Senior Project Proposal
Department of Computer Science
Calvin College

Title: Integrating Online Learning in 6-DoF Pose Estimation Algorithms for Surgical Machine Vision Applications
Author: Henri Prudhomme
Date: September 23, 2020
Mentor: IMT, Stryker
Honors Project: Yes

Vision and General Overview of Proposal

Background and Problem

A team at Stryker is working on replicating the work of Dr. David Joseph Tan on 6-DoF pose estimation and temporal tracking with machine learning for machine vision applications in his PhD Thesis, *"Learn to Track: From Images to 3D Data"* [1], and one of his more recent publications, *"Looking Beyond the Simple Scenarios: Combining Learners and Optimizers in 3D Temporal Tracking"*[2].

Dr. Tan's 3D tracking algorithm works as follows. To start, RGB-D cameras are used to capture frames of image data. The frames of data are composed of two types of images. The first type of image is a standard image which stores a red, green, and blue color value for each pixel. The other type of data is a depth image which stores a distance-from-the-camera value for each pixel. See Figure 1.

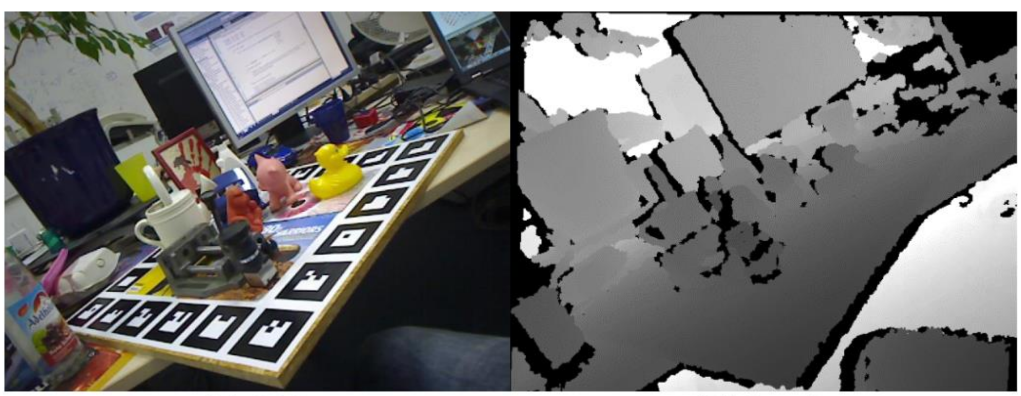

Figure 1: The different images captured in a RGB-D frame.

As these frames are captured, they are passed as input to the main tracking algorithm. Using the RGB image, an object region is extracted from the scene using an RGB edge detection algorithm. Then, using the depth data within the object region, projective point correspondence and registration algorithms are used to map the depth data and the 3D model into one 3D reference coordinate system. Random forest models are then used to compute the optimal 6-DoF transformation parameters to map the depth data to the 3D model. 6-DoF is an acronym used for the six degrees of freedom namely, x-axis, y-axis, and z-axis translation, as well as roll, pitch, and yaw rotation. The final computed 6-DoF parameters are then used to update the known position of the object, and the process repeats. Figure 2 shows how the algorithm can be used to overlay a rendered image precisely on the object(s) it is tracking.

