Endoscopic Reconstruction with Robust Feature Matching

Plan for CIS II Course Project

Xiang Xiang

Mentors: Dr. Dan Mirota, Dr. Greg Hager and Dr. Russ Taylor

Computer Science, JHU

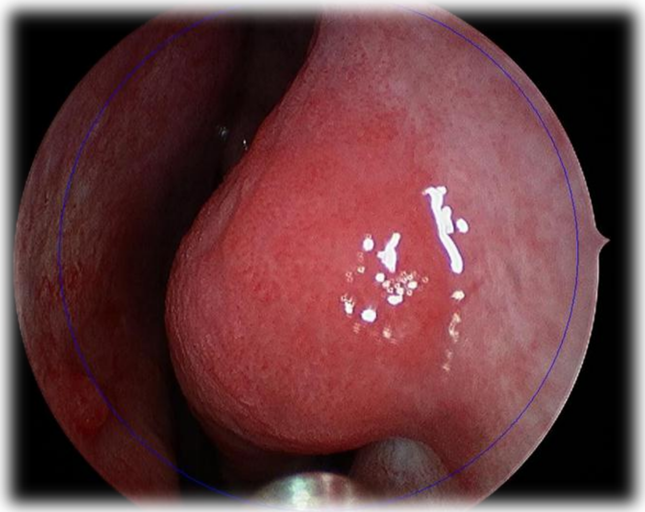
Feb. 28, 2013

Outline

- Background, Goal and Significance
- **Approach**
- Deliverables and Milestones

Background of endoscopic reconstruction

- > 200,000 functional endoscopic sinus surgeries (FESS).
- Surgical navigation systems to visualize critical structures that must not be disturbed during surgery.
- Sinuses contain structures that are smaller than a millimeter in size.
- Navigation can provide a qualitative sense of location.



http://www.youtube.com/watch?v=8CYSYH-8JjY

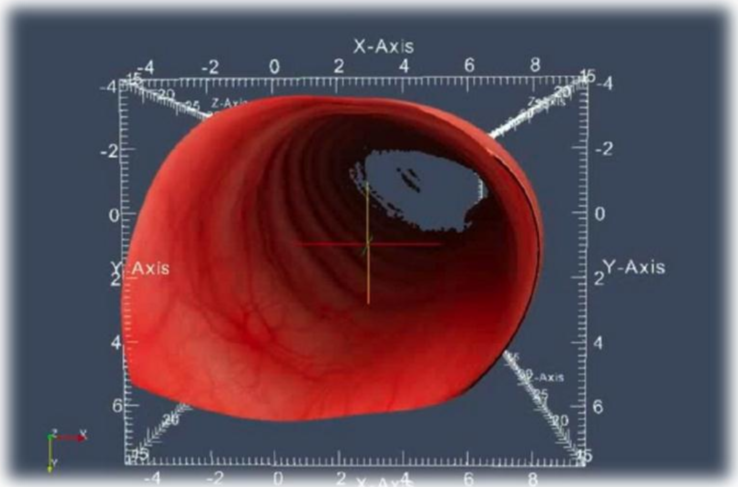
Background in 3D Vision

- Feature matching based 3D reconstruction is a standard technique in 3D Computer Vision.
- An natural extension is to reconstruct dynamic surfaces from videos.

http://www.youtube.com/watch?v=sz0UbHvEttI
http://www.youtube.com/watch?v=NiMlVgEu7mg

