Group 13: Software for an Intra-Operative Kinect With a Flexible Endoscope
Advanced Computer Integrated Surgery, Spring 2017

Seminar Presentation Paper Critical Review
Student: Elli Tian

Background

The increased availability of "depth cameras" such as the Microsoft Kinect has increased the popularity of research on real-time 3D reconstruction. By using a small camera and a laser fiber that emits a laser pattern similar to that of the Kinect, we are working to develop portable software that will allow for a real-time reconstruction of an intra-operative surgical space. In order to do this, we need to be able to simultaneously identify important features in the operative space and track the motion of the laser-camera setup. We hope to achieve this by adapting an implementation of the Simultaneous Localization and Mapping (SLAM) algorithm, which is discussed in the papers examined below.

Paper 1: A Solution to the Simultaneous Localization and Map Building (SLAM) Problem

The main goal of this paper was to prove both mathematically and experimentally that a solution to the SLAM problem exists. Given an autonomous vehicle or tracker in an unknown location in an unknown environment, it is possible to reconstruct a map of the environment while also identifying and keeping track of the location of the vehicle. This was shown by the proof of three results:

1. Uncertainty in the relative map estimates reduces monotonically.

The most popular approach to solving the SLAM problem is a Kalman filter-based approach, which allows for both a recursive solution and a means of determining consistent estimates for uncertainty in location measurements. In order to maintain consistency of the filter, one must account for the fact that estimates of relative landmark locations are correlated. This results from a common error in the estimation of the vehicle location. For the purposes of this proof, the authors assumed that the landmarks were stationary and that the movement of the vehicle and observation of landmarks could be modeled linearly and in discrete time; more realistic, non-linear movement could be modeled with the linearization of an extended Kalman filter. The following were defined:

- $x_v(k)$: state of the vehicle at time step k
- p_i: location of the ith landmark
- $z_i(k) = H_i x(k) + w_i(k)$: observation model for the ith landmark
- $P(i|j)$: the state covariance matrix

The determinant of the state covariance matrix is related to the uncertainty associated with any particular state estimate. Since the landmarks were assumed stationary and the algorithm is initialized using a positive semidefinite (*psd*) matrix, it is possible to prove that the state variance at any time step *k+1* is less than that at time step *k*. Therefore, with each additional measurement, the error in the estimate of landmark location decreases.

2. In the limit, uncertainty in the map estimates converges to zero.

The map covariance matrix at time step *k+1* can be written as the difference between the map covariance matrix at time step *k* and a *psd* matrix.

$$\begin{aligned}\mathbf{P}_{mm}(k+1|k+1) =& \mathbf{P}_{mm}(k+1|k) - \mathbf{M}_2\mathbf{S}_i^{-1}{\mathbf{M}_2}^{\mathrm{T}} \\ =& \mathbf{P}_{mm}(k|k) - \mathbf{M}_2\mathbf{S}_i^{-1}{\mathbf{M}_2}^{\mathrm{T}}\end{aligned}$$

One can then evaluate that the block columns of the map covariance matrix are linearly dependent, and therefore that as *k* approaches infinity, the determinant of the covariance matrix converges to 0. In this case, the map is fully correlated and the relative location between landmarks can be known with certainty.

3. Uncertainty in the vehicle and absolute map and landmark locations achieves a lower bound determined by the error in the initial vehicle location estimate.

