

The 3D Interaction Tool: A Pointing Device for Virtual Reality Applications

Lukas Gemar

March 26, 2016

Abstract

The 3D Interaction Tool is a hand-held device that enables a user to interact with virtual reality applications. Since both the position and orientation of the tool are tracked, user input with the tool accomplishes tasks such as selecting, translating, and rotating three-dimensional objects, or free-hand drawing in space. The tools position is tracked by a software application that filters observations from a camera attached to the virtual reality display, and the orientation is found by means of observations from an inertial measurement unit attached to the tool itself. Simple computer-aided design and drawing applications demonstrate the capabilities of the 3D Interaction Tool.

I. Nomenclature

Position vectors definitions in the inertial coordinate frame:

\vec{p}_s \equiv position of the spherical tracking target on tool rear

\vec{p}_r \equiv position of the tool rear

\vec{p}_f \equiv position of the tool front

\vec{p}_c \equiv position of the tracking camera

\vec{p}_d \equiv position of the virtual reality display

Rotated position vectors:

\vec{p}'_s \equiv position of the spherical tracking target on tool rear, camera coordinates

\vec{p}'_r \equiv position of the tool rear in display coordinates

\vec{p}'_f \equiv position of the tool front in display coordinates

Attitude matrices:

\mathbf{A}_c \equiv attitude of the tracking camera

\mathbf{A}_d \equiv attitude of the virtual reality display

\mathbf{A}_t \equiv attitude of the interaction tool

\mathbf{A}'_t \equiv attitude of the interaction tool in display coordinates

Body-referenced quaternion representations of attitude:

$\bar{\mathbf{q}}_c$ \equiv attitude of the tracking camera

$\bar{\mathbf{q}}_d$ \equiv attitude of the virtual reality display

$\bar{\mathbf{q}}_t$ \equiv attitude of the interaction tool

Position error covariances in inertial coordinate frame:

Σ_{ps} \equiv covariance of spherical tracking target position

Σ_{pr} \equiv covariance of the position of the rear of the interaction tool

Σ_{pf} \equiv covariance of the position of the front of the interaction tool

Σ_{pc} \equiv covariance of the position of the tracking camera

Σ_{pd} \equiv covariance of the position of the virtual reality display

Attitude error covariances in the inertial coordinate frame:

Σ_{qc} \equiv covariance of the tracking camera attitude quaternion error

Σ_{qd} \equiv covariance of the display attitude quaternion error

Σ_{qt} \equiv covariance of the interaction tool attitude quaternion error

Position error covariances in rotated body coordinates:

Σ'_{ps} \equiv covariance of the spherical tracking target position

Σ'_{pr} \equiv covariance of the position of the rear of the interaction tool

Σ'_{pf} \equiv covariance of the position of the front of the interaction tool

Attitude error covariances in the rotated body coordinates:

Σ'_{ac} \equiv covariance of the tracking camera attitude angular error

Σ'_{ad} \equiv covariance of the display attitude angular error

Σ'_{at} \equiv covariance of the interaction tool attitude angular error

II. Introduction

A. Motivation

As much as \$25 billion could be spent on virtual reality (VR) and augmented reality (AR) software applications by 2025 [1]. There are two primary modes of user-input that are possible for these applications: indirect and direct. Indirect input devices translate the user's physical movements into the motion of a virtual cursor by means of a fixed transformation. Common examples of indirect input devices are a mouse or joystick. Direct input devices also translate physical to virtual movements, but this transformation is one-to-one. For example, a touch screen is a direct input device because the physical location of the user's press is colocated with the virtual click.

With the advent of virtual reality applications the following question must be answered: which of these input paradigms – indirect or direct – will prove most useful for interacting with virtual reality applications? While indirect input devices have achieved widespread popularity for two-dimensional displays, the direct input paradigm for VR applications embraces the immersiveness of VR experiences by enabling intuitive interactions for users. Direct input devices allow users to reach out and interact with the objects of their virtual space, just as they would with objects in their physical space.

B. Background

B.1 Commercial Solutions

Recent market trends suggest that direct input devices are on the rise for consumers. Two of these direct input devices, the "Oculus Touch" controller and the HTC Valve controller, are achieving widespread publicity surrounding the commercial release of virtual reality headsets [?] [?] [?]. While originally scheduled for release alongside the Oculus VR headset at the end of March 2016, the Oculus Touch controller will be available in the second half of 2016 [?]. Meanwhile the Valve controllers have been reviewed in preparation for the release of the HTC Vive headset at the beginning of April, 2016, and their utility has been demonstrated with a three-dimensional sketching program by Google called TiltBrush [?]. These input devices track both their position and orientation, making them degree-of-freedom (6DOF) controllers. While the 6DOF controllers are reported to work well with their respective headsets, the headsets are expensive – both retail for more than \$600.