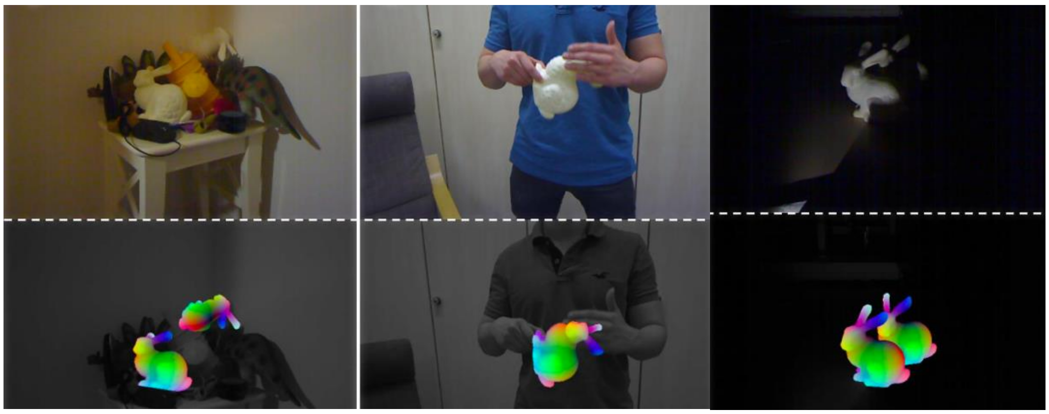

Figure 2: Examples of scene and tracking.

The random forest models which are used to compute the transformation are trained before operation of the live tracking. Each model is trained on an error vector obtained by taking the 3D CAD model, applying a random 6-DoF transformation and computing the objects' differences in position. Each model is trained at a unique camera viewpoint at a vertex of a geodesic sphere surrounding the object. See Figure 3.

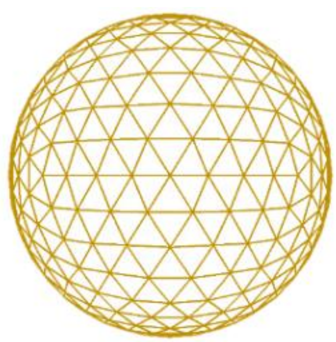

Figure 3: A geodesic sphere on which camera viewpoints are generated

There is a problem with this solution. A 3D CAD model is necessary to train the random forest models. The problem that I will be addressing is a problem that Dr. Tan states himself: "When tracking an object in real scenes, there are situations when its 3D model is not at hand, which makes model-based tracking impossible."

The IMT department at Stryker is concerned with developing and preparing technologies which don't yet have a home. In the case of the 3D tracking software, one of their goals is

> "to develop a machine vision toolbox that provides a generic set of capabilities that may be useful in a broad range of applications, then partner with specific divisions for procedure specific uses"
> – Tony Dowson

While the requirement of a 3D model is not necessarily a hindrance, there is still an opportunity for investigation of methods which do not require a 3D model. Hence, I will be investigating a proposed method for using online learning to dynamically train the random forest models used in the 3D tracking algorithms. A possible use case for such a technology may be field use for calibrating a tracker.

Brief Description of Solution Being Provided

To investigate general case tracking, I will be researching and implementing the proposed solution by Dr. Tan to "From one frame to the next, incrementally add new trees to the forest from different object viewpoints". This forms the basis for the online learning method.

Note: The process written out in this proposal is based on my understanding of the process at the time of writing; the process may be implemented or discovered to work differently at a later date.

The tracking algorithm has to have a ground truth from which it starts. This ground truth is a set of depth data within a 3D box which is obtained from the first RGB-D frame. The centroid of the object is assumed to be at the center of the bounding box, and the initial object transformation is the translation from the camera center to the centroid. The goal being that we don't define the object's initial location at the camera, but rather at the center of a geodesic sphere where it is actually located. Using this initial information, the depth data is used to generate and train *n* trees per parameter for the starting camera view. The method of training *n* trees per parameter is only used at the initial frame to establish extra stability before new random forests can be learned. After the initial frame, the process of computing transformations and tracking follows the original process described with exception that upon each new iteration, only one tree per parameter is learned. The geodesic sphere is used is used to avoid re-learning trees from similar viewpoints by finding the closest untrained vertex of the sphere from the camera location in the object coordinate system.

In summary, in contrast to starting from an initialized random forest, the proposed solution deploys 3D online tracking where the trees are adaptively generated from the initial 3D pose through successive frames. The model is learned on the fly.

Your Interest and Qualifications

I have an interest in this area particularly because it is a math intensive topic. It also involves algorithms and machine learning. Finally, I am interested because of this project's convenience to be both an internship project as well as a senior research project. I believe I am well qualified to pursue this project because of my soon to be completed BCS and Mathematics double major, as well as my history of work at Stryker.

