Next-Generation Robotic Surgery

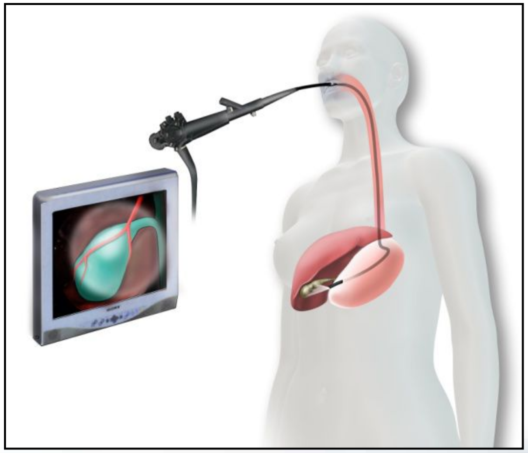
Peter K. Allen
Department of Computer Science
Columbia University

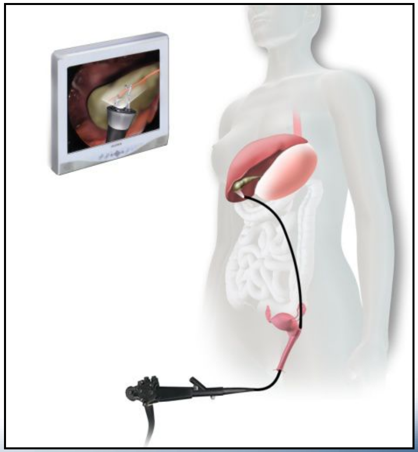
Talk Overiew

1. In-vivo imaging device (mono)
2. In-vivo imaging device (stereo)
3. Insertable Robotic Effector Platform (IREP)
4. Markerless Surgical Tool Tracking
5. Continuum Surgical Robot Pose Estimation using Vision

Surgical Robotics: Research Goals

- Create simple-to-use and cost-effective surgical robots
- Convert more "major access" operations to "minimal access" operations.
- Reduce the invasiveness of current minimal access interventions
  - SPA: **S**ingle **P**ort **A**ccess for laparoscopic surgery
  - NOTES: **N**atural **O**rifice **T**ranslumenal **E**ndoscopic **S**urgery
  - Use natural body openings with robotic platforms





Current Generation Robotic Surgery

Devices such as DaVinci®
Huge leap in robotics, but:

- Large footprint in the OR
- Cost is extremely high
- Requires multiple incisions
- Multiple assistants needed
- Uses traditional endoscope with limited mobility within body cavity
- Has not reduced the invasiveness of robotic MIS
- While this paradigm has been enormously successful, and has spurred development of new methods and devices, it is **ultimately limiting in what it can achieve**

Next-Generation Robotic Surgery

- Surgery will be radically different in the future
- New thrusts in computer & robotic technologies can make automated surgery, if not feasible, an approachable goal.
- Vision: **teams of insertable robots** performing surgical tasks in the body under both surgeon & computer control.
- Remotize sensors and effectors in the body cavity where they can perform surgical & imaging tasks **unfettered by traditional endoscopic instrument design.**

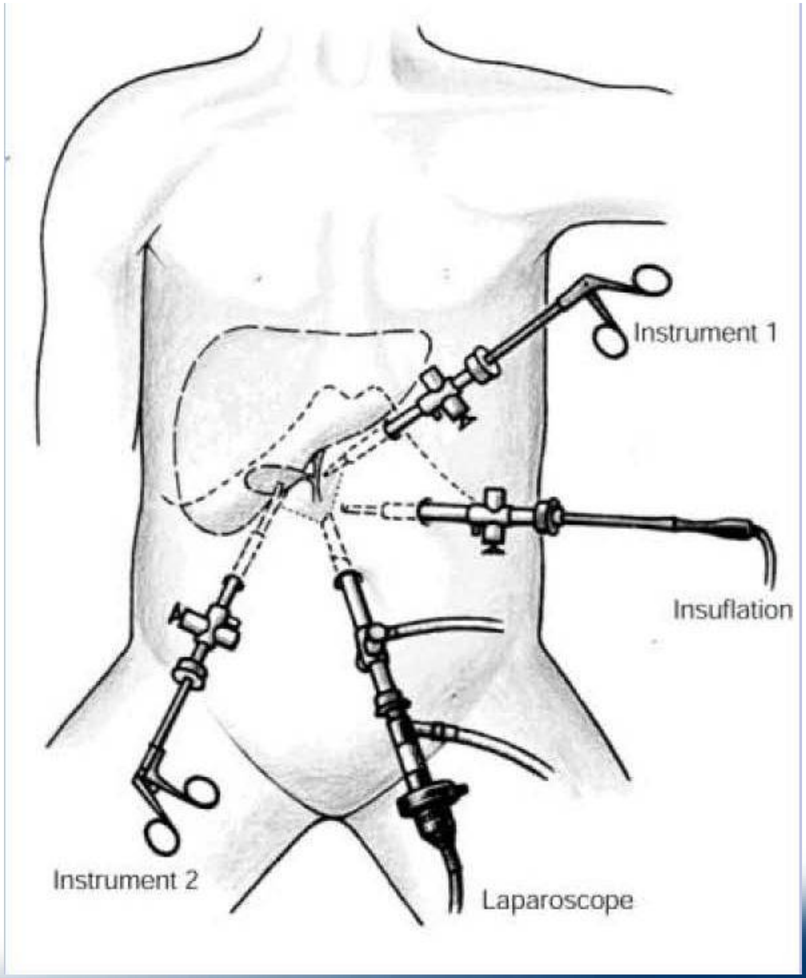
Building New In-vivo Devices

- Current minimal access surgery adheres to the Chopstick Paradigm:
  Pushing long sticks into small openings

Problems with Current Imaging Devices

Can we improve on the traditional laparoscope?

- Laparoscope Issues:
  - Narrow angle imaging
  - Limited workspace
  - Multiple incisions for camera placements
  - Counter intuitive motion for control
  - Trained assistants needed to control the camera
  - Multiple incisions for camera placements
  - Additional incisions needed for laparoscopic instruments.



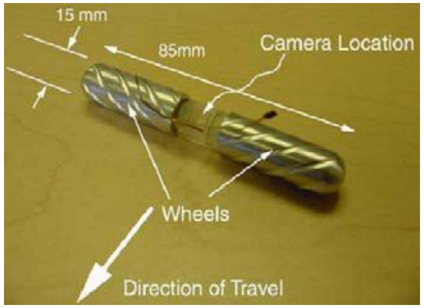
In-Vivo Imaging Devices

Flexible Endoscope

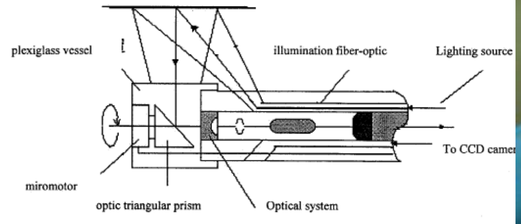
Karl Storz Endoscope

Pill camera

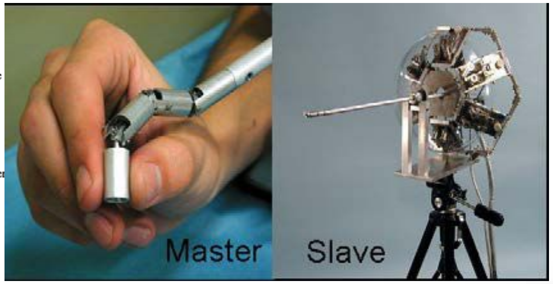
- Rod-lens by Hopkins and cold light source of fiber optics by Karl Storz.
- Flexible endoscope using fiber optics to delivery light and transmit image.
- Pill camera without locomotion.
- Endoscope with rotating mirror.
- Endoscope positioned by multilink arm with piezoelectric actuators.
- Mobile robots, U. Nebraska



U. Nebraska



Gao et al., 1998



Ikuta et. al., 2002

Columbia Imaging Device: Design Goals

- Device must be fully insertable into body cavity, leaving the insertion port free for other sensors and tooling
- Device diameter must be restricted to 15 mm diameter for use with standard trocars.
- Pan and Tilt degrees of freedom needed to increase internal imaging field of view
- Image Zoom function required
- Integrated lighting
- Simple intuitive control interface to operator
- Real-time computer control of DOF's to allow tracking and visual servoing
- User friendly 2D/3D display system
- Low cost and possible disposal use

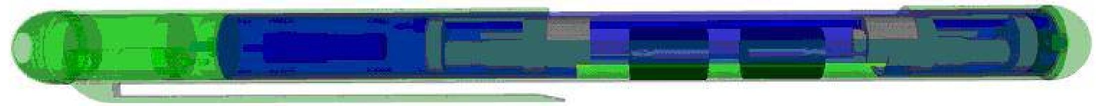
Columbia Imaging Device Overview

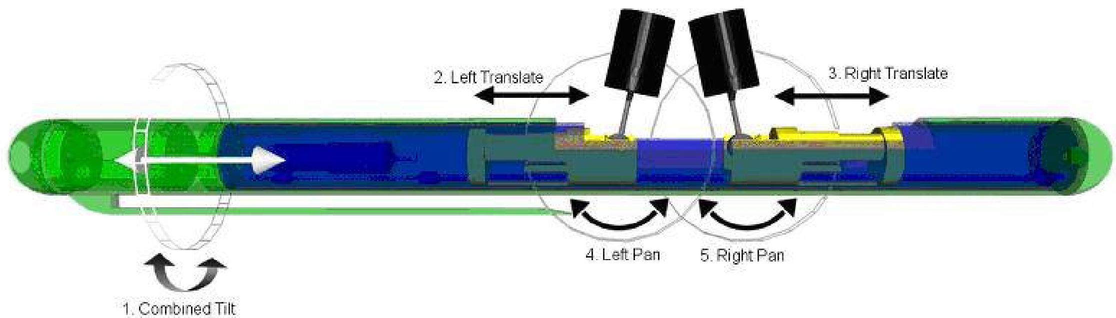
- Design 0: Paper design, 2 cameras, 5-DOF
- Device 0: Single camera prototype, 3-DOF, tested in surgical trainer
- Device I: Single camera, pan/tilt/lighting, tested in animals
- Device II: Single camera, pan/tilt/zoom, tested in animals
- Device III: Stereo cameras, pan/tilt, tested in animals

Columbia Robotics

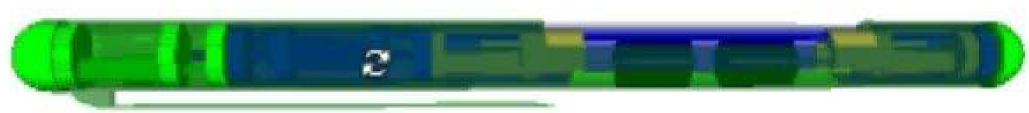
Design 0

- Stereo cameras with 6 DOF are desirable – full mobility
- Difficult to achieve in small actuated package
- Compromise – 3 DOF per camera
  - Cameras share tilt axis (1 DOF)
  - Independent translation (2 DOF)
  - Independent pan (2 DOF)



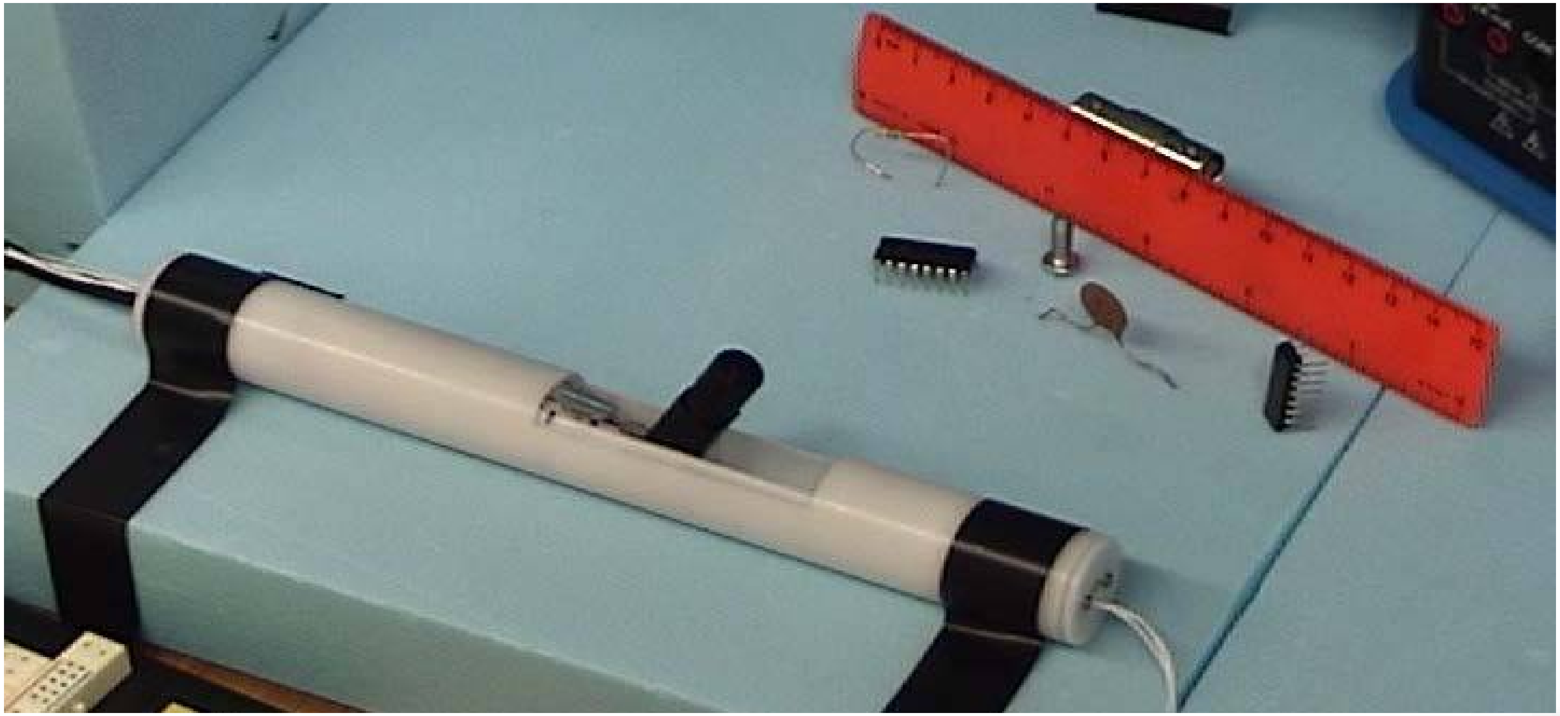


Design of 5-DOF insertable camera device



Device I: Single Camera

Diameter: 22 mm; Length: 190 mm
Camera opening: 58 mm
3 DOF: Pan: 120°; Tilt: 130°; Translation: 50 mm



Video

Initial Testing and Validation

Does new imaging device improve surgery visualization?

