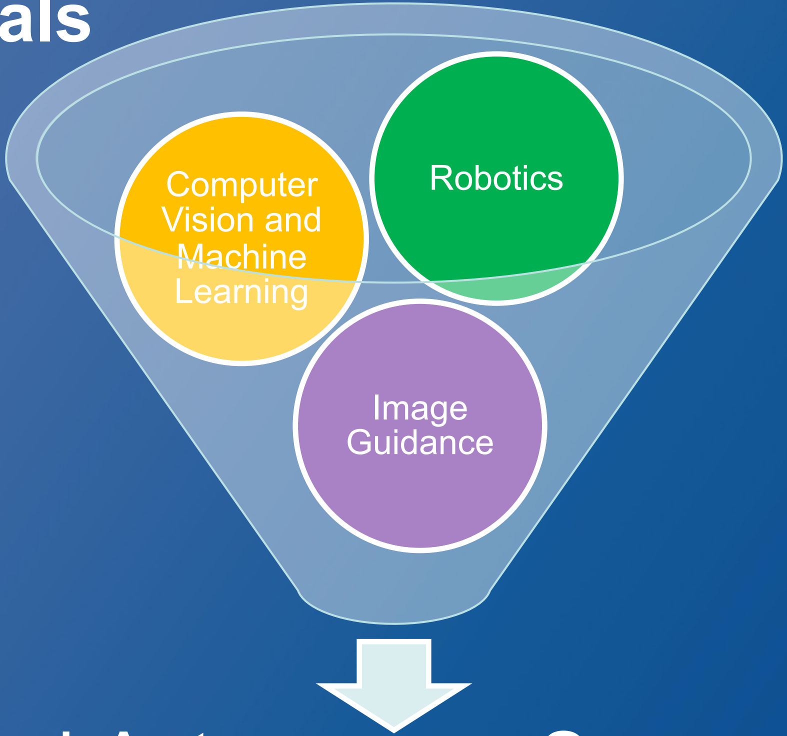
Robotics

?

Image Guidance

Surgery

Research Goals

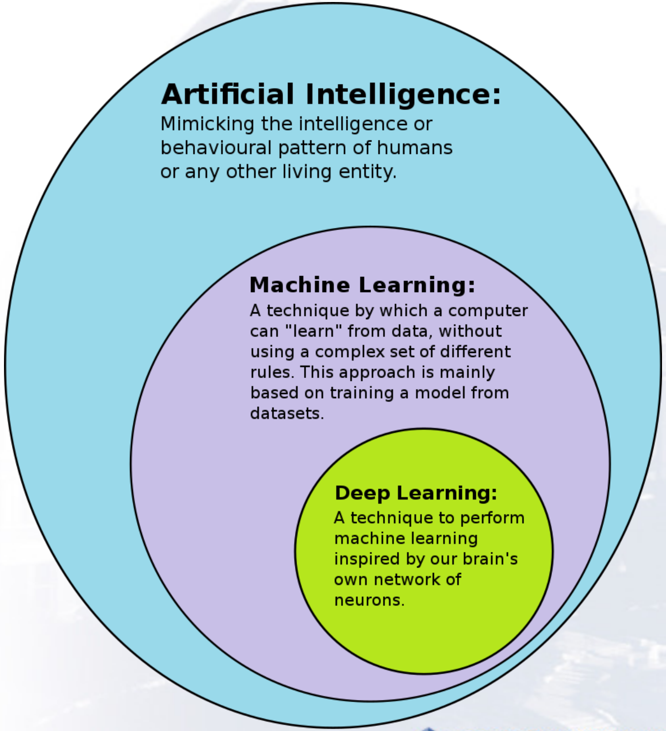


Semi-Autonomous Surgery of the Skull Base

What is Machine Learning?

- The study of computer algorithms that improve automatically through experience
  - Uses "training data” to make predictions or decisions without being explicitly programmed to do so

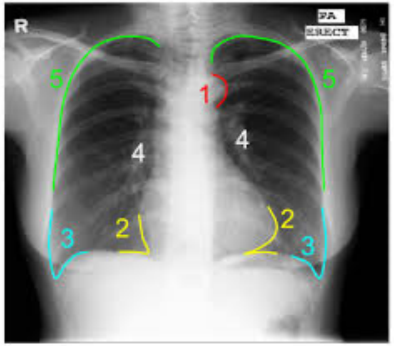
- Deep Learning
  - A subset of machine learning that uses Neural Networks



What is Computer Vision

- Computer vision is an interdisciplinary scientific field that deals with how computers can gain high-level understanding from digital images or videos

- Attempts to understand and automate tasks that the human visual system can do



What is Computer Vision

- Computer vision is an interdisciplinary scientific field that deals with how computers can gain high-level understanding from digital images or videos

- Attempts to understand and automate tasks that the human visual system can do

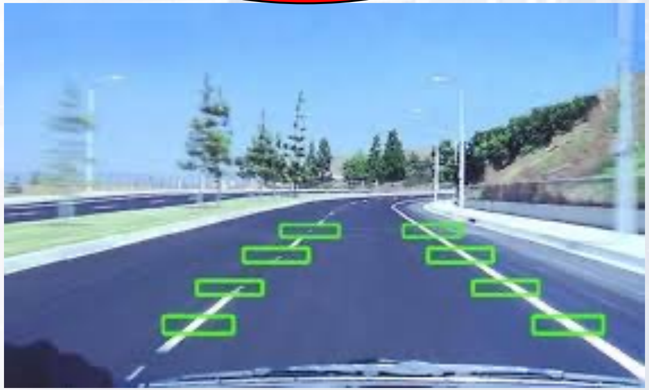
Semi-Autonomous vs. Autonomous Surgery

- Autonomous ~ “TESLA”
  – System needs no input from operator to perform task

- Semi-Autonomous ~ “LEXUS”
  – System relies on operator to perform task, but augments and improves operator performance

Why Semi-Autonomous?

- MUCH MUCH MUCH EASIER!!!!!!
  - Takes advantage of a surgeon's inherent knowledge and skill
  - Don’t have to program and design for every eventuality



Why Semi-Autonomous?

- Public Perception
  - – Patients are not comfortable with fully autonomous systems

Tesla on autopilot had steered driver towards same barrier before fatal crash, NTSB says

The man told his family about the problem prior to the fatal crash.

By **Catherine Thorbecke**
February 12, 2020, 1:16 PM • 5 min read

Another Tesla Autopilot Crash Has Wrecked A Model 3 In Greece

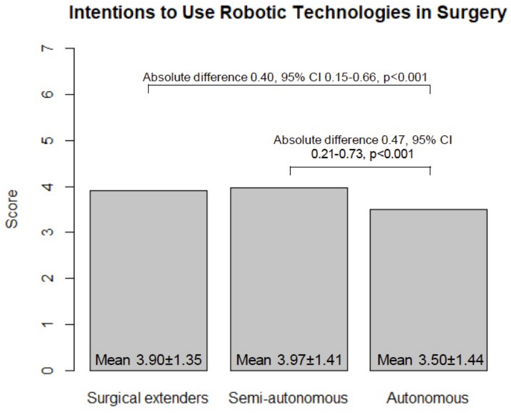
In what's believed to be the first case of an active Autopilot crash on European soil, a Tesla Model 3 has been wrecked after suddenly swerving off a motorway and into a crash barrier

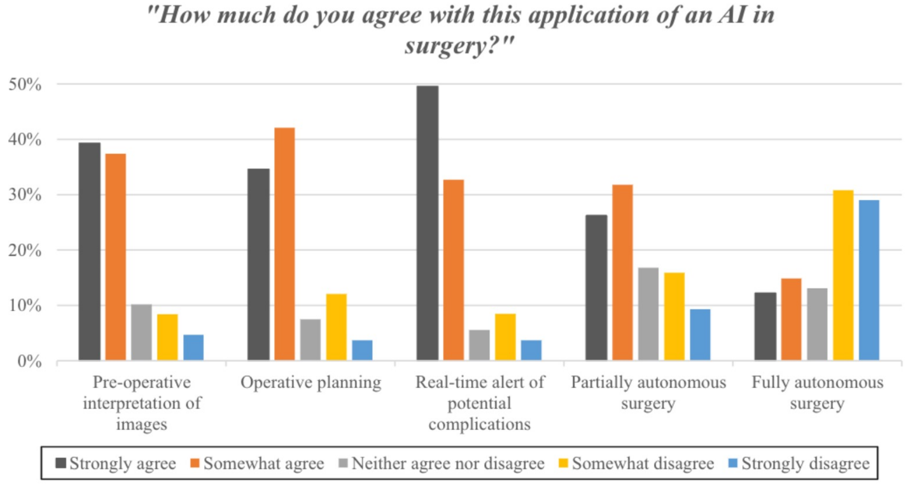
NTSB Releases Report On 2018 Fatal Silicon Valley Tesla Autopilot Crash

Brad Templeton Senior Contributor
Transportation
I cover robocar technology & previously worked on Google's car team.

Why Semi-Autonomous?

- Public Perception
  - Patients are not comfortable with fully autonomous systems





Palmisciano P, Jamjoom AAB, Taylor D, Stoyanov D, Marcus HJ. Attitudes of Patients and Their Relatives Toward Artificial Intelligence in Neurosurgery. World Neurosurg. 2020 Jun;138:e627-e633. doi: 10.1016/j.wneu.2020.03.029. Epub 2020 Mar 14. PMID: 32179185.

Barriers to a Semi-Autonomous Surgical Platform

JOHNS HOPKINS MEDICINE

- Better Robot:
  - Needs to work in Parallel with the Surgeon, allowing us to take advantage of surgeon's skill
- Better Image Segmentation:
  - Need to automatically identify critical structures so the robot knows what to avoid
- Better Image Navigation
  - Even best available systems are between 1-2mm accuracy in clinical setting
  - Skull Base Surgery requires submillimeter accuracy

Cooperative Control vs Master/Slave Robots

- Master/Slave manipulators
  - Robot holds and physically moves the instruments; surgeon commands the robot remotely – “DA VINCI”

