VISION-BASED ANALYSIS OF CONVENTIONAL SURGICAL PROCEDURES

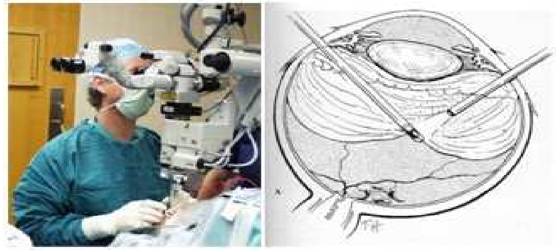
Oliver Ruepp and Darius Burschka

Machine Vision and Perception Group (MVP)

Department of Computer Science
Technische Universität München

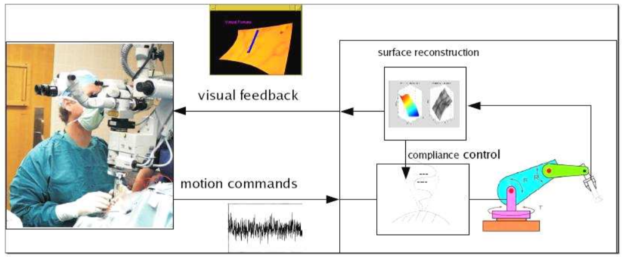
Safety in Surgical Procedures

Virtual Fixtures (prevention of organ injuries)



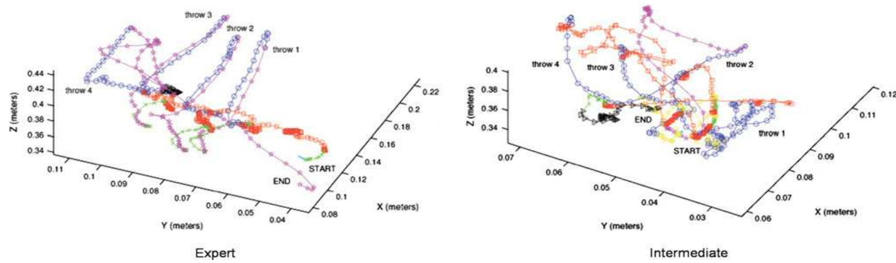
Example:
eye surgery

Task dependent **Motion Constraints** (compensation of human imperfections)



Hand tremor

Skill Evaluation of the surgeons (certification of skills as estimate of success rate for a given procedure)



Source: Surgical Endoscopy, July 2010

All of these tasks (Virtual Fixtures, Motion Constraints, and Skill Evaluation) rely heavily on perception:

- **Virtual Fixtures**:
  Relies on 3D Reconstruction and Motion Capturing.
- **Motion Constrains:**
  Relies on 3D Reconstruction, Motion Capturing and Procedure Modeling.
- **Skill Evaluation:**
  Relies on Motion Capturing.
- There are three major building blocks to discuss:
  - Motion Capturing
  - Procedure Modeling
  - 3D Reconstruction

Our Contributions

- ***Motion capturing:*** common systems are based on robot kinematics from robots like daVinci (e.g. the JHU system).
  Problem: What about standard procedures? They are still most common, thus should not be neglected.
  (→*Motion Constraints*, *Skill Evaluation*)
- ***Procedure Modeling:*** movement constraints in a single step of the procedure are of interest. Operation-relevant properties and actions can be described by means of Object container, Location Map and Functionality Graph.
  (→Motion Constraints)
- ***3D Reconstruction:*** since the anatomy details vary between patients, an accurate reconstruction to map the procedure into the given environment is required.

  Problem: many procedures still use monocular equipment, which generates low-quality image data. Photogrammetric enhancements to the reconstruction process are necessary. View planning algorithms can be used to improve accuracy.

  (→ *Virtual Fixtures*, *Motion Constraints*)