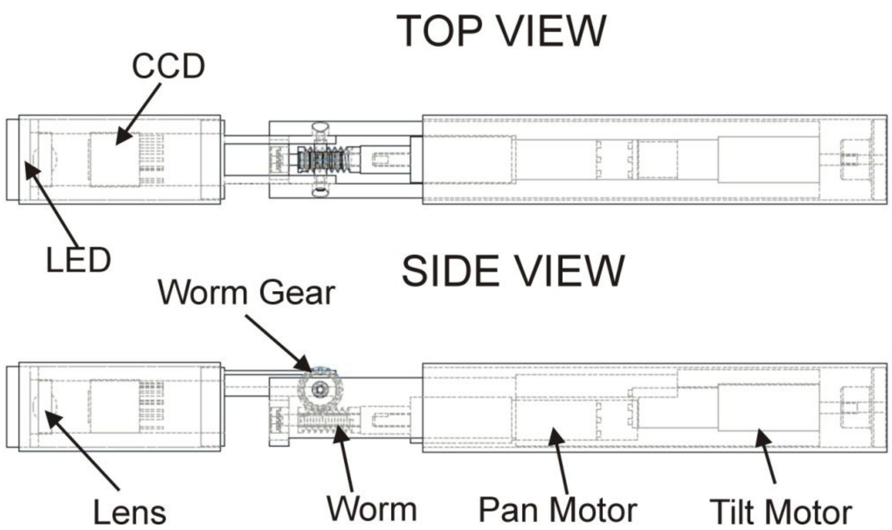
- 6 fellows & surgeons performed MISTELS* tests with standard laparoscope and the new robotic camera
- 5 of 6 subjects showed no significant difference in MISTELS task performance with the robotic camera compared to the standard laparoscope
- Mean score of 999 +/- 69 using a laparoscope
- Mean score of 953 +/- 68 for the robotic camera: statistically insignificant difference

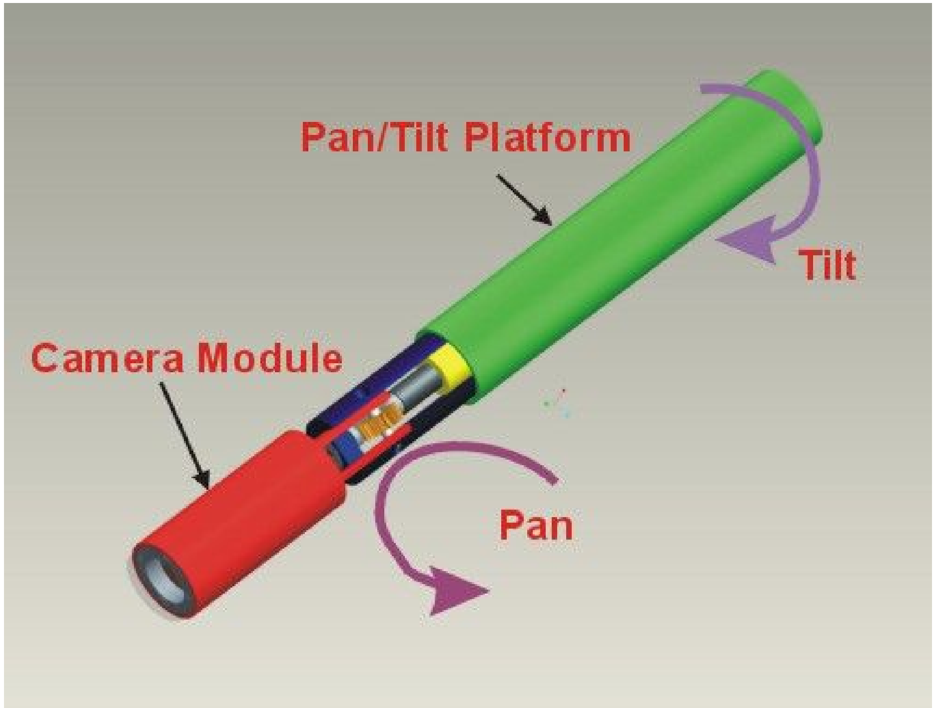
**McGill Inanimate System for the Training and Evaluation of Laparoscopic Skill*

Device I: Design Goals

- Need to reduce size to fit 12mm trocar
- Motors are major determinant of device size
- Removing a camera reduced motor count by 2
- Translation DOF is least useful. Removing this also reduced motor count by 1
- Include integrated light source
- Make imaging head modular
- Tradeoff: Degrees-of-freedom for compactness

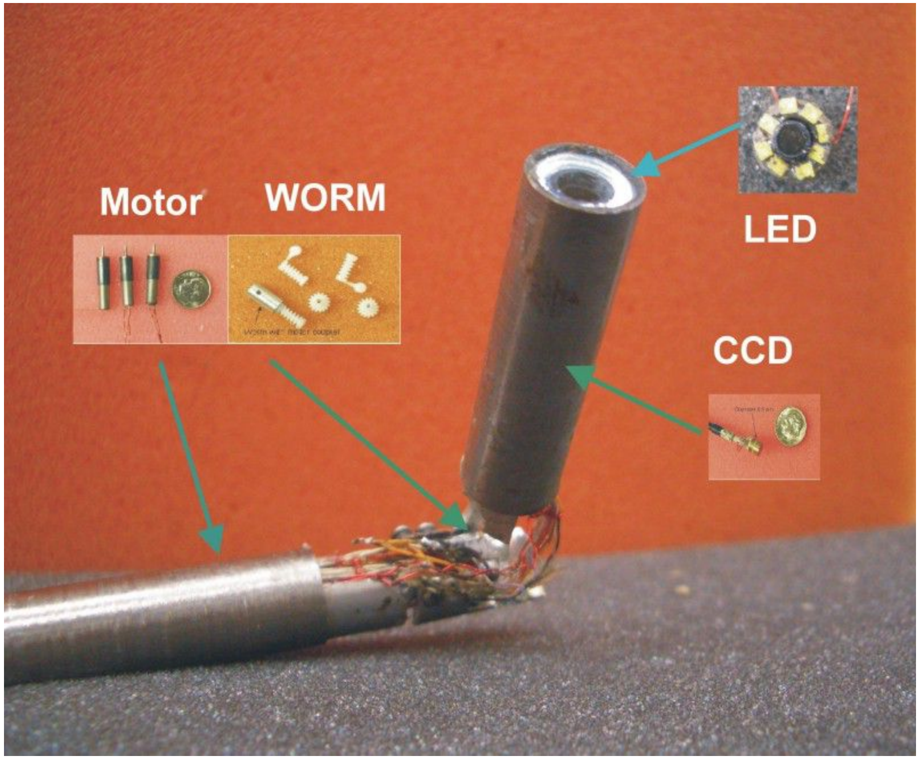
Device I: Single Camera





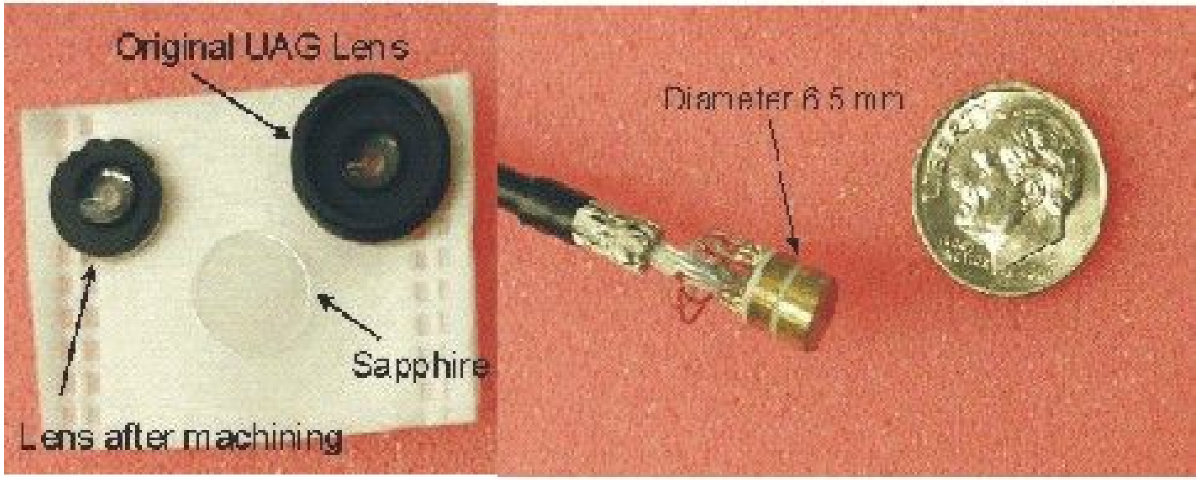
Device I: Single Camera*

- 110 mm in length and 11 mm in diameter.
- 130 degree Pan, 90 degree Tilt.
- Integrated 8 LED light source.
- 6.5 mm CCD sensor.
- Fully sealed camera head.
- Joystick control.



*Tie Hu, Peter K. Allen, Nancy Hogle and Dennis Fowler Surgical Imaging Device with Pan, Tilt, Zoom, and Lighting, Intl. Journal of Robotics Research, 2009

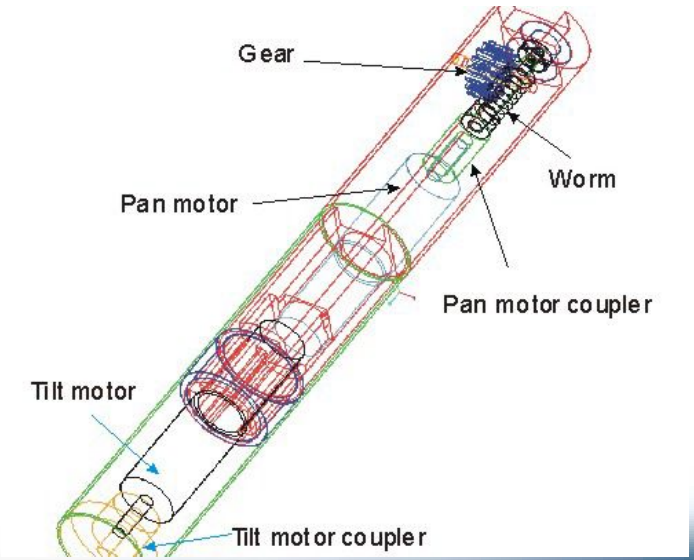
Lens and Camera Unit



- Pin hole lens
  - Focal length 5.0 mm, F number 4.
  - Angle of view D-H-V(85.4-68.3-50.9).
- 6.5 mm CCD camera sensor.
  - 450 TV lines in horizontal resolution and 420 TV lines in vertical resolution.
- Fully sealed package to isolate body fluid and moisture.

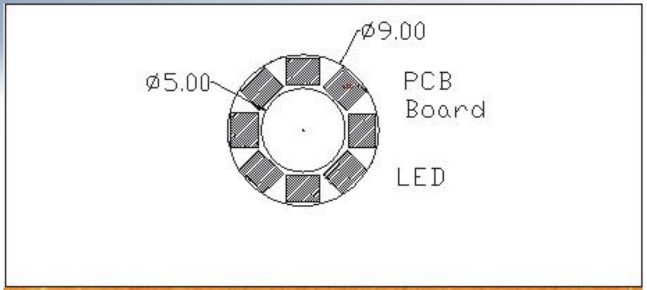
Pan/Tilt Mechanism

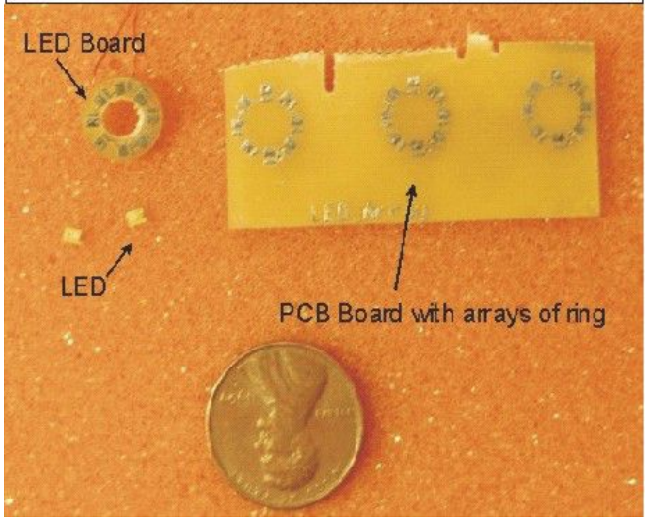
- Miniature Brushless DC motor (0513G, Faulhaber Group).
    - 25mNm torque.
    - 5.8 mm in diameter.
- Miniature worm gear (Kleiss Gear Inc.)
    - gear ratio 16:1.

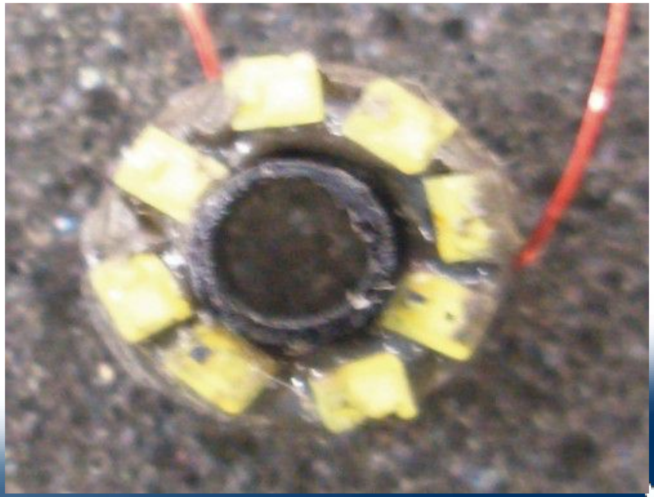


LED Light Source

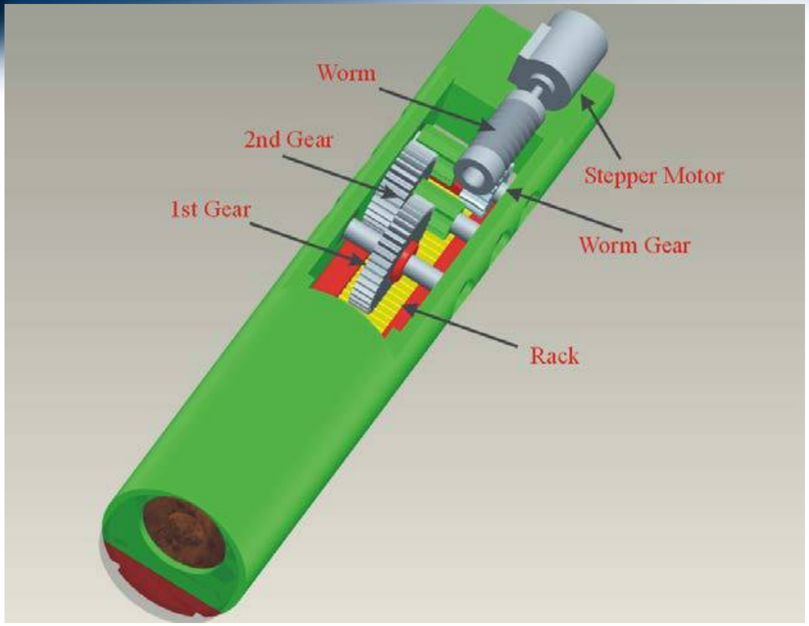
- Light-emitting diode (LED) as a light source in laparoscopy:
  - Lower power
  - Higher efficiency
  - Compact package
  - Longer lifespan
  - Lower cost
- Luxeon portable PWT white LED(LXCL_PWT1)
  - 2.0 X1.6 X 0.7 mm
  - 26 lumens of light at 350 mA
- 8 PWT LED in a printed circuit board with 9mm diameter.
  - 208 lumens light at 8.4 w