Review of Relevant Design Norms

This project has the potential to affect human life and health and thus raises questions on the reliability of the proposed technology. An online tracking algorithm that uses dynamic computations introduce the problem of having limitless input possibilities, and thus potentially unknown behaviors. This project will address these issues by considering the trust norm from the section titled "Redemption and Responsible Technology" from Schuurman's paper [3].

Mentor Selection, Expert User and Collaboration

Tony Dowson and Rich DeLuca of Stryker, IMT will be the technical mentors for this project. Keith VanderLinden will be my logistical advisor. This project will be supported and provided by Stryker.

Research Question

The research question will be "How do I implement and integrate Dr. Tan's online learning proposal into Stryker's current adaptation of his work?" The main challenges that come with this question include finding the proper interpretation of the wording used to describe the proposed process as well as discovering the proper implementation and ideas for any other holes or ambiguities that will inevitably appear with deep study of the reference articles.

Development Approach

Work required to research and implement this proposed solution will include studying online learning, basic robotic coordinate systems concepts, matrix calculus, MatLab programming syntax, and of course the publication resources. Work will also require writing a research paper which describes the implementation process and results, as well as programming a demonstration.

I anticipate the development approach to involve a period of 4 to 6 weeks of topics discussed, which will transition into a balance of programming and more research throughout the rest of the project. Final deliverables will include a research paper and a proof of concept demo.

Order of priority for final deliverables include:

1. Research Paper describing the mathematics of the approach
2. Proof of Concept Demo Implementation
3. Integration of demo system into Stryker's Machine Vision Application

Meetings will be conducted with Stryker one or more times each week to provide research mentorship and programming code reviews. Biweekly synchronization meetings will be conducted with Professor VanderLinden.

Quality Assurance

Critical Delivery Dates

November 1	Literature review and updated project description
December 1	Draft project website Project status report - Preliminary mathematical/computational writeup. - Initial prototype
March 1	Second draft - Second draft of mathematical/computational writeup. - Updated prototype
April 1	Performance report of demo system
May 1	Final Deliverables - Research paper - Completed demo project - Final presentation

Reviews

Once a development of a demo begins, weekly code reviews will be conducted with the Stryker team.

Testing

This research project will include unit/integration tests. If integrated with Stryker systems this project will be put through Stryker's Quality Management System. If Stryker is able to complete development of a model-based implementation, performance comparisons between the demo system and the Stryker system will be done. Additionally, performance comparisons between the demo system and Dr. Tan's published results will also be conducted.

Resources

Programming resources: MatLab.

Initial research resources:

1. D. J. Tan. “Learn to Track From Images to 3D Data”. [Unpublished doctoral dissertation]. University of Munich.

2. D. J. Tan, N. Navab and F. Tombari, "Looking Beyond the Simple Scenarios: Combining Learners and Optimizers in 3D Temporal Tracking," in IEEE Transactions on Visualization and Computer Graphics, vol. 23, no. 11, pp. 2399-2409, Nov. 2017, doi: 10.1109/TVCG.2017.2734539.

3. Schuurman, Derek (2017) "Technology and the Biblical Story," Pro Rege: Vol. 46: No. 1, 4 - 11.

Risk Analysis

Risk	Exposure Analysis	Mitigation Strategy
Do you have a dependency on others completing work for your project to be a success?	Yes. Stryker may decide to sideline the project.	Focusing on a proof of concept demo and research will address this.
Is there any doubt about the availability of financial resources?	No	N/A
Do you have a dependency on an expert user to provide advice and who may not always be available at critical times?	Yes and No	I have enough resources that I will not have an issue getting help when I need it.
If success depends on testing by an outside source, are there any barriers to completing testing?	No	N/A
Will this project involve new skills for you?	Yes. Knowledge of math required for multiple coordinate systems. Matrix calculus concepts for specific registration algorithms, MatLab programming syntax, online learning implementation, deeper understanding of random forest algorithms	Studying
Will there be anything preventing you from investing at least six hours a week on this at a minimum?	No	N/A

Is there any potential of physical resources you have listed of not being available?	No	N/A