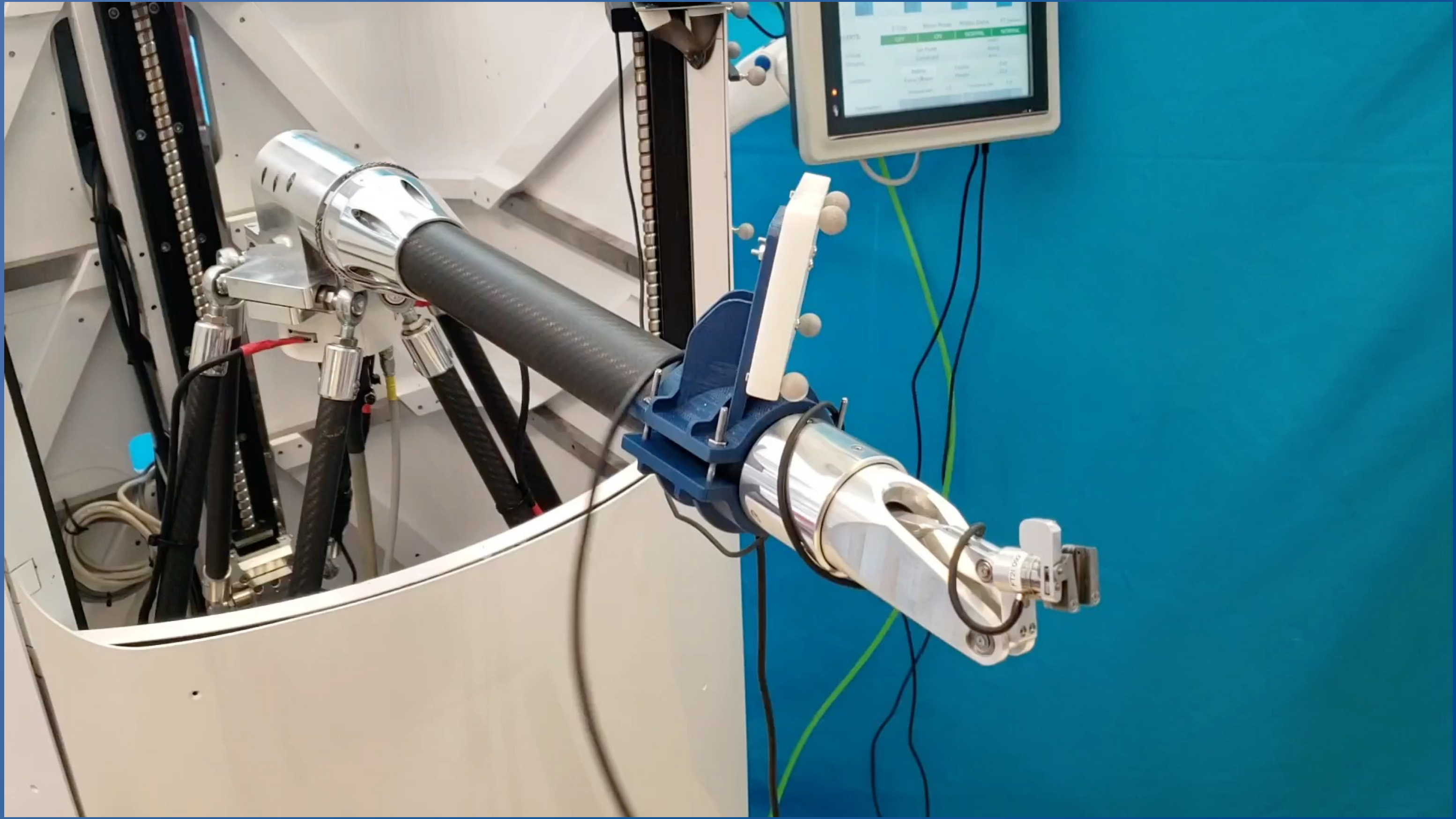
- Cooperative Control manipulators
  - Robot and surgeon hold instrument in parallel, surgeon moves the instrument while robot monitors and augments that motion

Background- *REMS*

JOHNS HOPKINS MEDICINE

- The *Robotic ENT Microsurgery System (REMS)*
- Cooperative control
- Conventional instruments
  - Custom adaptors
- 6 DOF
  - Delta (X,Y,Z)
  - Roll/Tilt stages
  - Unactuated tool rotation

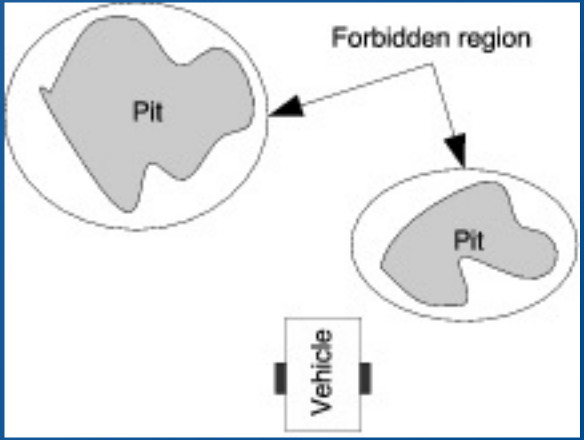
REMS



Virtual Fixtures

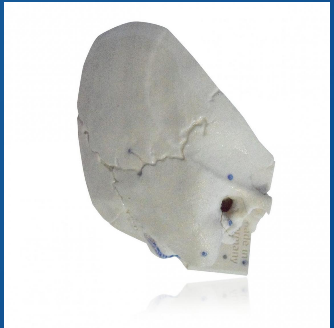
JOHNS HOPKINS MEDICINE

- Augmented sensory overlay onto a user's real perception of the environments
- Can great virtual, robotically enforced “No go zones”



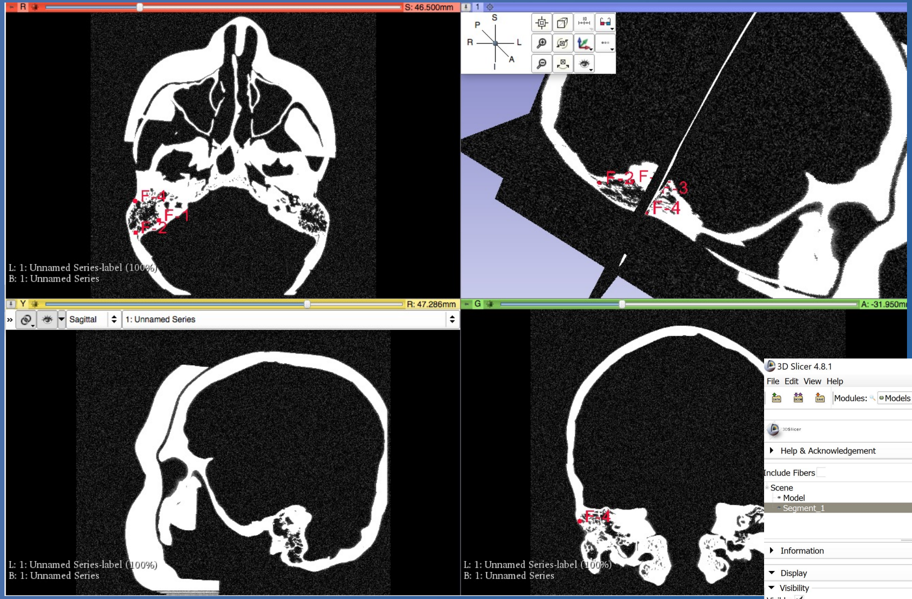
Methods

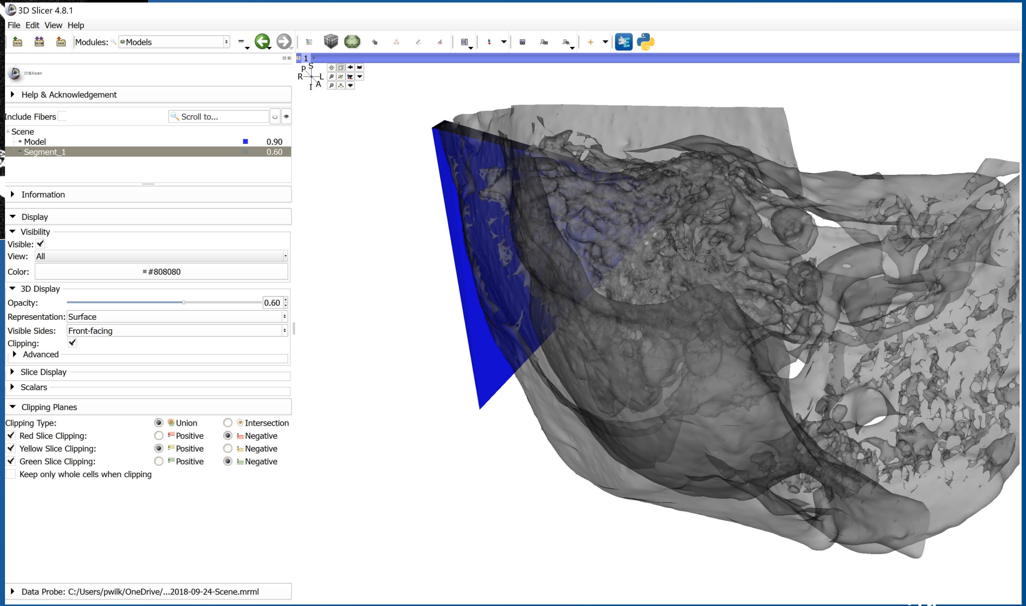
JOHNS HOPKINS MEDICINE

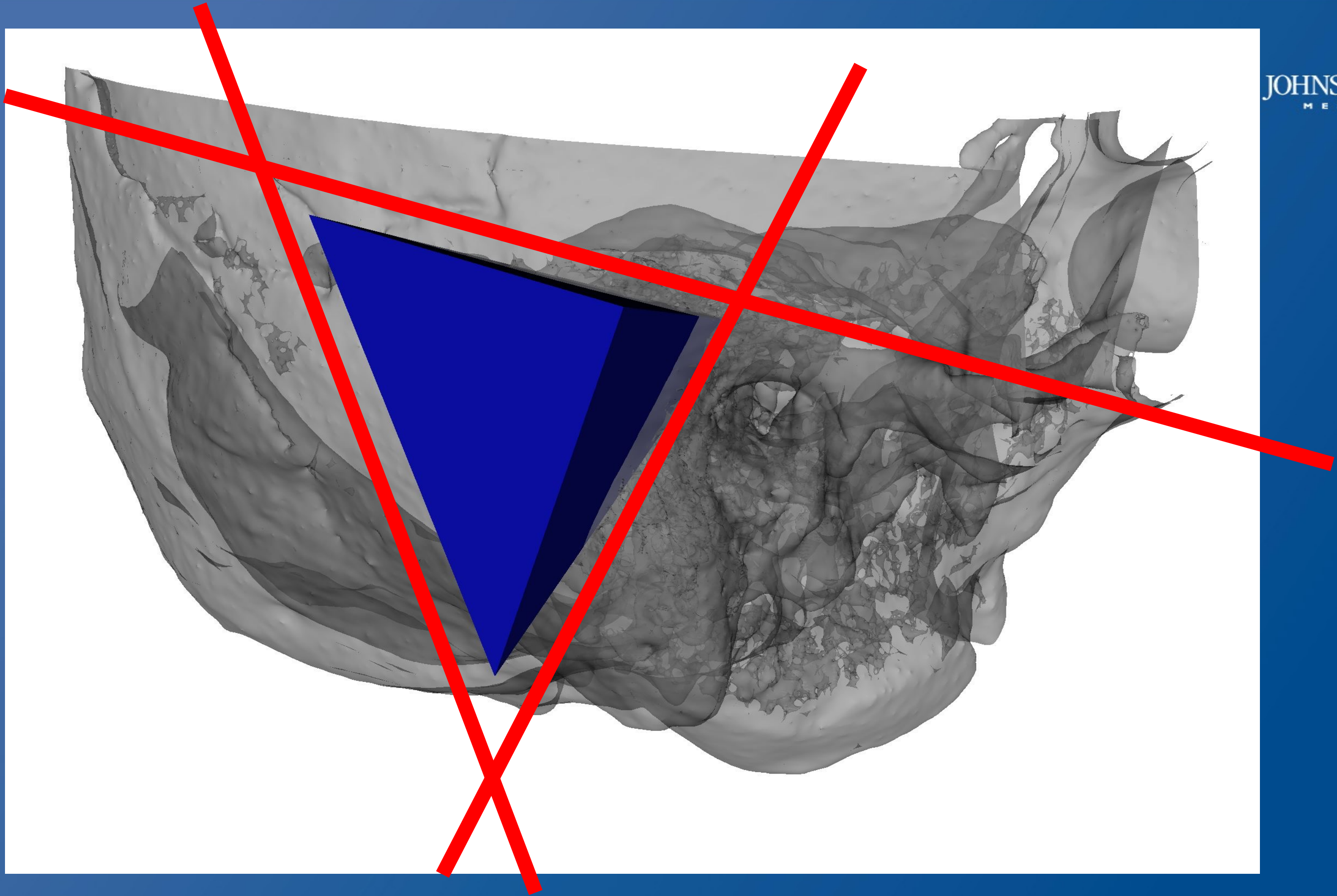


- Commercially available temporal bone models
  - Phacon (Leipzig, Germany)
  - R temporal bone with accompanying CT
- 3 Planar virtual fixtures defined on CT
  - Resulting volume between planes approximates cortical mastoidectomy
  - Phacon/CT registered to *REMS*
- Computer Engineer with no prior knowledge of mastoidectomy instructed to drill away all material within allowable working space
  - Performed on 5 identical models
- 3rd generation research version of the technology