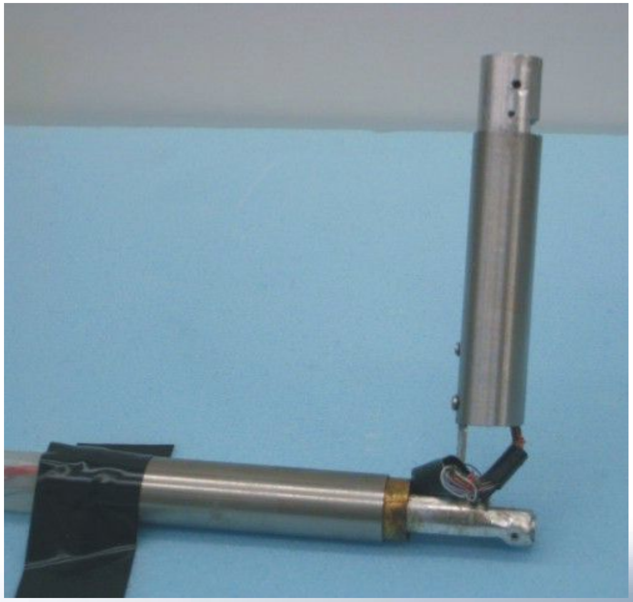




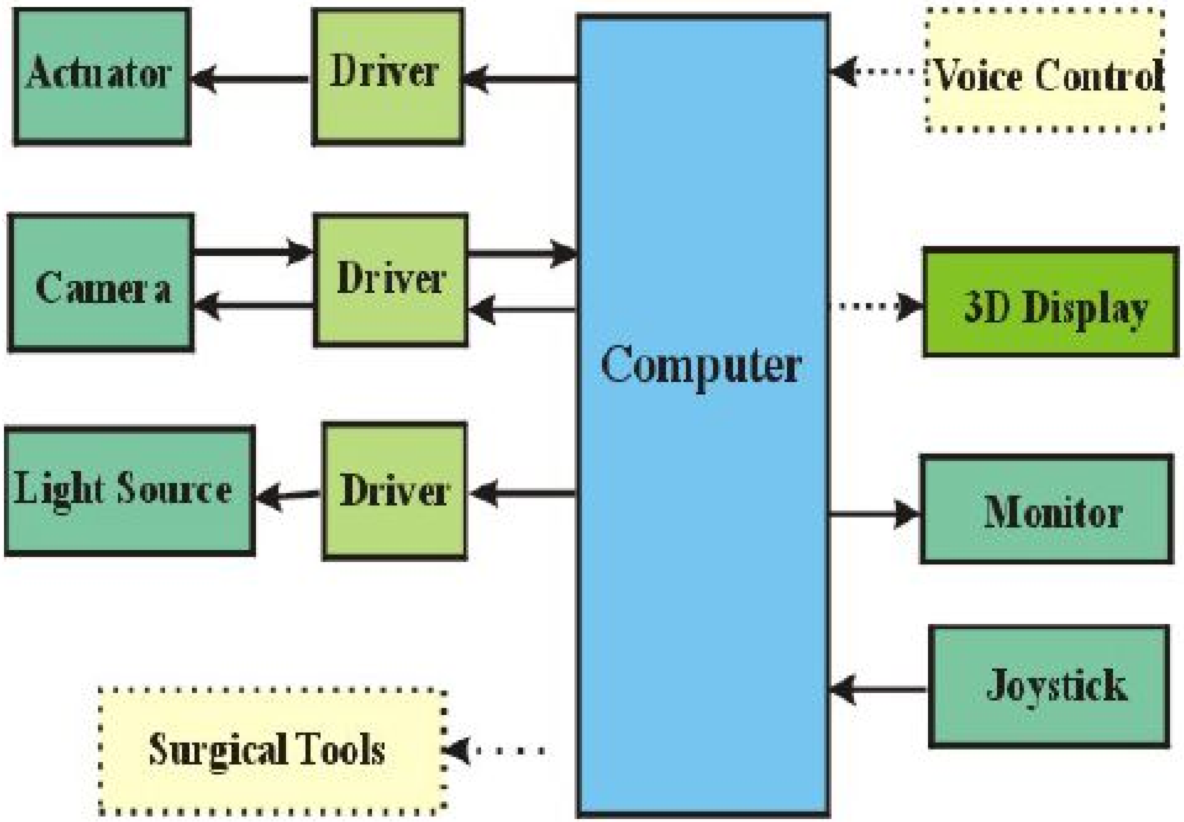
<u>Device II:</u>
Pan, Tilt, Zoom



- Mechanical zoom: linear motion of camera head
- Stepper motor drives rack and pinion mechanism
- Can only achieve ~ 2x zoom



System Architecture

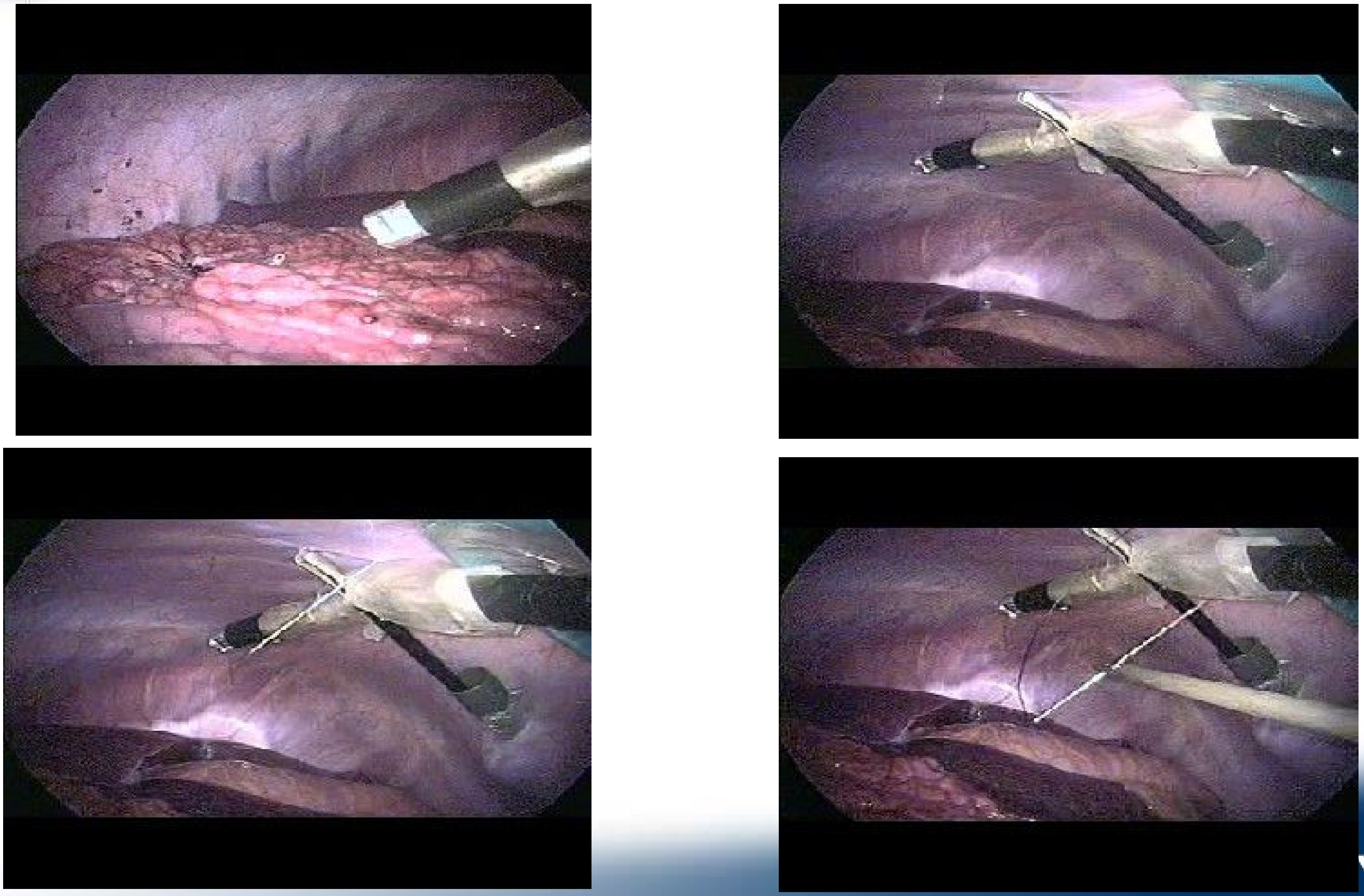


No expensive console needed, just a standard PC!

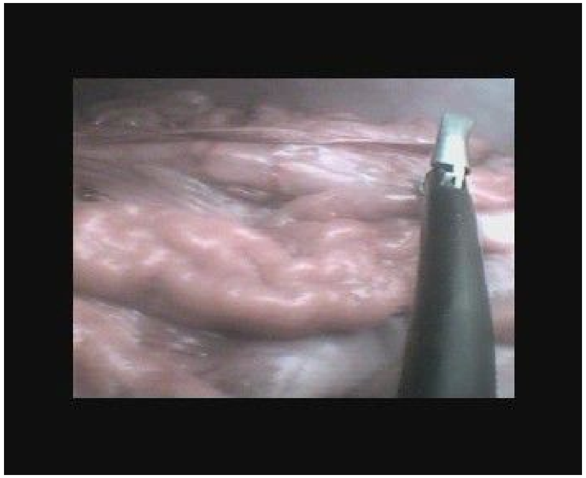
Mounting the Camera

- Camera attached to insuflated abdominal wall
- Attachment methods:
  - Suturing: small stitch through abdomen
  - Magnets
  - “Fish Hook” which grabs the abdominal wall
  - Intelligent trocar for attachment

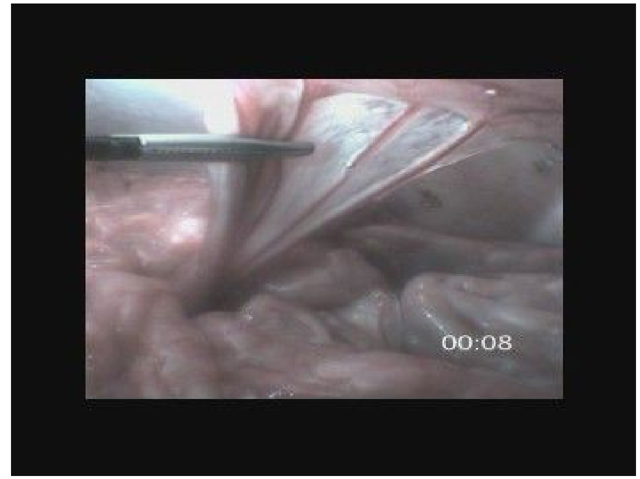
Suturing the Camera



In-Vivo Animal Experiments

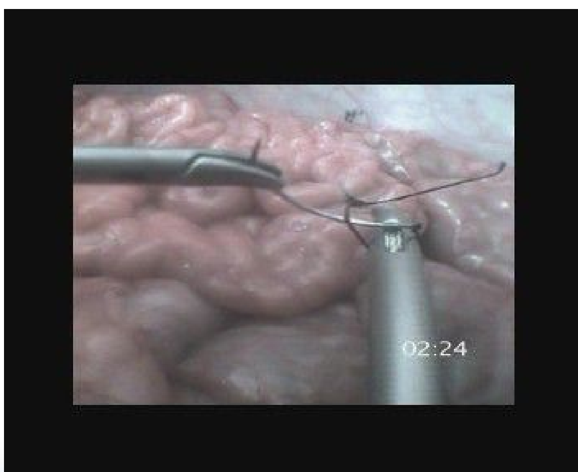


Bowel Running



Appendectomy

Video



Suturing

Nephrectomy

Procedure Timings

Procedure	Device	Time (min)
Running Bowel	Laparoscope	4:20
Running Bowel	Robot	3:30
Appendectomy	Laparoscope	2:20
Appendectomy	Robot	2:20
Suturing	Laparoscope	5:00
Suturing	Robot	4:00
Nephrectomy	Laparoscope	18:00
Nephrectomy	Robot	21:00

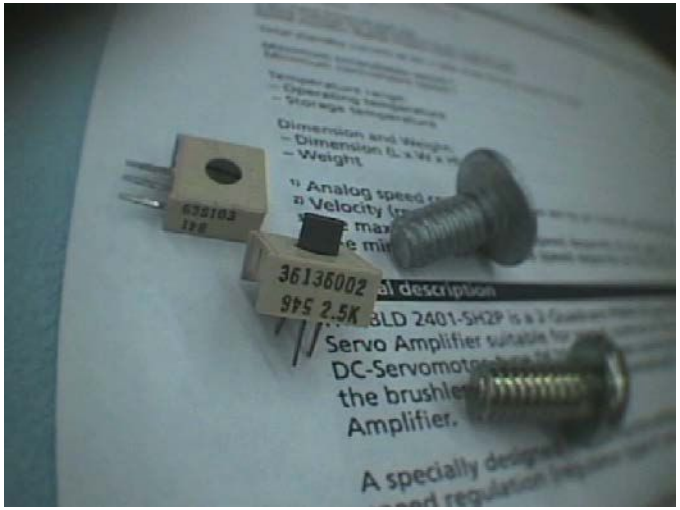
Intelligent Software

- Position/Velocity control of axes
- Intuitive Joystick Control
- Real-Time Image Processing:
  - Digital Zoom
  - Image rotation/stabilization
  - Distortion Correction
  - Picture-in-Picture
  - Visual Servoing/Tracking
  - 3D Stereo output

Columbia Robotics

Image Processing

Zoom :



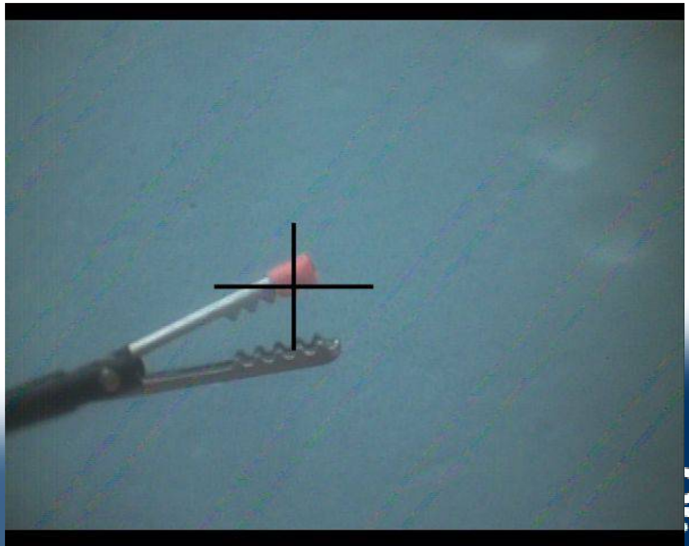
Picture in Picture :

Rotation :

Distortion Correction :

Tracking Instruments using Color Markers

- Place colored marker on instrument
- Convert RGB to HSV space
- Hue value of a pixel is much less susceptible to lighting changes
- Record hue value of marker to be tracked
- Search entire image for hue values within epsilon range
- Centroid of matched pixels gives position of tracker in the image
- If target is detected, localize search to a smaller neighborhood
- Tracking performed in real-time at 25 fps



