# Scale-Invariant Registration of Monocular Stereo Images to 3D Surface Models

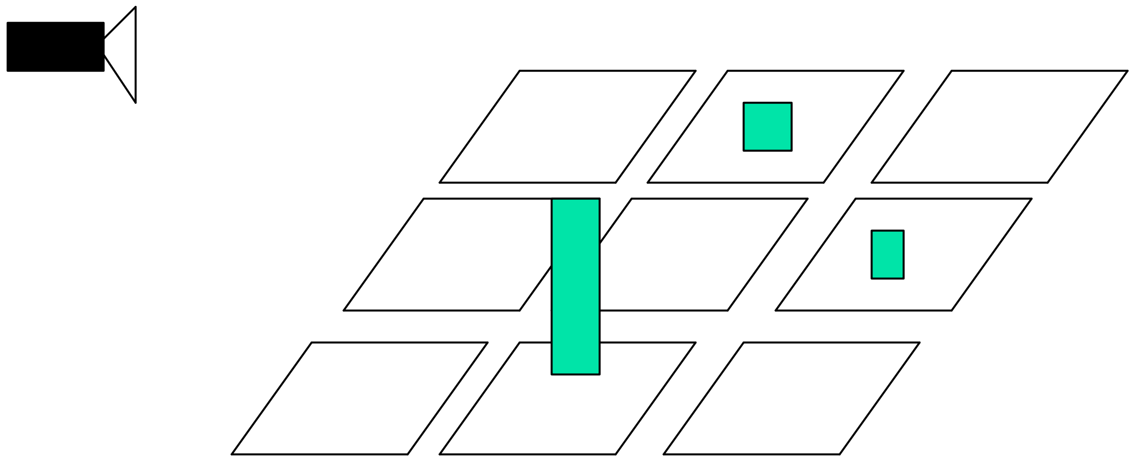
*Darius Burschka[1], Ming Li[2], Russell Taylor[2] and Gregory D. Hager[1]*

[1]**Computational Interaction and Robotics Laboratory (CIRL)**

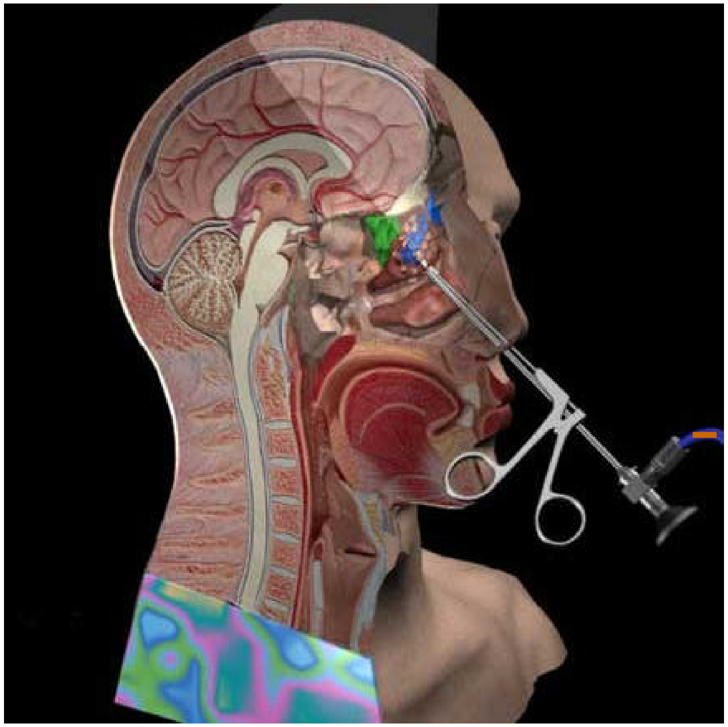
[2]**Computer Integrated Surgical Systems and Technology (CISST)**

**Johns Hopkins University, Baltimore, USA**

# Generic Problem
# Surface Structure Registration



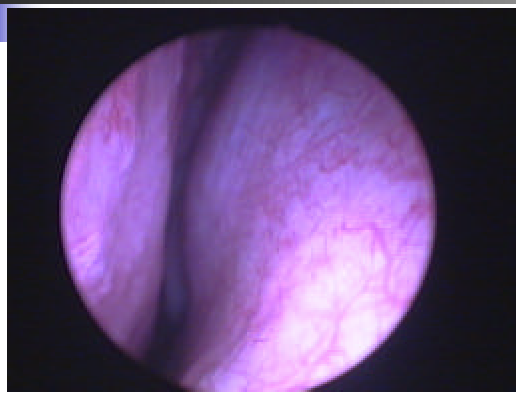
# Our Example Application



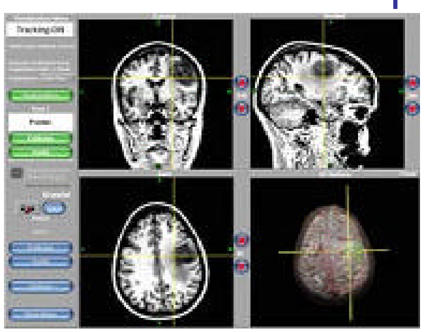
Endoscopic sinus surgery is a procedure used to remove blockages in the sinuses (the spaces filled with air in some of the bones of the skull). These blockages cause sinusitis, a condition in which the sinuses swell and become clogged, causing pain and impaired breathing.

The sinuses are physically close to the brain, the eye, and major arteries, always areas of concern when a fiber optic tube is inserted into the sinus region → accurate navigation is important!

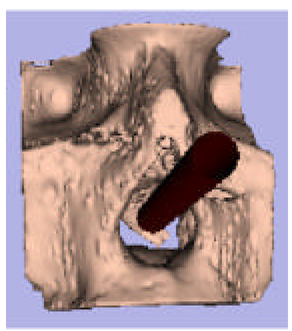
# Goal: Vision-Based Endoscope Navigation



