Articulated Object Tracking by Rendering Consistent Appearance Parts

International Conference on Robotics and Automation '09

Zachary Pezzementi, Sandrine Voros, and Gregory D. Hager

Computational Interaction and Robotics Lab
Laboratory for Computational Sensing and Robotics
Johns Hopkins University

May 16, 2009

Title piece by piece

Articulated Object Tracking
by Rendering Consistent Appearance Parts

Problem Context

- Want to estimate configuration of a target object

Title piece by piece

Articulated Object Tracking
by Rendering Consistent Appearance Parts

Problem Context

- Want to estimate configuration of a target object
- Flexible object model and configuration space

Title piece by piece

Articulated Object Tracking
by Rendering Consistent Appearance Parts

Problem Context

- Want to estimate configuration of a target object
- Flexible object model and configuration space
- Have geometric model of the target

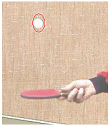
Title piece by piece

Articulated Object Tracking
by Rendering Consistent Appearance Parts

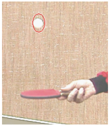
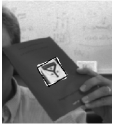
Problem Context

- Want to estimate configuration of a target object
- Flexible object model and configuration space
- Have geometric model of the target
- Learn appearance model

		Model	
		Simple	Complex
Applications	General		
	Specific		

	Model	
	Simple	Complex
General	Kernel[1] 	
Applications		
Specific		

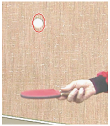
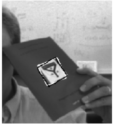
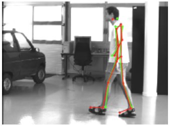
[1]Comaniciu et al, PAMI '03.

	Model	
	Simple	Complex
Applications: General	Kernel[1] 	Template/SSD[2] 
Applications: Specific		

[1] Comaniciu et al, PAMI '03. [2] Hager & Belhumeur, PAMI '98.

Overview

Model

Applications	Simple	Complex
General	Kernel[1] 	Template/SSD[2] 
Specific	Faces[3] 	People[4] 

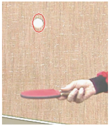
[1]Comaniciu et al, PAMI '03. [2]Hager & Belhumeur, PAMI '98. [3] Birchfield, CVPR '98. [4], Z Lu et al, NIPS '07.

Model

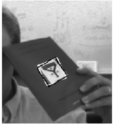
Simple — Complex

Applications: General — Specific

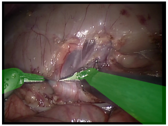
Kernel[1]



Template/SSD[2]

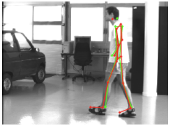


This Work



Faces[3]

People[4]



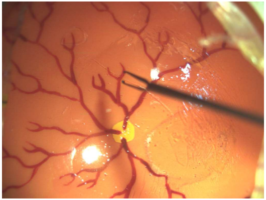
[1]Comaniciu et al, PAMI '03. [2]Hager & Belhumeur, PAMI '98. [3] Birchfield, CVPR '98. [4], Z Lu et al, NIPS '07.

General Approach

- Estimate appearance probability map
- Render object model in candidate configurations
- Maximize correlation b/w rendering and probability map

General Approach

- Estimate appearance probability map
- Render object model in candidate configurations
- Maximize correlation b/w rendering and probability map

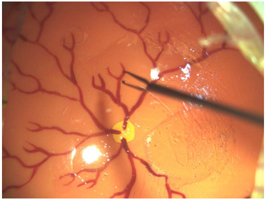


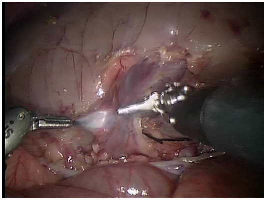
Retinal surgery

- Monoscopic
- 2D tracking of rigid tool mask
- 3 DOF

General Approach

- Estimate appearance probability map
- Render object model in candidate configurations
- Maximize correlation b/w rendering and probability map



