CIS II Paper Review

Group 15: Mingyi Zheng
04/24/2019

My Project Goal

My project is HMD-based Navigation for Ventriculostomy. Ventriculostomy is a neurosurgical procedure that involves creating a hole (stoma) within a cerebral ventricle for drainage. The goal of the project is to introduce image guidance via augmented reality on HoloLens to reduce the chance. The image guidance is AR overlay of ventricle model from CT image and catheter guide include insertion depth and angle overlay.

Paper Selection and Relevance

The paper I've selected to review is *A fast, low-cost, computer vision approach for tracking surgical tools* by R. Dockter, R. Sweet and T. Kowalewski. The reason why I selected this paper is that it provided a fast and accurate tracking algorithm in near real-time for surgical tool tracking. It is applicable to the catheter tracking part of our project since the surgical tool in the paper shown in Figure 1 looks similar to the catheter shown in Figure 2. Also, the criteria (fast and accurate) of the tracking algorithm in the paper matches our project requirement.

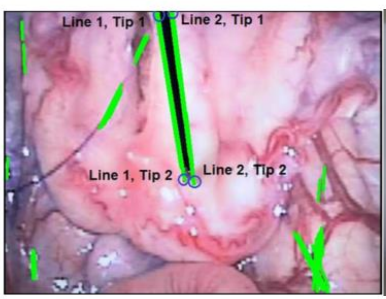

Figure 1: Surgical tool in the paper

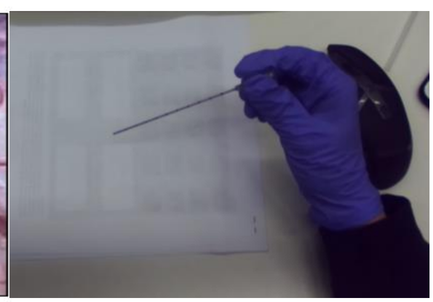

Figure 2: Catheter

Paper Objective

The authors indicated that the demand of skill assessment based on motion metrics such as path length, motion smoothness, and response orientation during Robot-Assisted Minimally Invasive Surgery(RMIS) training curricula to provide information to discriminate expert from novice surgeon. In the consideration of the fact that the motion metrics require a fast and accurate tool tracking method with reasonable cost, the authors aimed to design a fast, accurate, low-cost and platform independent computer vision-based tracking system in order to increase the chance of adoption.

Hardware Design

There are two camera units used for the algorithm design. The first one is an experimental stereo-camera unit comprised of two Microsoft webcams and a 3d-print camera mount, which provide 30 FPS with interocular distance of 29.1mm and costs only $100. The second unit is a surgical endoscope from a da Vinci surgical system, which is mainly used for tracking algorithm evaluation. The endoscope is capable of resolution up to 1920x1080 and provides a framerate up to 100 FPS with the interocular distance of 5.1mm.

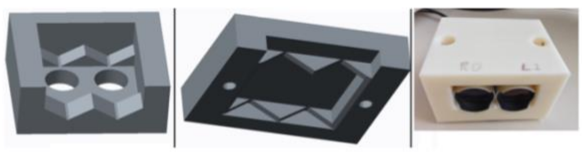

Figure 3:Experimental Webcam Unit

Figure 4: Endoscope Da Vinci Camera Unit

Software and Algorithm Design

The algorithm can be broken down into 3 steps as shown in Figure 5. The first step is 2D tool tips detection. And the algorithm for step 1 is in Figure 6.

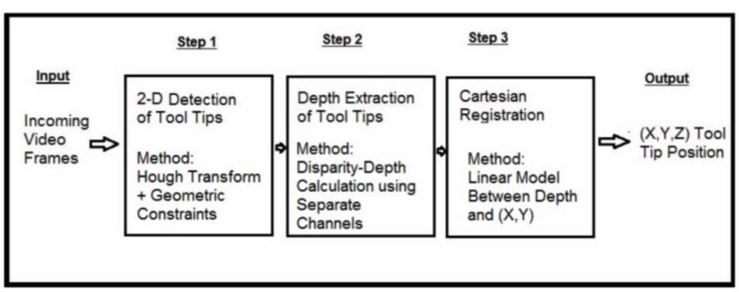

Figure 5:The 3-step approach for object tracking

Data: Frame
Result: Cartesian Tool Location
Convert frame to grayscale;
Blur grayscale image;
Sobel edge detection;
Dynamic edge gradient threshold;
Probabilistic hough transform;
for *Each line in PHT line array* **do**
    Extract unique line pair;
    Compute line parameters;
    if *Parameters best match constraints* **then**
        Save line;
    else
        Extract 2 new lines;
    end
end
Calculate endpoints closest to prior location;
Save tool tip location;