Limited information from the endoscope

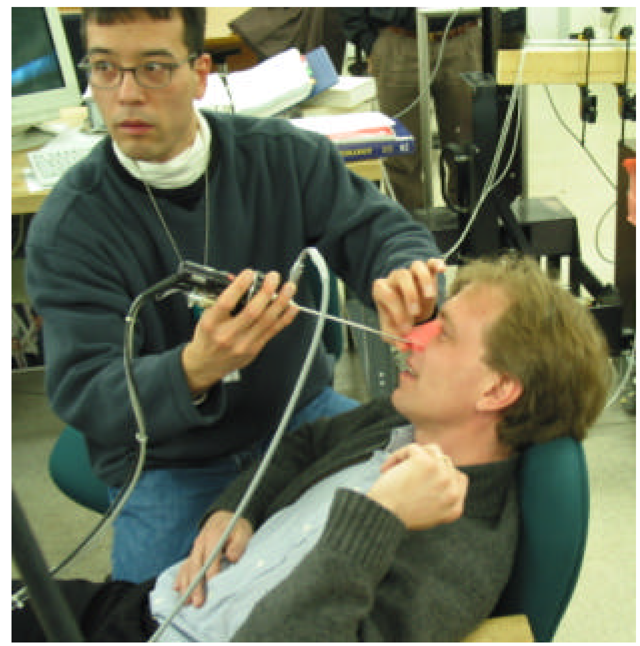


IGS system from GE

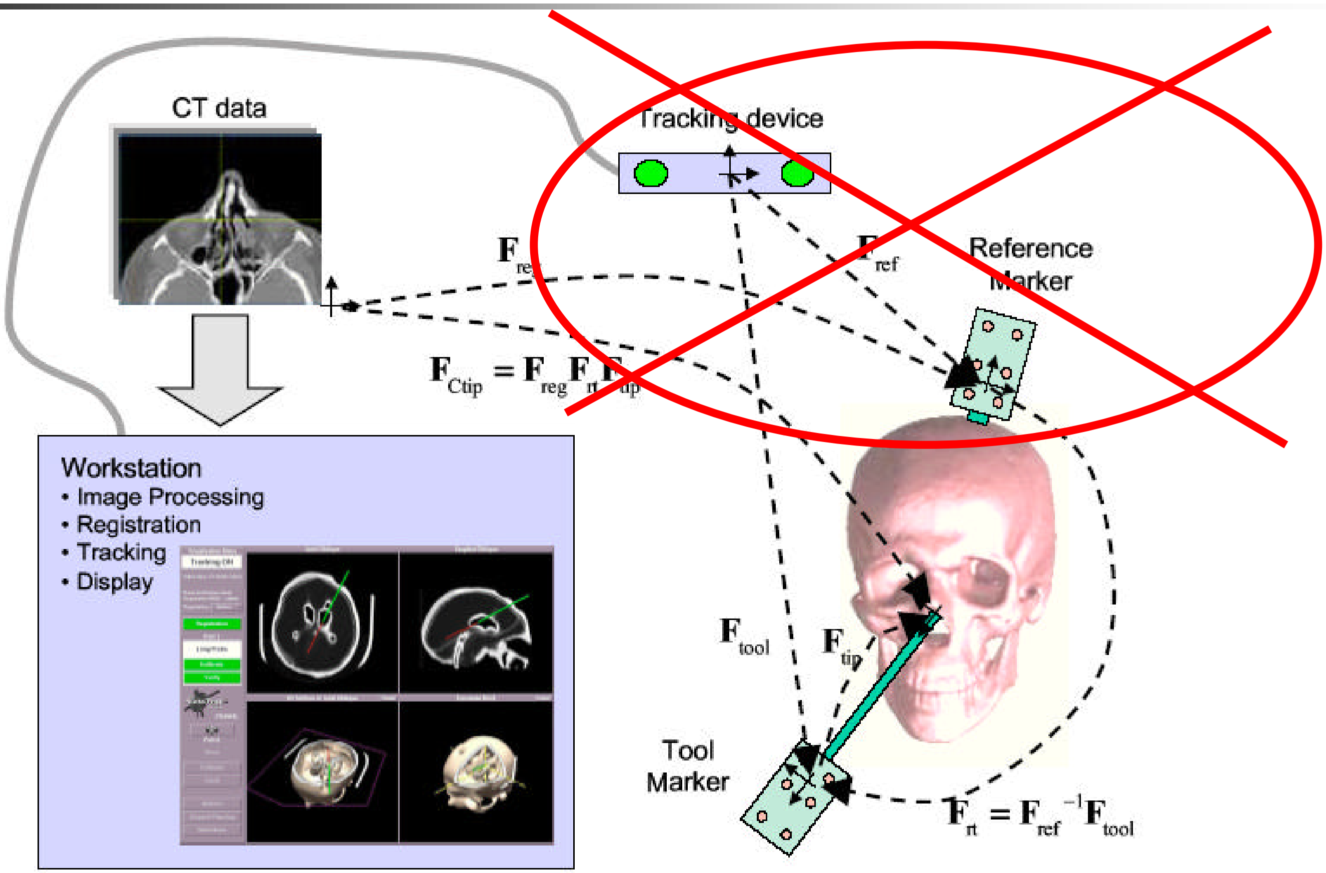


VTK-based 3D-visualization

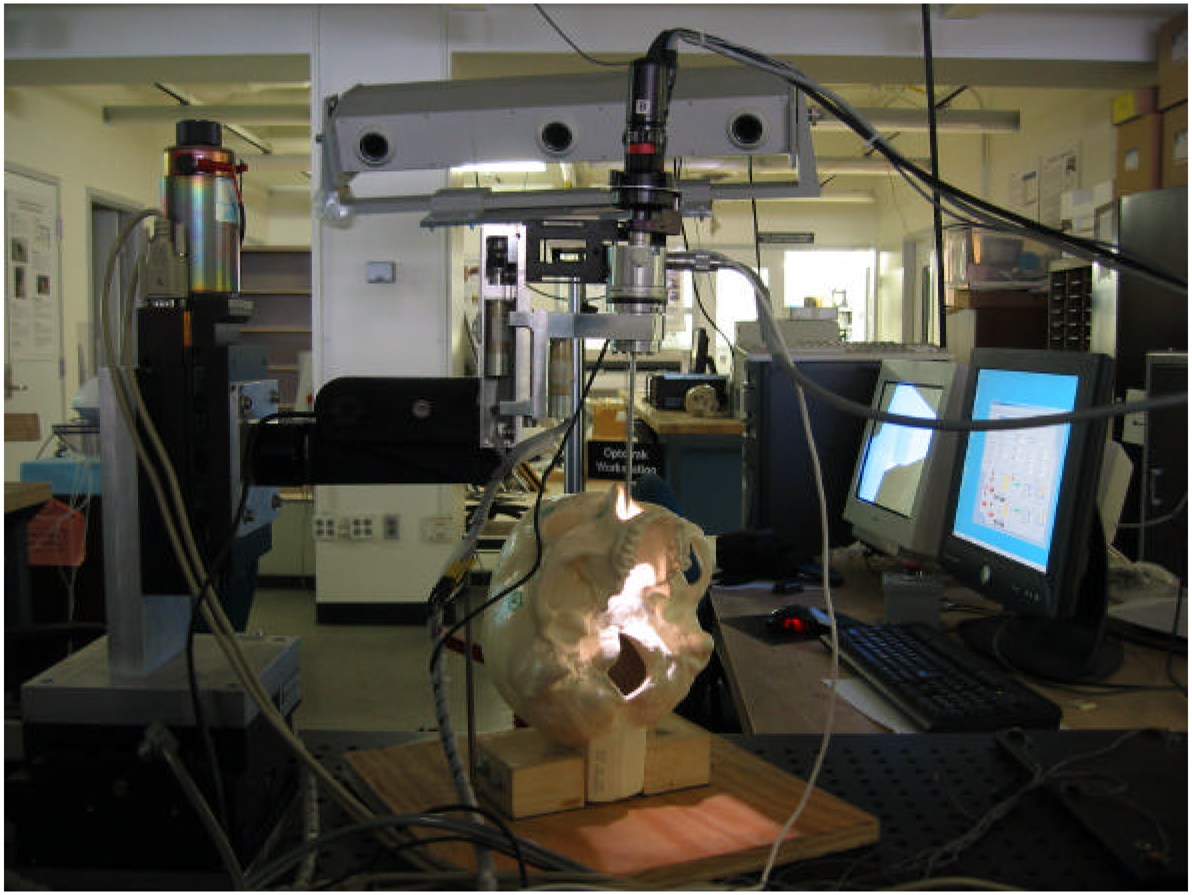
Goal: simplification of the navigation



# Typical Surgical Navigation System



# Experimental Setup



# Related Work

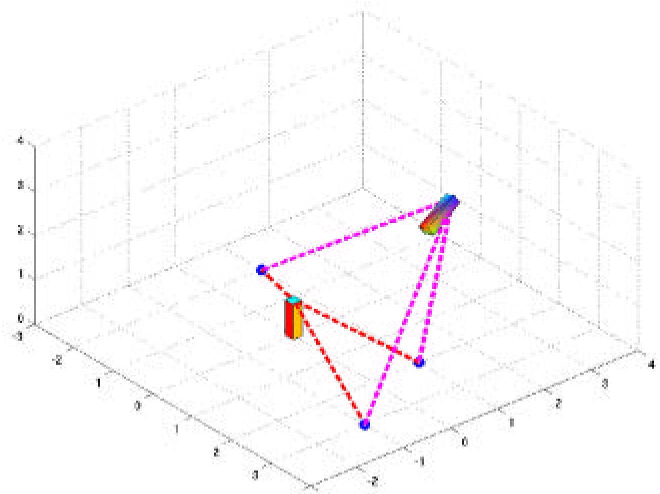

W. M. Wells, P. Viola, and R. Kikinis. Multi-model volume registration by maximization of mutual information. In *MRCAS 97*, pages 55–62, 1997.

D. Dey, P. Slomka, D. Gobbi, and T. Peters. Mixed reality merging of endoscopic images and 3d surfaces. In *Medical Image Computing and Computer-Assisted Interventions*, volume 1935, pages 796–80. Springer Verlag, 2000.