Motion Capture Systems

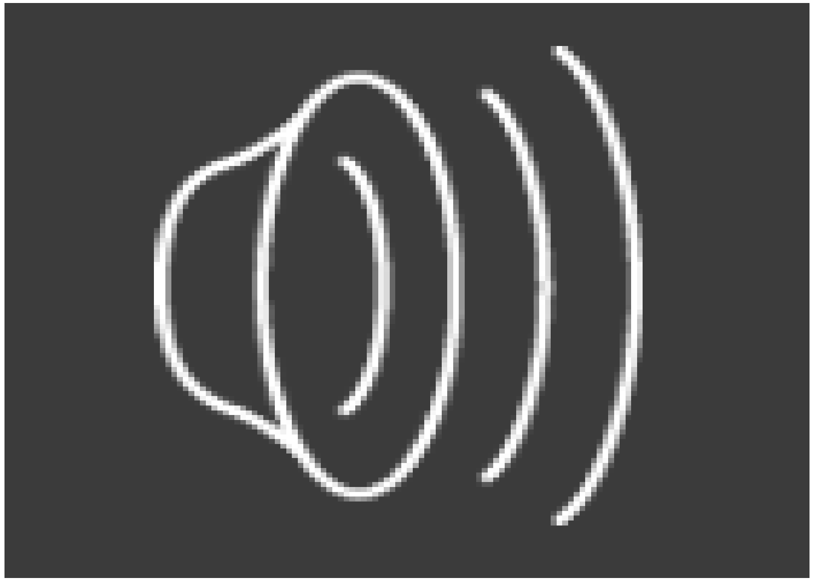
conventional

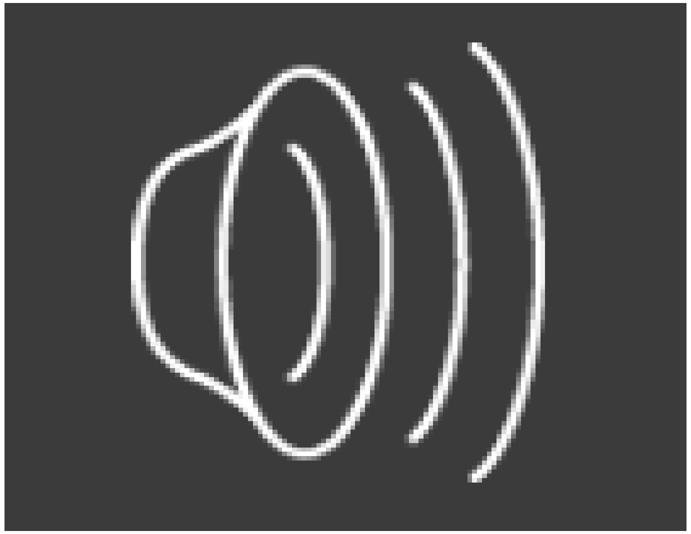
Source: MedGadget

Kinematics-based

We are interested in providing similar motion information about tools as from systems working with kinematics data

A new trend in computer-assisted surgery is usage of IMUs in videoendoscopes.





Problem: high dynamics of the human hand is too fast for the low frame-rates of standard video cameras

Solution: coupling of inertial sensors IMU with visual perception
(Mair et al. IROS 2009)

Template-based Tool Tracking
in image space
Application: guidance of the
camera

(work with G.Hager, JHU)

6DoF pose estimation of tools
manipulated in the scene
The system estimates in real-time
the pose of the moved object

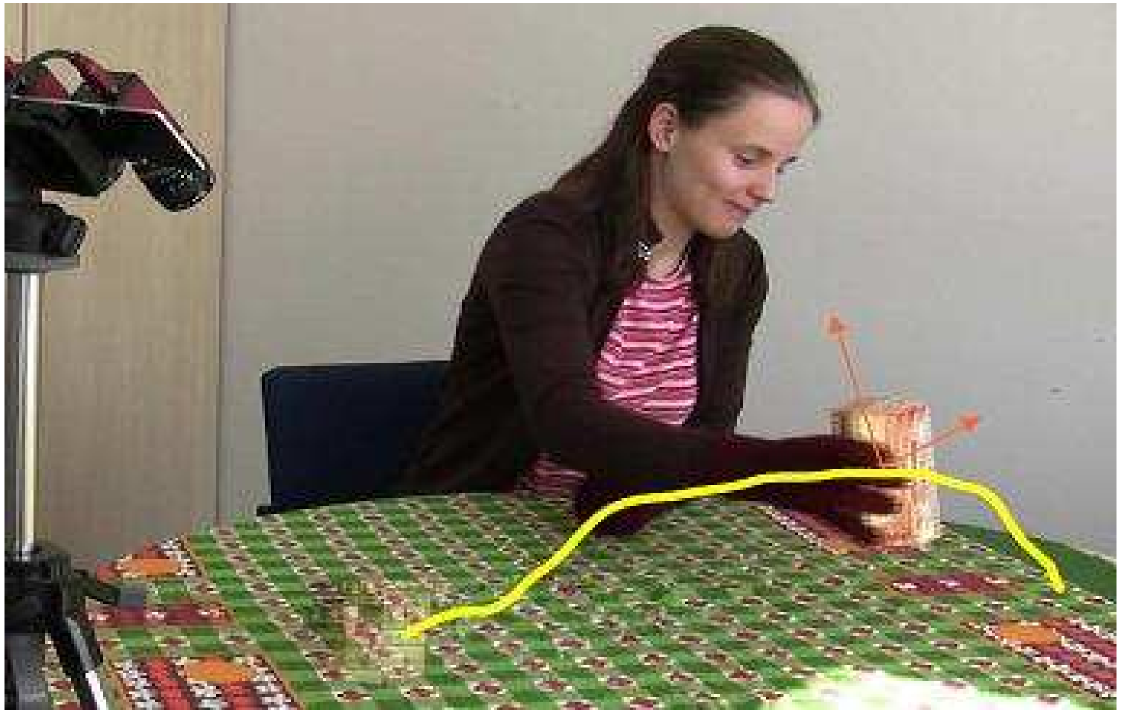
Our Contributions

- ***Motion capturing:*** common systems are based on robot kinematics from robots like daVinci (e.g. the JHU system).
  Problem: What about standard procedures? They are still most common, thus should not be neglected.
  (→*Motion Constraints*, *Skill Evaluation*)
- ***Procedure Modeling:*** movement should be task-, object- and location-specific. Manipulation-relevant properties and actions can be described by means of Object container, Location Map and Functionality Graph.
  (→Motion Constraints)
- ***3D Reconstruction:*** since the anatomy details vary between patients, an accurate reconstruction to map the procedure into the given environment is required.

  Problem: many procedures still use monocular equipment, which generates low-quality image data. Photogrammetric enhancements to the reconstruction process are necessary. View planning algorithms can be used to improve accuracy.

  (→*Virtual Fixtures*, *Motion Constraints*)

Learning of Action Attributes through Observation of Human Actions



http://mvp.visual-navigation.com/

Representation of Manipulation-Relevant Object Properties and Actions

- Observation of human demonstrations
- Extraction of knowledge:
  - New knowledge
  - Relevance
- Knowledge:
  - Object properties
  - Object functionalities in the environment