Figure 6: Algorithm for 2D Detection of tip

The algorithm preprocesses the image by converting frame to grayscale and blurring grayscale image to reduce noise. Once the sobel edge detection is implemented to find the edges, the dynamic edge gradient threshold is set by computing the number of edge pixels in the image. If the number of edge pixels is too high or too low, dependent on resolution, the gradient threshold value is then increased or decreased correspondingly.

Probabilistic hough transformation(PHT) with geometry constrain is then utilized to find the end points of the tool shaft. By using PHT instead of hough transformation, it can improve frame rate while finding the endpoints of line. PHT is also helpful to limit the loss of lines due to occlusions along the shaft.

After lines and endpoints found by PHT, geometric constraint is applied to those lines and points to determine endpoint of tool shaft. There are four constraints shown in Table I. For the constraint, tool endpoint is chosen to be the closest to the center of image for initialization. After initialization, endpoint is taken as last known location. And the midpoint of the two endpoints is defined as wrist. If there are multiple tools, the line endpoint is sorted according to previous constraints and separated according to prior spatial information and tool shaft angle.

Constraint	Condition
Lines along the tool shaft are parallel	$\Delta\theta < \theta_{threshold}$
Endpoints of the lines should be near to each other.	$d(p(i)_{1,1}, p(i)_{2,1}) < d_{threshold}$
Length of two lines should be longer than any other set of parallel lines	$d(p(i)_{1,1}, p(i)_{1,2}) > length_{max}$
The resultant endpoints should be 'near' the last known location.	$d(p(i)_{1,1}, p(i-1)_{1,1}) < d_{near}$

TABLE I. Geometric constraints and their implementation in code. d() represents the euclidean distance formula and $p(k)_{m,n}$ represents a point (x,y) within frame k, line m, index n

The next step is depth extraction and cartesian registration. Since a full stereo correspondence is computational expansive, a single tool tip disparity calculation is simpler and used. Disparity is calculated by Euclidean distance formula, and it will then be used to determine the depth to the tool tip and cartesian coordinates. In order to compute depth in world coordinate Zw, the depth model is determined by non-linear curve fitting with calibration data collected on a planar board.

Once the depth model is found by calibration, correlation between (x,y,disparity) and Zw can be modelled as the following equations. The equation (1) is for experimental webcam setup, and the cartesian Z model for the endoscope is the third order polynomial in order to offset disparity in equation (2). The offset disparity is then used to compute depth Zw in equation(3).

$$Z_w = a_1(disp^{a_2}) + a_3 x_p + a_4 y_p + a_5 \quad (1)$$

$$disp_o = \begin{bmatrix} b_0 & \dots & b_n \end{bmatrix} * \begin{bmatrix} 1 & x & y & x^2 & xy & y^2 & x^3 & x^2y & xy^2 & y^3 \end{bmatrix}^T \quad (2)$$

$$Z_w = c_0 e^{c_1(disp - disp_o)} \quad (3)$$

Experimental Design

The experiment is design by using a 3D-printed board with known trajectory shown in Figure 6. And three metrics are evaluated in the experiment, which are time, accuracy, and noise. The time is computed by subtracting the timestamp of receives and returns data. The accuracy is evaluated by moving the tool tip around a known fixed trajectory and compute the error between world coordinates and world trajectory. Lastly, the noise is the deviation of a fixed surgical tool over time.