Goal of this course project

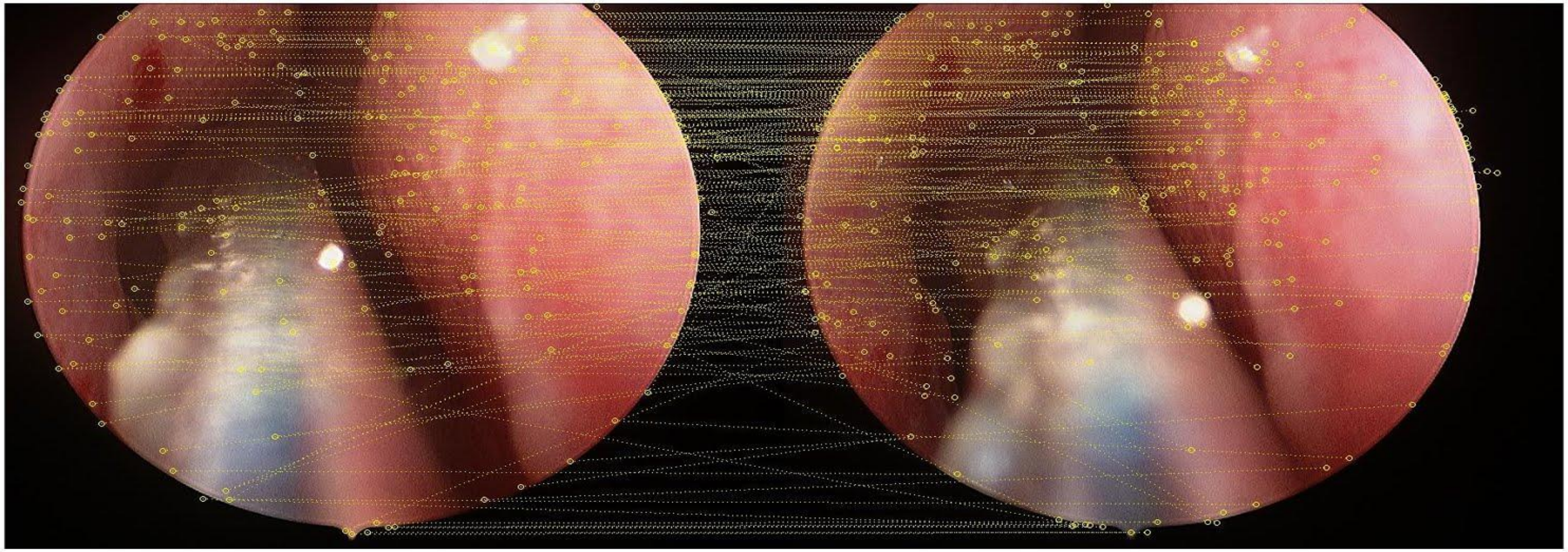
- To develop methods for surface reconstruction from endoscopic videos.
  - To build up the 3D endoscopic reconstruction pipeline.
  - To validate the pipeline's performance under a baseline design.
  - To test the pipeline's performance with improved components such as more robust feature matching.



A full 3D reconstruction of a pediatric airway from video imagery acquired with a tracked endoscope. [*Image from a NIH-funded project proposal with permission.*]

Significance of this course project

- Since the camera is moving and the surfaces are more or less deformable, feature matching is not always satisfactory.
- We will employ a state-of-the-art feature description and matching strategy called Hierarchical Multi-Affine (HMA) for endoscopic feature representation.



Approach: the pipeline

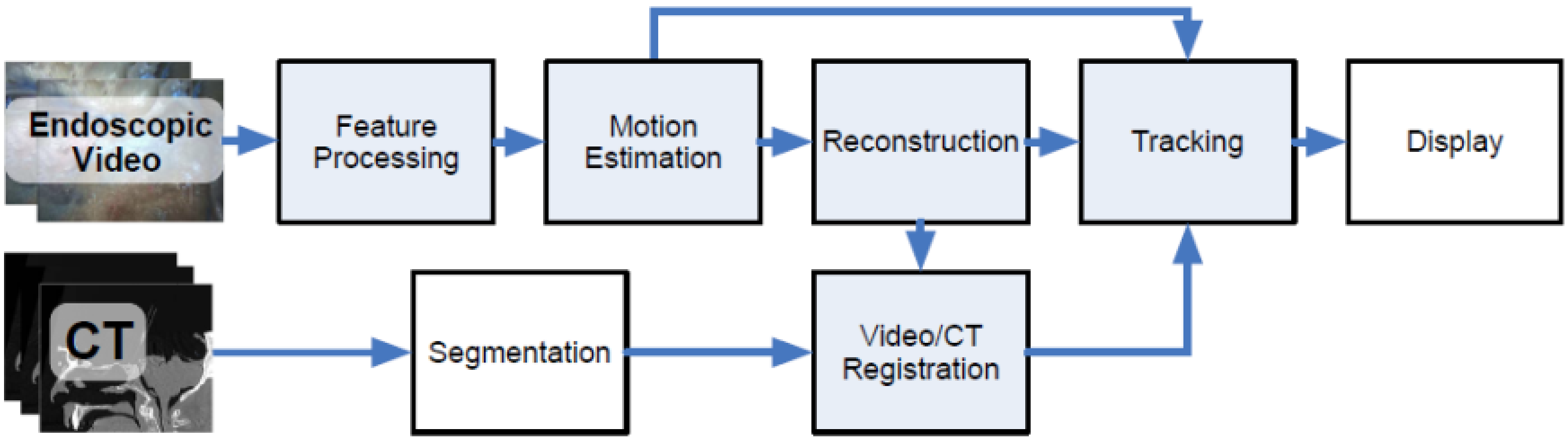


Figure from [*Mirota etal 2012*]