JOHNS HOPKINS
MEDICINE

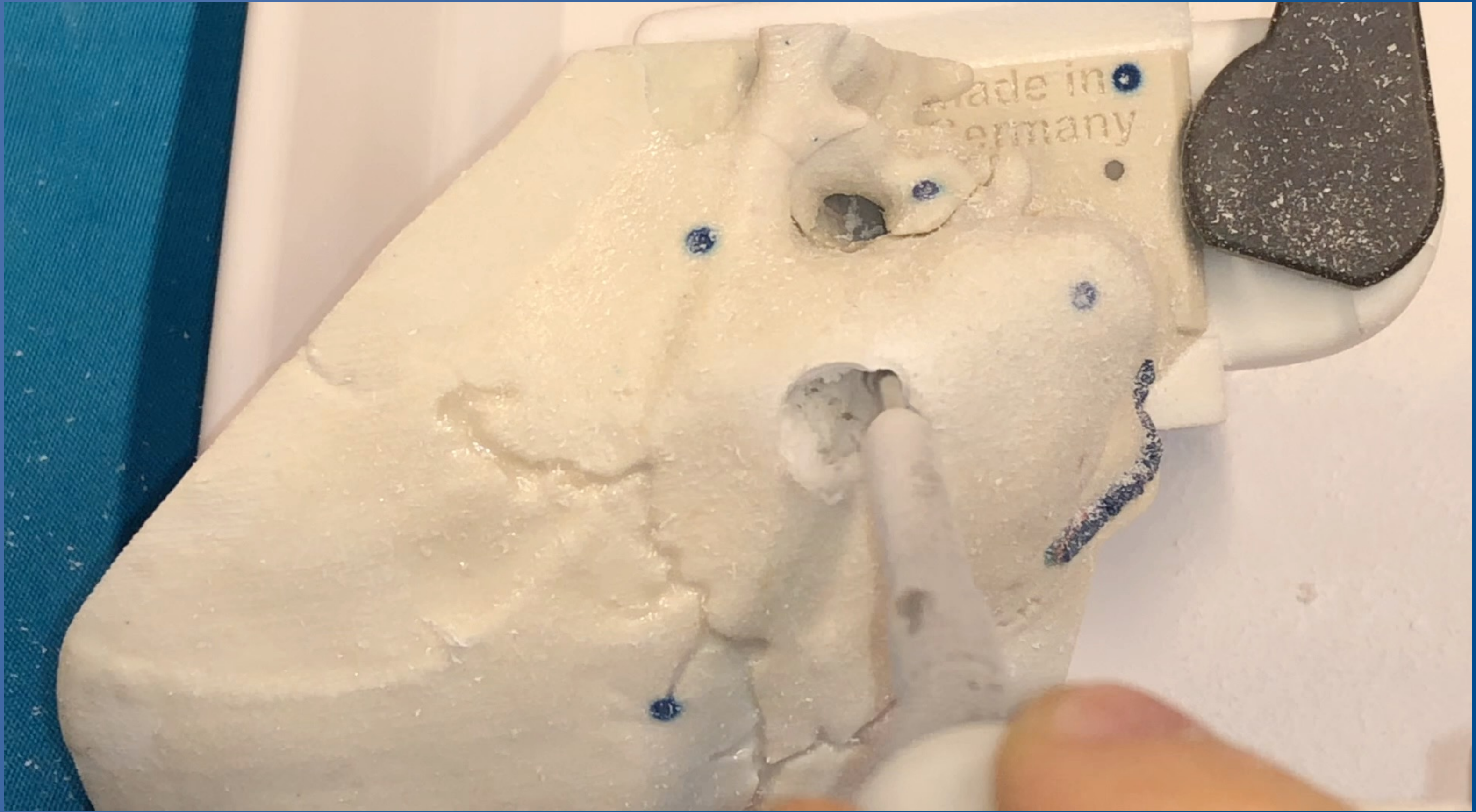






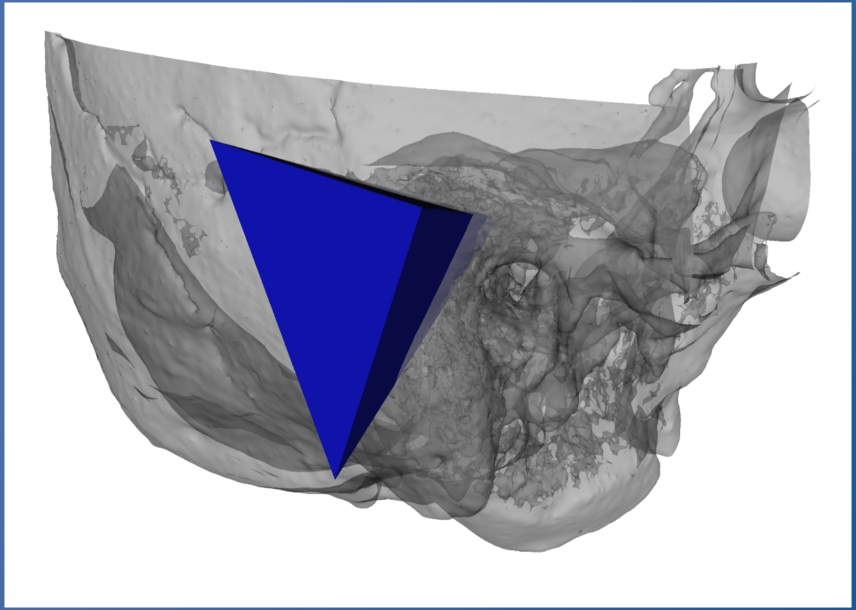
JOHNS HOPKINS
MEDICINE

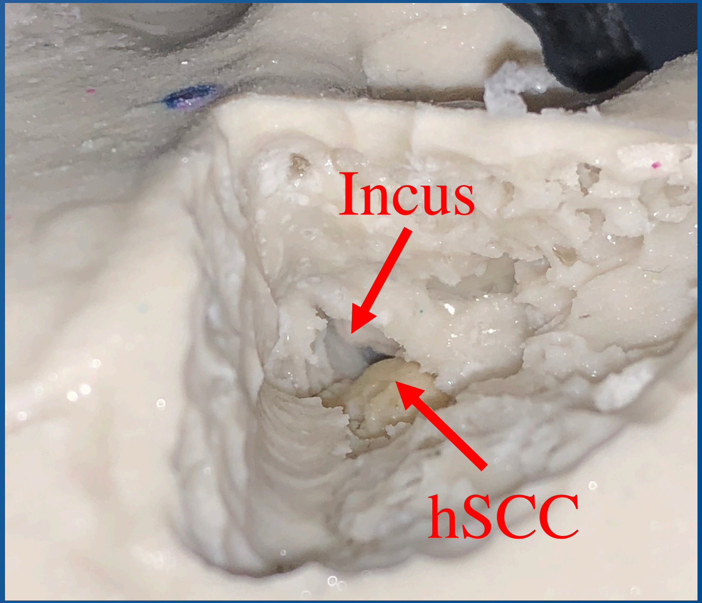
REMS-Assisted Mastoidectomy



Results

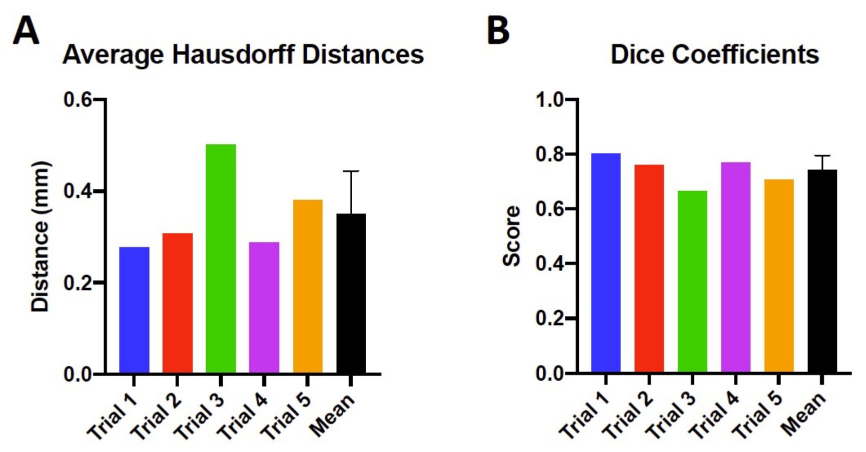
- Virtual Fixtures 3D Reconstruction:

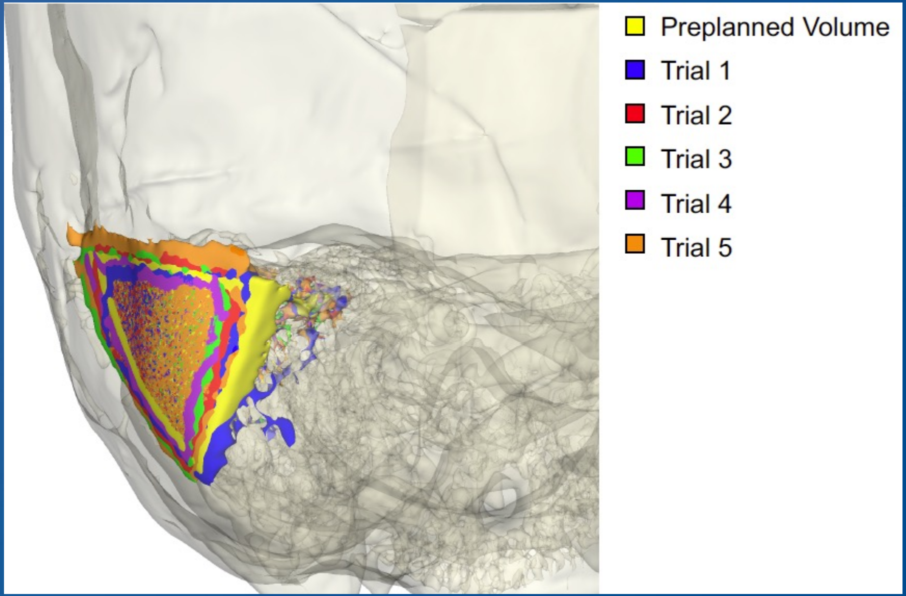




Results

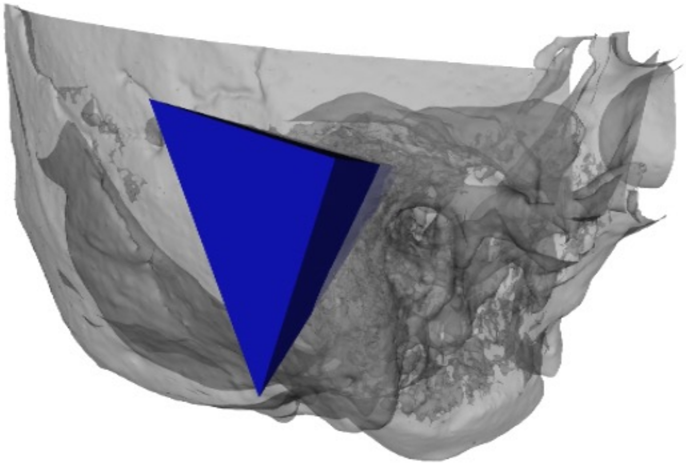
- Mean time to completion: 221 +/- 35 seconds (3.6 min)
- Average Hausdorff Distance ~0.3mm