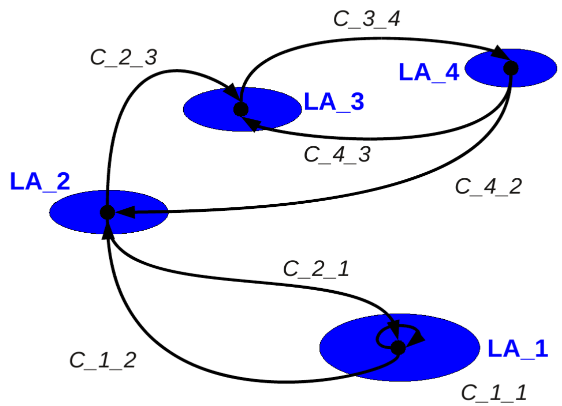
Knowledge Representation

Object container

shape	parametric description	grasp points	actions and handling
			

Functionality Map

Locations Area **LA_i**



C_i_j: connection properties

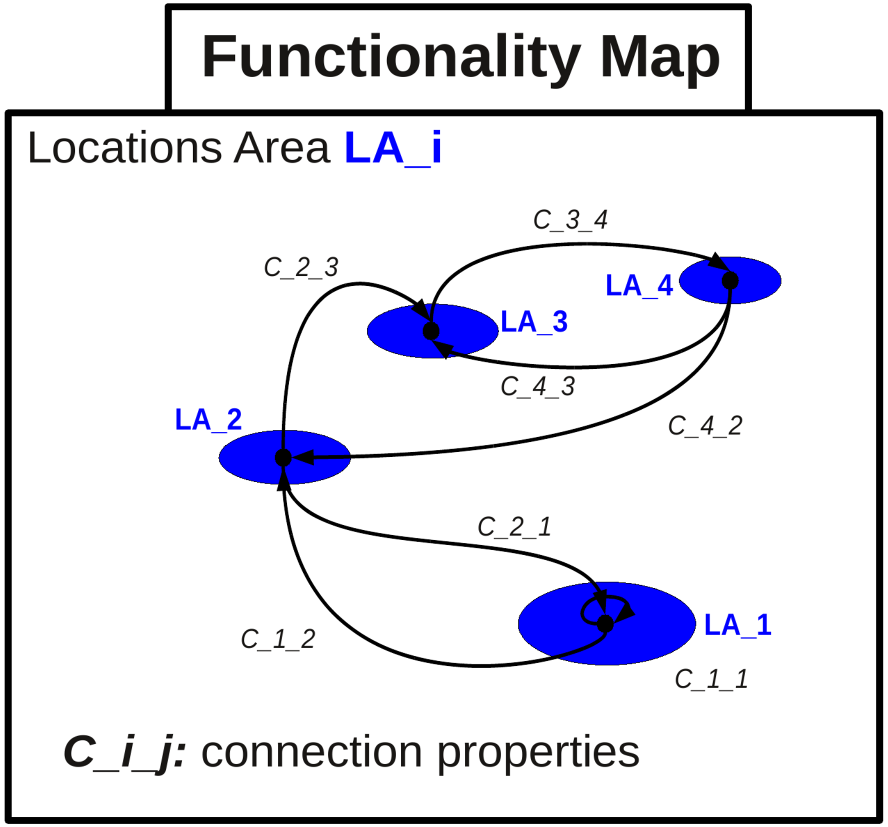
Each tool used in the procedure has its own container describing its shape, handling properties etc.

(Petsch/Burschka IROS2011)

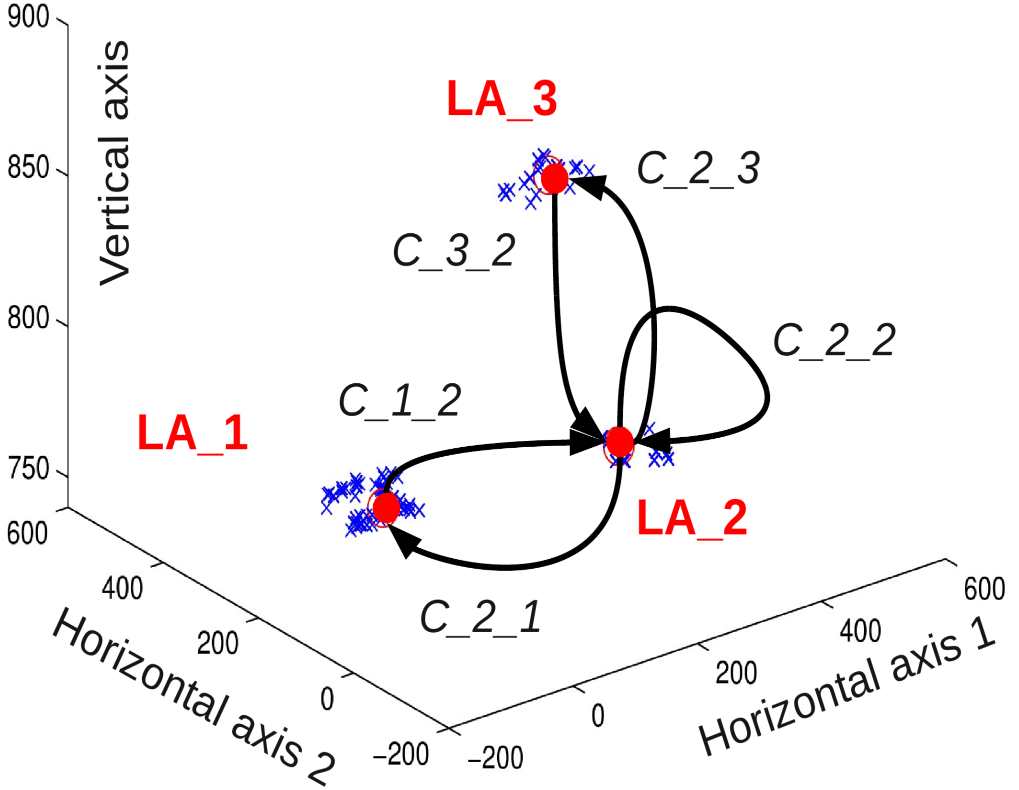
In addition, we store a functionality map for a specific procedure that describes the way how the tool was used during the procedure while moved between points in the body

Functionality Map

- Location Areas
- Connection Properties:
  - Arbitrary movement vs. constrained trajectory
  - Connection relevance
  - Velocity constraint during pick-up
  - Grasp taxonomy