Following from the previous conclusion, one can conclude that the limiting covariance of any single landmark estimate is obtained when the vehicle is stationary. Thus, the state covariance matrix can be written as the following:

$$\mathbf{P}(k|k) = \begin{bmatrix} \mathbf{P}_{0v} & \mathbf{P}_{0v}{\mathbf{H}_v}^{\mathrm{T}}\left[{\mathbf{H}_{p1}}^{\mathrm{T}}\right]^{-1} \\ \mathbf{H}_{p1}^{-1}\mathbf{H}_v\mathbf{P}_{0v} & \mathbf{H}_{p1}^{-1}\mathbf{H}_v\mathbf{P}_{0v}\mathbf{H}_{p1}^{-1}{\mathbf{H}_v}^{\mathrm{T}} + \mathbf{Y}_3 \end{bmatrix}$$

In the limit, this matrix gives the lower bound of a single landmark state estimate variance as the following:

$${\mathbf{P}_{p_i,p_i}}^{\infty} = \min_{i\in[1,N]}\left\{\mathbf{H}_{pi}^{-1}\mathbf{H}_v\mathbf{P}_{0v}\mathbf{H}_{pi}^{-1}{\mathbf{H}_v}^{\mathrm{T}}\right\}$$

This value is determined only by the initial covariance in the vehicle location estimate. The landmark uncertainty can be reduced if more than one landmark is observed.

These mathematical results were also validated with a physical experiment, in which a test vehicle was fitted with millimeter wave radar (MMWR), a linear variable differential transformer (LVDT), and a differential GPS system. The testing environment contained both radar reflectors and natural landmarks, which were assumed to be stationary and whose locations were estimated using an implementation of the SLAM algorithm. Landmark detection quality was validated using an algorithm that labeled “tentative” landmarks when they were first identified and “confirmed” landmarks when a sufficiently accurate measurement was obtained.

A survey of the true landmark locations showed that the absolute accuracy of the vehicle path was approximately 5 cm. In addition, the uncertainty in location landmark estimates decreased monotonically, as shown by the graph below of standard deviation of the estimates over time:

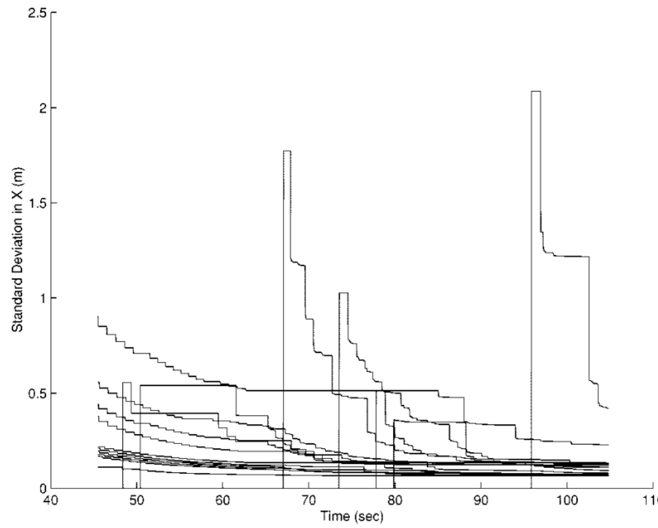


Fig. 12. Decreasing uncertainty in landmark location estimates.

The authors indicated that while they were able to develop a rigorous proof for the existence of a solution to the SLAM problem, an ideal implementation of such a solution would require the maintenance of a full state vector that contained all of the states of both the vehicle model and every landmark in the map. This would be particularly unmanageable for both large environments with many features and confined spaces in which landmarks are not well defined, such as the intra-operative space. Thus, for our situation, the effectiveness of a SLAM algorithm depends greatly on the existence of a reliable, efficient, and accurate method for detecting and matching features in a confined, deformable, and variably lighted environment.

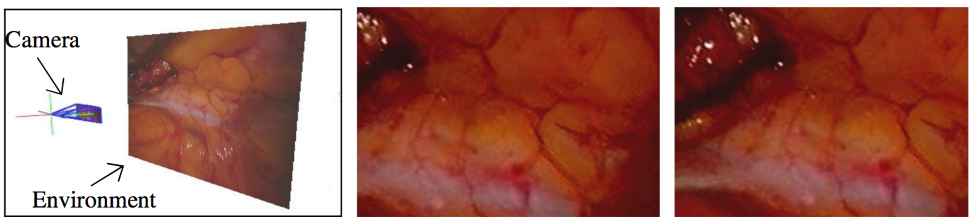
Paper 2: Simultaneous Stereoscope Localization and Soft-Tissue Mapping for Minimal Invasive Surgery