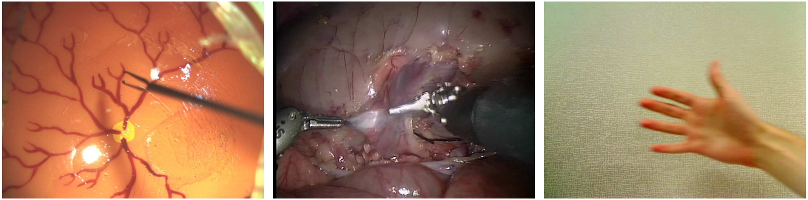


Da Vinci laparoscopic surgery

- Stereoscopic
- 3D tracking of 2 articulated tools
- 18 DOF

General Approach

- Estimate appearance probability map
- Render object model in candidate configurations
- Maximize correlation b/w rendering and probability map

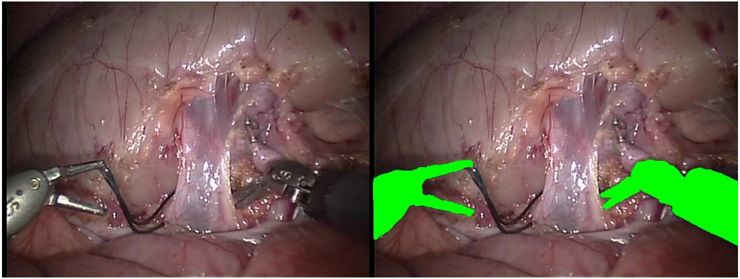


Human hand tracking

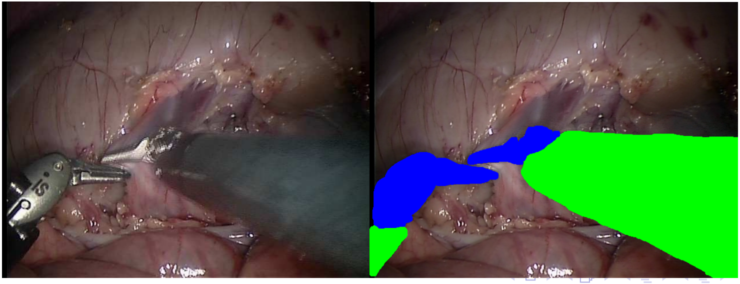
- Monoscopic
- 3D tracking of simplified hand model
- 18 DOF

Training

Hand-segmentation with 1 foreground class

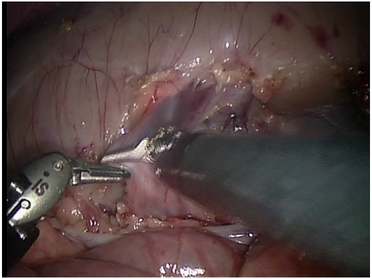


Hand-segmentation with 2 foreground classes

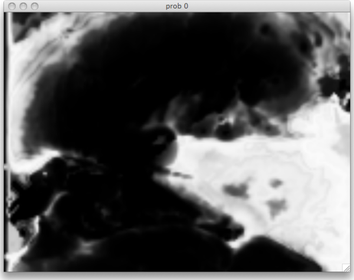


Output

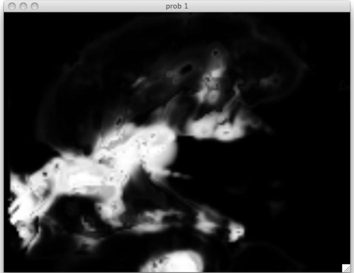
Input



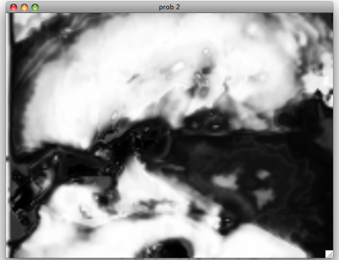
Tool shaft class



Tool tip class



Background class



Appearance Modeling Approach

- Separate object into consistent appearance parts and train
- Extract color and texture features
  - RGB and HSV blocks around current pixel
  - Haralick (GLCM) features: contrast, correlation, energy and homogeneity
- LDA to reduce dimensionality
- Fit GMM to each class