Multi-Affine: basic idea

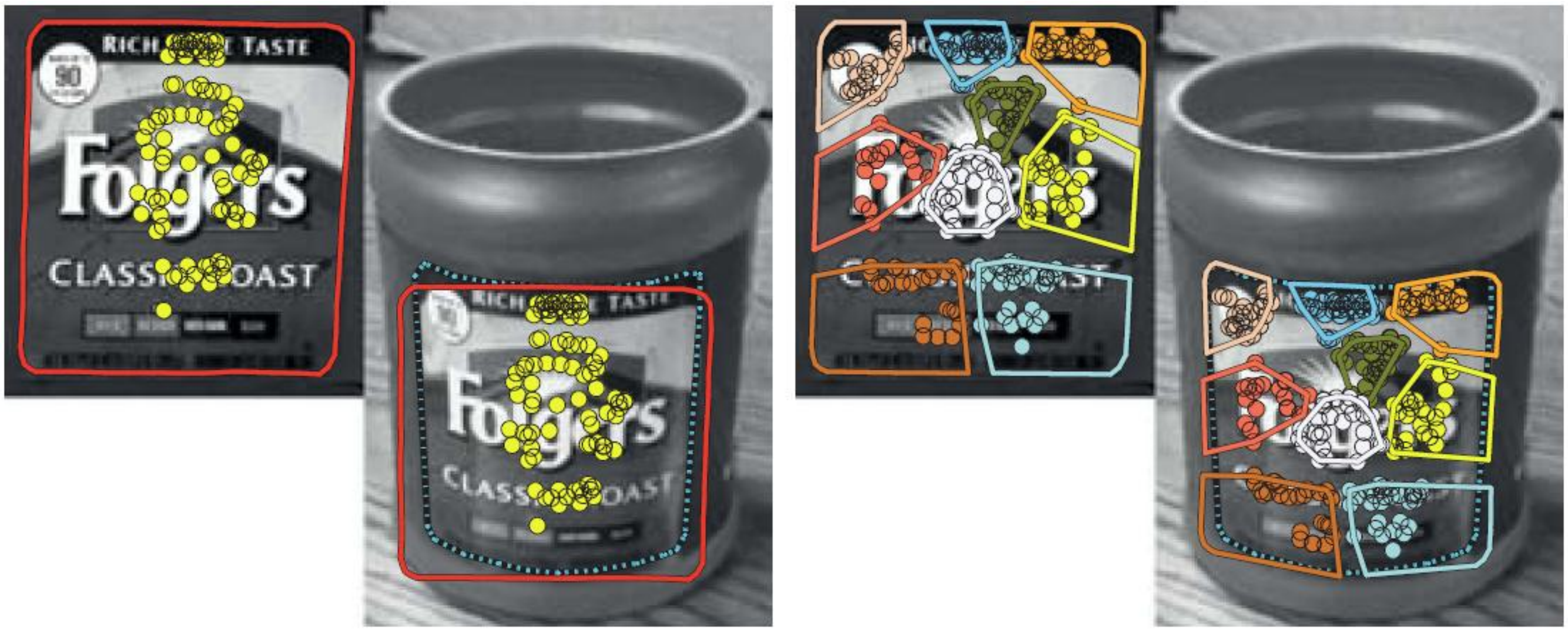


Figure from [Puerto-Souza *etal 2011*]