This paper presented a means of using SLAM to create a 3D map during minimally invasive surgery (MIS). Traditionally, MIS procedures can be difficult for a surgeon to navigate manually due to the small operating space and limited camera view, resulting in a loss of 3D vision and navigation. Using a stereoscopic laparoscope and a sequential vision only approach, the authors were able to develop an algorithm that accurately maps soft-tissue structures that may be encountered during surgery.

A Kalman filter approach was used, and the following were defined:

- $\hat{x}_v$: state of the camera
- $\hat{y}_i$: states of the features that make up the map (landscape)
- P: a full feature covariance matrix, representing the first order uncertainty of all quantities in the state vector

The stereoscopic laparoscope was pre-calibrated before the surgical procedure, assuming that the center of the camera system was located at the left camera. The camera motion was controlled so that inter-frame pixel motion was no more than 20 pixels, reflecting actual expected movement, and a single plane with a realistic image rendition was used to model the *in vivo* space as shown below:



(a) (b)

Fig. 3. An illustration of the simulation environment used to generate a stereo-laparoscopic video with known ground truth data for camera motion. A 3D rendition of the virtual world is shown in **(a)** and an example stereo pair taken from the virtual cameras is shown in **(b).**

The algorithm was also implemented on video footage, which was then reversed to assess the accuracy of camera tracking. The number of visible features was thresholded to reduce reliance on weak detections, and features were matched between left and right stereo images using a normalized sum-of-squares difference and epipolar geometry. The state of the camera and the covariance matrix were then updated by two steps.

1. Prediction step

This step uses a statistical motion model to calculate camera movement, taking into account the unknown intentions of the camera operator. Firstly, a deterministic element was imposed, assuming that on average the camera moves with constant velocity and angular velocity. Secondly, the uncertainty in movement was modeled stochastically with a Gaussian profile, assuming that on average the camera acceleration remains small. The covariance matrix *P* can be designed to reflect the expected movement of the camera.

2. Measurement step

This step describes how new input can reduce uncertainty in position measurements predicted by SLAM. Given the position of a 3D feature relative to the camera, one can calculate the predicted position of the feature in world coordinates and then search around this predicted position to find the actual feature. Thus, the predicted SLAM features can be compared with the visual input from the stereoscope.

Using the proposed technique, the authors were able to reconstruct a SLAM map of the environment and track the motion of the laparoscope in a way that was robust to large changes in acceleration and in appearance. In addition, tracking on reversed camera footage was able to accurately follow the camera back to its starting position. This particular algorithm and application is not yet efficient in real time, but it is possible to increase its accuracy by incorporating all available information about the structure, behavior, and expected deformations of the intra-operative space.

Conclusions

Both of these papers describe how solutions to the SLAM problem can be used to produce an effective and accurate reconstruction of a 3D environment. Our experimental setup, using a single camera and the laser fiber, differs slightly from the implementations discussed in these papers as we are not able to obtain stereo images and will also be able to construct a depth map from the observed laser light pattern. However, a variation on the SLAM algorithm would still allow us to map key features in the intra-operative space while tracking the movement and location of the laser-camera setup.

References

1. Gamini Dissanayake, M.W.M., Newman, P., Clark, S., Durrant-Whyte, H.F., and Csorba, M. "A Solution to the Simultaneous Localization and Map Building (SLAM) Problem." *IEEE Transactions on Robotics and Automation* 17.3. (2001): 229-241. Print.
2. Moutney, P., Stoyanov, D., Davison, A., and Yang, G.Z. "Simultaneous Stereoscope Localization and Soft-Tissue Mapping for Minimal Invasive Surgery." *Medical Image Computing and Computer-Assisted Intervention* 4190. (2006): 347-354. Print.