Tracking Data: Location Areas



Legend:

Location Area i with its corresponding center

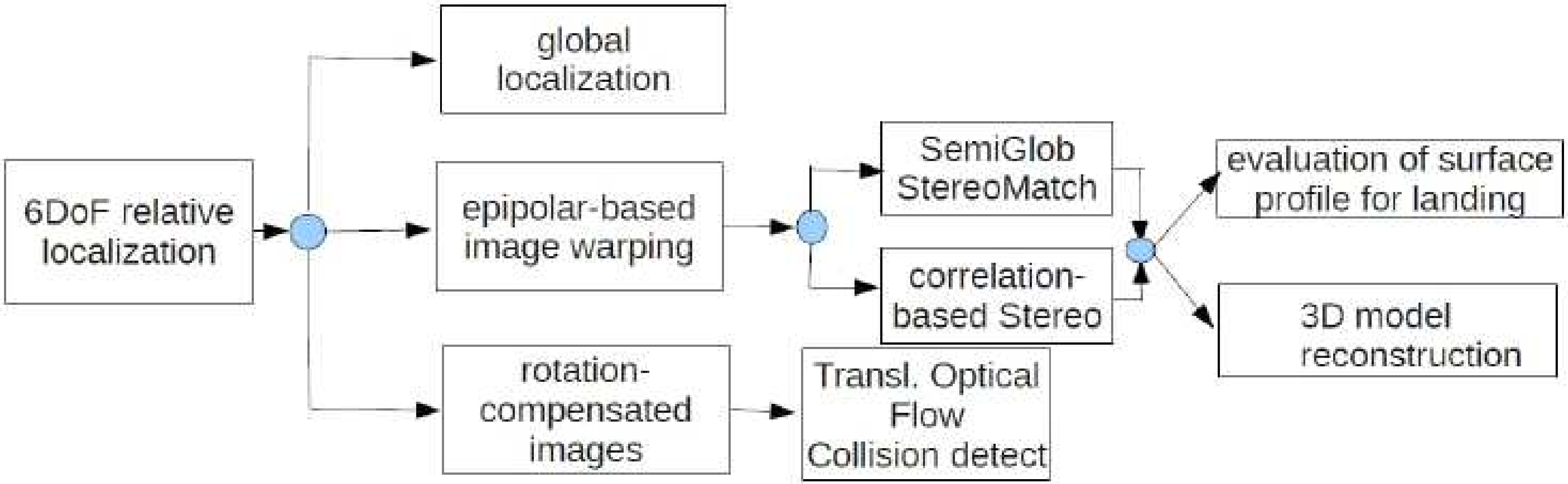
C_i_j Connecting edge from Location Area i to j

Our Contributions

- ***Motion capturing:*** common systems are based on robot kinematics from robots like daVinci (e.g. the JHU system).
  Problem: What about standard procedures? They are still most common, thus should not be neglected.
  ($\rightarrow$ *Motion Constraints*, *Skill Evaluation*)
- ***Procedure Modeling:*** movement constraints in a single step of the procedure are of interest. Operation-relevant properties and actions can be described by means of Object container, Location Map and Functionality Graph.
  ($\rightarrow$ Motion Constraints)
- ***3D Reconstruction:*** since the anatomy details vary between patients, an accurate reconstruction to map the procedure into the given environment is required.

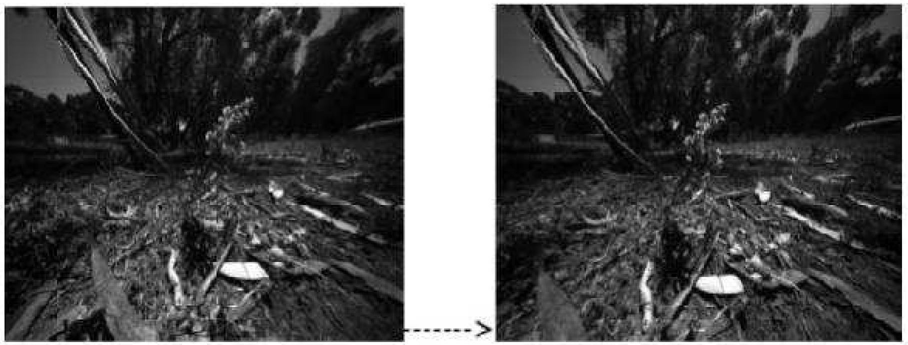
  Problem: many procedures still use monocular equipment, which generates low-quality image data. Photogrammetric enhancements to the reconstruction process are necessary. View planning algorithms can be used to improve accuracy.

  ($\rightarrow$ *Virtual Fixtures*, *Motion Constraints*)