MA feature matching

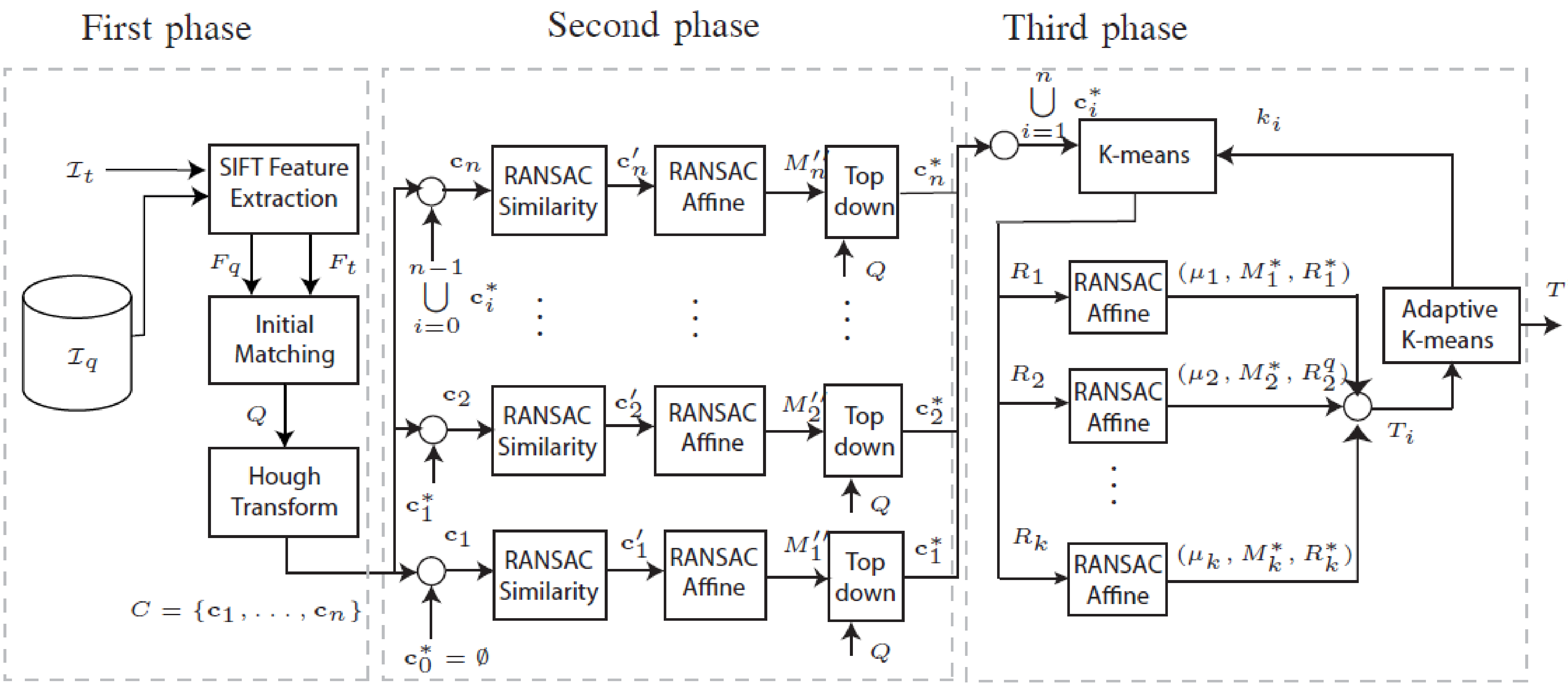


Figure from [Puerto-Souza *etal 2011*]

Hierarchical MA: basic idea

- quantizes matches according to their spatial position on the object's surface

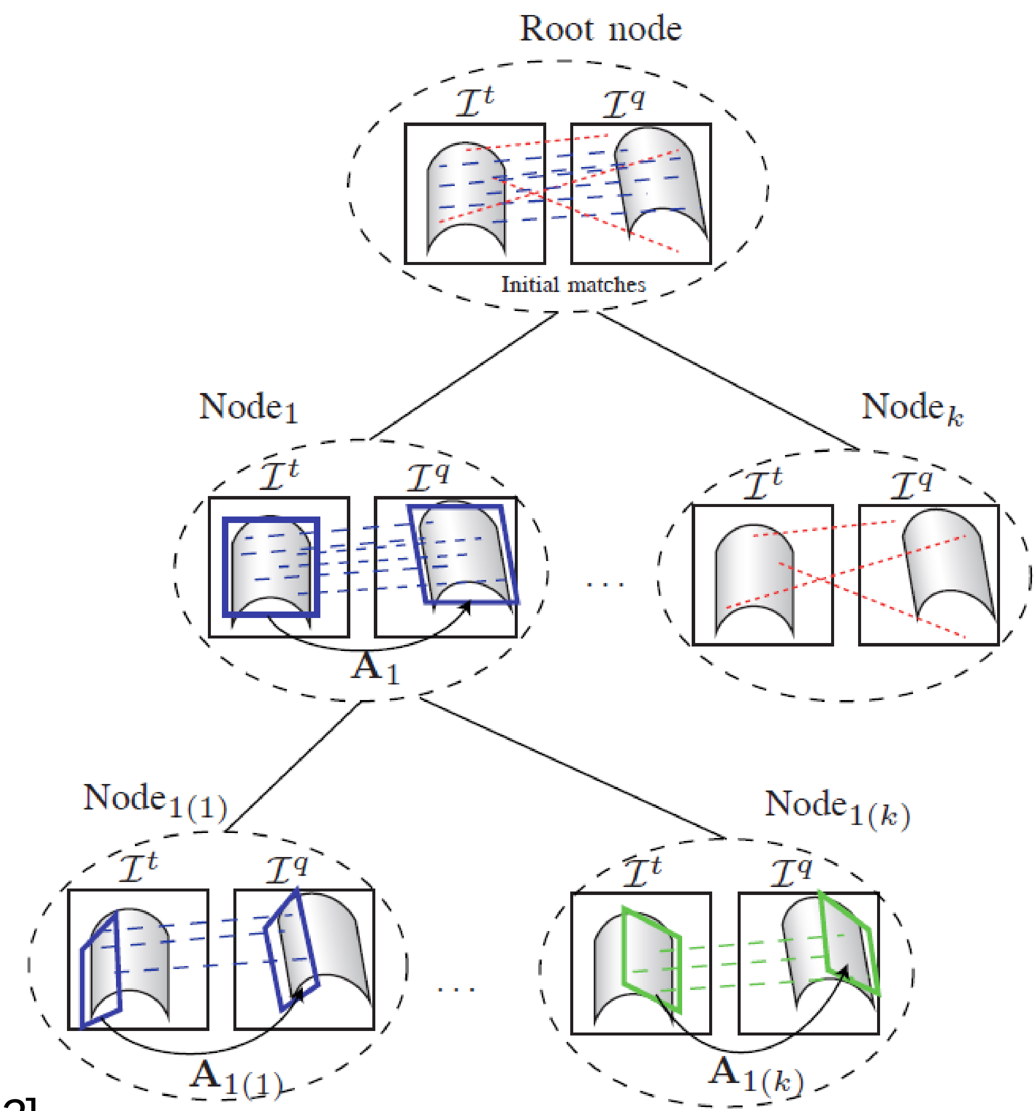


Figure from [Puerto-Souza *etal 2012*]