K. Mori, D. Deguchi, J. Hasegawa, J. Toriwaki Y. Suenaga, H. Takabatake, and H. Natori. A method for tracking the camera motion of real endoscope by epipolar geometry analysis and virtual endoscopy system. In *Medical Image Computing and Computer-Assisted Intervention*, volume 2208, pages 1–8. Utrecht: Springer, 2001.

F. Delgianni, A. Chung, and G. Yang. Pq-space based 2d/3d registration for endoscope tracking. In *Medical Image Computing and Computer-Assisted Intervention*, pages 311–318, 2003

# What do we try to solve?



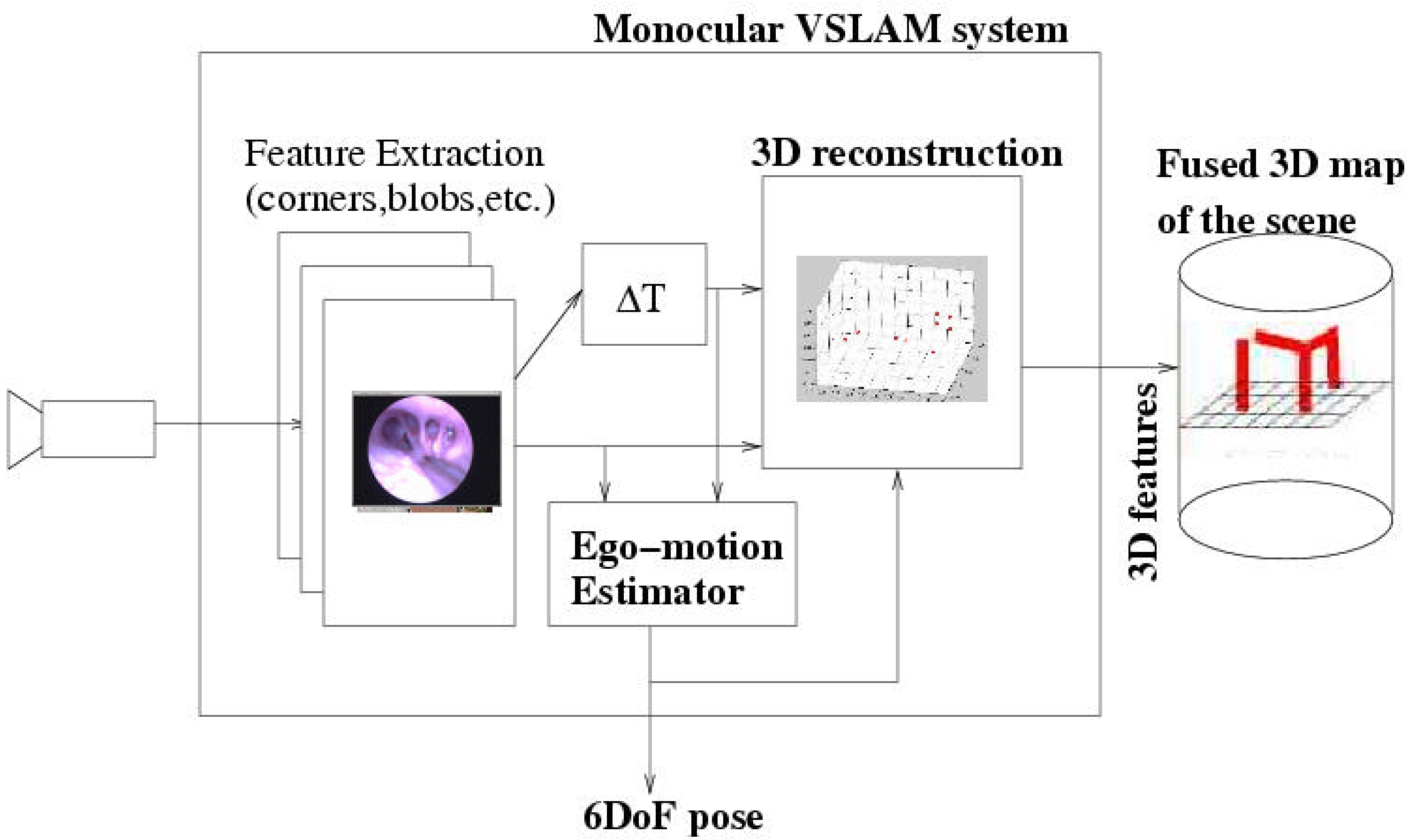
How to estimate the relative translation T and the rotation R between two camera positions

⇒ Camera Ego-Motion

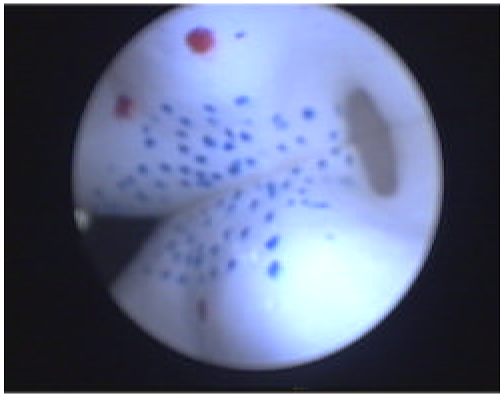
Problem: monocular camera projection reduces the space by one dimension, therefore, external reference (a model) is necessary for 6DoF pose estimation