Gaussian Mixture Model (GMM) fit to each class, using LNKnet

Symbols

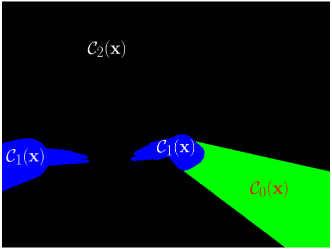
v	pixel feature vector
$\mathcal{C}_j$	class j
α	mixture weight
π	LDA projection
μ	mean
Σ	covariance

$$P(v \mid \mathcal{C}_j) = \sum_{i=1}^{\mathcal{C}n} \alpha_i P(\pi v \mid \mu_i, \Sigma_i)$$

with $P(\pi v \mid \mu_i, \Sigma_i) = \frac{1}{\sqrt{2\pi}|\Sigma_i|^{\frac{1}{2}}} e^{-\frac{1}{2}(\pi v - \mu_i)^T \Sigma_i^{-1} (\pi v - \mu_i)}$

Symbols

$\mathbf{I}$	Image
$\mathbf{x}$	Configuration
θ	Set of appearance classes
$\mathcal{C}_j(\mathbf{x})$	Pixels in class j

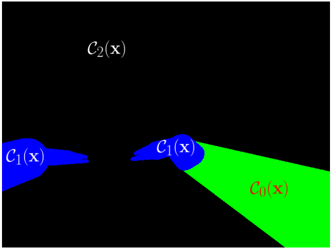


Visualization of $\theta(\mathbf{x})$

Want to maximize $P(\mathbf{I} \mid \mathbf{x}, \theta) = \prod_{\mathcal{C}_j \in \{\theta\}} \prod_{v \in \mathcal{C}_j(\mathbf{x})} P(v \mid \mathcal{C}_j)$

Symbols

$\mathbf{I}$	Image
$\mathbf{x}$	Configuration
θ	Set of appearance classes
$\mathcal{C}_j(\mathbf{x})$	Pixels in class j

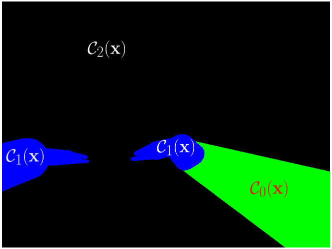


Visualization of $\theta(\mathbf{x})$

Want to maximize $P(\mathbf{I} \mid \mathbf{x}, \theta) = \prod_{\mathcal{C}_j \in \{\theta\}} \prod_{v \in \mathcal{C}_j(\mathbf{x})} P(v \mid \mathcal{C}_j)$

Use Nelder-Mead downhill simplex

Symbols	
$\mathbf{I}$	Image
$\mathbf{x}$	Configuration
θ	Set of appearance classes
$\mathcal{C}_j(\mathbf{x})$	Pixels in class j