HMA: speed-up

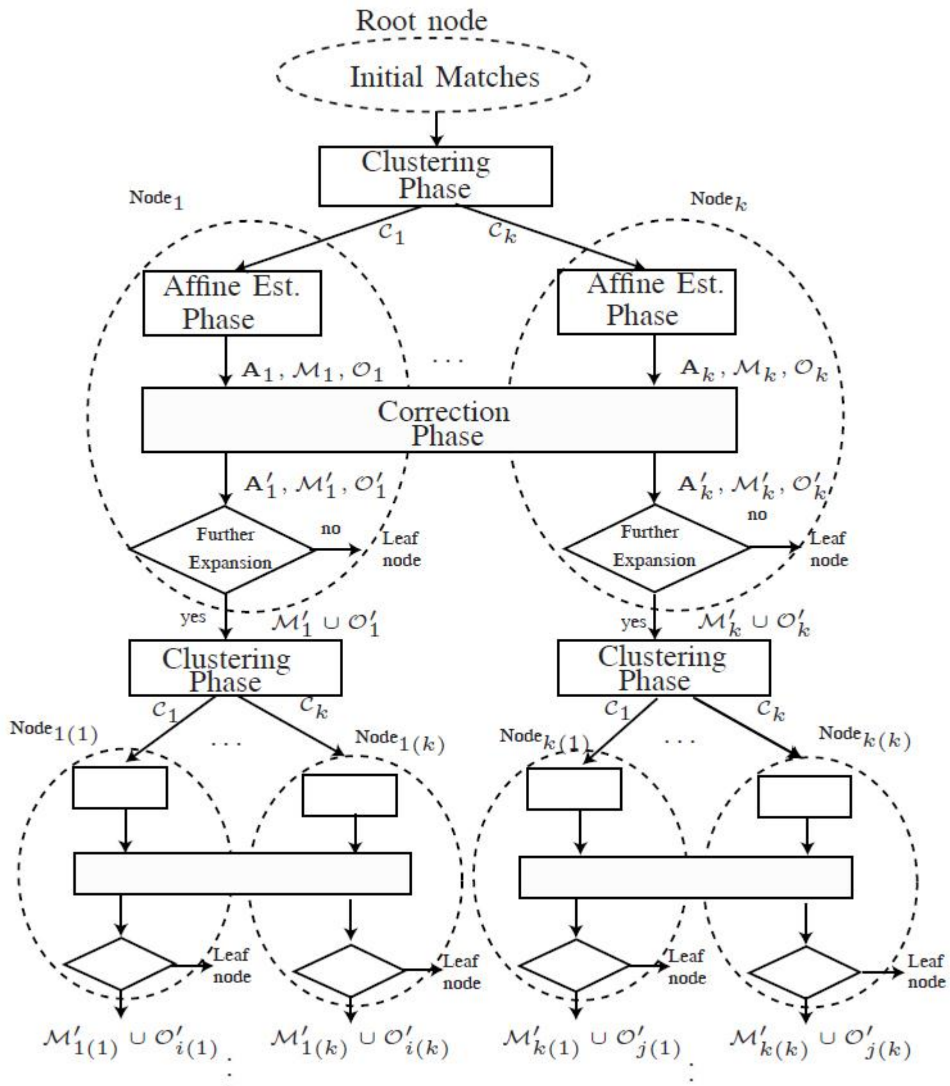