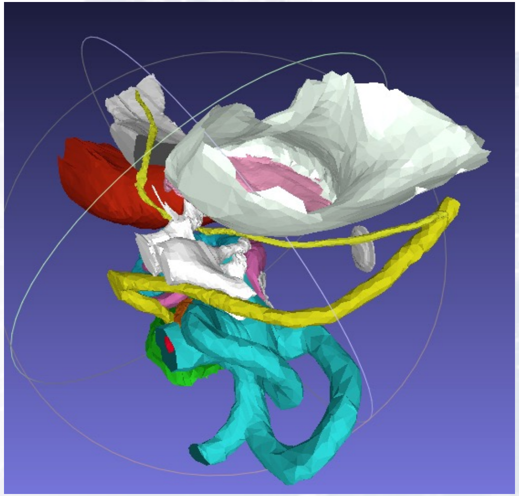


Future Directions

Define What to Drill

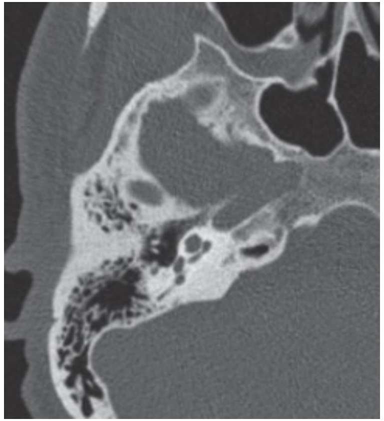


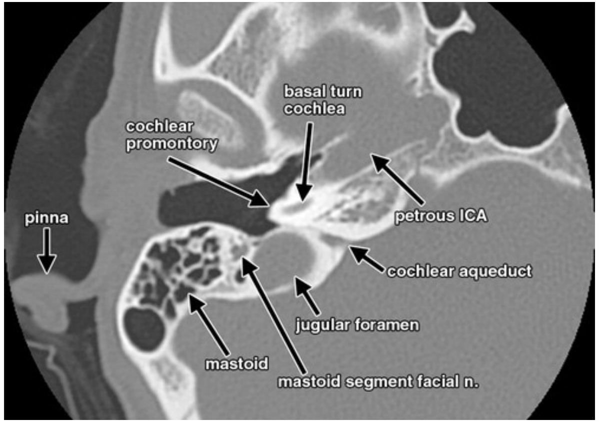
Define What not to Drill

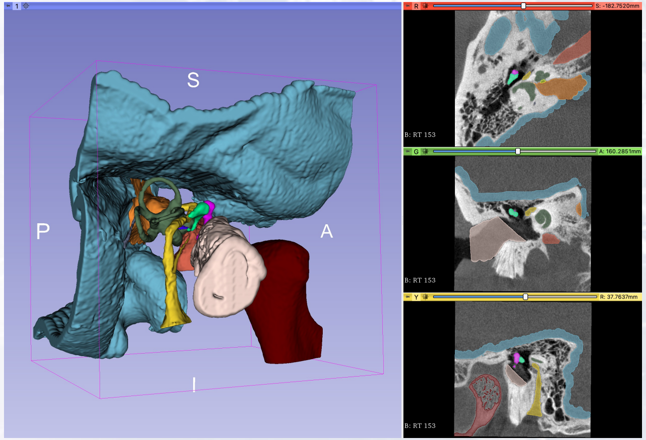


Need for Automated Segmentation

- To allow a robot to use image navigational information we need to be able to tell the robot anatomical information about that image

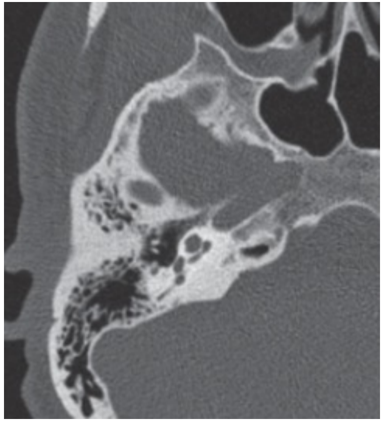


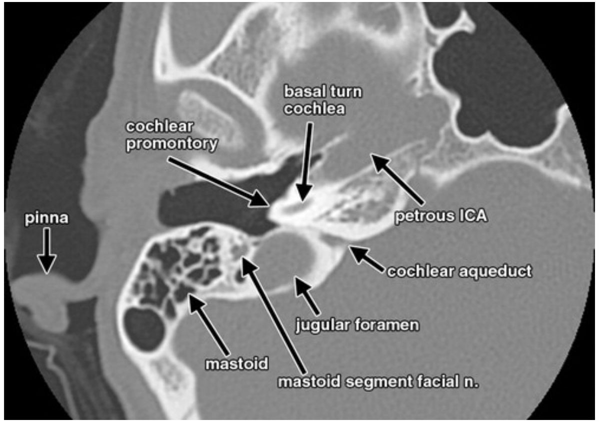


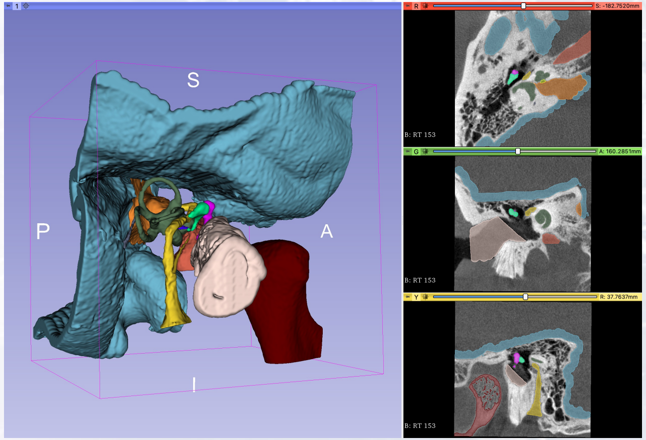


Need for Automated Segmentation

- To make this clinically feasible, this process needs to be automated





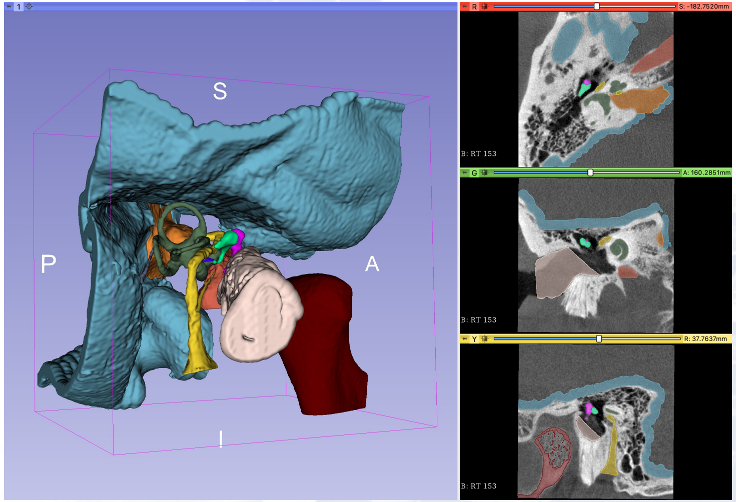


Development of an Automated Segmentation Algorithm

1. Manual Segment 42 Temporal Bone CT scans
2. Create Statistical Shape Models of the structures of the temporal bone and Skull Base to create an Average skull base
3. Use this average skull base as a template for segmenting new CT scans
4. Overlay template onto new CT scan
5. Use deformation fields obtained from SSMs to non-rigidly deform template to match new CT scan

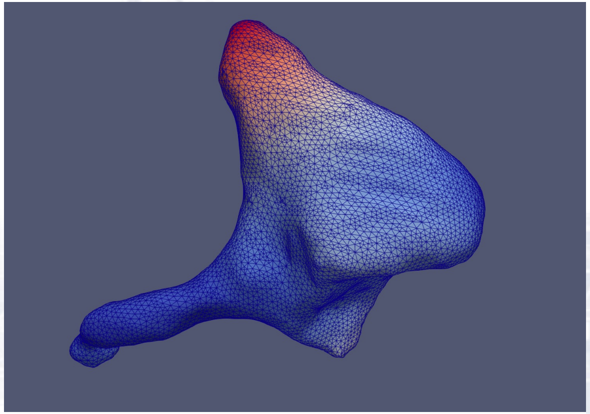
Development of an Automated Segmentation Algorithm

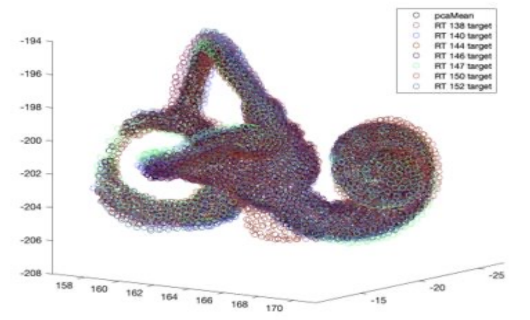
1. **Manual Segment 42 Temporal Bone CT scans**
2. Create Statistical Shape Models of the structures of the temporal bone and Skull Base to create an Average skull base
3. Use this average skull base as a template for segmenting new CT scans
4. Overlay template onto new CT scan
5. Use deformation fields obtained from SSMs to non-rigidly deform template to match new CT scan



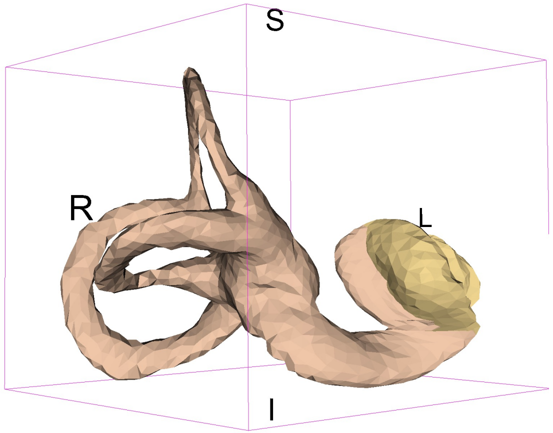
Development of an Automated Segmentation Algorithm

1. Manual Segment 42 Temporal Bone CT scans
2. **Create Statistical Shape Models of the structures of the temporal bone and Skull Base to create an Average skull base**
3. Use this average skull base as a template for segmenting new CT scans
4. Overlay template onto new CT scan
5. Use deformation fields obtained from SSMs to non-rigidly deform template to match new CT scan