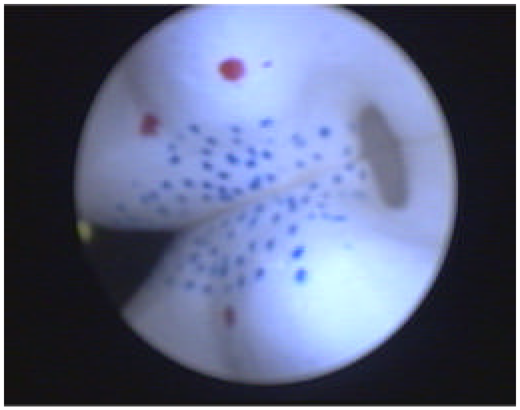
→ SLAM

# VSLAM System



# Initialization Method I
## *Essential Matrix*





$$p_i^* \tilde{\mathbf{E}} p_i = 0$$

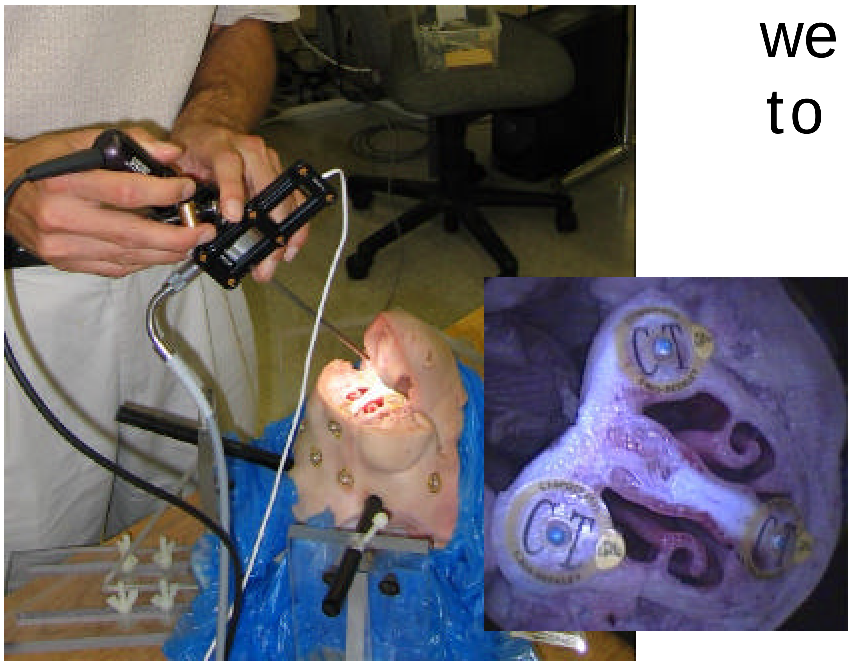
$$\tilde{\mathbf{E}} = \tilde{\mathbf{R}} \cdot \mathrm{sk}(\mathbf{T}),$$

$$\text{with} \quad \mathrm{sk}(\mathbf{T}) = \begin{pmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{pmatrix}$$

Using rigid body assumptions we are able to recover the rotation *R* and a scaled version of translation *T*

⇒ Essential Matrix

# Initialization II
## *Known Reference Pattern*



Using known reference structures, we can calculate the initial distances to the observed points

$$\bar{G} = \frac{1}{n}\sum_{i=1}^{n} G_i, \quad v_1 = G_1 - G_2, \quad v_2 = G_3 - G_2,$$

$$\text{normal vector to} \quad \mathcal{E}: \quad n_e = \frac{v_1 \times v_2}{\|v_1 \times v_2\|} = \begin{pmatrix} n_{ex} \\ n_{ey} \\ n_{ez} \end{pmatrix}$$

$$n_\perp = \frac{(0\ n_{ez}\ -n_{ey})^T}{\|(0\ n_{ez}\ -n_{ey})^T\|} \Rightarrow \tilde{\mathbf{R}}_e = ((n_\perp \times n_e)\ n_\perp\ n_e)$$

$$\pi(G_i^*) = \begin{pmatrix} u_i \\ v_i \end{pmatrix} = \frac{f}{z_i}\begin{pmatrix} x_i \\ y_i \end{pmatrix}$$

Points $G_i$ define the initial distances and their projections

# Scale Preserving Handoff-Process

$$\left(\tilde{\mathbf{R}}\mathbf{n_x} \quad -\mathbf{n_x^*}\right)\begin{pmatrix}\lambda_x \\ \lambda_x^*\end{pmatrix} = \tilde{\mathbf{R}}\lambda_1\mathbf{n_1} - \lambda_1^*\mathbf{n_1^*} \qquad (7)$$

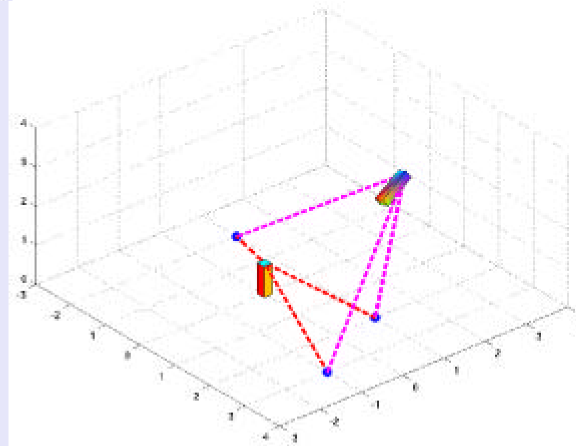
or in a more robust way from 3 frames to (8)

$$\begin{pmatrix}\tilde{\mathbf{R_1}}\mathbf{n_x} & -\mathbf{n_x^*} & 0 \\ \tilde{\mathbf{R_2}}\tilde{\mathbf{R_1}}\mathbf{n_x} & 0 & \mathbf{n_x^{**}}\end{pmatrix}\begin{pmatrix}\lambda_x \\ \lambda_x^* \\ \lambda_x^{**}\end{pmatrix} =$$

$$= \begin{pmatrix}\tilde{\mathbf{R_1}}\lambda_1\mathbf{n_1} - \lambda_1^*\mathbf{n_1^*} \\ \tilde{\mathbf{R_2}}\tilde{\mathbf{R_1}}\lambda_1\mathbf{n_1} - \lambda_1^{**}\mathbf{n_1^{**}}\end{pmatrix}. \qquad (8)$$

# Recursive Ego-Motion Estimation

$$P_i^* = \lambda_i \cdot \begin{pmatrix} \cos\beta_i^* \cos\alpha_i^* \\ \sin\beta_i^* \\ \cos\beta_i^* \sin\alpha_i^* \end{pmatrix} = \lambda_i \cdot \vec{n}_i^*, \quad P_i^* = \lambda_i \cdot n_i^* = \tilde{R} \cdot P_i + \vec{T}$$