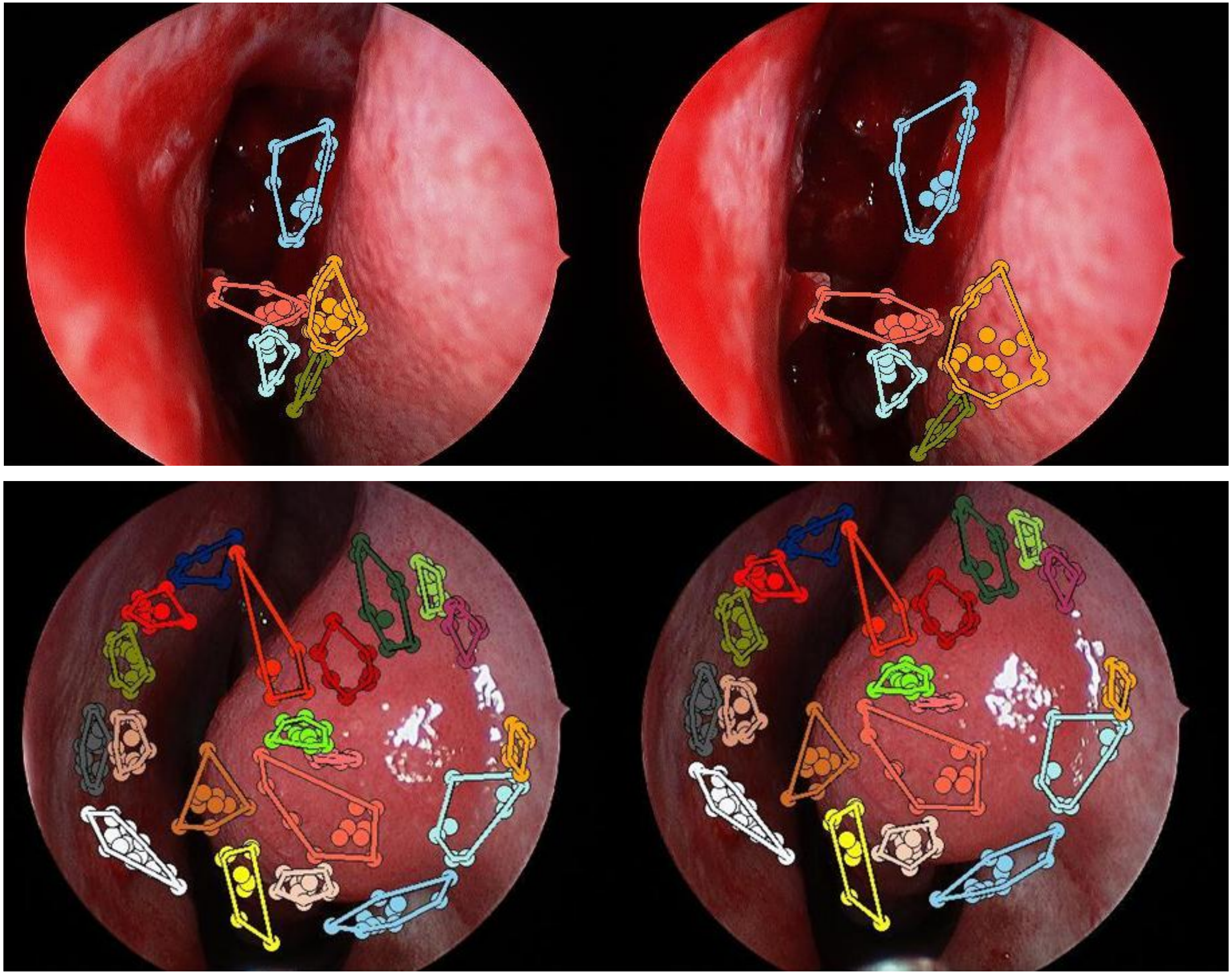
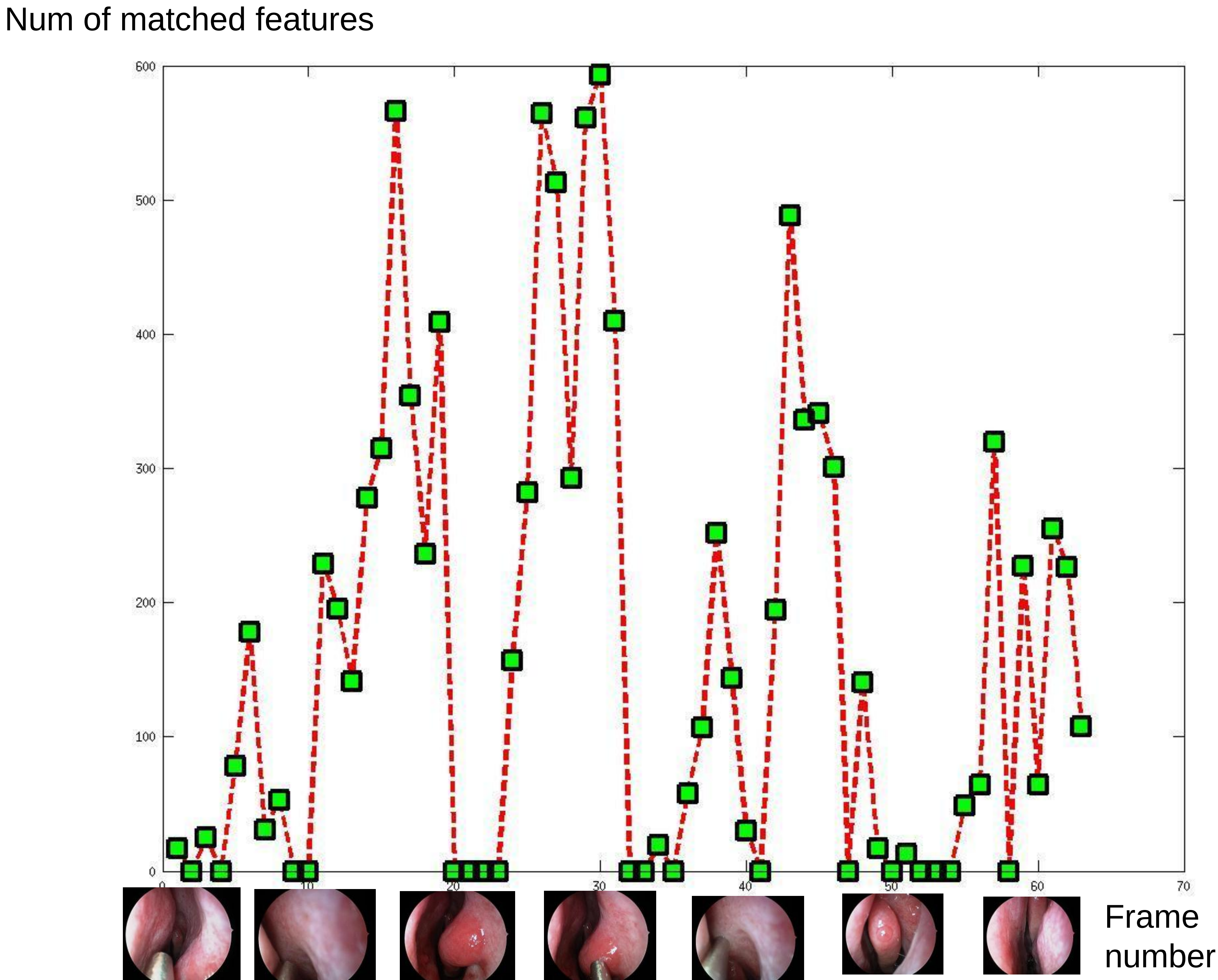