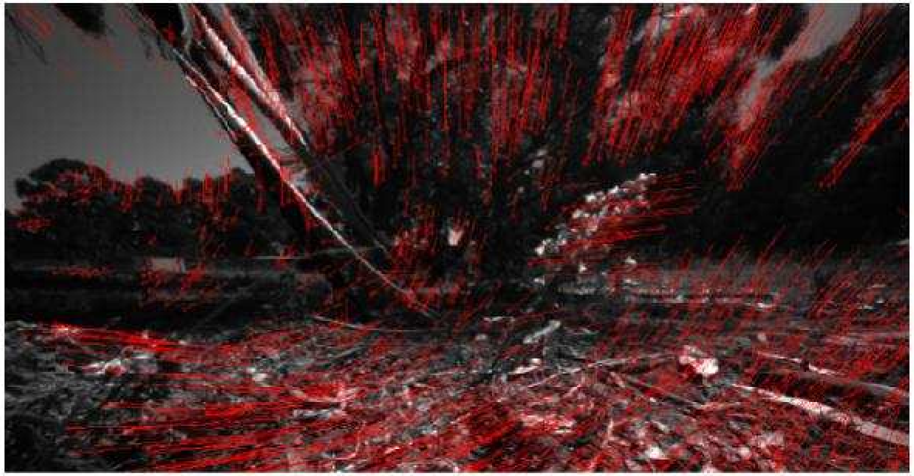


Modular reconstruction pipeline for an easy exchange of modules

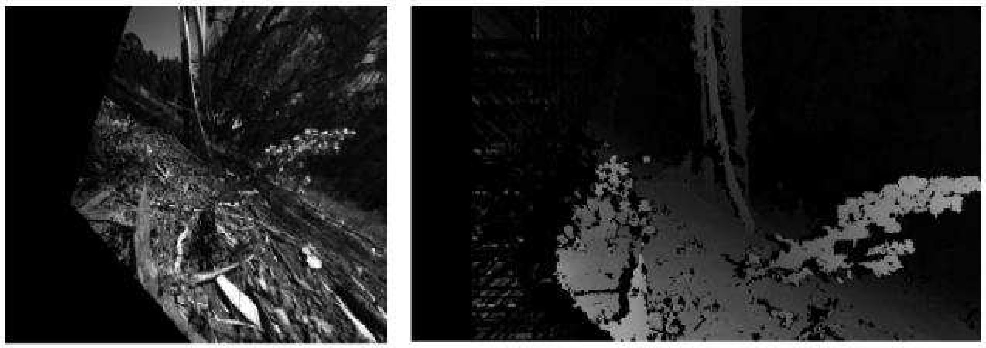
Reconstruction Pipeline (ICRA2012 submission)



Initial images



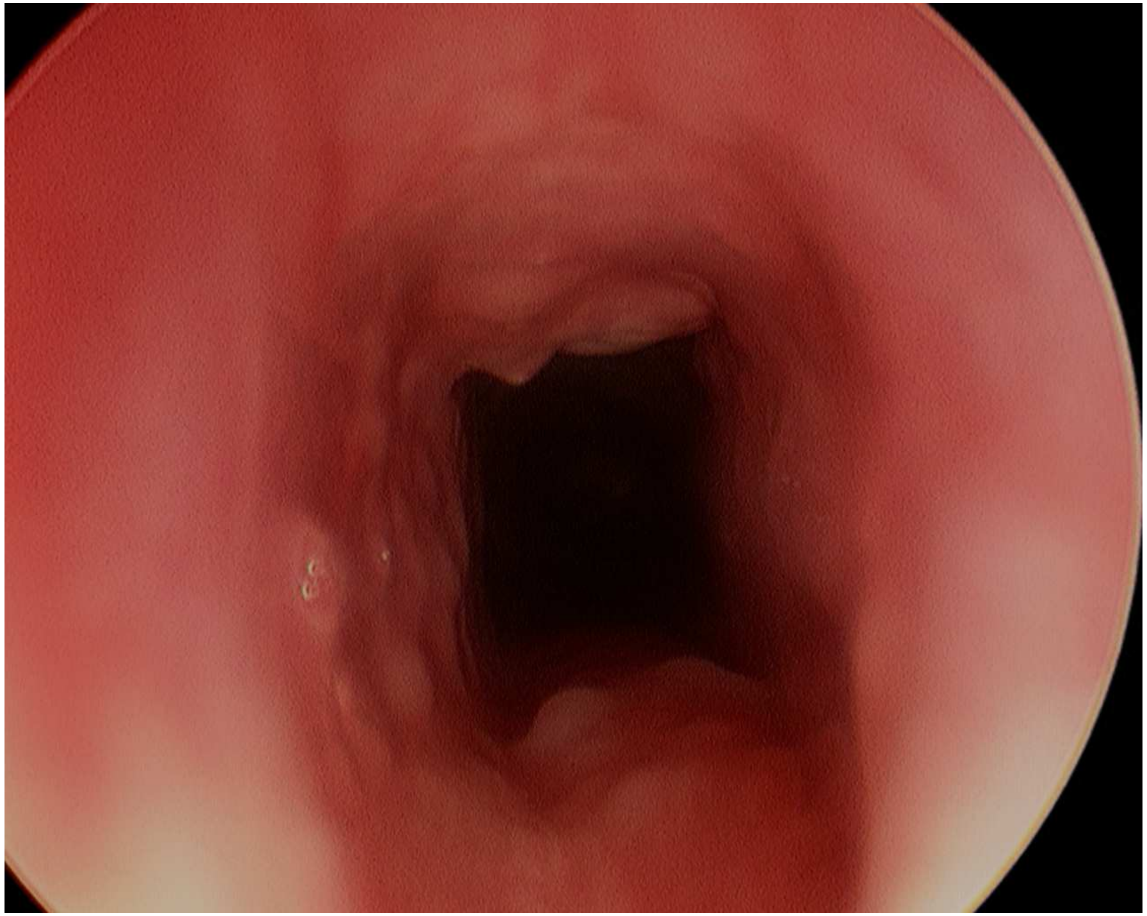
Motion estimation



Motion correction
and reconstruction
result

Scene and Motion Recovery from Monocular Images

- A well-known problem in Computer Vision
- Algorithms for “simple” cases are abundant
- But: Videoendoscopy is difficult!
- Problems:
  - Difficult lighting conditions
  - Bad image quality
  - Very little image structure