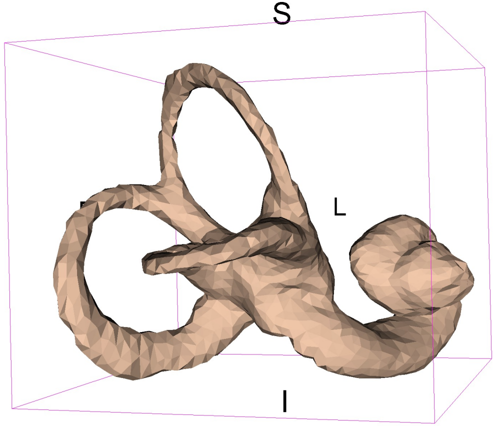




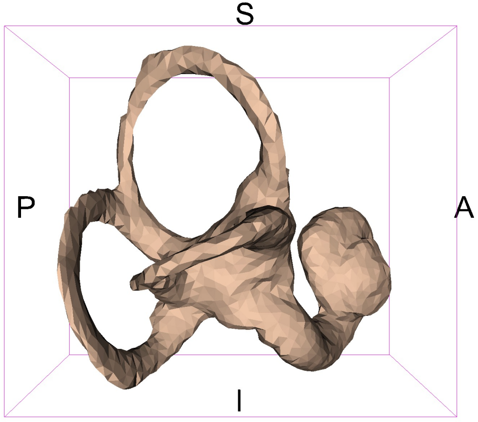
Bony Labyrinth SSM PCA



Principal Component 1



Principal Component 2



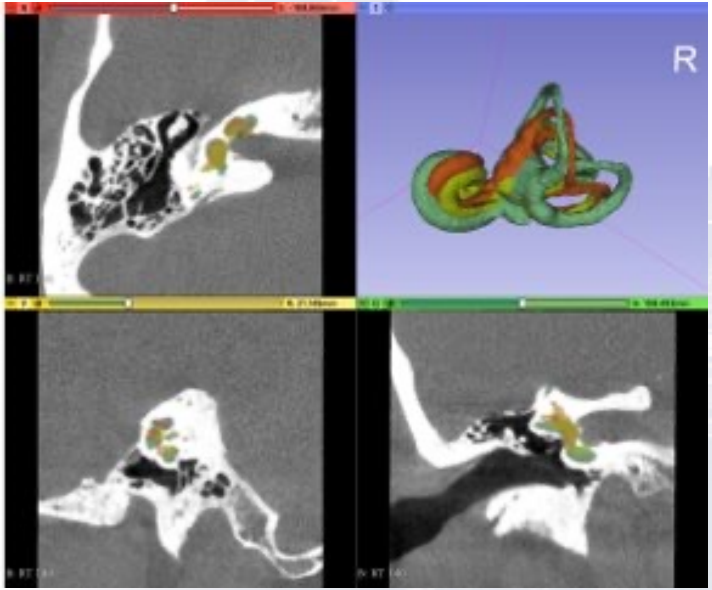
Principal Component 3

Development of an Automated Segmentation Algorithm

1. Manual Segment 42 Temporal Bone CT scans
2. Create Statistical Shape Models of the structures of the temporal bone and Skull Base to create an Average skull base
3. **Use this average skull base as a template for segmenting new CT scans**
4. Overlay template onto new CT scan
5. Use deformation fields obtained from SSMs to non-rigidly deform template to match new CT scan

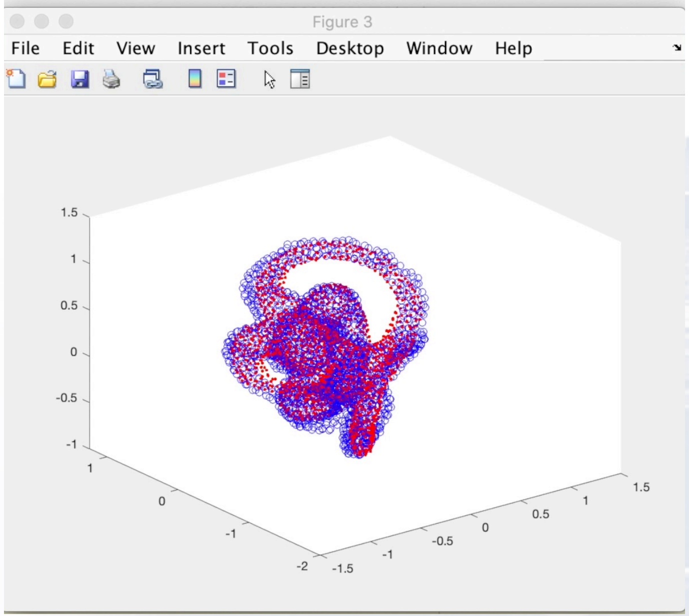
Development of an Automated Segmentation Algorithm

1. Manual Segment 42 Temporal Bone CT scans
2. Create Statistical Shape Models of the structures of the temporal bone and Skull Base to create an average skull base Atlas
3. Use this average skull base as a template for segmenting new CT scans
4. **Overlay template onto new CT scan**
5. Use deformation fields obtained from SSMs to non-rigidly deform template to match new CT scan



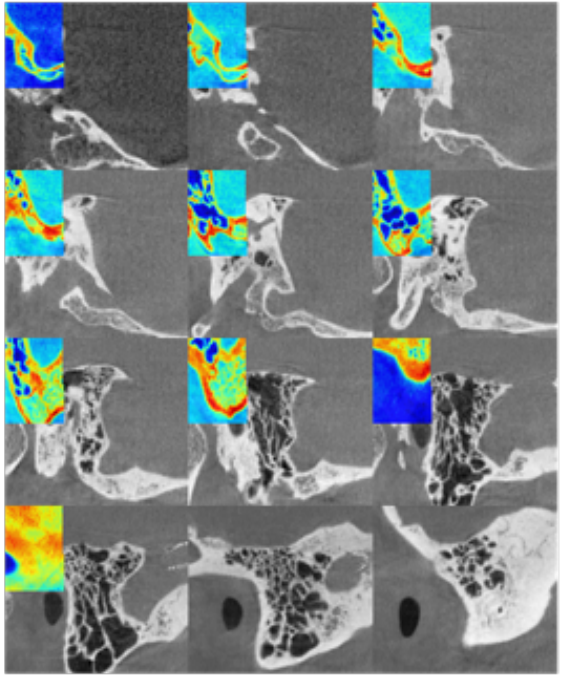
Development of an Automated Segmentation Algorithm

1. Manual Segment 42 Temporal Bone CT scans
2. Create Statistical Shape Models of the structures of the temporal bone and Skull Base to create an Average skull base
3. Use this average skull base as a template for segmenting new CT scans
4. Overlay template onto new CT scan
5. **Use deformation fields obtained from SSMs to non-rigidly deform template to match new CT scan**

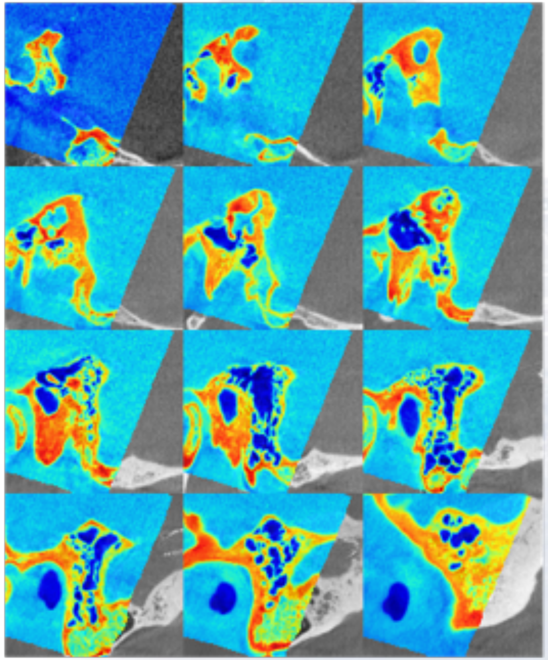


Development of an Automated Segmentation Algorithm

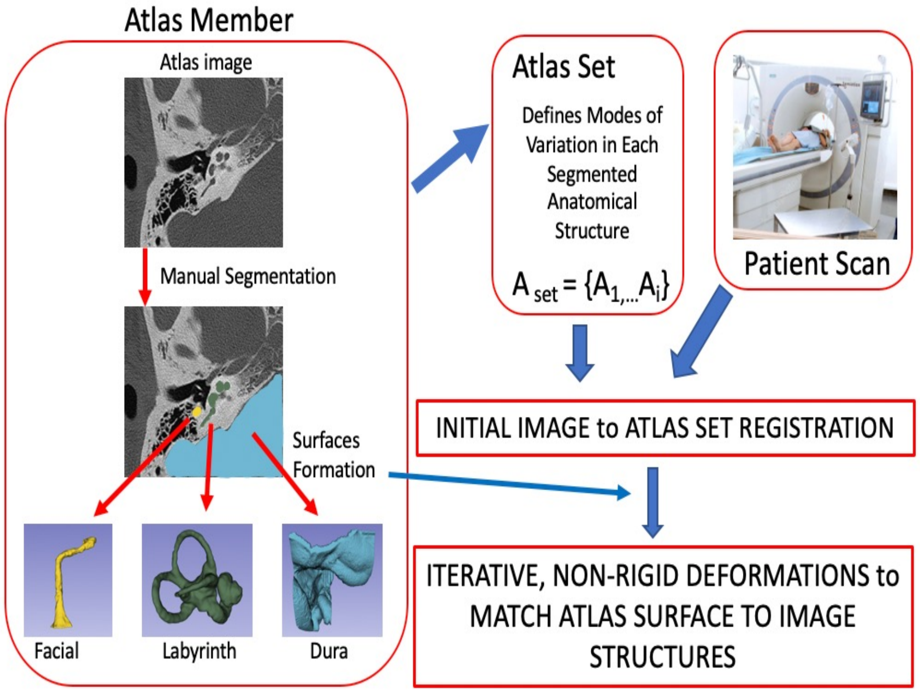
Before Registration



After Registration



Development of an Automated Segmentation Algorithm



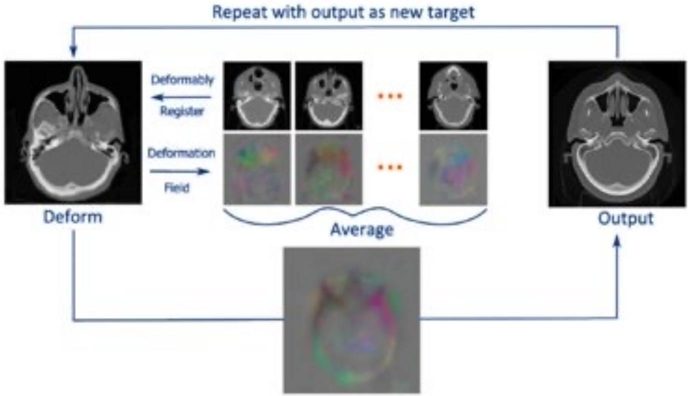


Figure 2.1: Template creation pipeline: all input images are deformably registered to one target image, which is then deformed by the mean of the deformation fields resulting from the registrations. The colors in the deformation fields represent the direction of the deformation vectors, whereas the intensity of the colors indicates the magnitude of the vectors. Deforming the target image by the mean deformation field takes the target image towards the mean of the input images. This process is iterated with the output image as the new target image. Individual variation from the initial target image decreases with every iteration, and the resulting output moves closer to the mean of the input set of images.