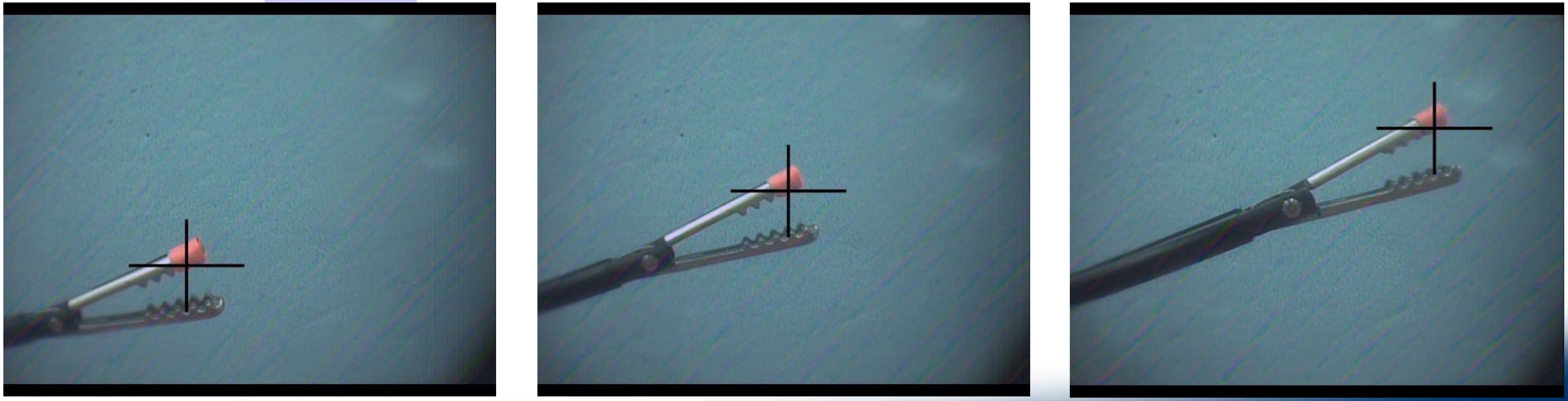
Visual Servoing

- Allows shared autonomy with surgeon
- The feedback from the tracker can be used to drive motors to keep the tool in the center of the image
- PD controller used
- (E_x , E_y): offset error of tracker from center of image

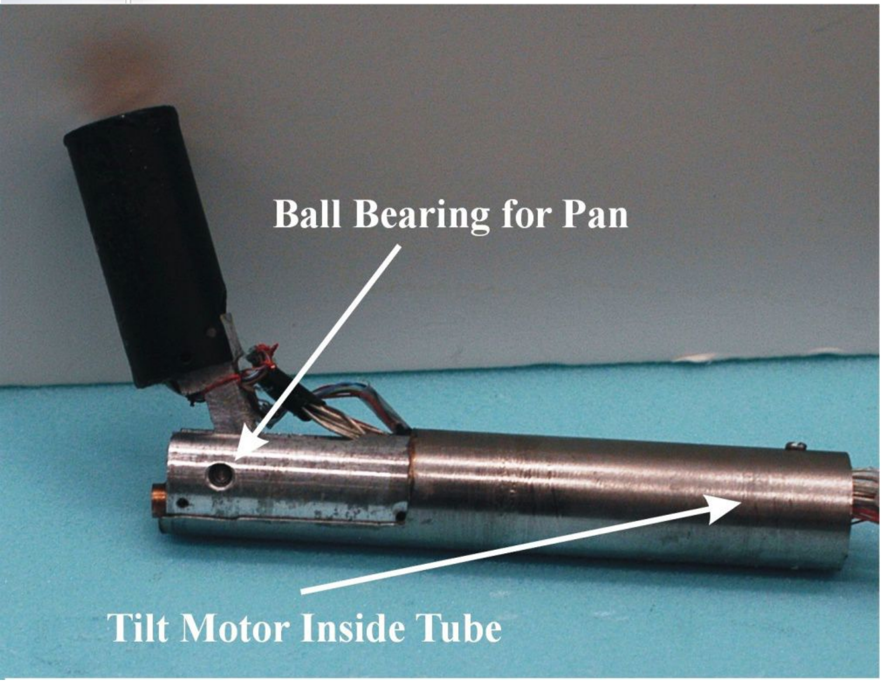
$$\text{Pan speed} \propto (\alpha_x * E_x) - (\beta_x * dE_x/dt)$$

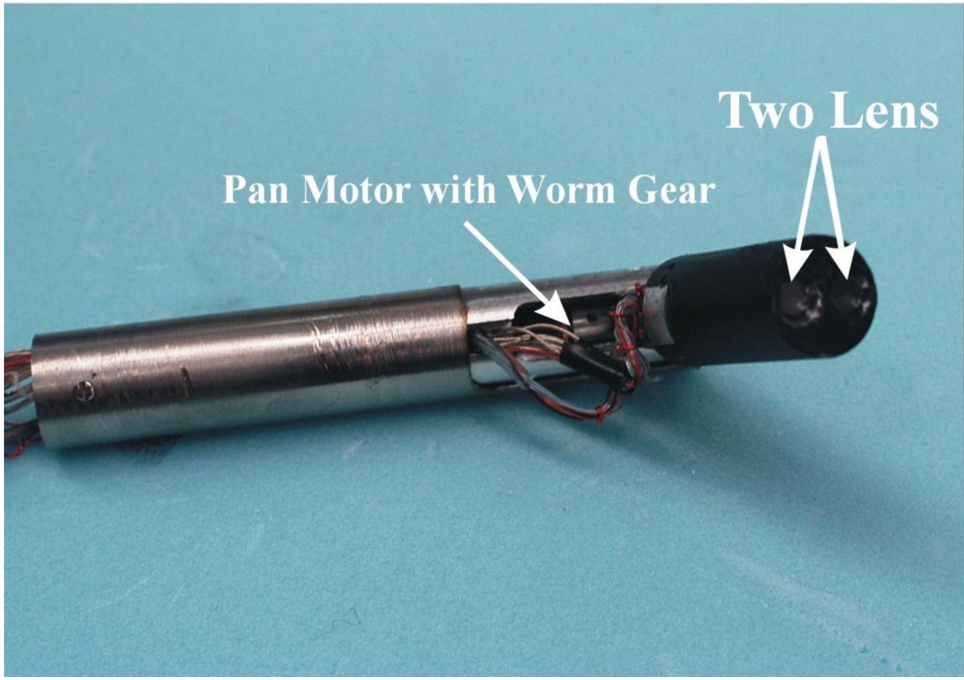
$$\text{Tilt speed} \propto (\alpha_y * E_y) - (\beta_y * dE_y/dt)$$

- Video



Device III: Stereo Imaging*

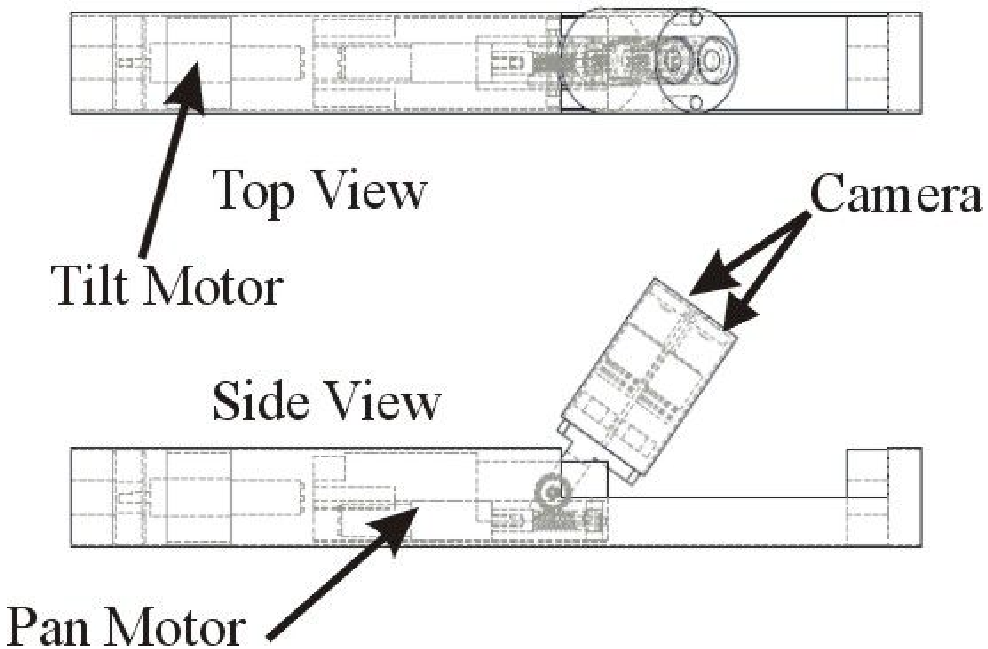


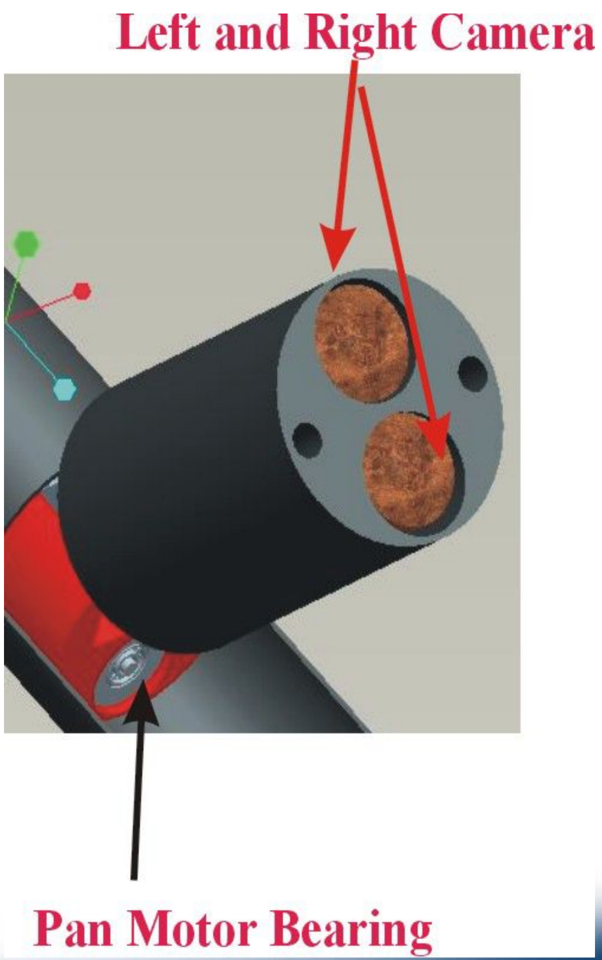


- A stereo imaging device with similar mechanical design.
- 15 mm in diameter and 120 mm in length.
- 6.5mm Inter-Pupillary Distance (IPD)

*T. Hu, P. Allen,, T. Nadkarni, N. Hogle, D. Fowler, *Insertable Stereoscopic 3D Surgical Imaging Device*, IEEE BIOROB 2008