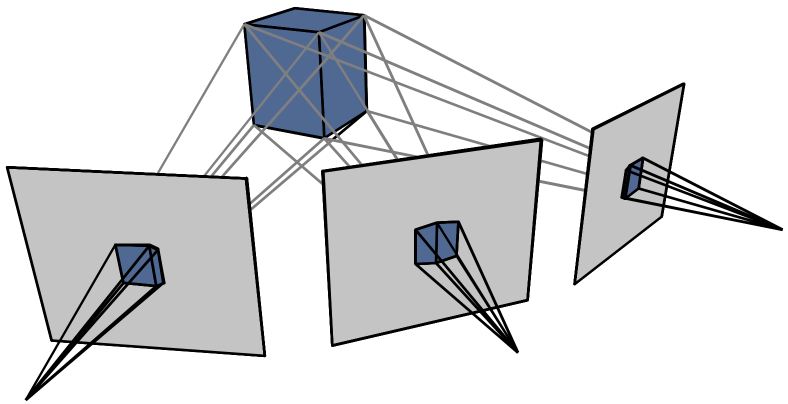


Typical difficulties: image blur, image noise, almost no texture

The Basic Idea

- Bundle adjustment is a well-known algorithm for structure and camera position recovery
- Feature-based and intensity-based variants
- Toy example for feature-based reconstruction:
- Nonlinear optimization minimizes reprojection error of pixel coordinates.

$$\min_{\boldsymbol{a},\boldsymbol{b}} \sum_{i=1}^{n} \sum_{j=1}^{m} d\left(\boldsymbol{Q}\left(\boldsymbol{a}_j, \boldsymbol{b}_i\right), x_{i,j}\right)$$



Intensity-based methods

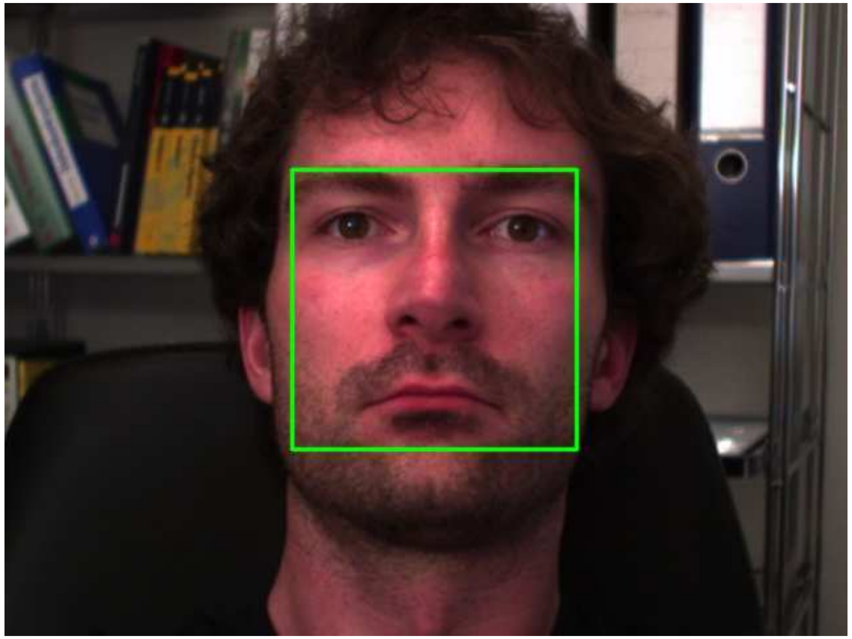
- In videoendoscopy, reliable feature detection is nearly impossible
- We use an intensity based approach
- Price to pay:
  - Computationally expensive
  - Less detailed models (regularization)
- Advantage:
  - Reconstruction is possible where feature-based methods fail!

Reconstruction with constant lighting

- First idea: Apply a multi-view intensity-based reconstruction method
- Important advantage: Intensity-based methods use as much information as there is
- This approach uses the brightness constancy assumption
- Idea: Find parameters that explain images by nonlinear optimization

$$\min_{\boldsymbol{a},\boldsymbol{b}} \sum_{i=1}^{n} \sum_{j=1}^{m} d\left(I_j\left(\boldsymbol{Q}\left(\boldsymbol{a}_j,\boldsymbol{b}_i\right)\right), I_0\left(\boldsymbol{p}_i\right)\right)$$

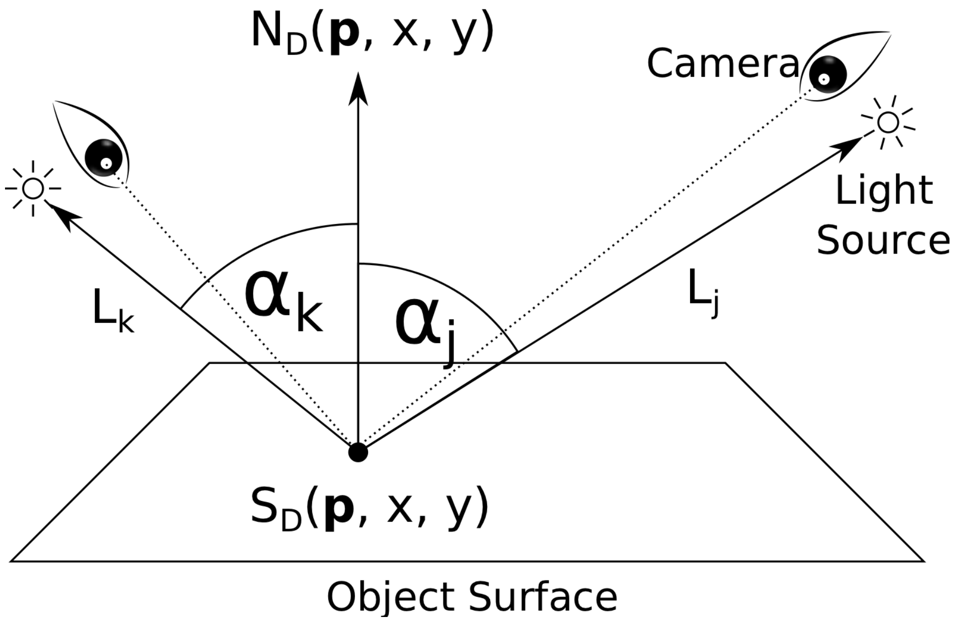
- Works well for Lambertian surfaces



Ruepp and Burschka. Fast recovery of weakly textured surfaces from monocular image sequences. ACCV 2010