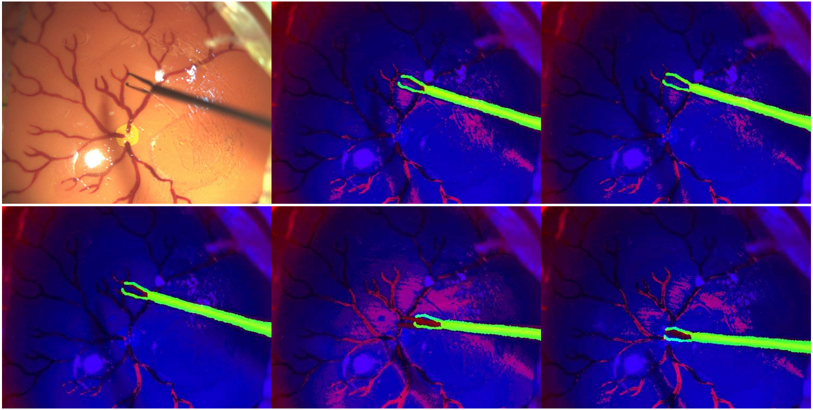


Visualization of $\theta(\mathbf{x})$

Want to maximize $P(\mathbf{I} \mid \mathbf{x}, \theta) = \prod_{\mathcal{C}_j \in \{\theta\}} \prod_{v \in \mathcal{C}_j(\mathbf{x})} P(v \mid \mathcal{C}_j)$

Use Nelder-Mead downhill simplex

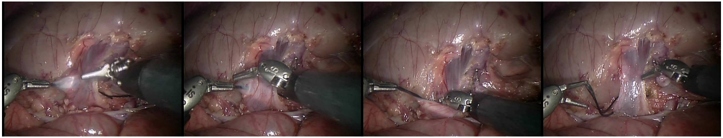
For stereo, maximize $P(\mathbf{I}_1, \mathbf{I}_2 \mid \mathbf{x}, \theta)$ and $\mathcal{C}(\mathbf{x})$ may span both images.



Blue channel - greyscale of the original input
Red channel - foreground probability map
Green channel - rendering of best configuration estimate

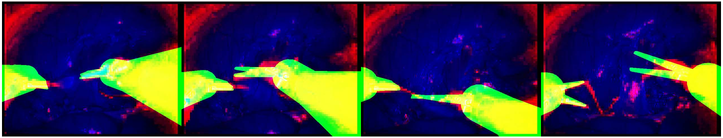
Red channel - greyscale of the original input
Blue channel - foreground probability map
Green channel - rendering of best configuration estimate

Laparoscopic Surgery



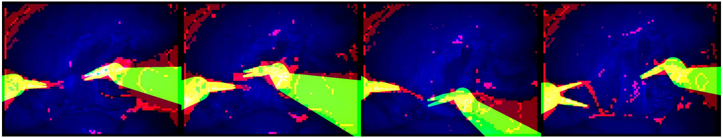
Input

Blue channel - input greyscale, Red channel - foreground prob., Green channel - config. estimate



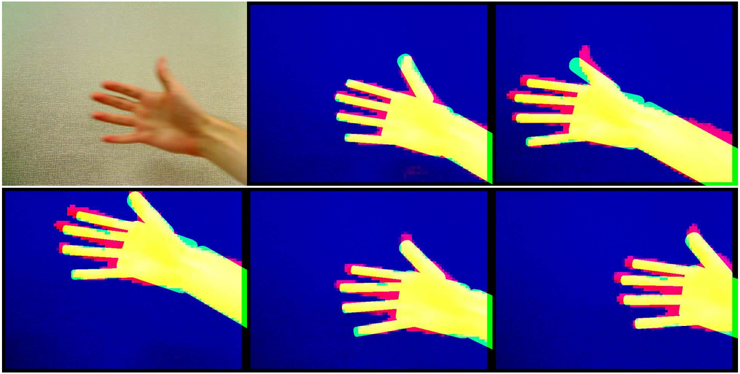
Single-class

Blue channel - input greyscale, Red channel - highest prob. class, Green channel - config. estimate



Multi-class

Blue channel - greyscale of the input
Red channel - value corresponding to highest probability class
Green channel - configuration estimate



Blue channel - greyscale of the original input
Red channel - foreground probability map
Green channel - rendering of best configuration estimate

Future Work

Potential extensions:

- Parallel implementation (GPU)
- Improved traversal of configuration space
- Adding edge information
- Dynamically updating appearance model

Acknowledgements

Thanks to NSF Grants IIS-0205318 and R21 EY016359 and our reviewers.

Thank you for your attention.

Questions?