A solution to the following least-squares problem

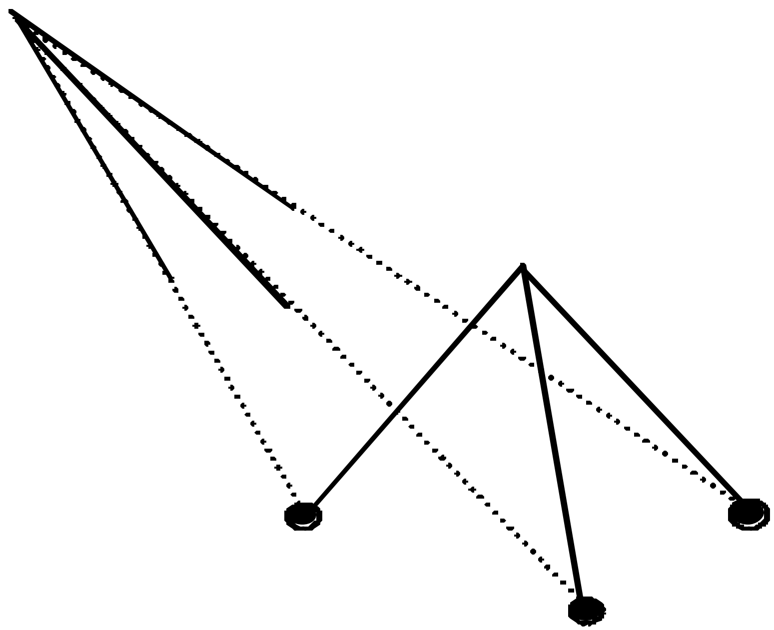
$$\min_{\tilde{R},\vec{T}} \sum_{i=1}^{n} \|\tilde{R}P_i + \vec{T} - P_i^*\|^2, \quad \text{subject to } R^T R = I.$$

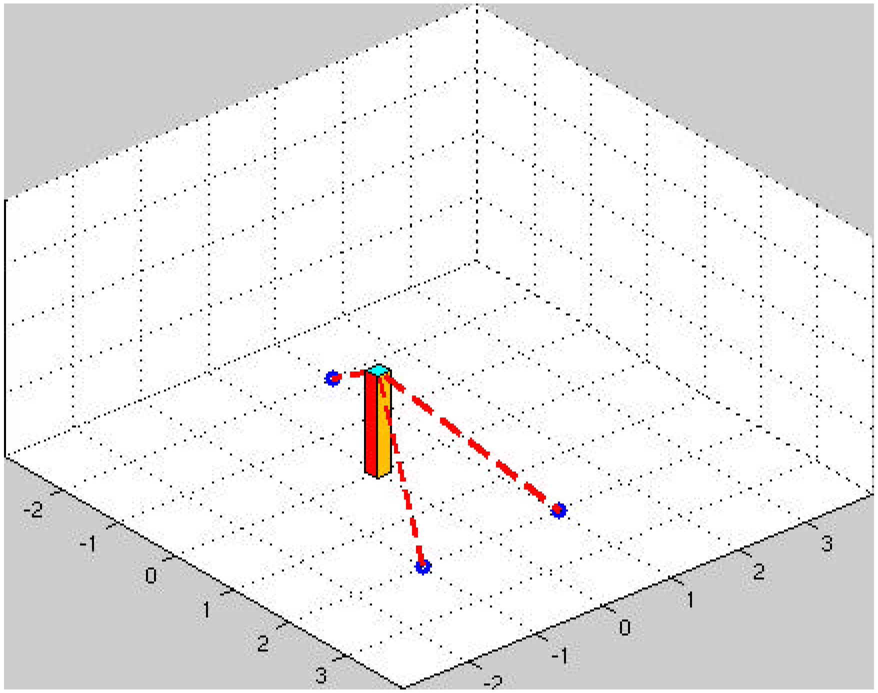
is

$$M = \sum_{i=1}^{n} P_i'^* P_i'^T, \quad \text{with } P_i' = P_i - \frac{1}{n}\sum_{i=1}^{n} P_i, \quad P_i'^* = P_i^* - \frac{1}{n}\sum_{i=1}^{n} P_i^*$$

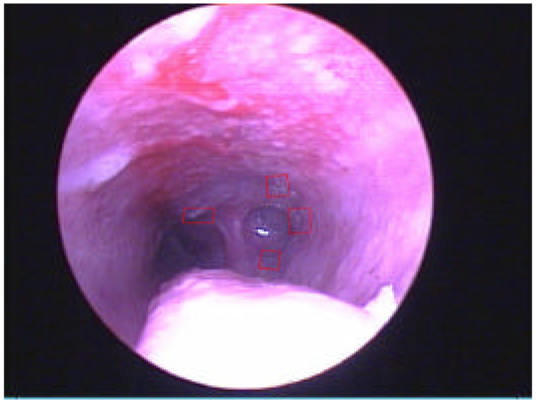
$$(U,\Sigma,V) = \text{svd}(M) \Rightarrow R^* = VU^T, \quad \vec{T}^* = \bar{P}^* - R^*\bar{P}$$

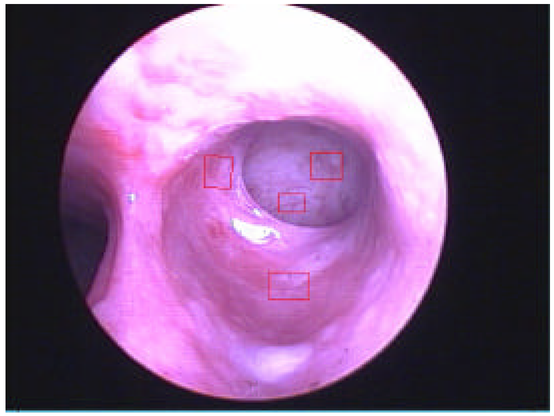
# Convergence Visualization

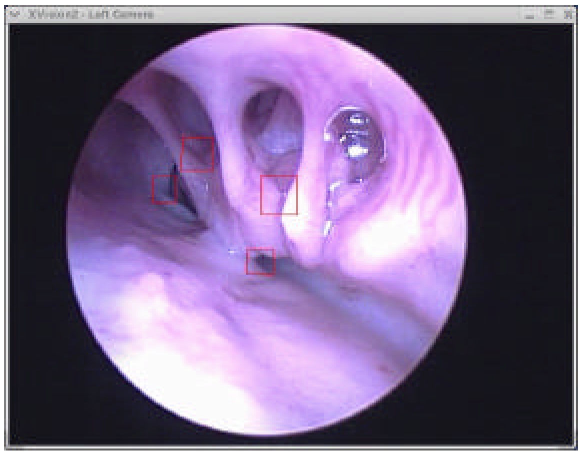


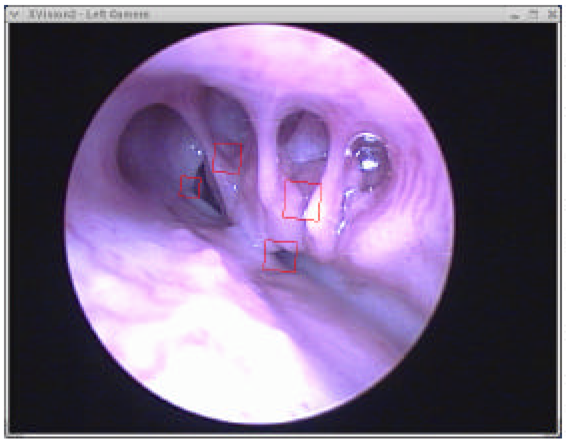


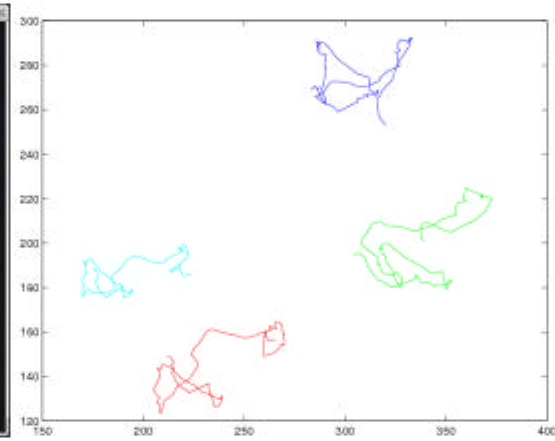
# Tracking Primitives in Real Sinus (preliminary results)