Figure from [Puerto-Souza *etal 2012*]

HMA: preliminary results




Num of matched features
600
500
400
300
200
100
0
0
10
20
30
40
50
60
70
Frame number

Motion estimation

- Epipolar geometry
- Basically [R, t] denotes the motion
  - Can be decomposed from Essential matrix E = sk(t)*R
  - Camera calibration
    - Convert image coordinates to world coordinates.
- Essential matrix from 5/7/8 point algorithm
  - 8 point: classic Louguet-Higgins
- Robust statistics
  - Variants of Least Squares: LMedS and many.
  - Iterative refinement (EM style): RANSAC and variants.
  - Scale estimation: ASKC and several.

Point reconstruction

- Epipolar geometry
- Basically minimizing re-projection error
- Group parameter estimation by bundle adjustment

http://www.youtube.com/watch?v=sz0UbHvEttI
http://www.youtube.com/watch?v=NiMlVgEu7mg

Video-CT registration

- Trimmed ICP
- EM style
- Trim: robustness
  - Sort errors

TrICP $\big(X = \{x_i\}_m, Y = \{y_i\}_m, F = [R, t]\big)$ where x_i, y_i are point vectors.

Guess F, initialize $\varepsilon = 0$.
repeat
  for $i = 1...m$ **do**
    Solve $y_i = argmin_{y_j \in Y} \|y_j - Fx_i\|^2$ by searching in kd-tree.
  end for
  Get $\Pi = \big\{(x_i, y_i)\big\}_m$.

  Sort Π by $\|y_i - Fx_i\|^2 \nearrow$ to get $\Pi' = \big\{(x_{\phi(j)}, y_{\phi(j)})_j\big\}_m$.

  Solve $n = argmin_n \frac{\sum_{i=1}^n \|y_i - Fx_i\|^2}{n}$ $s.t.$ $\mu m \leq n \leq m$ by brute-force search.
  Get $\Pi'' = \big\{(x_{\phi(j)}, y_{\phi(j)})_j\big\}_n$.

  Solve $F' = argmin_F \sum_{i=1}^n \|y_i - Fx_i\|^2 = [R', T']$ by
    Centralize X and Y to get ΔX and ΔY by
      $\Delta x_i = x_i - mean(X) = x_i - \frac{\sum_{i=1}^n x_i}{n}$
      $\Delta y_i = y_i - mean(Y) = y_i - \frac{\sum_{i=1}^n y_i}{n}$
    Compute cross-covariance of ΔX and ΔY:
      $H = \Delta X \cdot \Delta Y \big(= \sum_{i=1}^n \Delta x_i \cdot \Delta {y_i}^T\big)$
    Peform SVD to H:
      $H = U \Lambda V'$
    Solve $R' = argmin_R \sum_{i=1}^n \|\Delta y_i - R\Delta x_i\|^2$ by $R' = VU^T$ if $\det(R') = 1$.
    Get $t' = mean(Y) - R' \cdot mean(X)$.
    Update $F' = [R', T']$.