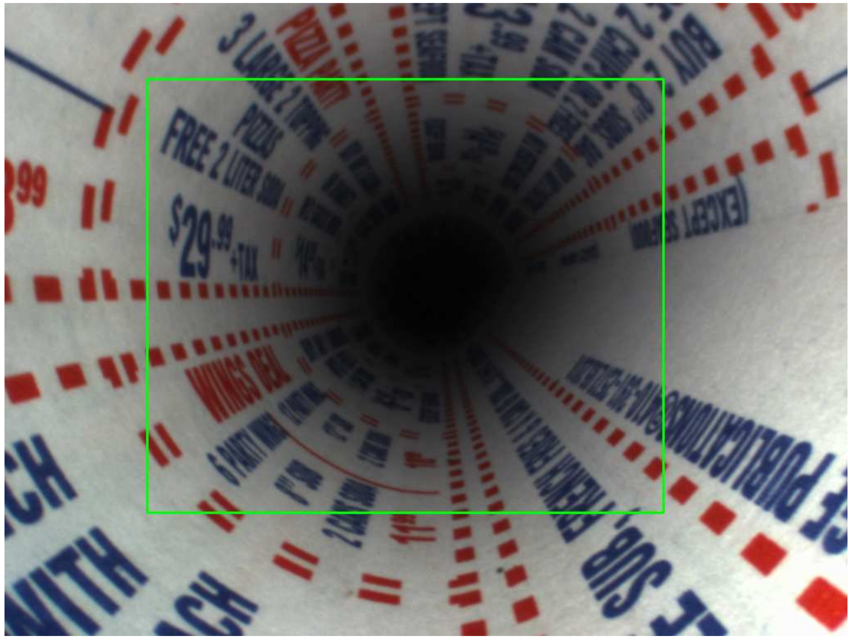
Under changing lighting conditions

- In videoendoscopy, the brightness constancy assumption is violated!
- The light source is moving with the camera
- This can be modeled by the reconstruction algorithm



http://mvp.visual-navigation.com/

Reconstruction results



Current and Future Research

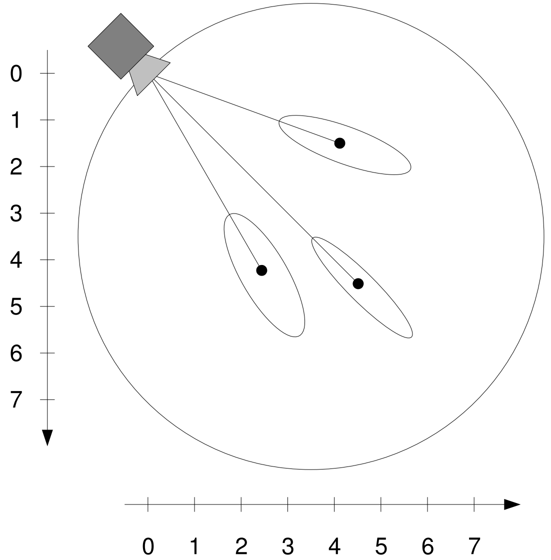
- The algorithm makes idealizing assumptions that lead to suboptimal results:
  - Linear camera response, no vignetting
  - Singular point light source
  - Uniform light source intensity
  - Light source coincident with camera position
- A calibration method is needed!
- Also: Performance can be improved
- Deal with partial occlusions

View planning

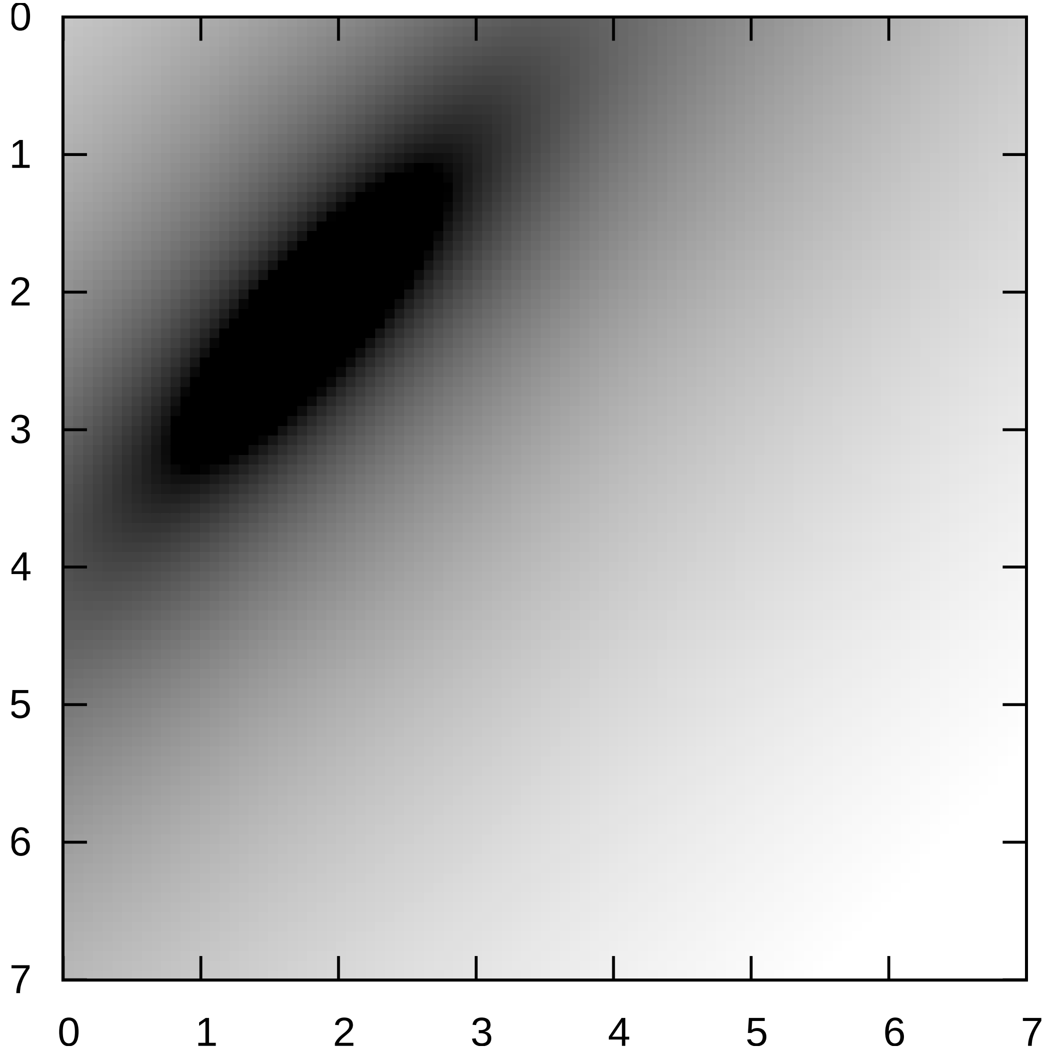
- Reconstruction accuracy is crucial for safety.
- How do we evaluate the accuracy?
- How can we improve it?
- View planning algorithms help to determine camera movement so as to maximize accuracy