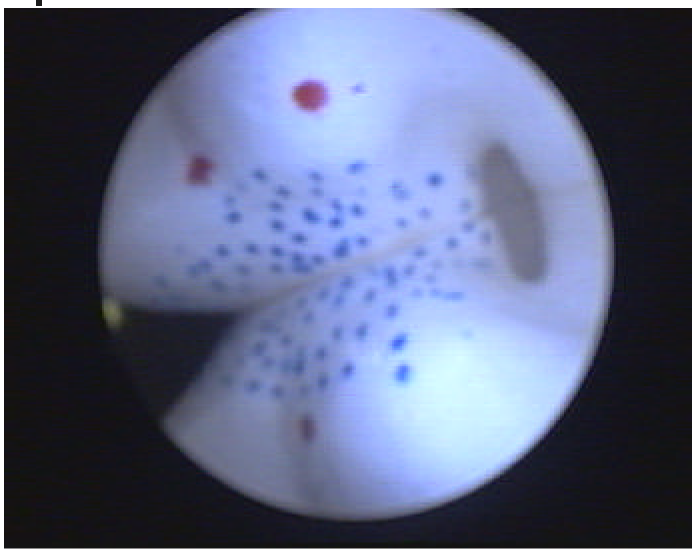


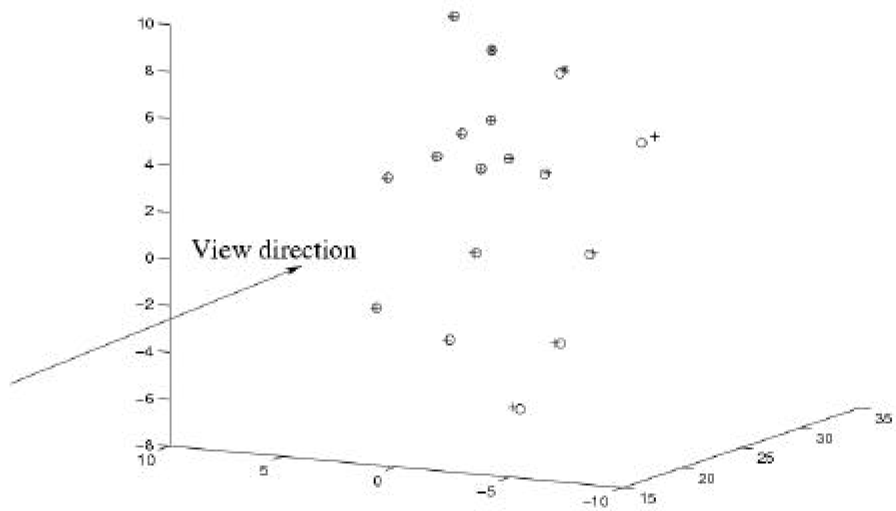






# 3D Reconstruction

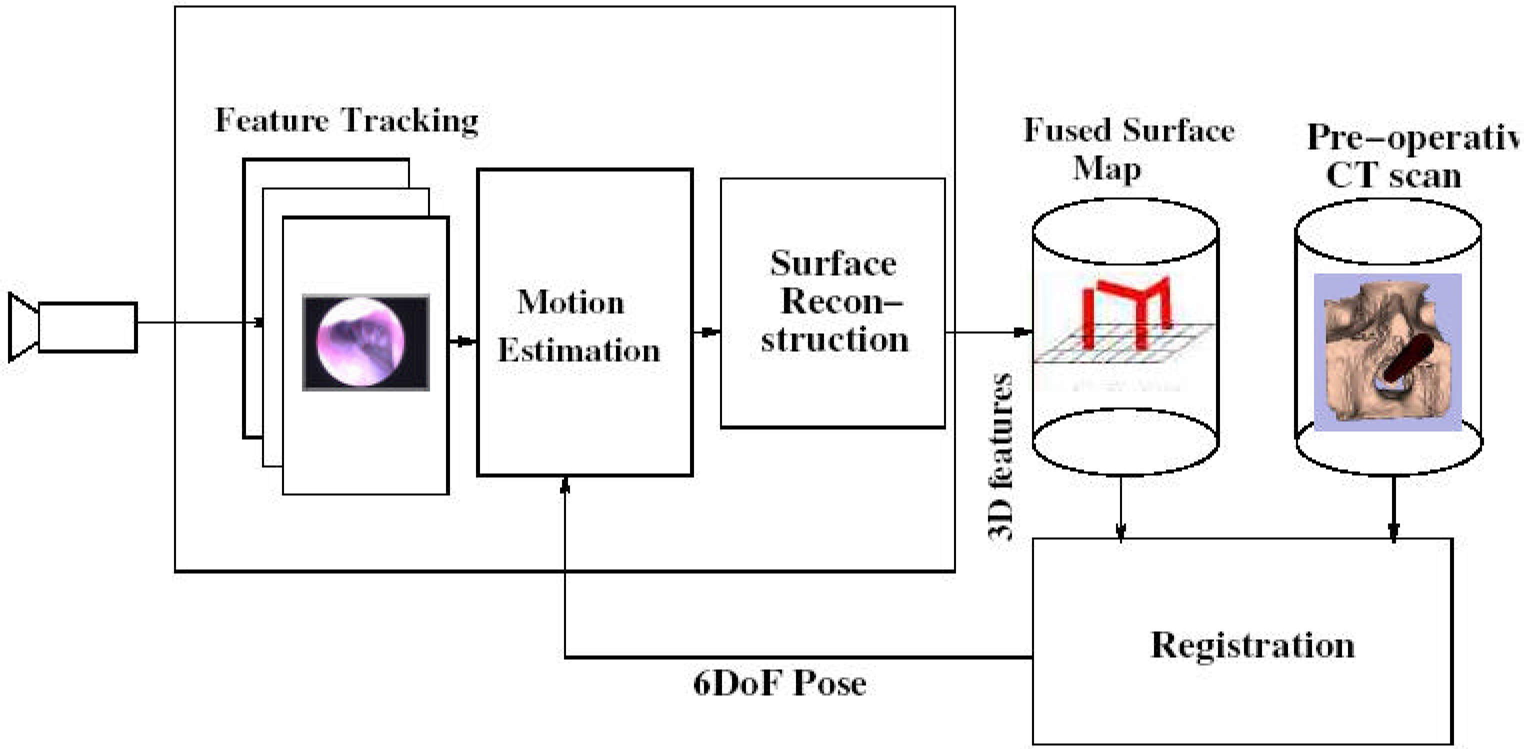




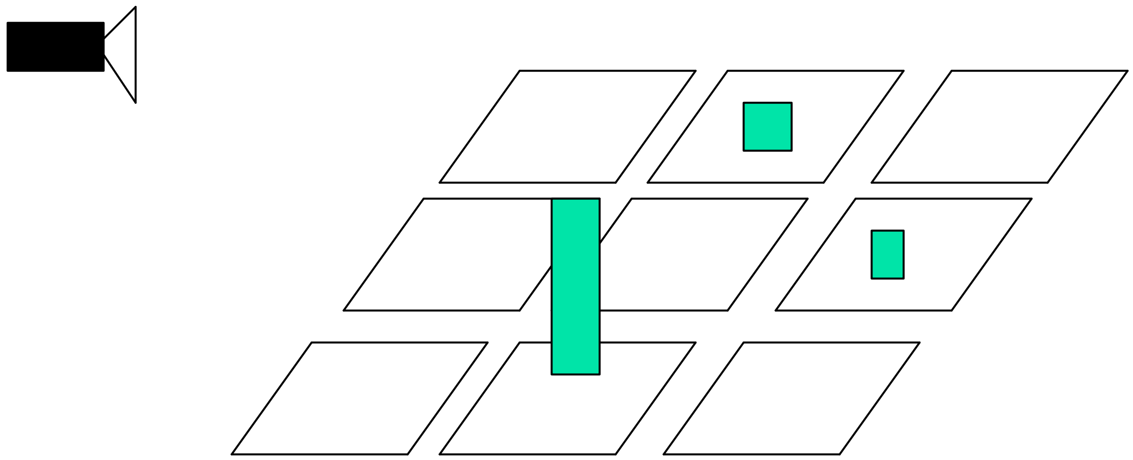

$$\frac{D_i^*}{m}\mathbf{n}_i^* = \tilde{\mathbf{R}} \cdot \frac{D_i}{m}\mathbf{n}_i + \frac{\mathbf{T}}{m} \quad \Rightarrow \quad \begin{pmatrix} \lambda_i^* \\ \lambda_i \end{pmatrix} = \left(\mathbf{n}_i^* \quad -\tilde{\mathbf{R}}\mathbf{n}_i\right)^{-1} \cdot \frac{\mathbf{T}}{m}$$

*Can be used for omnidirectional camera as well!*