Stereo Camera



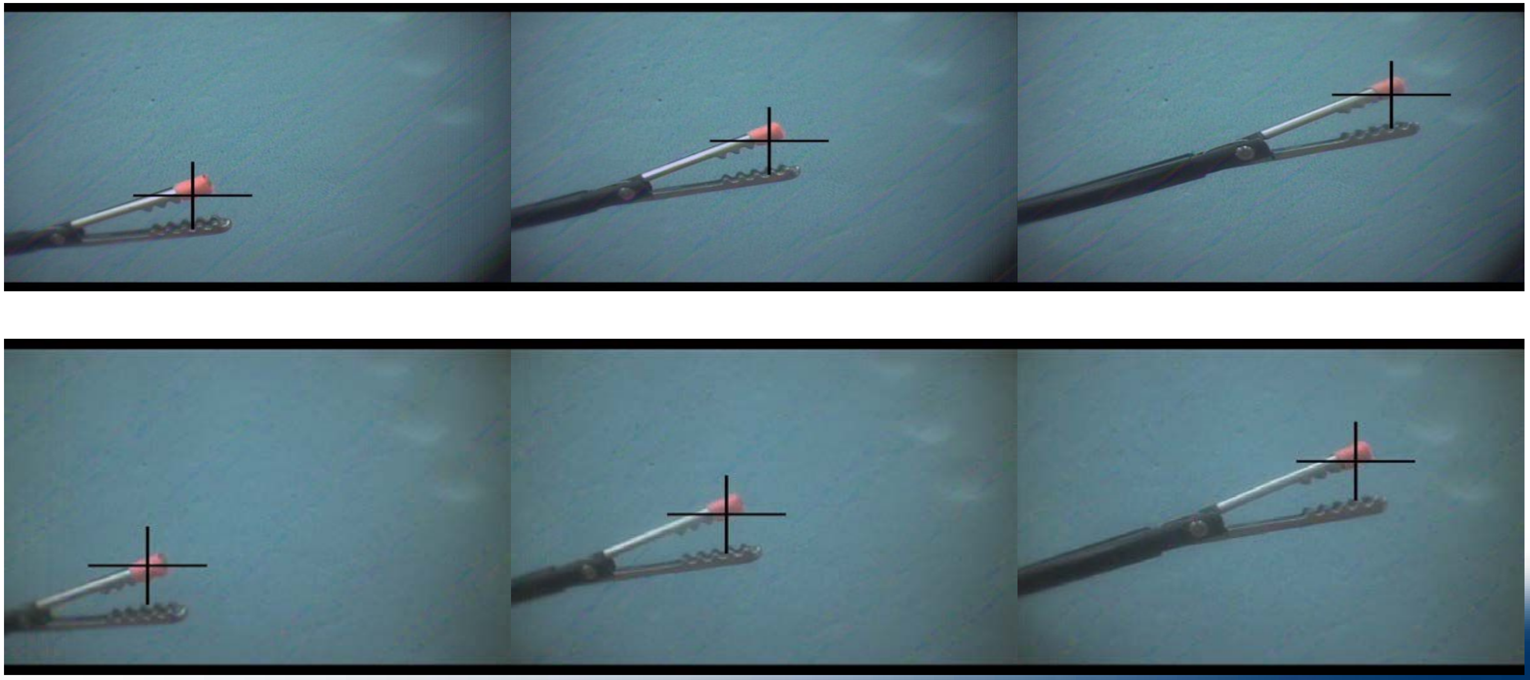


Visual Servoing with Stereo

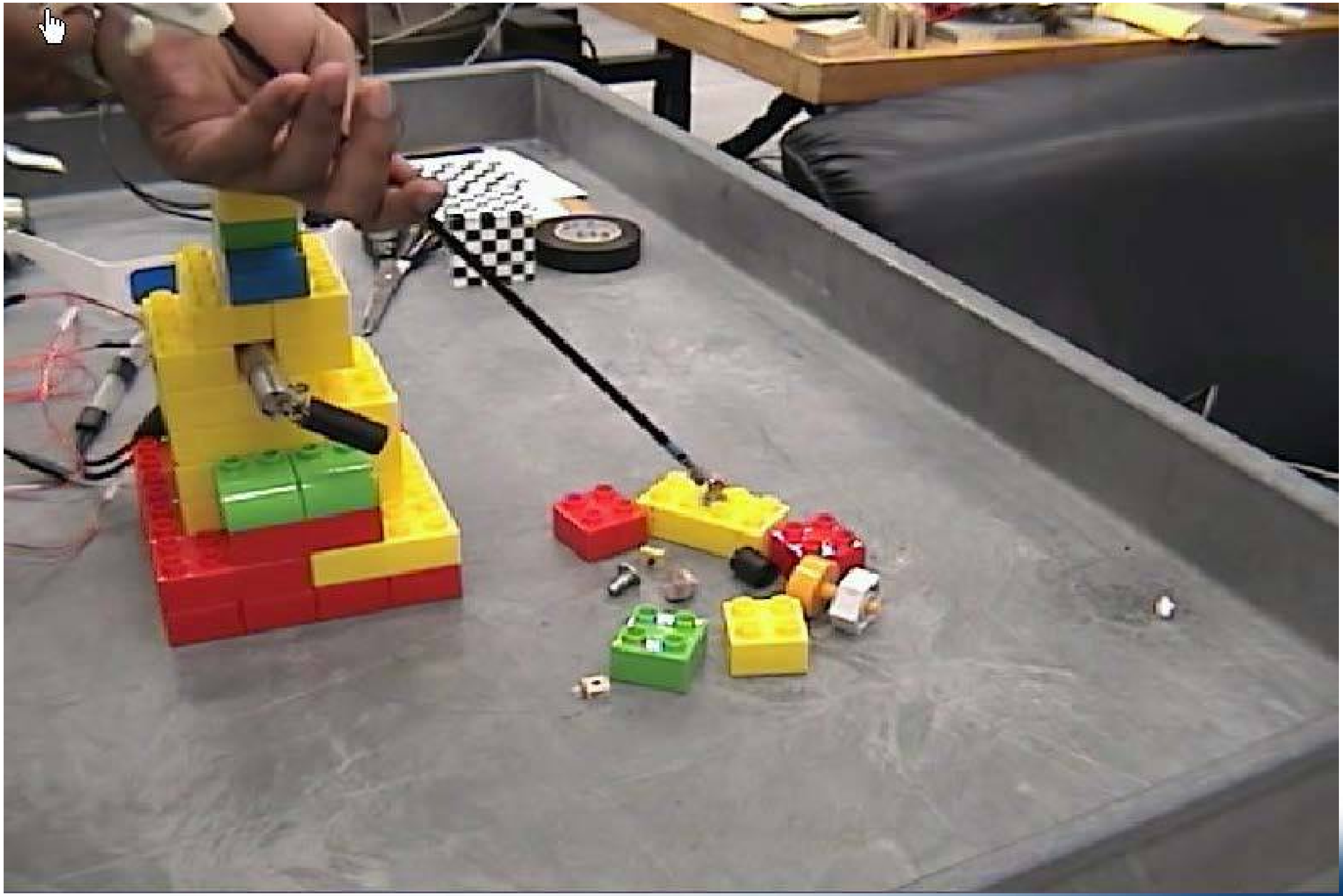
- When using stereo cameras the pixel disparity E_p between stereo images is used to damp the motors

  Speed Damping $\propto$ (γ * E_p)

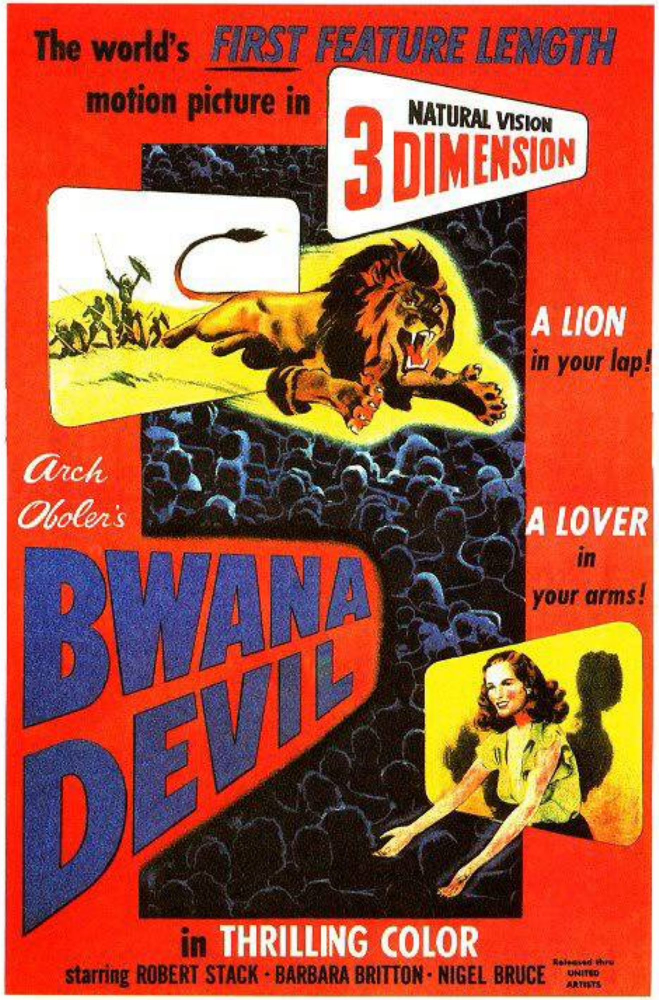
- Damping is applied to both Pan and Tilt motors
- Prevents the motors from oscillating when instrument is too close to camera



Device III: Stereo Imaging



Caution: 3D viewing ahead !

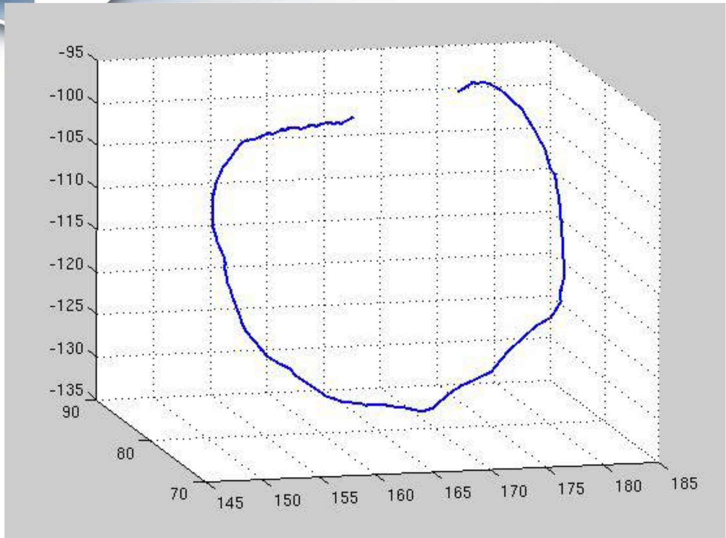


Video 1 Video 2

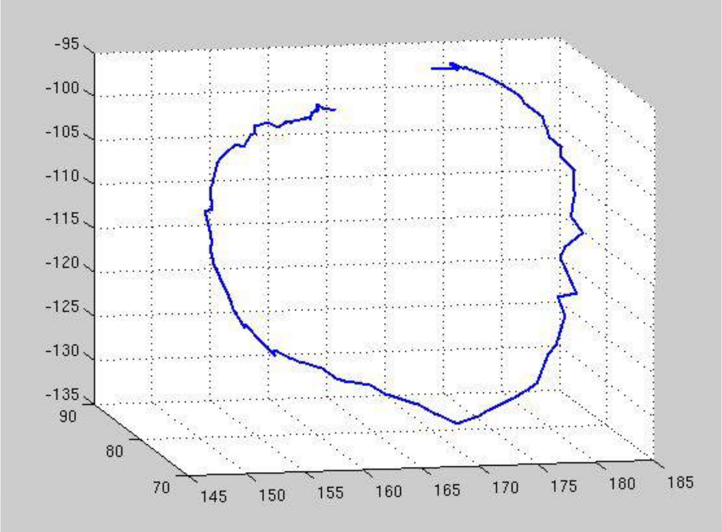
Device III (Stereo) Timings

Procedure	Device	Time (min)
Running Bowel	Laparoscope	5:35
Running Bowel	Robot	3:14
Appendectomy	Laparoscope	1:57
Appendectomy	Robot	1:38
Suturing	Laparoscope	4:30
Suturing	Robot	2:12
Nephrectomy	Robot	9:59

Trajectory Reconstruction



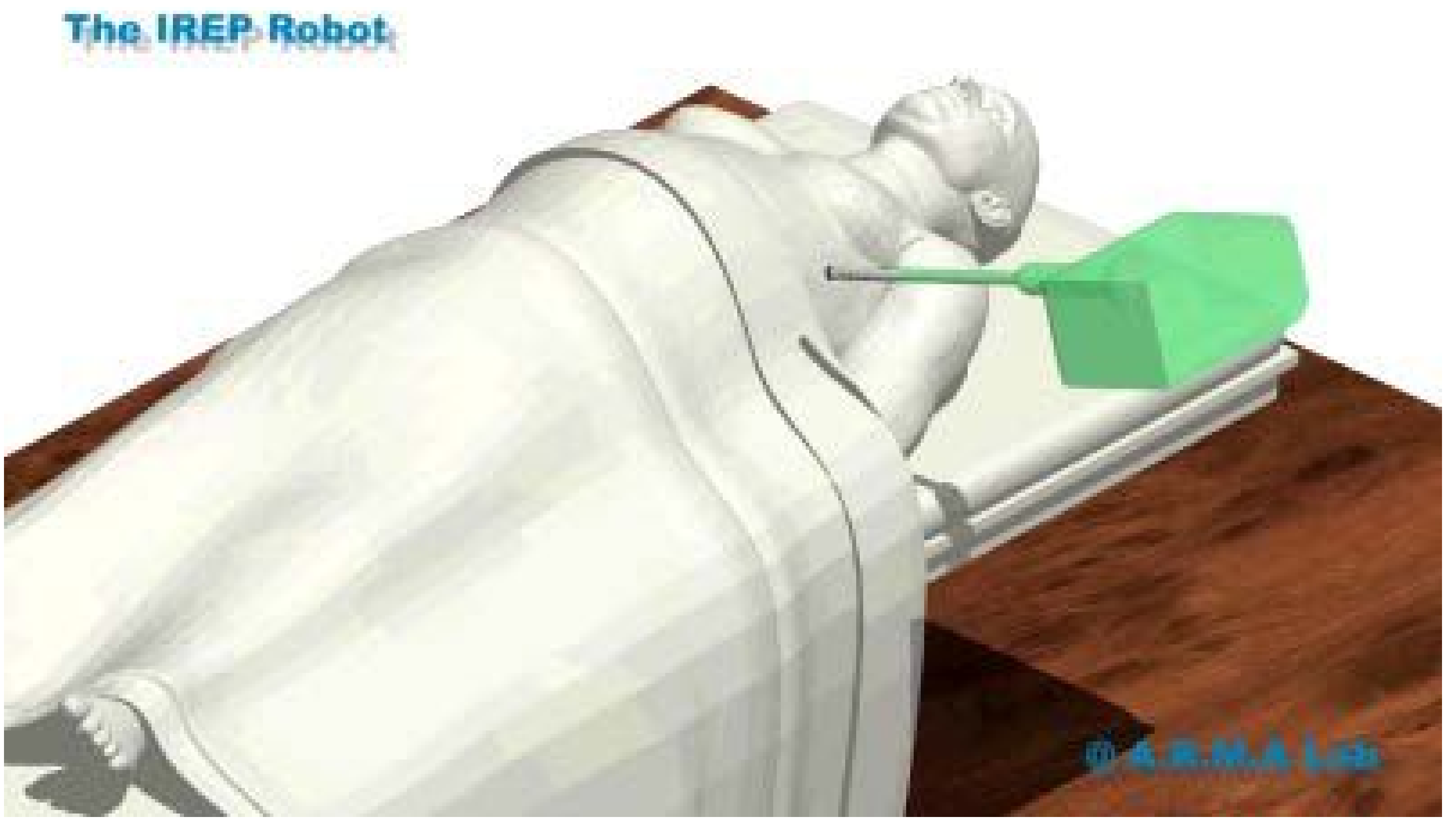
FOB Sensed Trajectory



Stereo Reconstruction of Trajectory

- We traced a trajectory in 3D space using the FoB sensor
- At the same time the sensor was being tracked by our stereo cameras
- The tracking results were used to predict the 3D position of the sensor
- Using this data we plotted the trajectory of the sensor
- average reprojection error ~3mm

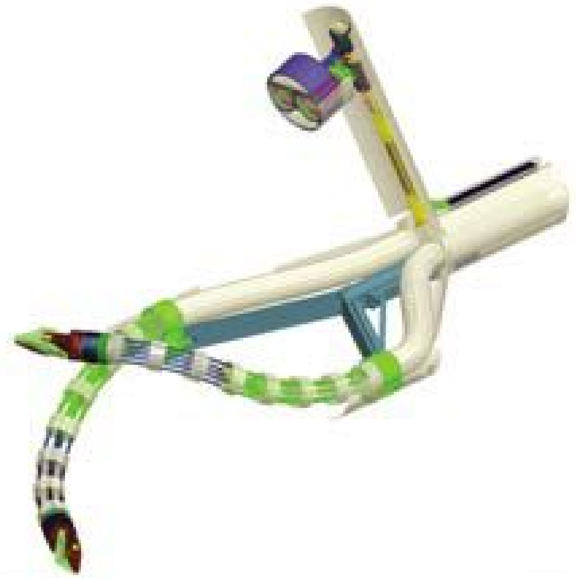
Insertable Robotic Effector Platform



K. Xu, R. Goldman, J. Ding, P. Allen, D. Fowler and N. Simaan, System Design of an Insertable Robotic Effector Platform for Single Port Access (SPA) Surgery, IROS 2009

Vision for In-Vivo Surgical Platforms

- IREP Platform integrates vision and tooling: Cameras, Graspers, Dissectors, Scissors, Energy sources
- Vision needed for:
  - Instrument tracking
  - Kinematic control
  - 3D measurement/reconstruction
- Vision system is key part of HCI
- Surgeon is focused on the task, not controlling the camera images



In-Vivo Tool Tracking*

- Tracking in in-vivo surgical environments:
  - Very dynamic environment
  - Rapid/Erratic movements
  - Significant pose changes
  - Deformable surfaces
  - Inconsistent lighting, liquid occlusion
  - Small 3D movements = large 2D non-linear pixel displacements
  - Frequent broken tracks, lost targets
- What features are best for tracking?
  - Color? Texture? Geometry?
- Markers or markerless?