Segmentation Evaluation — Dice and Hausdorff Distances

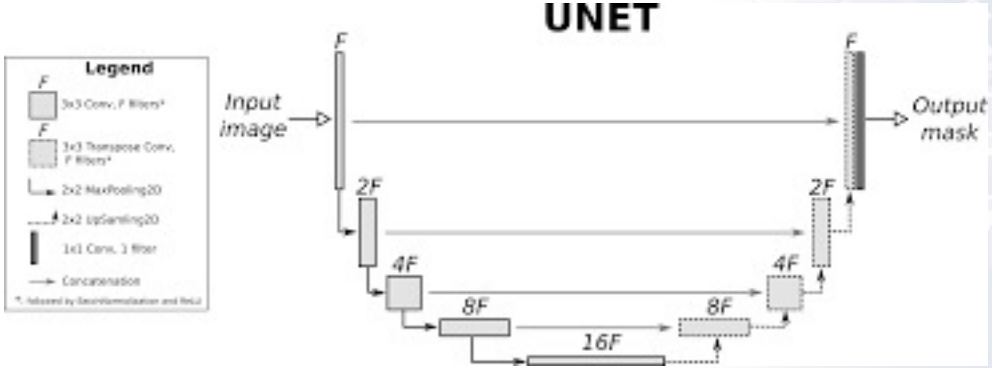
Segment	mean	std dev	Max Hausdorff Distances 138	152	147	146	144	143	142
Malleus (mm)	0.645	0.268	0.4660	1.2128	0.6253	0.4120	0.5791	0.6943	0.5234
Incus (mm)	0.870	0.247	1.1278	1.2071	0.8756	0.9324	0.6661	0.5100	0.7730
Stapes (mm)	0.946	0.402	1.8123	0.7766	0.7888	0.9018	0.9267	0.5508	0.8674
Vestibule + Cochlea (mm)	1.509	0.735	[illegible].5664	1.0930	2.0624	2.4749	2.1653	1.4024	0.8003
Facial Nerve (mm)	3.892	1.763	[illegible].6530	2.6300	4.5432	2.8950	7.3989	4.4492	2.6733
Chorda Tympani (mm)	12.819	8.000	1[illegible]6473	18.4033	21.2440	2.9788	19.2804	4.6203	5.5602

Segment	mean	stdev	Mean Hausdorff Distances 138	142	143	144	146	147	152
Malleus (mm)	0.106	0.007	0[illegible]142	0.1082	0.1073	0.1000	0.0942	0.1082	0.1118
Incus (mm)	0.129	0.029	0[illegible]884	0.1147	0.1004	0.1159	0.1346	0.1359	0.1154
Stapes (mm)	0.257	0.214	0[illegible]322	0.1874	0.1555	0.2660	0.1623	0.1416	0.1506
Vestibule + Cochlea (mm)	0.175	0.081	0[illegible]158	0.1184	0.1353	0.1969	0.2405	0.3196	0.1008
Facial Nerve (mm)	0.761	0.413	0[illegible]202	0.5074	0.7860	1.4298	0.5708	1.2277	0.3845
Chorda Tympani (mm)	1.035	0.783	1[illegible]232	0.9625	0.8941	0.4473	0.7252	0.3874	2.7049

Segment	mean	stdev	Dice 138	142	143	144	146	147	152
Malleus	0.836	0.014	0.8242	0.8550	0.8423	0.8424	0.8414	0.8307	0.8135
Incus	0.846	0.029	0.7835	0.8722	0.8497	0.8424	0.8608	0.8552	0.8561
Stapes	0.353	0.140	0.0659	0.4203	0.4107	0.3016	0.4270	0.3563	0.4904
Vestibule + Cochlea	0.847	0.063	0.8651	0.8827	0.8754	0.8645	0.8364	0.7106	0.8979
Facial Nerve	0.551	0.101	0.6323	0.5786	0.5209	0.4907	0.5829	0.3738	0.6799
Chorda Tympani	0.099	0.177	0.0006	0.0650	0.0991	0.4907	0.0028	0.0360	0.0000

Future Directions

- Develop deep learning models to improve on segmentation accuracy and speed



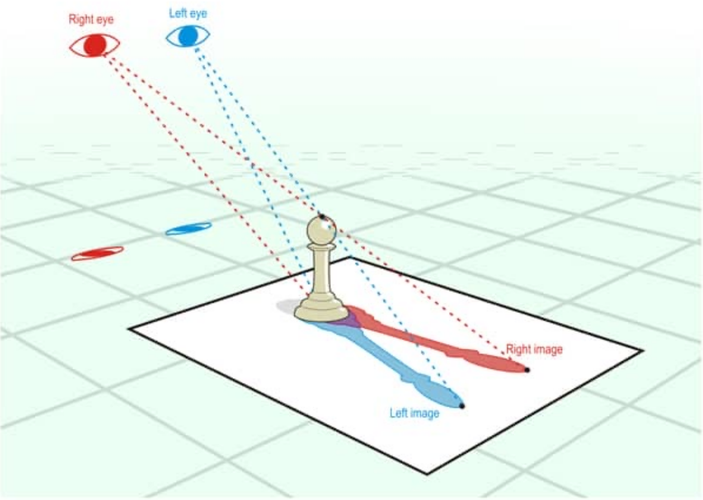
Limitations of Current Image Navigation Systems

- EM and Optical Trackers
  - Rely on fiducial markers, bone anchored fiducials or surface scanning
  - All inhibit the surgical workflow
  - Prone to errors (clinically reliable to 1-2mm)
  - Often work well at the surface but lose fidelity as you proceed deeper in the skull

- <u>No currently available method to update and improve registration intraoperatively as bone is removed</u>

Can you teach a microscope to detect and register anatomy?

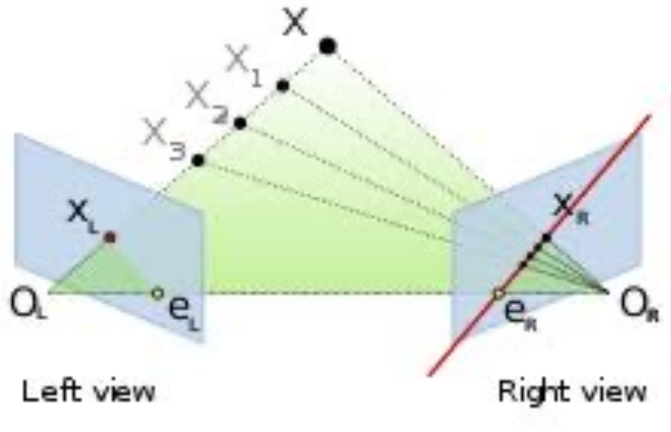
- Traditional Microscopes
  - Allow only **monovision video** recording through single eye pieces
- New fully digital microscopes
  - Present digital image from two eye pieces
  - Allows for **stereoscopic video**



JOHNS HOPKINS MEDICINE

Can you teach a microscope to detect and register anatomy?

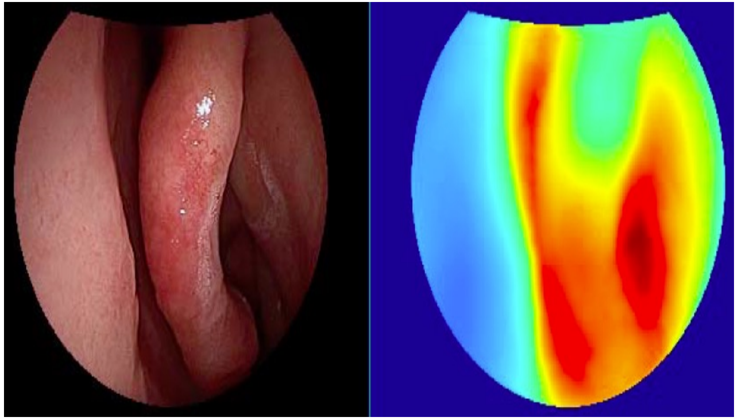
- Stereovision Video
  - **Allows us to take advantage of epipolar geometry**
  - Creates the potential to determine image depth and real world 3-D location of a point on an image
  - A deep learning network could use this to correlate the stereoscopic surgical image to the preoperative image



Can you teach a microscope to detect and register anatomy?

- Previously we have shown you can do a similar calculation with an endoscope

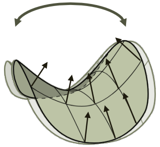
Reconstruction of Sinus Anatomy from Monoscope Endoscope Video



Xingtong Liu, *et al.*, "Self-supervised Dense 3D Reconstruction from Monocular Endoscopic Video", MICCAI 2019

Deformable Registration to Statistical Model

f_{shape} = likelihood of deformation in $\Psi(\mathbf{s})$

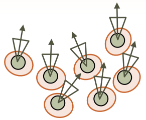


Statistical shape model with shape parameters, $\mathbf{s}$

f_{match_i} = likelihood of $\mathbf{Fx}_i$ matching $\mathbf{y}_i \in \Psi(\mathbf{s})$