# System Structure



# Generic Problem
# Surface Structure Recovery



# Scale Recovery (PCA)

Smallest Eigenvalues ($E_{ct}$,$E_{3d}$) describe the depth variation. We use them for estimation of m

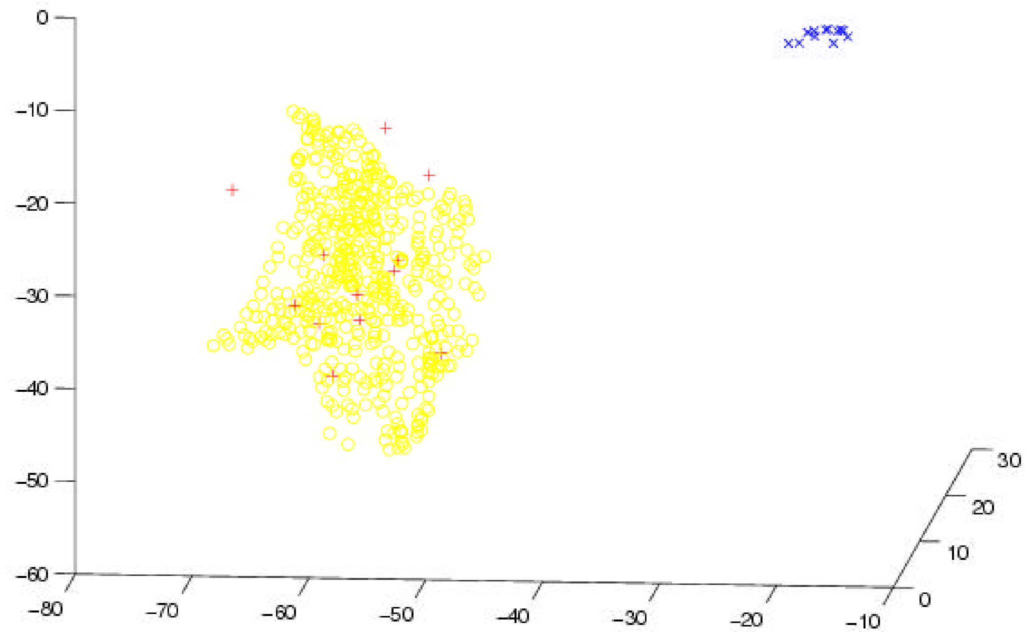
$$m = \frac{\sqrt{E_{ct}}}{\sqrt{E_{3d}}},$$

$$V_{p\in\{CT,3D\}} = (V_{px}\ V_{py}\ V_{pz})^T,$$

$$V_{n-p} = \frac{(0\ V_{pz}\ \ -V_{py})^T}{||(0\ V_{pz}\ \ -V_{py})^T||}$$

$$\tilde{R}_{ct} = \left(\ (V_{n-ct} \times V_{ct})\quad V_{n-ct}\quad V_{ct}\ \right),$$
$$\tilde{R}_{3d} = \left(\ (V_{n-3d} \times V_{3d})\quad V_{n-3d}\quad V_{3d}\ \right),$$

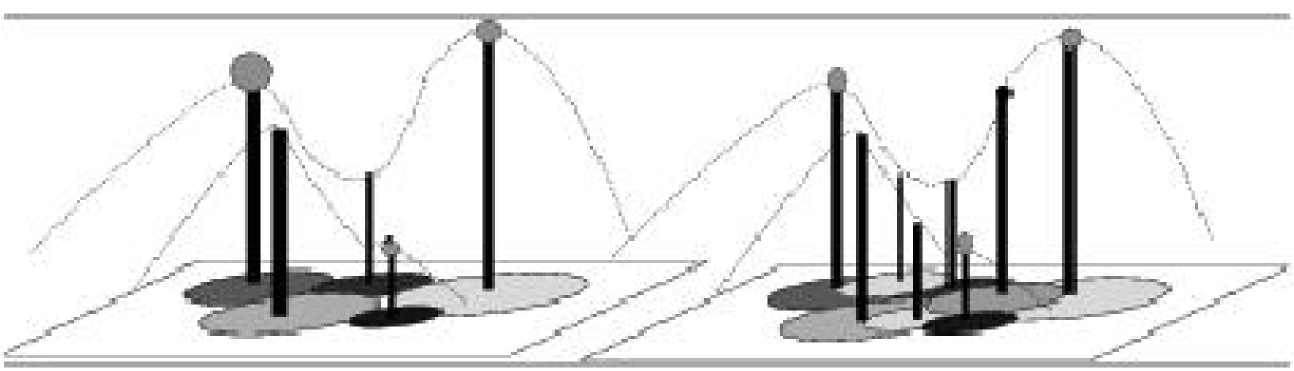
$$\tilde{R}_{tot} = \tilde{R}_{3d} \cdot \tilde{R}_{ct}^T$$