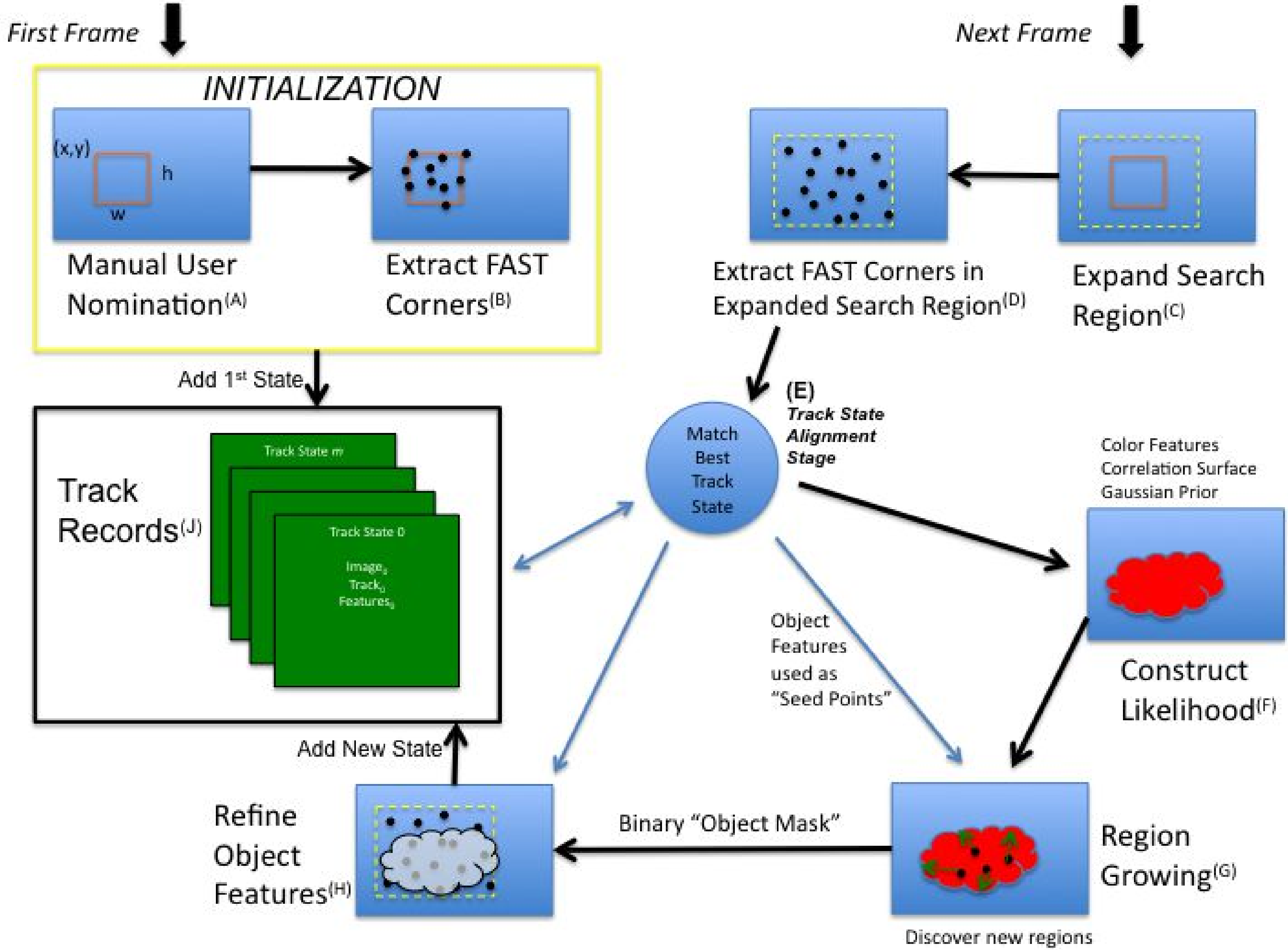
*Austin Reiter and Peter Allen, *An Online Approach To In-Vivo Tracking Using Synergistic Features*, IROS 2010

Previous Work: Tool Tracking

- Wei, Arbter, Hirzinger 1997; Groeger, Arbter, Hirzinger 2008 (German Aerospace) – *Custom Color Markers*
- Doignon, Nageotte, de Mathelin 2004; Doignon, Graebling, de Mathelin 2005; Doignon, Nageotte, de Mathelin 2006 (Univ of Strasbourg) – *Use color without markers*
- Krupa, et. al. 2003 – *Laser Pointing Assistive Device*
- Voros, Long, Cinquin 2007 – *Prior geometry for confinement*
- Pezzementi, Voros, Hager 2009 – *Offline learning of multiple features*
- Mountney, Yang 2008 – *Online feature adaptation for tissue tracking*

Our Method: Feature Flow

- It's **essential** to *learn new features online*
- Over time, features change as tool moves, lighting changes, occlusions occur – feature set has to evolve and adapt
- Need to automatically discover new parts of your object to track as it moves and turns
- Key Ideas:
  - Single feature is not enough for robust tracking in-vivo
  - Synergy: Features working together are better than features working alone
  - Learning features online is important, when lacking any prior information - even more than offline training
  - Use likelihood/probability maps to define most promising regions for matching



First Frame
Next Frame
INITIALIZATION
(x,y)
h
w
Manual User Nomination(A)
Extract FAST Corners(B)
Extract FAST Corners in Expanded Search Region(D)
Expand Search Region(C)
Add 1st State
Track Records(J)
Track State m
Track State 0
Image
Track
Features
(E)
Track State Alignment Stage
Match Best Track State
Color Features
Correlation Surface
Gaussian Prior
Object Features used as "Seed Points"
Construct Likelihood(F)
Add New State
Refine Object Features(H)
Binary "Object Mask"
Region Growing(G)
Discover new regions

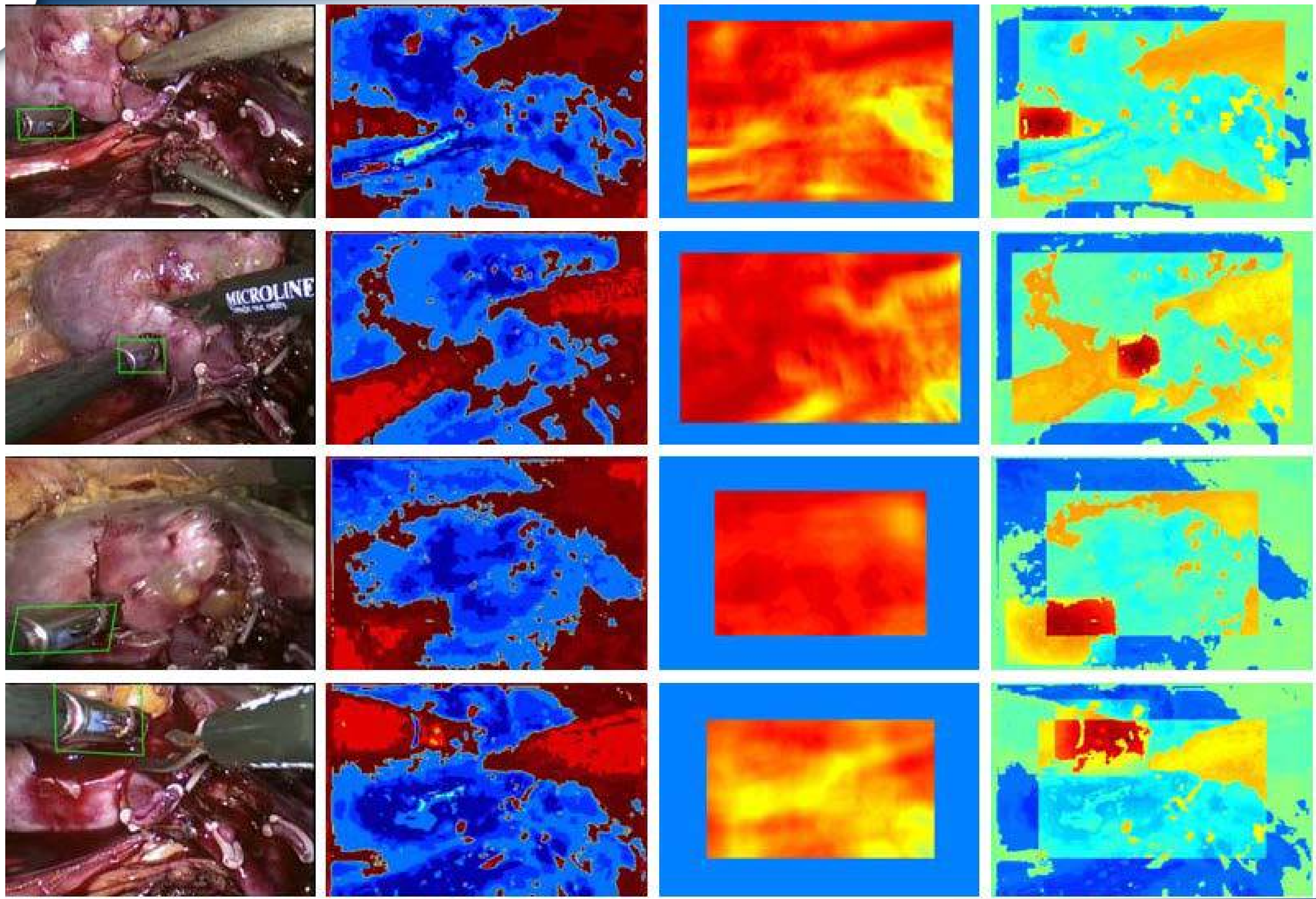
Algorithm

- Initial manual nomination (x,y,width,height)
  - Extract corners and add first entry to database
- On each successive frame:
  - Expand search region from previous location (don't want to process entire frame)
  - Extract corners
  - Feature matching – normalized cross correlation (NCC) of feature patches
  - Feature alignment via M-SAC (like RANSAC, but better) – affine model
  - Evaluate alignment quality – warp proposed track state to current frame and correlate to get similarity measure
  - Choose *best matching* track state based on alignment metric and add to database
- Note that this part is completely parallelizable

Combining Features/Likelihoods

- Combine multiple features into a composite likelihood map
- Defines which pixels are likely to be our object
- Use a weighted sum of individual probability maps
- We tested three features:
  - A Gaussian prior, representing where the object alignment estimates the object to be. This will bias us to correct general location.
  - Normalized Cross Correlation(NCC) surface using a **warped** track frame. Addresses severe pose and deformation changes
  - Color features: linear combinations of RGB color space, learned from the current environment (Collins et. al. 2005)
- Other features possibility: Region Covariance, Histogram of Oriented Gradients (HOG) or SIFT

Columbia Robotics



Track Frame | Color feature | Correlation | Combined

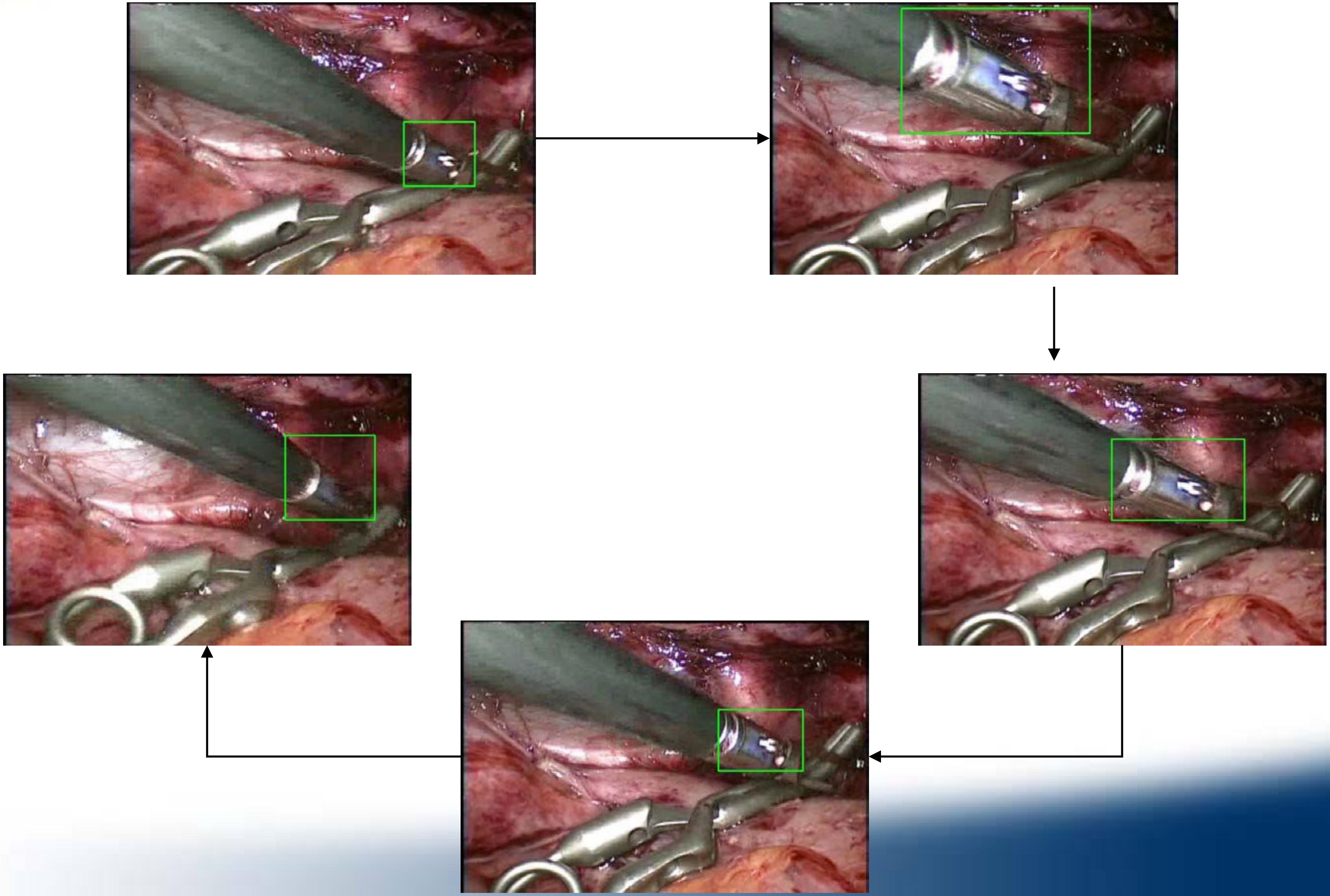
Feature Flow Results: In-Vivo

Improvements with Different Feature sets

Features	Track Time
No Likelihood Discovery	1:01
NCC (No Warp)	1:16
Color (Alone)	1:44
NCC (With Warp)	1:55
Color and NCC (With Warp)	2:17