  $\varepsilon' = \frac{1}{n}\sum_{i=1}^n \|y_i - F'x_i\|^2$.
  $\Delta\varepsilon = \varepsilon' - \varepsilon$.
  Set $\varepsilon = \varepsilon'$.
  Set $F = F'$.
until $\Delta\varepsilon <$ threshold or maxiters

Deliverables

- **Minimum:** (Expected by 7th April - 14th April)
  - Matlab program for robust feature matching by HMA algorithm.
  - Feature matching validation experiments, analysis and documentation.
  - C++ Program for motion estimation by RANSAC and 5 point algorithm.
  - Motion estimation validation experiments, analysis and documentation.
  - C++ program for video-CT registration by Trimmed ICP algorithm.
  - Video-CT registration validation experiments, analysis and documentation.
  - Program for 3D reconstruction.
  - 3D reconstruction validation experiments, analysis and documentation
- **Expected:** (Expected by 21st April - 28th April)
  - Program for visual display of 3D reconstruction results.
- **Maximum:** (Expected by 9th May)
  - Experiments for the holistic pipeline and documentation.

Milestones

- Milestone 1: Program for robust feature matching by HMA algorithm.
  - Planned Date: 28th February. Expected Date: 7nd March
- Milestone 2: Program for motion estimation by RANSAC and 5 point algorithm.
  - Planned Date: 14th March. Expected Date: 14th March
- Milestone 3: Program for video-CT registration by Trimmed ICP algorithm
  - Planned Date: 21st March. Expected Date: 28th March
- Milestone 4: Program for 3D reconstruction
  - Planned Date: 11st April. Expected Date: 14th April
- Milestone 5: Program for visual display of 3D reconstruction results
  - Planned Date: 25th April. Expected Date: 28th April
- Milestone 6: Experiments for the holistic pipeline
  - Planned Date: 7th May. Expected Date: 9th May

Reference

- D. Mirota, H. Wang, R. H. Taylor, M. Ishii, G. L. Gallia and G. D. Hager. A System for Video-Based Navigation for Endoscopic Endonasal Skull Base Surgery. IEEE Trans. Med. Imaging, 31(4), 963-976, 2012.
- G. Puerto, M. Adibi, J. Cadeddu1 and G. L. Mariottini, Adaptive Multi-Affine (AMA) Feature-Matching Algorithm and its Application to Minimally-Invasive Surgery Images. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 2371 - 2376, Sept. 25-30, San Francisco, California, 2011.
- G. A. Puerto-Souza and G. L. Mariottini. Hierarchical Multi-Affine (HMA) algorithm for fast and accurate feature matching in minimally-invasive surgical images. 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems October 7-12, 2012. Vilamoura, Algarve, Portugal.
- D. Chetverikov , D. Svirko , D. Stepanov and Pavel Krsek. The Trimmed Iterative Closest Point Algorithm, ICPR, 2002.

Thank you! Comments!

An concurrent comparison: Learning Continuous CRF with Geometric Priors for Stereo from Endoscopic Videos

- Feature matching + Least Square routine $d(x,y,k) = argmax_d \quad C(x,y,d) = argmax_d \sum_k \left(\tilde{I}(x,y,d,k) - I_r(x,y)\right)^2$
- Powerful MRF-based global optimization $E(\mathbf{x},\mathbf{y}) = \mu \sum_i U(x_i,\mathbf{y}) + (1-\mu) \sum_{i\sim j} V(x_i,x_j,\mathbf{y})$
  - Heuristic MRF objective design
  - Free parameter estimation by learning CRF $P(\mathbf{X}=\mathbf{x} \mid \mathbf{y}) = \frac{1}{Z(\mathbf{y})} \exp\left(-E(\mathbf{x},\mathbf{y})\right)$
    - Learning on discriminative CRF from ground-truth labeled data
    - Scalable in the large and complex such as stereo from video streams $Z(\mathbf{y}) = \sum_{\mathbf{x}} \exp\left(-E(\mathbf{x},\mathbf{y})\right)$
- Our concern is how the learning performs given limited training data, due to the difficulty of obtaining labels. Some hopes have been put on embedding reliable priors and structure prediction for abundant testing data.
- Our attempts will include five aspects.
  - To formalize correctly matched features as data priors (i.e., soft constraints) into the pixel-level MRF/CRF optimization.
  - To test if distribution-free data density estimation induce better performance than heuristic assumptions such as Gaussian family like Mixtures of Gaussian.
  - To test if higher-order smoothness priors induce better performance than the standard pairwise ones.
  - To design a weakly supervised learning workflow for stereo matching from endoscopic videos.
  - To conduct an empirical analysis on the tradeoff complexity and performance between learning-based methods and pure geometric estimation methods.