Initial shape, $\Psi(\mathbf{s}_0)$

Initial transform, $\mathbf{F}_0$



Observed features X with noise model

$\mathbf{X} = \{\mathbf{x}_i\}$

Current transform, **F**, current deformed shape $\Psi(\mathbf{s})$

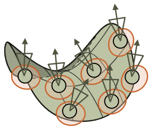
Correspondence Phase

Find matches between **FX** and $\Psi(\mathbf{s})$ that maximize f_{match}

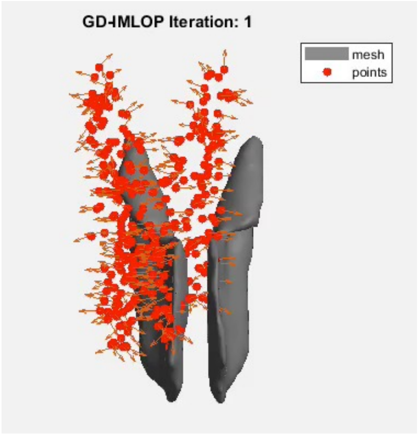
Registration Phase

Find **F** and **s** that maximize

$$f_{match\_deformable} = \left(\prod_i f_{match_i}\right) \cdot f_{shape}$$

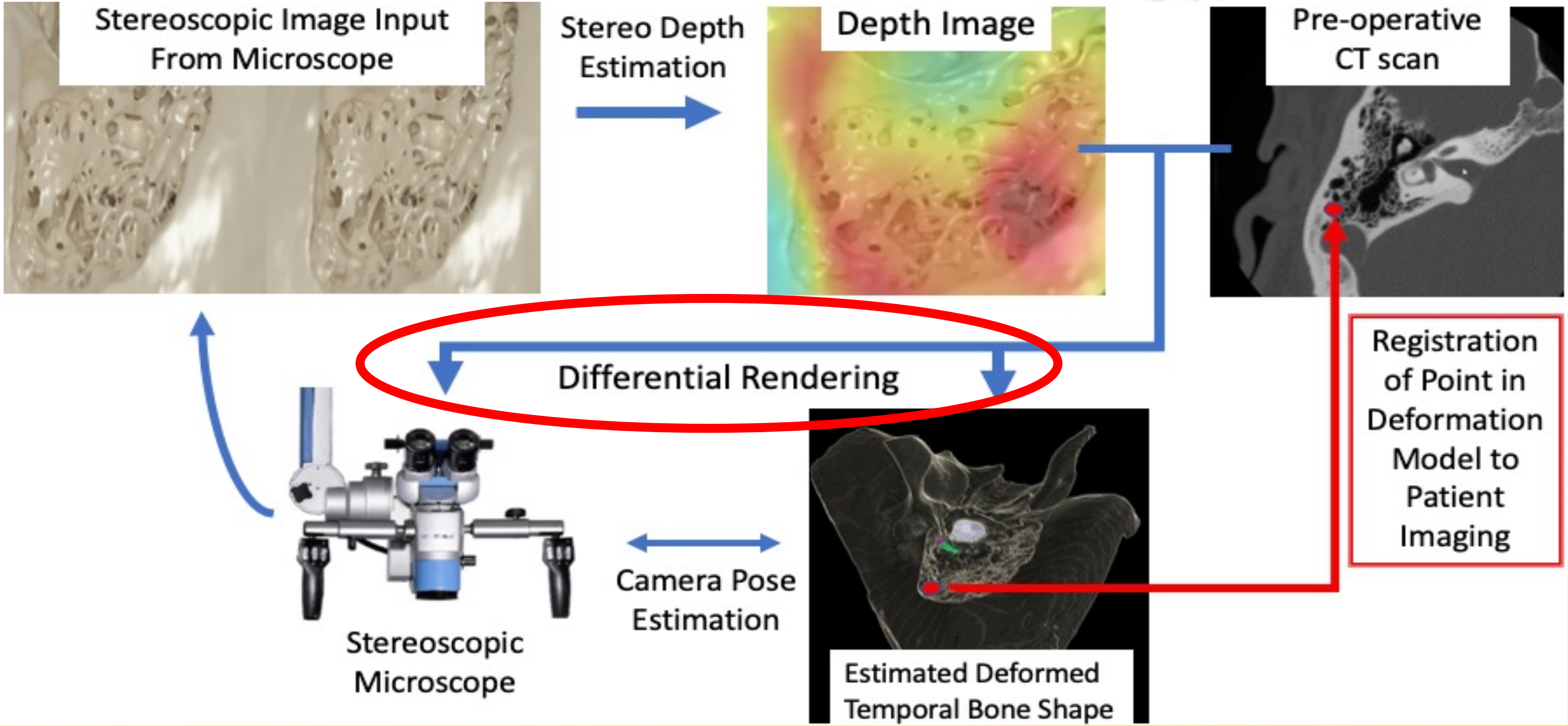


Final registration and deformed shape

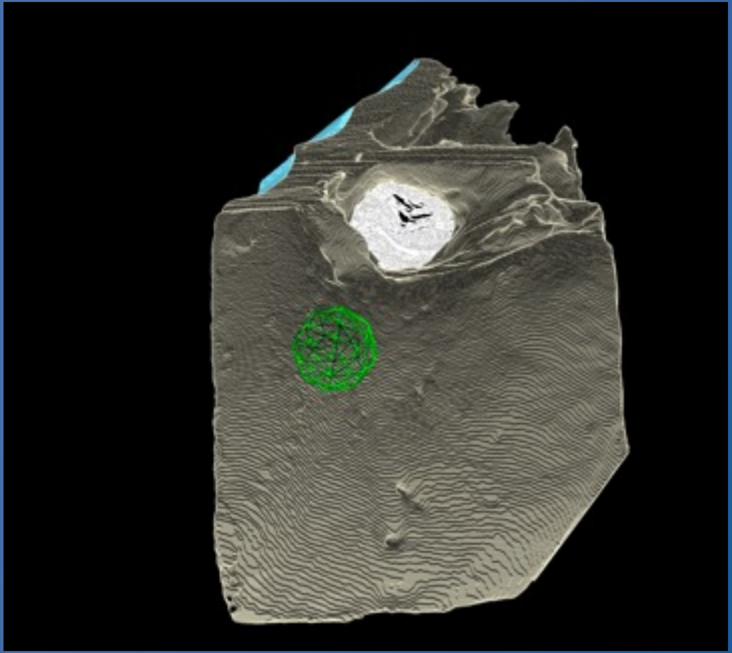


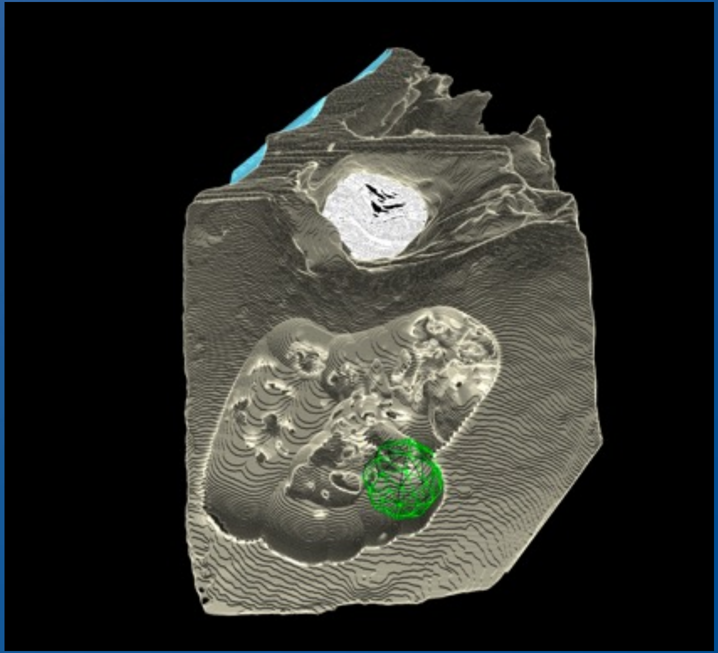
A. Sinha, S. D. Billings, A. Reiter, X. Liu, M. Ishii, G. D. Hager, and R. H. Taylor, "The deformable most-likely-point paradigm", Medical Image Analysis, vol. 55-, pp. 148-164, July, 2019.

