## A. Statement of Purpose

Minimally invasive surgery in which an endoscope is used can offer great benefits to both patients and surgeons. Such system grants surgeons field of view of the surgical scene and for patients, less trauma and minimal incisions. Despite such advantages, there is only so much the nature of endoscopes allows for visualization; mainly, the narrow and restricted field of vision compared to other imaging systems. The purpose of this paper is to provide a framework for reconstructing 3D organ surfaces from endoscopic videos. Hence, no tracking system will be needed to obtain a visual representation of the surgical scene. Such 3D information could also be used for pre-operative planning and image guided surgery.

## B. Technical Summary

3D reconstruction of surface in minimally invasive surgery only using endoscopic images is an exciting concept but often times can be vulnerable to camera noise, missing data values, and badly tracked outliers, hindering robust visualization of the scene. This paper describes on the basis of a projective reconstruction which excludes the assumption that the camera internals are same throughout, a feature-tracking algorithm using image alignment for accurate feature extraction, an outlier removal procedure using trifocal tensor with three images, and a reconstruction algorithm.

### B.1 Feature Tracking

First step to reconstruct 3D structures is to track feature points along with the movement of the endoscope. One of the well-known methods that is used here is the Lucas-Kanade tracking algorithm. The main idea is to minimize the sum of squared differences between two images or frames of a video, I_k and I_k+1 which is warped back onto the frame of the first image I_k. In short, the purpose is to align an input image I_k+1 to the template image, I_k in this case. The optimization fuction is given as:

$$\min \sum_{\mathbf{x}} [I_{k+1}(W(\mathbf{x};\mathbf{p}+\Delta\mathbf{p})) - I_k(\mathbf{x})]^2$$

Here, W(x;p) represents warping of the feature pixel at **X** by the warping parameter **p** which is the optical flow in this case. This expression can be linearized using first order Taylor expansion.

$$\min \sum_{\mathbf{x}} \left[ I_{k+1}(W(\mathbf{x};\mathbf{p})) + \nabla I_{k+1} \frac{\partial W}{\partial \mathbf{p}} - I_k(\mathbf{x}) \right]^2$$
,

where $\nabla I_{k+1} = gradient\ descent$

$$\frac{\partial W}{\partial p} = Jacobian\ of\ warping\ function$$

Upon minimization, a closed form solution for the warping parameter **p** can be obtained which provides information about movement of the feature. It is possible that some features are only detected and tracked for few frames, which can cause complication in reconstruction process. Such data have to be recovered. Using epipolar geometry and through camera calibration, camera matrix of an image containing internal and external parameters and respective projective depth of the feature points in the image can be obtained: this can be formulated as such.

$$\underbrace{\begin{bmatrix} \lambda_1^1 \mathbf{x}_1^1 & \cdots & \lambda_n^1 \mathbf{x}_n^1 \\ \vdots & \ddots & \vdots \\ \lambda_1^m \mathbf{x}_1^m & \cdots & \lambda_n^m \mathbf{x}_n^m \end{bmatrix}}_{\mathbf{M}} = \underbrace{\begin{bmatrix} \mathbf{P}^1 \\ \vdots \\ \mathbf{P}^m \end{bmatrix}}_{\mathbf{P}} \underbrace{\begin{bmatrix} \mathbf{X}_1 & \cdots & \mathbf{X}_n \end{bmatrix}}_{\mathbf{X}}$$

Then, the solution is to solve for P and X that minimize the residual, a squared difference, between PX and M. However, due to missing data values, an operation of fitting an unknown matrix to the noisy matrix M must be accomplished using incomplete submatrices of M, linear independent columns of M, as they can provide stronger constraints.

## B.2 Outlier Removal

The authors describe an outlier removal method using a trifocal tensor with three images. The trifocal tensor captures the internal geometry and motion information of three independent views which can provide stronger constraints to image points and allow for better detection of outliers. It must be noted that there has to be substantial motion between frames as well as overlapped regions so that such method delivers a reliable result.

Given three poses characterized by their camera (projective) matrices, the points of interest in each image can be defined as X for first image, X' for second image, and X'' for third image in a homogenous frame. Then obtaining camera projections, such as V' and V'' which are defined as projection of first camera into second and third camera, can yield a compact expression of the trilinear constraints across the three perspectives in terms of the trifocal tensor T.

$$\mathbf{T}_i^{jk} = \mathbf{v}'^j \mathbf{b}_i^k - \mathbf{v}''^k \mathbf{a}_i^j, \quad i,j,k = 1,2,3$$

, where a = frame from camera 1 to 2
b = frame from camera 1 to 3

A method for detecting outliers using this approach is also explained in the literature. The authors implemented an outlier detection system based on the geometric error. In a more mathematical sense, it measures the sum of squared differences between the image points and the corrected points using trilinear constraint given above, and the measure of error is used to detect outliers.

$$R = \sum_{i=1}^{n} R_i = \sum_{i=1}^{n} d(\mathbf{x}_i, \hat{\mathbf{x}}_i)^2 + d(\mathbf{x}'_i, \hat{\mathbf{x}}'_i)^2 + d(\mathbf{x}''_i, \hat{\mathbf{x}}''_i)^2$$

To apply this method to more than three images for outlier detection, the authors implemented a method to process an entire sequence of images both sequentially and non-sequentially:
(j,j+1,j+2) AND (j,j+1,j+3) etc.