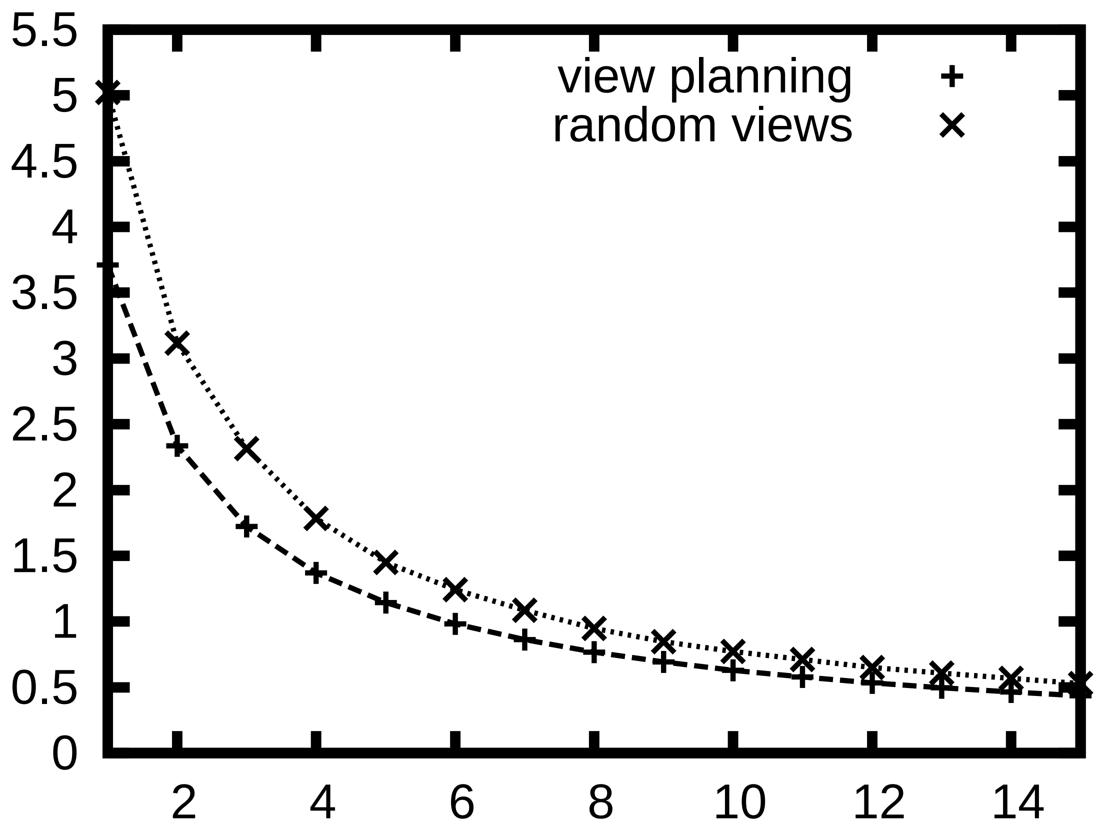
Schematic Overview

- Reconstruction results are typically affected by uncertainty
- Viewing direction should be orthogonal to maximum uncertainty
- This leads to an evaluation function for camera positions



Camera Position Evaluation





Especially in the first few steps, uncertainty is reduced significantly

Future Research

- Until now: Camera movement restricted to a circular path around target
- Needs to be adapted to videoendoscopy, where movement is more restricted
- Deal with partial occlusions

Thank you for your attention!

http://mvp.visual-navigation.com/

Our Contributions

- ***Motion capturing:*** common systems are based on robot kinematics from robots like daVinci (e.g. the JHU system).
  Problem: What about standard procedures? They are still most common, thus should not be neglected.
  (→*Motion Constraints*, *Skill Evaluation*)
- ***Procedure Modeling:*** movement constraints in a single step of the procedure are of interest. Operation-relevant properties and actions can be described by means of Object container, Location Map and Functionality Graph.
  (→Motion Constraints)
- ***3D Reconstruction:*** since the anatomy details vary between patients, an accurate reconstruction to map the procedure into the given environment is required.

  Problem: many procedures still use monocular equipment, which generates low-quality image data. Photogrammetric enhancements to the reconstruction process are necessary. View planning algorithms can be used to improve accuracy.

  (→*Virtual Fixtures*, *Motion Constraints*)