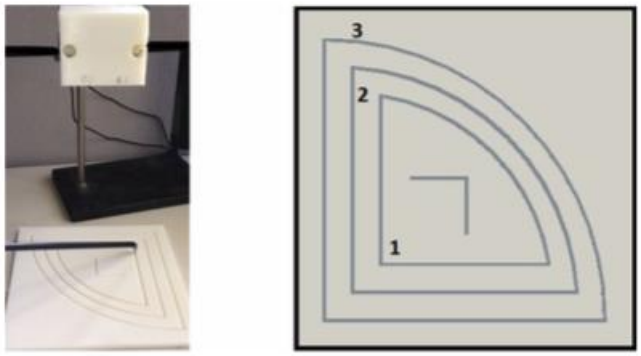

Figure 6: Known trajectory board below stereo camera and model with numbered arcs

Result

As shown in Table II, the result of Webcam setup is very promising, especially the result of 99.4% successful tool localizations with 25.86 FRS. Other noticeable result is that endoscope has larger tracking error overall than Webcam, but it is less than tool shaft diameter. From Figure 7, authors also showed the result that background and lighting doesn't negative affect the tool detection algorithm using Webcam setup.

Performance Metric	Webcam	Endoscope
Computation Time (ms)	39.9	33.99
Frame Rate (FPS)	25.86	26.98
Depth Reprojection Error (mm)	4.09	7.89
Localization Noise (Total) (mm)	1.29	7.92
Average 3D Error (mm)	3.05	8.68
Average 2D Error (mm)	1.59	1.88
95^{th} Percentile Error (mm)	5.47	15.06
Percent Within 8 mm	99.4 %	55.22 %

TABLE II. Performance metrics for the webcam setup and the endoscope setup.

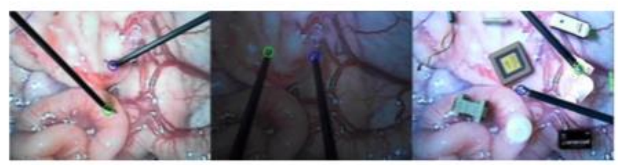

Figure 7: Tool tracking configurations using the webcam setup

Conclusion and Limitation

Tool tracking algorithm is invariant to end-effector type and lighting since it only tracks tool shaft and is independent on color. The worse performance in Endoscope is due to the narrow interocular separation as well as the uncertainties in camera optics. The range of disparities found in this camera setup was found to be about 10 pixels for a depth range of 200 mm. This results in a high signal to noise ratio. The authors believed the further improvement could include Geometric computer vision methods for 3D coordinate extraction, Epipolar geometry to constrain search space for 2nd stereo channel, and incorporation to real surgical applications.

This tracking algorithm is not able to track the configuration of tool wrist and tool with occlusion by smoke, blood, other tools, and movement outside FOV.

Assessment

Overall, the paper provides a fast and accurate tool tracking algorithm in 3D space as the title descripted. And the authors did very well in explaining the algorithm in 2D tool tip detection as well as each advantage of choosing certain method.

From the design of software and algorithm, I can see that the authors really take speed and accuracy as design criteria. First of all, choosing the method of PHT instead of HT with geometric constraint is an efficient way to detect tool shaft endpoint compared to other method like feature-based method. Also, it's a wise choice to use a single tool tip disparity calculation to reduce computation. The idea of calibration data based depth model from disparity is also a clever choice to minimize the computation while providing accurate result.

However, the paper is inadequate in terms of explaining depth extraction model and cartesian registration as well as experimental setup.

The authors just directly gave depth model equations for Webcam and Endoscope and stated it's from calibration data but didn't show the data or explain how it is derived. Also, there is no description about how to map from (x,y,disparity) to (Xw,Yw) in Cartesian registration section.

Moreover, some clarification was missing in the experimental design section. For instance, the distance of Webcam to the 3D printed board is unclear and the setup for endoscope test is unknown. Figure 7 shown in the paper to illustrate the tracking algorithm is invariant to lighting and background, but there is no description about this experiment setup.

Reference

R. Dockter, R. Sweet and T. Kowalewski, "*A fast, low-cost, computer vision approach for tracking surgical tools*," *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Chicago, IL, 2014, pp. 1984-1989. doi: 10.1109/IROS.2014.6942826