**B.3 Reconstruction**

In order to effectively choose a selection of sub-matrices which provide stronger constraints for image points, the authors describe a method of competitive evolutionary agents to solve reconstruction problem with missing data and outliers. The agents represent the subset of the columns used to construct the sub-matrices in feature tracking section. The agents behave in an evolutionary way, such as reproduction and diffusion, to search for good candidates from a vast pool of samples.

$$Agent = \left\langle \mathbf{V},\ a,\ F_{fitness},\ fml,\ Rep,\ Diff,\ Die \right\rangle$$

where V indicates the chosen column from the measurement matrix of the feature points.

The values a, F_fitness, and fml represent the internal state of an agent and Rep, Diff, Die represent the external. *Rep* denotes reproduction of finite number of offspring agents, *Diff* search for new position in the environment for offspring agents upon fitness comparison, and *Die* vanishing process of agents.

Upon recovery of a complete measurement matrix, the authors used Han and Kanade's method to update a projective reconstruction into Euclidean reconstruction. In short, if measurement matrix M and image points X are given in projective reconstruction, other reconstruction is of the form ({MH^-1},{HX}) and optimization problem is set up to solve for a matrix H.

**C. Results**

The authors used both synthetic and phantom data of the heart, and the proposed method retains a relatively good accuracy even when the measurement matrix contains big noise.

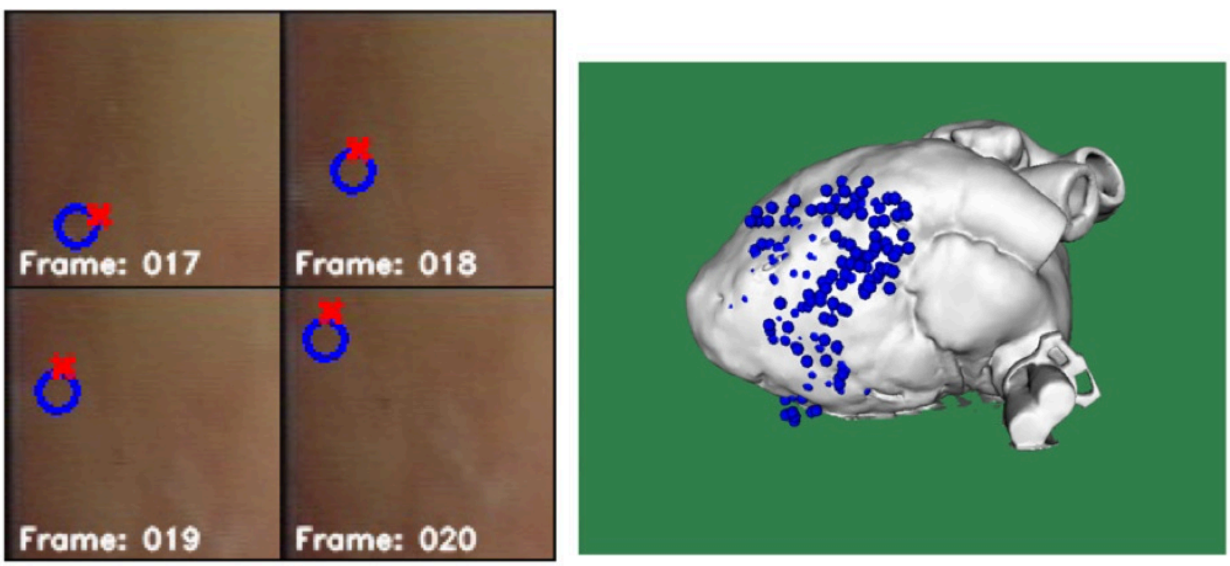


Figure 1. L) The blue dots indicate detected points using feature tracking method. Red crosses represent filled in data points. R) Image of registered 3D point clouds on to CT heart model through ICP.

**Table 4**
Proposed method: number of features and RMS error at different stages of 3D reconstruction.

| | Features | | | RMS error (pixel) | | |
|---|---|---|---|---|---|---|
| | Tracked features | Missing data level | Outliers removed | Without outlier removal | With outlier removal | Final refinement |
| Phantom data test | 401 | 48.8% | 59 | 2.41 | 1.69 | 1.22 |
| *In vivo* data test | 397 | 45.8% | 78 | 3.98 | 1.42 | 0.89 |

## D. Importance

- The proposed method is achieved solely using 2D endoscopic images, not requiring additional imaging systems or tracking devices.

- The proposed method is robust to measurement noises, missing data (up to 50%), and outliers.

- The proposed method provides further application for surgical planning and image-guided surgery.

## E. Relevance

| | Our project | Reviewed project |
|---|---|---|
| Similarities | Minimally Invasive Surgery | |
| | For improvement of surgeon's field of view during surgery. | |
| | Use of 2D endoscopic images to reconstruct a 3D scene. | |
| Differences | Robo-ELF is monoscopic | Recommends stereo data |
| | Area of interest: Airway | Area of interest: Heart |

## F. Critique

| Strengths | Weaknesses |
|---|---|
| Problem statement of the project is highly relevant to our research. | It is not clear how data filling procedure is propagated. |
| Provides insights in optimization in data selection prior to reconstruction. | Mathematically complex and challenging. |
| Covers synthetic data, phantom data, and in vivo data. | |

**G. Reference**

M.X. Hu et al. 2012. Reconstruction of a 3D surface from video that is robust to missing data and outliers: Application to minimally invasive surgery using stereo and mono endoscopes. In: Medical Imaging Analysis 16, pp. 597-611.