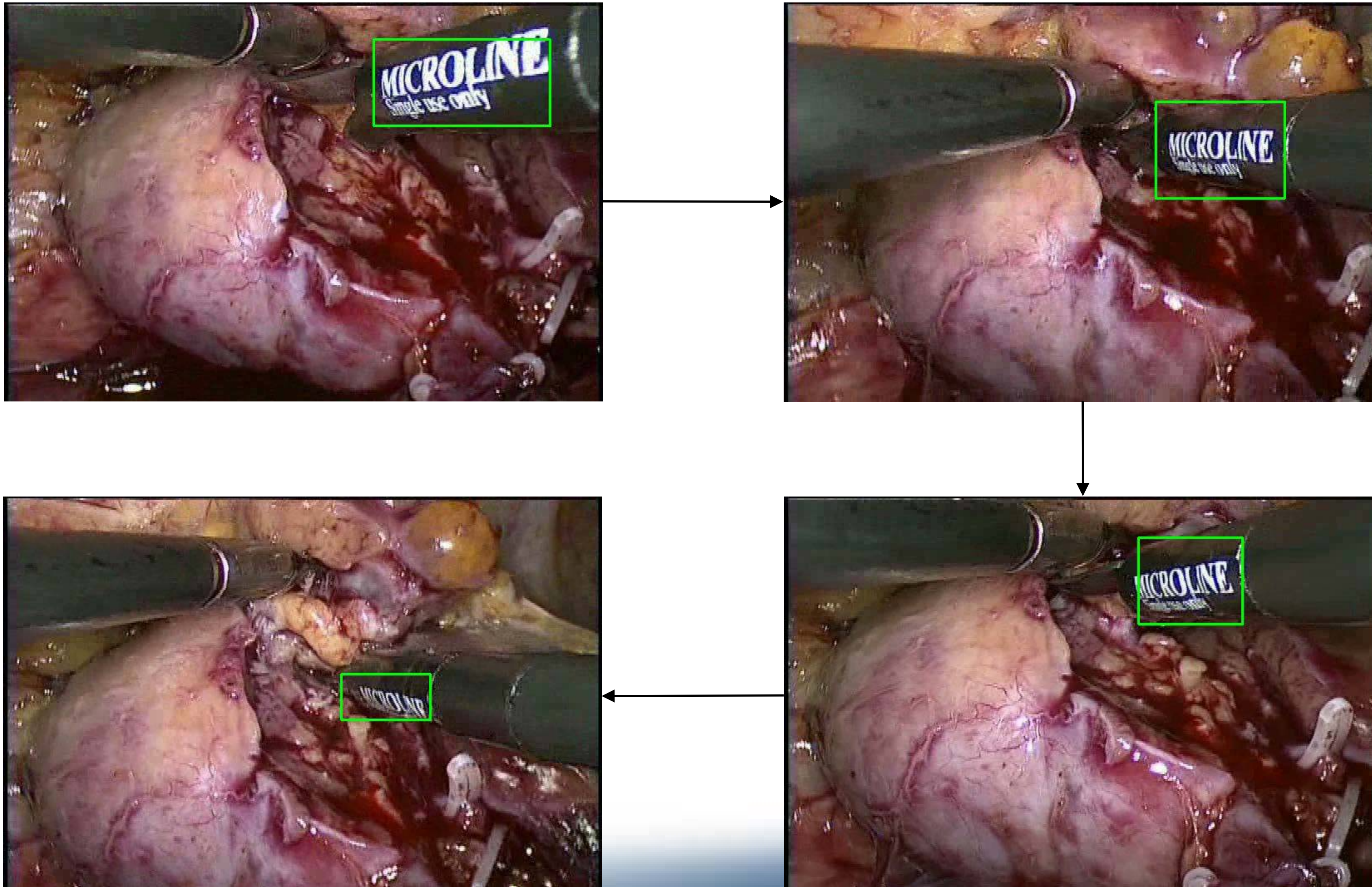
Real-time performance: 12fps

Size and Pose Changes

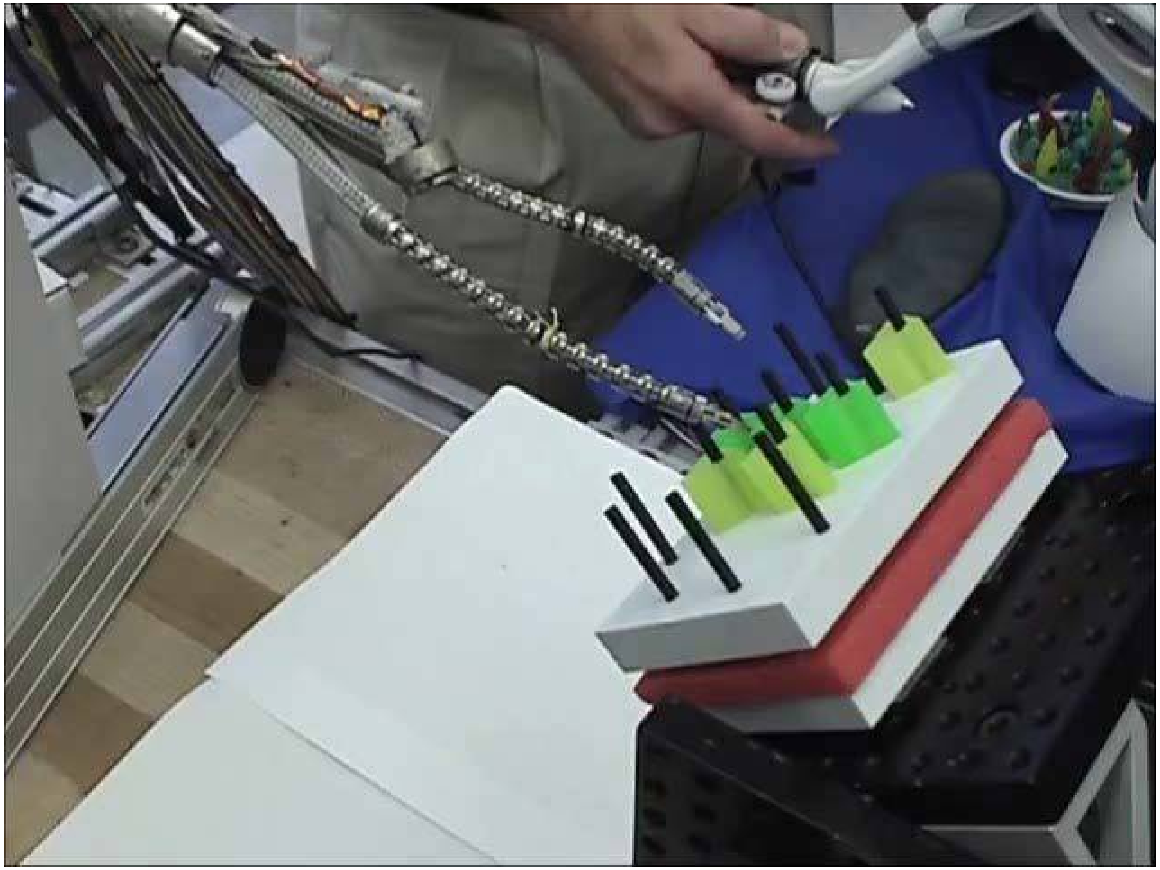


Columbia Robotics

A Different Tool – Text Label



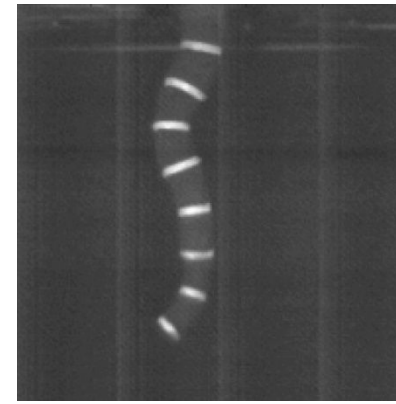
Continuum Robot Pose Estimation

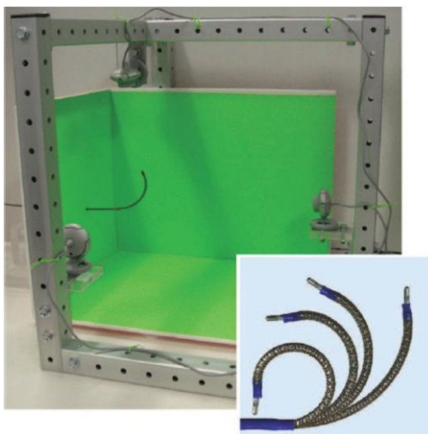


A Learning Algorithm for Visual Pose Estimation of Continuum Robots,
A. Reiter, R. Goldman, A. Bajo, K. Iliopoulos, N. Simaan, P. Allen, IROS 2011

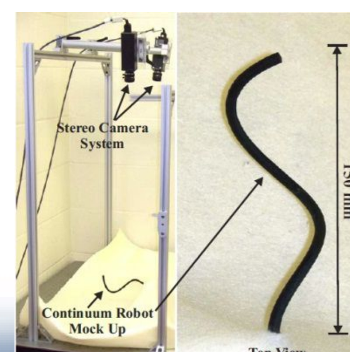
Previous Work

Hannan & Walker (2003)
- Extract vertebrae
- Fit successive circles
- Curvature comes from change in segment length due to bending



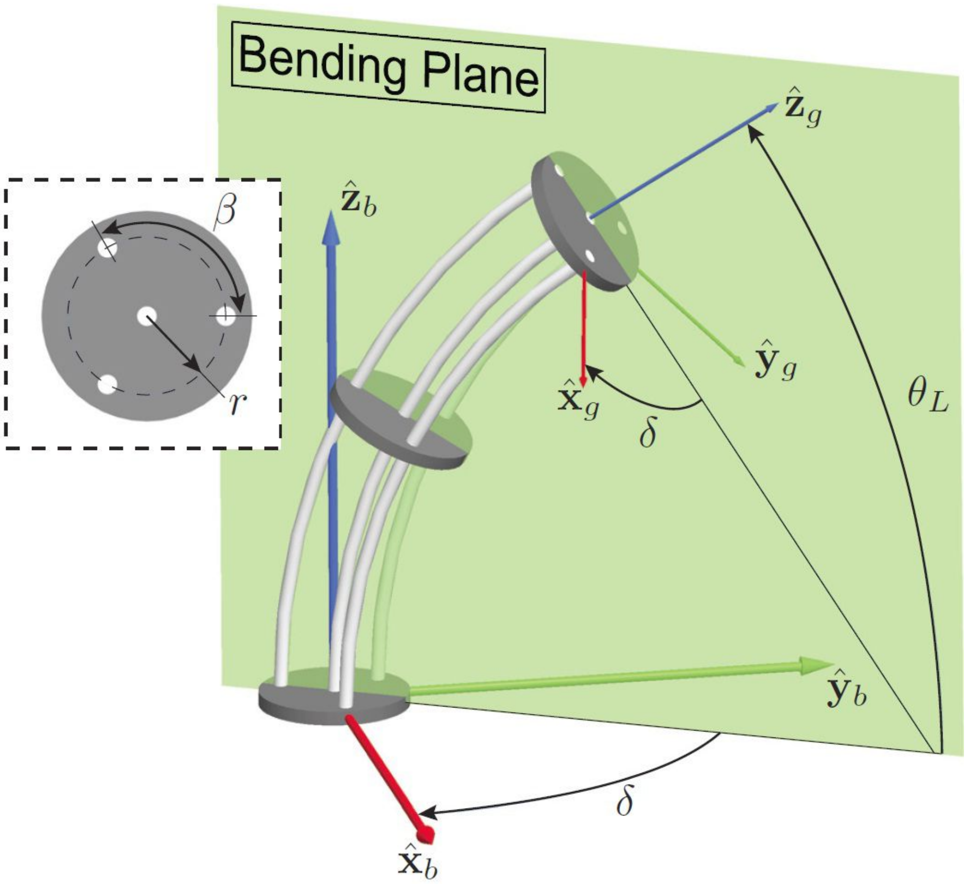


Camarillo, et. al. (2008)
- Voxel carving
- 3 orthogonal cameras
- High positional precision of flexible manipulator



Croom, et. al. (2010)
- Stereo SOMs
- No fiducials
- Detected shape of continuum robot

Continuum Segment M0del



$\psi = [\theta_L, \delta]^{\mathrm{T}}$

$\theta_0 = \pi/2$

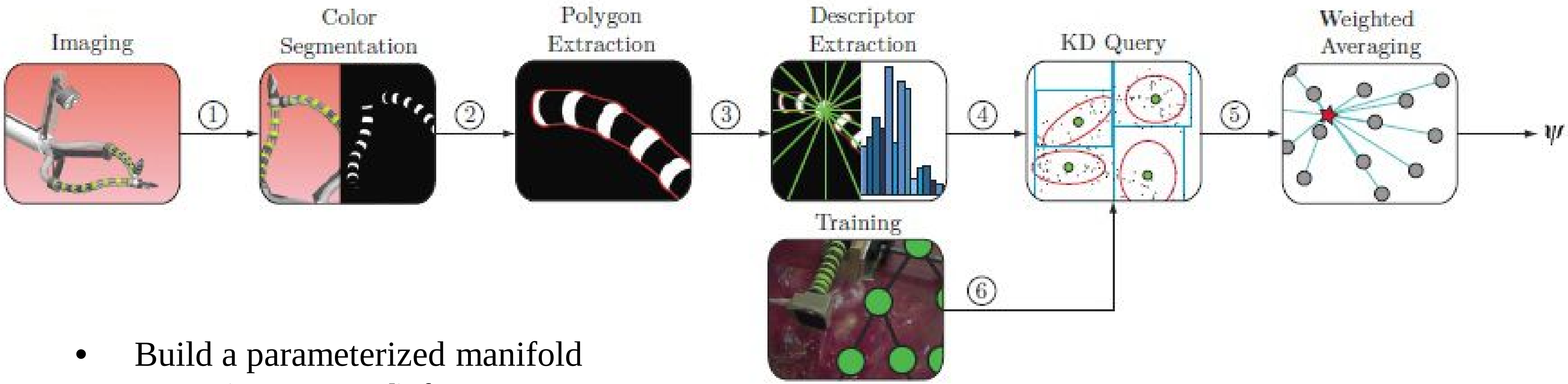
$$\theta_L = \theta_0 - \operatorname{atan2}\left(\sqrt{R_{13}^2 + R_{23}^2}, R_{33}\right)$$

$$\delta = -\operatorname{atan2}\left(R_{23}, R_{13}\right)$$

$\mathbf{R} = \mathbf{R}_z \mathbf{R}_y \mathbf{R}_z^{\mathrm{T}}$

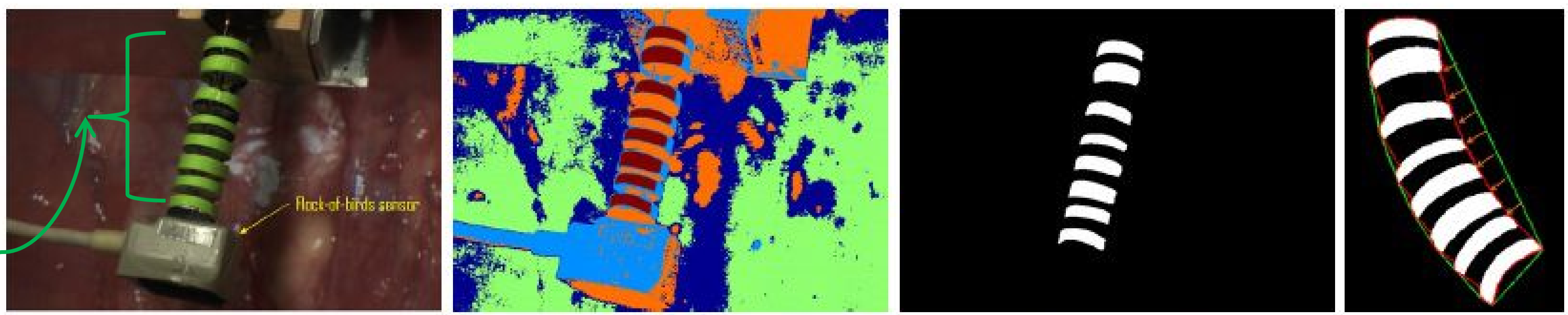
Orientation of the end disk

Visual Learning Overview



- Build a parameterized manifold using (compressed) feature descriptors
- Manifold is parameterized by robot DOFs
- Unknown configurations looked-up using weighted average of nearest neighbors in manifold