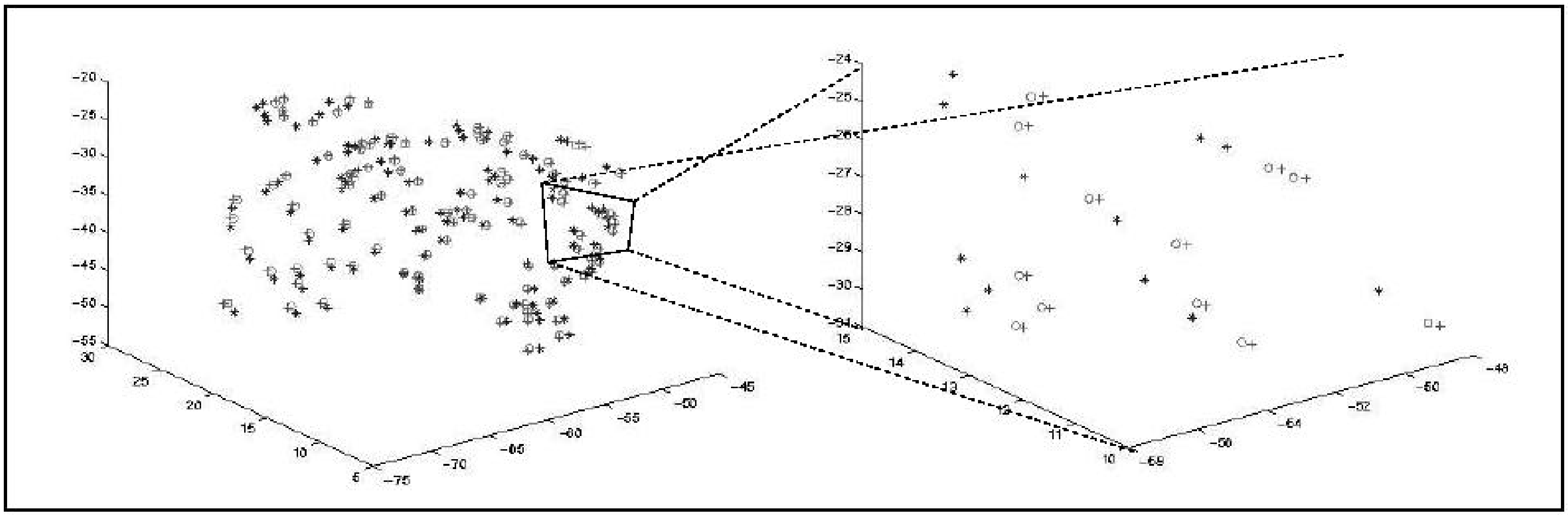

Patches are aligned up to an unknown rotation in the plane of the surface and translation within it

# Fine alignment within the surface

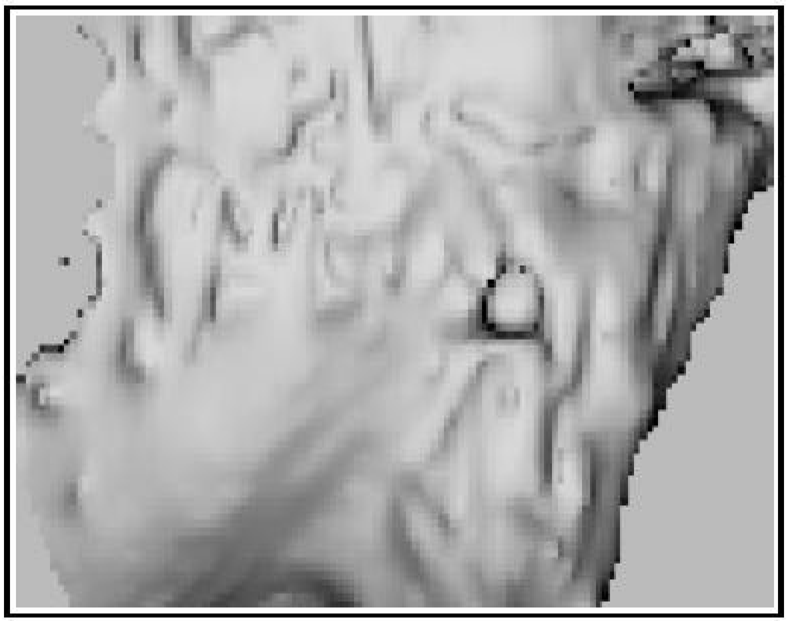
We consider the depth variation as pseudo-images and use standard image processing techniques (SSD) to find the rotation within the plane and the shift

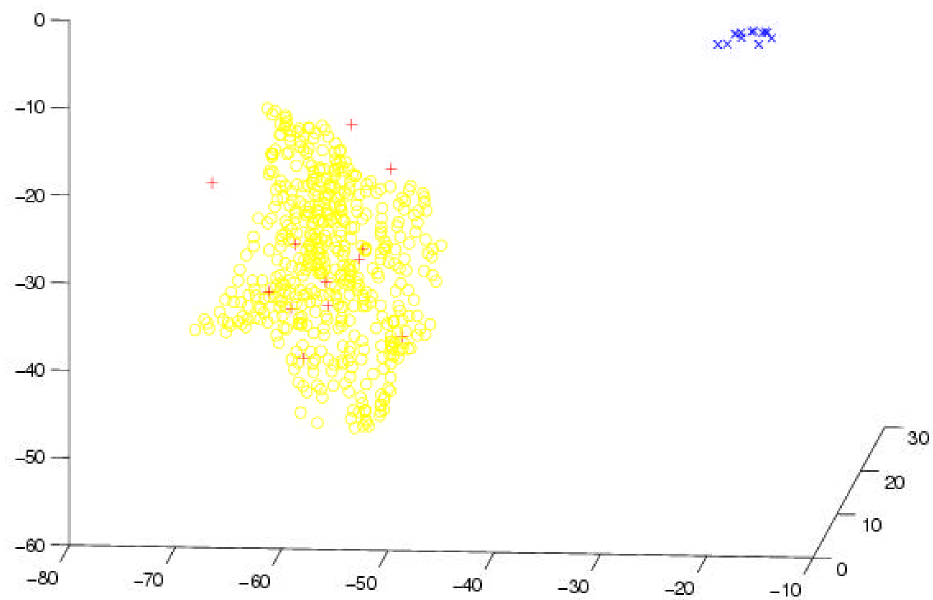


# Reconstruction Result after ICP



# Patch





# Conclusions

- The presented system allows an accurate reconstruction of 3D surface points and their registration to 3D surface data from CT scans or laser-range finder reconstructions.
- After ICP alignment, the average distance error for the sample points is around 0.65mm. This compares favorably to the fiducial-based registration, whose residual error is around

  0.40mm for four fiducials that are attached to the surface of the skull., because the target residual error (TRE) calculated from fiducial residual error (FRE) is around 1.25mm.

# Future Work

- We are currently investigating the feature type that can be used for a robust estimation and tracking of our *point features* in real endonasal images obtained in a preliminary experiment from a pig's head.
- Automatic Feature Selection in real endonasal images is another important future work