Stereoscopic Microscope Navigation Work-flow

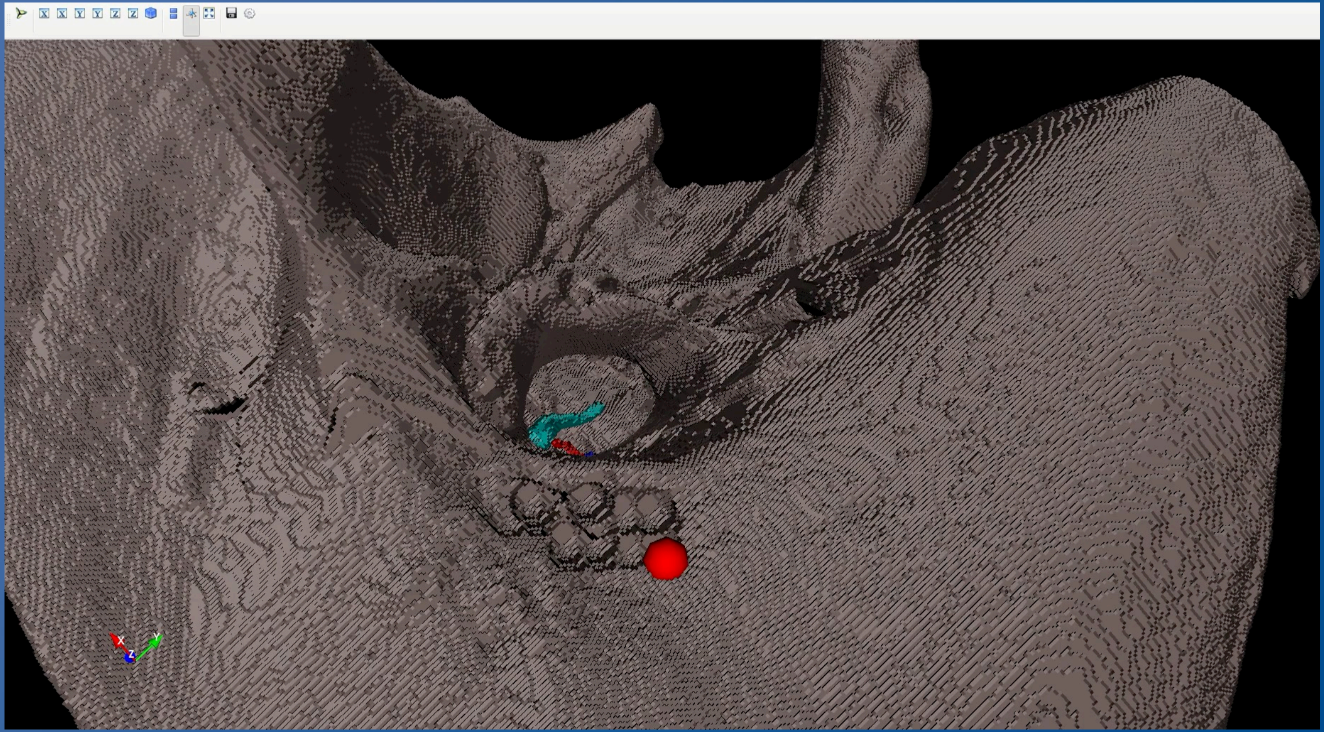


Develop synthetic training set:
Teach the microscope what to expect to see



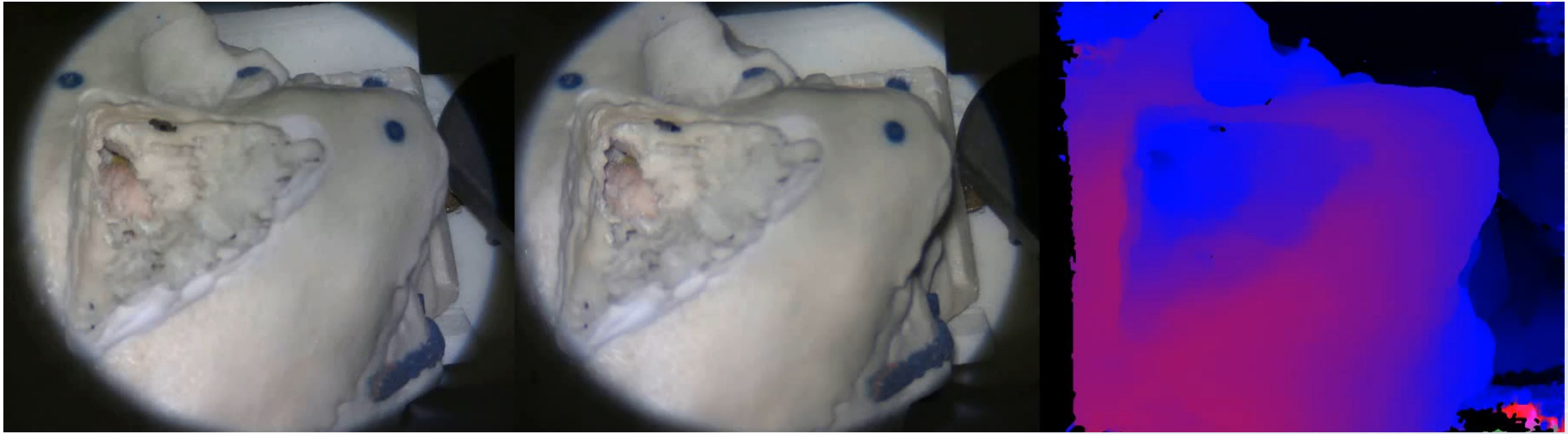


Develop synthetic training set

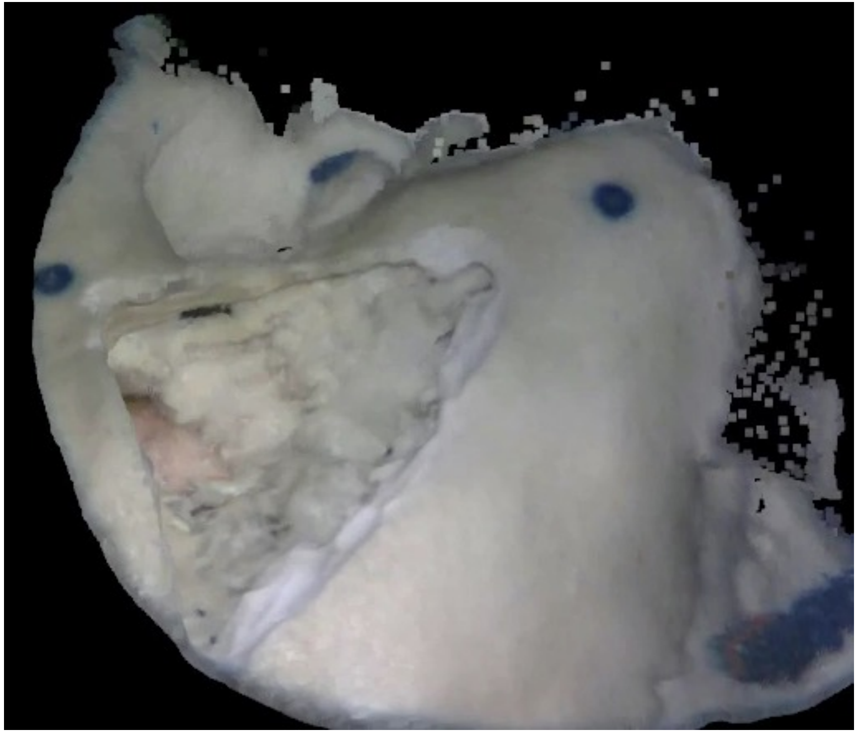


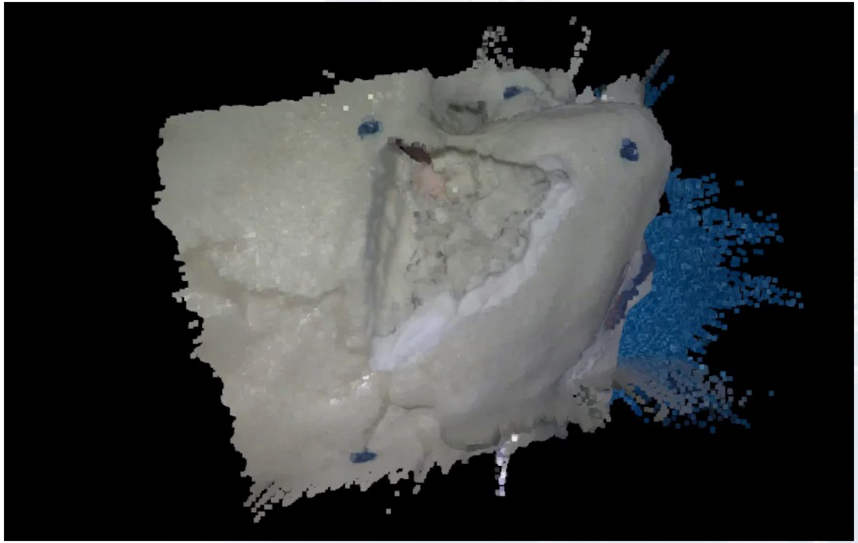
Develop synthetic training set

Using Stereovideo we can determine depth estimations

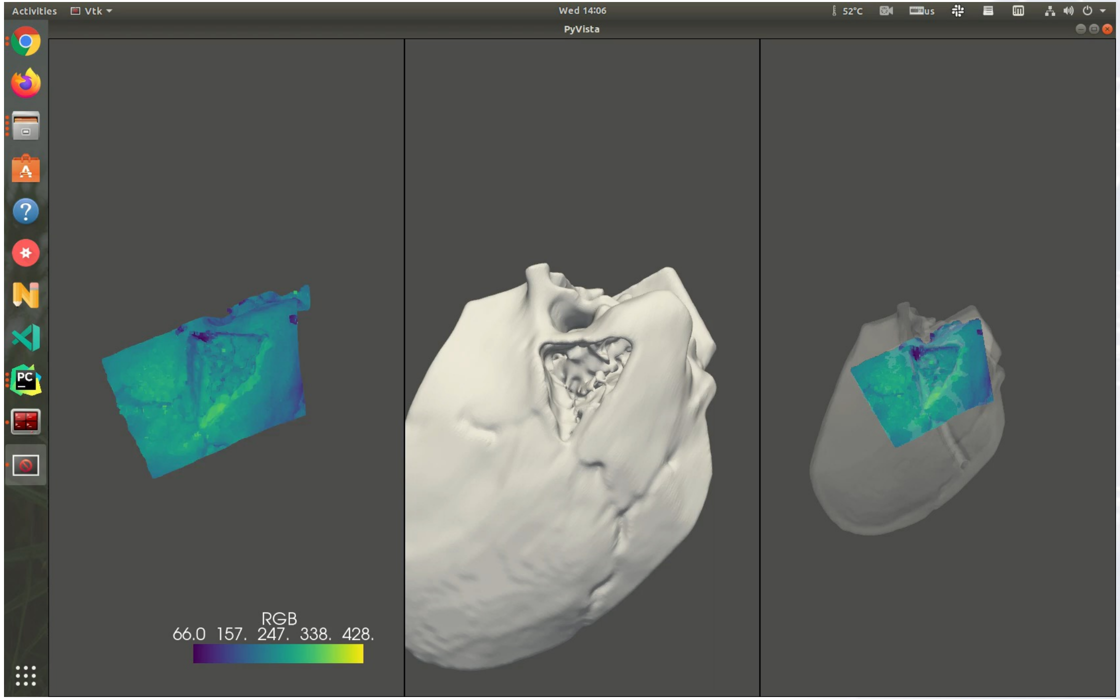


Incorporating this with training data and calibration data we can reconstruct 3D shapes from video

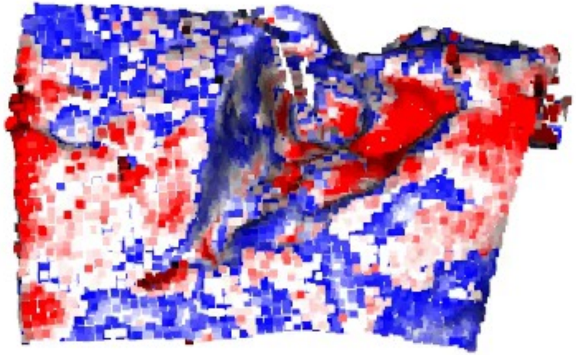




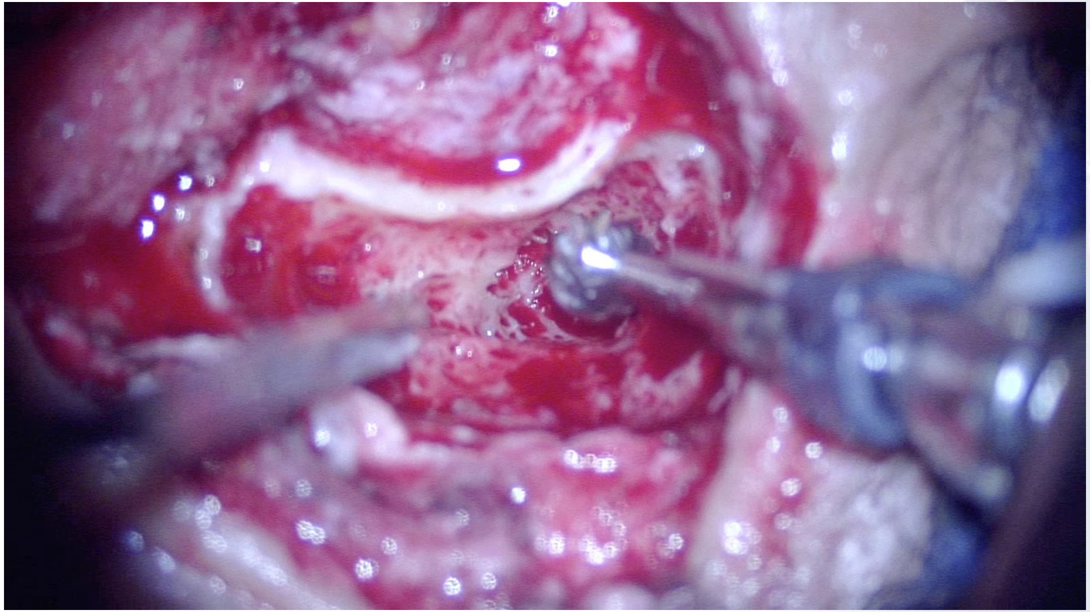
This data can then be merged with CT imaging to register the microscope to the patient



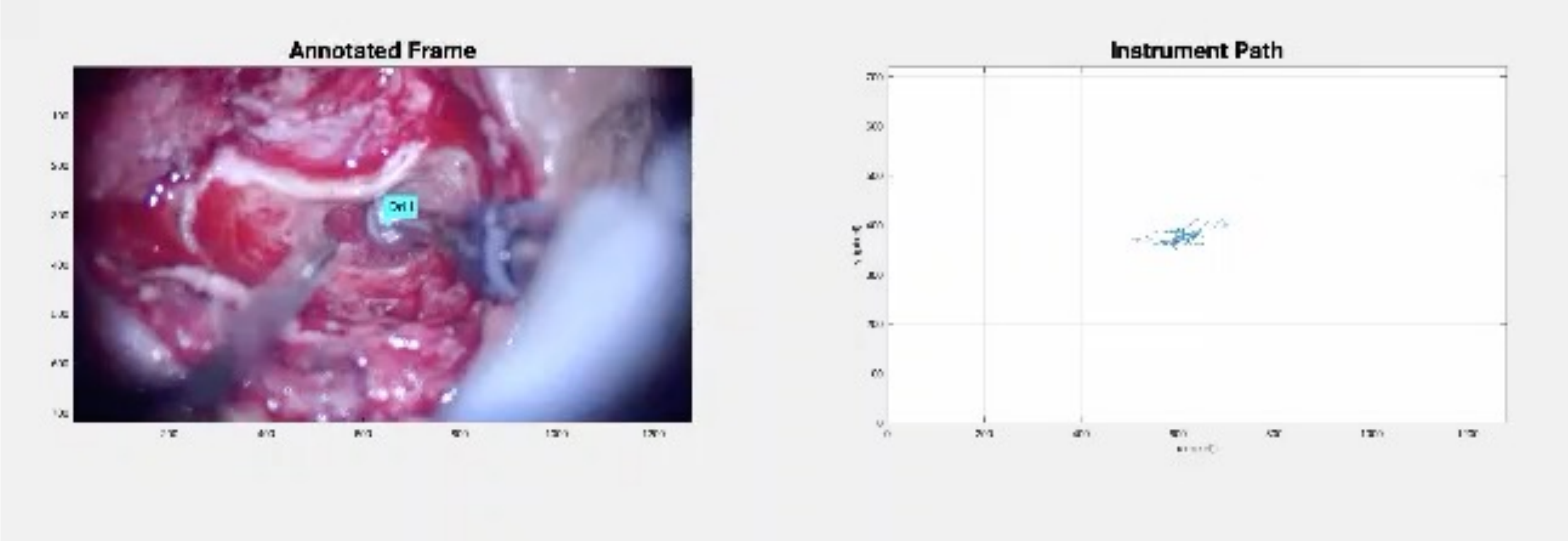
	Inlier RMSE	# of Correspondence
Direct generalization	1.152 mm	6946
Self supervision	1.147 mm	6928



Next Steps: Incorporate Instrument Detection and Tracking



Next Steps: Incorporate Instrument Detection and Tracking



LABORATORY FOR Computational Sensing + Robotics

I-STAR

Russ Taylor, PhD

Mathias Unberath, PhD

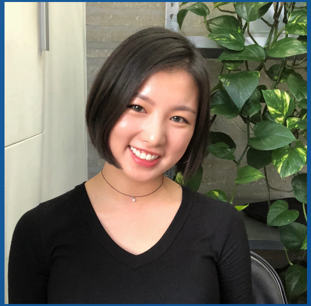
Deepa Galaiya, MD

Jeff Siewerdsen

Max Li

Andy Ding

Alex Lu



Sue Min Cho

Questions?