Image Segmentation



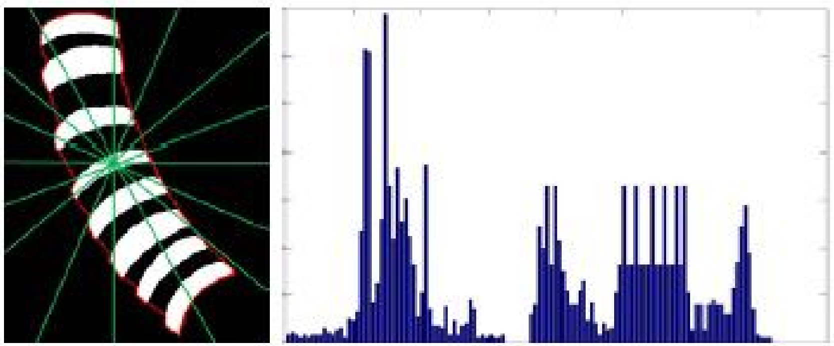
(a) Color Image (b) Cluster Labels (c) Segmented Image (d) Contour

K-Means Clustering in CIELAB-color space

Lime Green Fiducials (want to ***stick out*** in surgical imagery)

Feature Descriptor

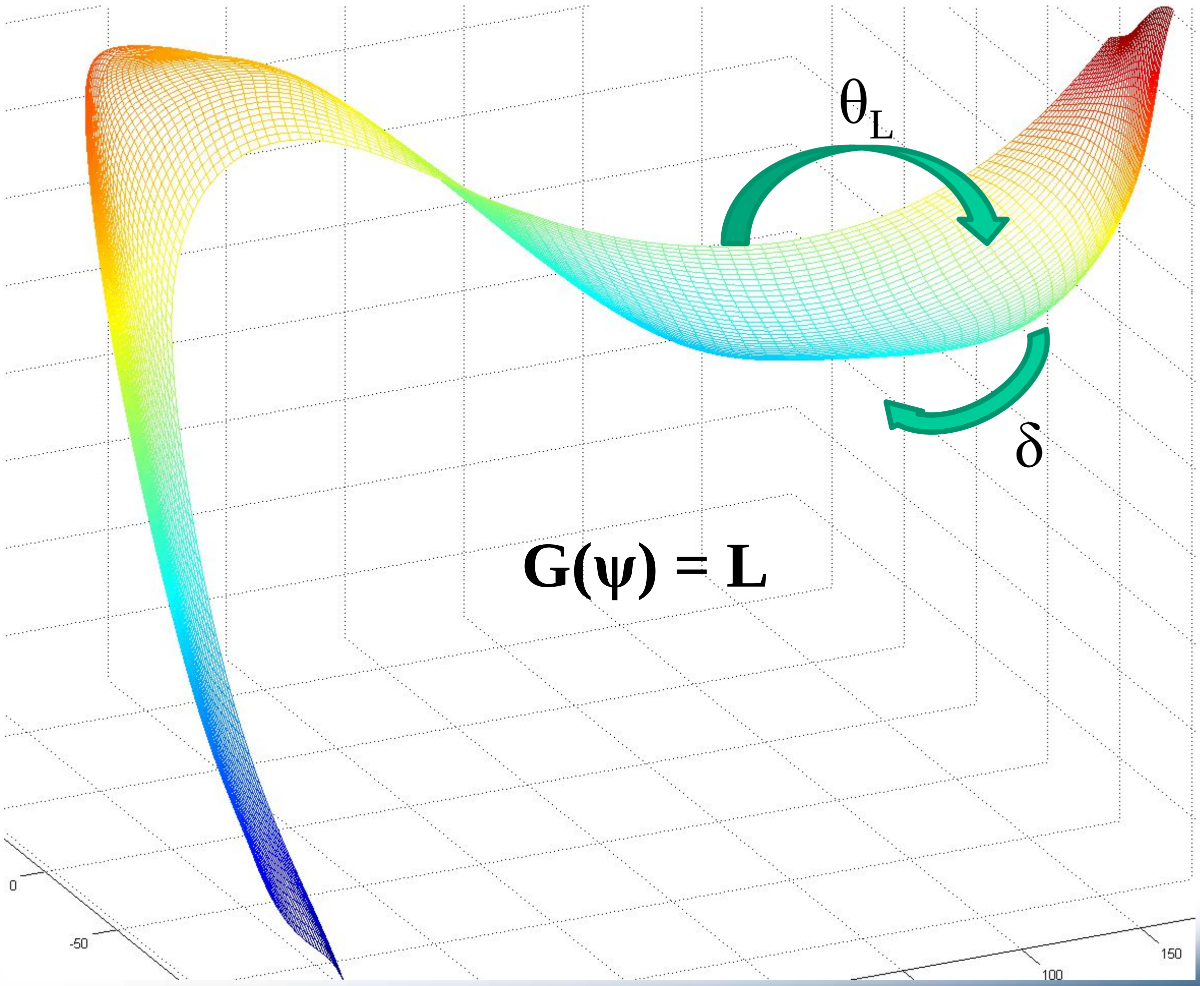
- Locate center of binary mask using median of marked pixels
- Compute angle of each point along red contour w.r.t. center
- Fill histogram
    - Experimented with different bin sizes: 72 (5°), **120 (3°)**, and 360 (1°)
- Extract separately from stereo images and combine as composite feature vector



Training Data

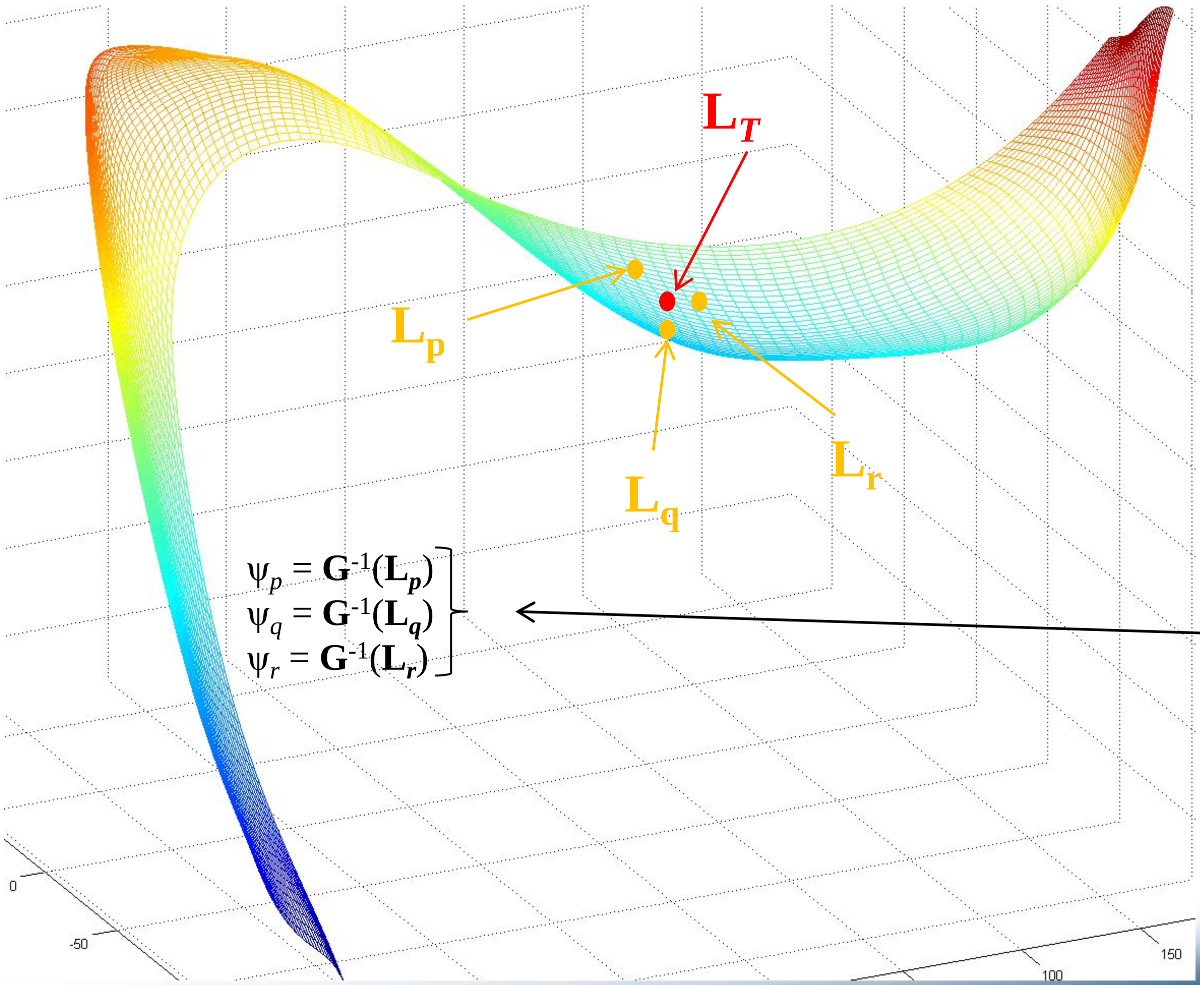
- Pre-processing
  - Reduce high-dimensional stereo composite feature using PCA, called *eigen-features*
  - Experiment with different variance recoveries (65%, 85%, **90%**, 95%) with different bin sizes
- Interpolate smooth manifold (spline) using compressed vectors
  - Parameterize on ψ-angles (2 DOFs), which yields a surface
    - More DOFs would lead to a volume, etc...

Descriptor Manifold



- Surface of *eigen-features* (2 DOFs)
- 2x120-bins reduced to 16 dimensions
- Only first 3 dimensions of eigen-projections shown, for convenience

Configuration Query

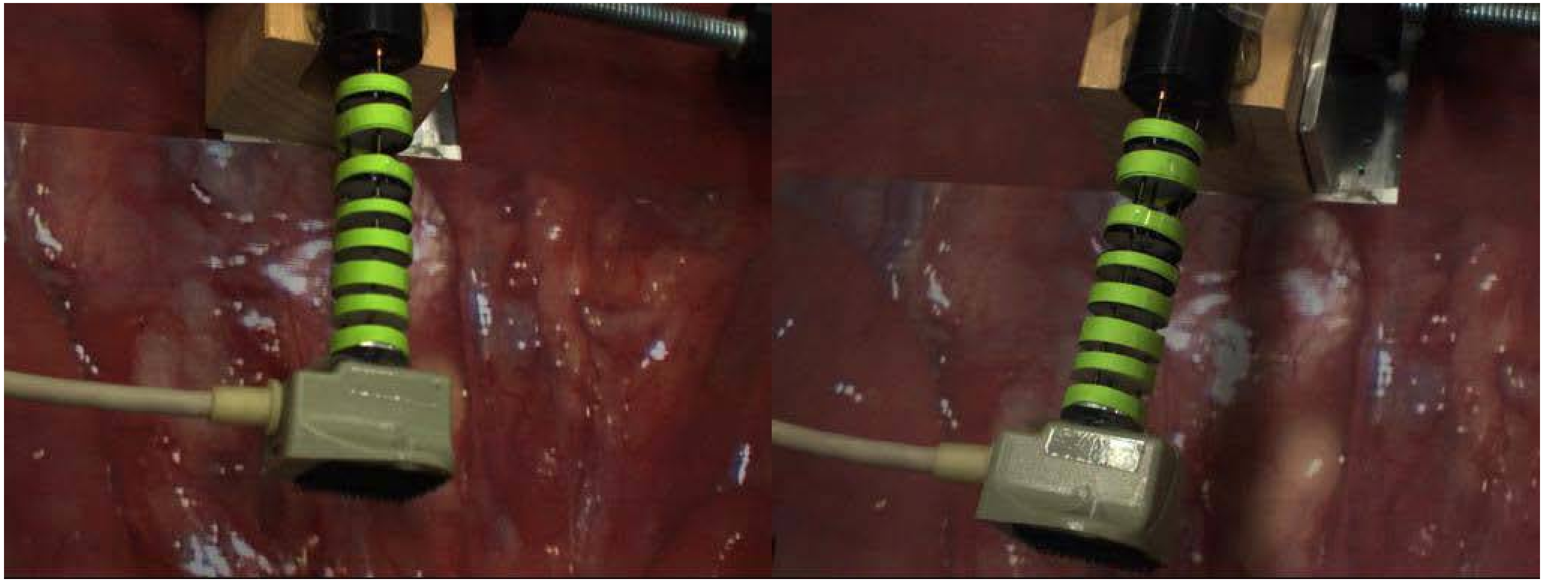


$\mathbf{L}_T$: query descriptor (unknown configuration)

$\mathbf{L_p}$, $\mathbf{L_q}$, $\mathbf{L_r}$: nearest-neighbor matches from manifold (created by training data)

- Recover configuration parameters which created those interpolated values
- Weighted average of nearest neighbors based on distance on manifold

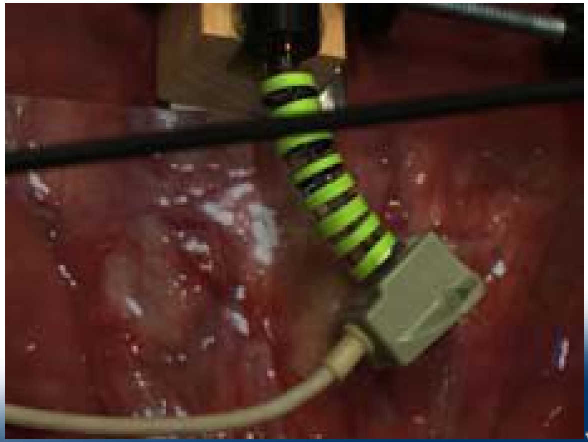
Experimental Setup



- Stereo camera system made up of 2 Point Grey Dragonfly2 cameras (1024x768, color)
- Manually-actuated single-segment snake arm
  - Lime green markers for segmentation
  - Flock-of-Birds at tip for pose ground truth
  - Background of printouts from laparoscopic procedure for more realistic setting

Accuracy

- Different combinations of histogram bin-size and variance recovery
  - 360, 120, and 72-bins (each *x2*)
  - 65%, 85%, 90%, 95%
  - Best tradeoff performance from 120-bin histograms (stereo feature $\in \mathbb{R}^{240}$) reduced to 90% variance $\in \mathbb{R}^{16}$
- Exp. 1: no occlusions, testing data **not** included in training data
  - ~ 800 frames
  - [$\varepsilon_\delta = 0.98°$, $\varepsilon_{\theta_L} = 1.28°$], 1.16° median error
- Exp. 2: occlusion from laparoscopic tool, testing data **not** included in training data
  - ~ 600 frames
  - [$\varepsilon_\delta = 1.04°$, $\varepsilon_{\theta_L} = 2.06°$], 1.46° median error



Summary

Columbia In-vivo Camera:

- Easier and more intuitive to use than a standard laparoscope.
- Insertion port available for tooling
- Joystick operation requires no specialized operator training.
- Pan/Tilt functions provide large imaging volume
- Time to perform procedures was better or equivalent
- Automatic Tracking and Visual Servoing assist surgeon
- 3D vision improves the visualization and depth perception
- Cost effective, perhaps single-use or modular replacement

IREP Robot:

- Integrated platform for imaging and effectors
- Vision system can be used to close the control loop of effectors

MarkerlessTool Tracking:

- Long term tracking with recovery
- Adaptive online feature learning
- Real-time performance

Pose Estimation:

- Learning-based method is accurate for recovering pose

Acknowledgements

- Austin Reiter
- Dennis Fowler
- Tie Hu
- Andrew Miller
- Tejas Nadkarni
- Nancy Hogle
- Nabil Simaan
- Kai Xu
- Roger Goldman
- Jienan Ding
- Andrea Bajo
- Konstantinos Iliopolous

This work was supported by NIH grants 1R21EB004999-01A1 and 5R21EB007779-02