B.2 3D Interaction Techniques

Three-dimensional interaction techniques can be separated into several categories. First, there are interactions that manipulate objects. These interactions accomplish tasks such as rotating, translating, or scaling three dimensional objects. Second, there are interactions that manipulate 3D viewpoint, allowing the user to adjust their field of view. Lastly, there are interactions that control an application. Application control interactions include navigating menus or widgets associated with the application logic. The first set of tasks, those that manipulate objects, have largely been the focus of three-dimensional input device design [?].

B.3 Tracking System Design

C. Design Objectives

There are two broad classes of objectives in 3D input device design: usability and performance.

C.1 Usability

The usability requirements presented in Figure 1 must be met for the tool to succeed as an *input* device for VR software applications.

Usability		
Objective	Justification	Target
Good portability	Target scenarios are seated applications for home and workplace.	The interaction tool setup should be easily translatable between uses.
Low configuration overhead	----	Limited setup time.
Natural form factor	The interaction tool should not require the user to learn a new interaction paradigm.	Form factor similar to a pen or stylus.
Large tracking volume	The user should be able to use the interaction tool over their full range of movement.	Tracking volume of about 1m ³ .
Low cost	----	Less than \$500.

Figure 1: Usability design objectives

C.2 Performance

The performance requirements presented in Figure 2 must be met for the tool to succeed as a *direct* input device for VR software applications.

Performance		
Objective	Justification	Target
High Accuracy	Tracking apparatus must translate user movements into direct input signals.	10cm or less.
Very High Precision	High precision ensures that cursor movements are repeatable.	1-2cm.
Low Latency	Too much discrepancy between physical and virtual motion of the tool induces simulator sickness.	Less than 60ms.
High Robustness	The system should be usable in non-ideal environments and use-cases.	Robust to tool velocity.

Figure 2: Performance design objectives

III. Design

A. Requirements

In order to enable direct input with the interaction tool, the tool’s tracking system must estimate both the position and orientation of the interaction tool. Additionally, rendering a virtual representation of the tool requires measuring its position and orientation relative to the virtual reality display orientation, given by the attitude matrix \mathbf{A}_d . To track the tool’s absolute trajectory, its position is determined relative to the world – that is, relative to a non-accelerating, inertial coordinate system in the user’s physical environment. The position and orientation of the tool in the world coordinates of the inertial frame are \vec{p}_f and \mathbf{A}_t , respectively. The position of the tool in the coordinate frame of the virtual reality display is denoted by a ‘prime’ sign: the tip or front of the tool in display coordinates is \vec{p}'_f . Thus, there are four tracking targets: \vec{p}_f , \mathbf{A}_t , \vec{p}'_f , and \mathbf{A}_d ; these are shown in Figure 3.

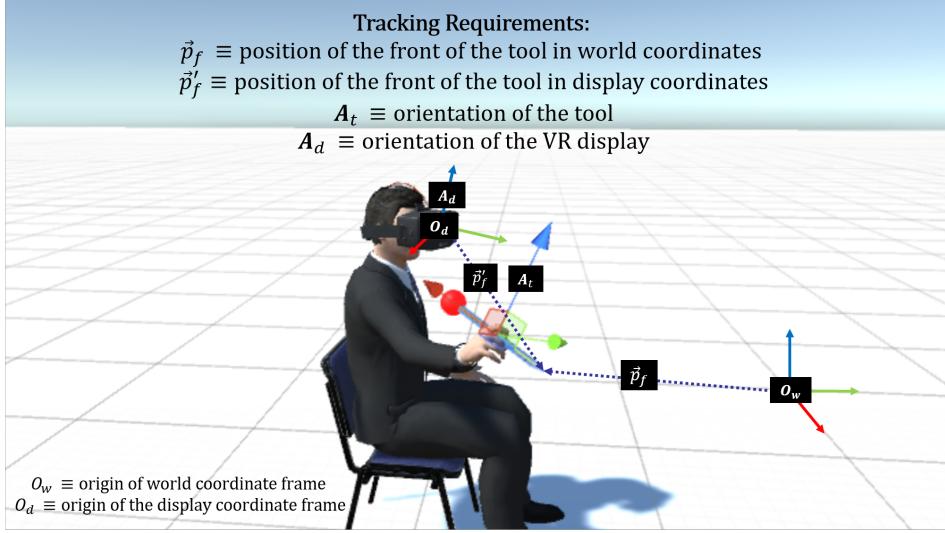


Figure 3: Position and orientation tracking requirements

B. Geometry

The geometry of the tool tracking system affects both the usability and the performance of the interaction tool. The display-referenced geometry is chosen because it eliminates the burden on the user to pre-calibrate their environment before providing direct input to VR applications. At the same time, this geometry is compatible with VR applications with which a user interacts while seated.

B.1 Option 1: World-referenced geometry

The world-referenced tracking geometry is shown in Figure 4.

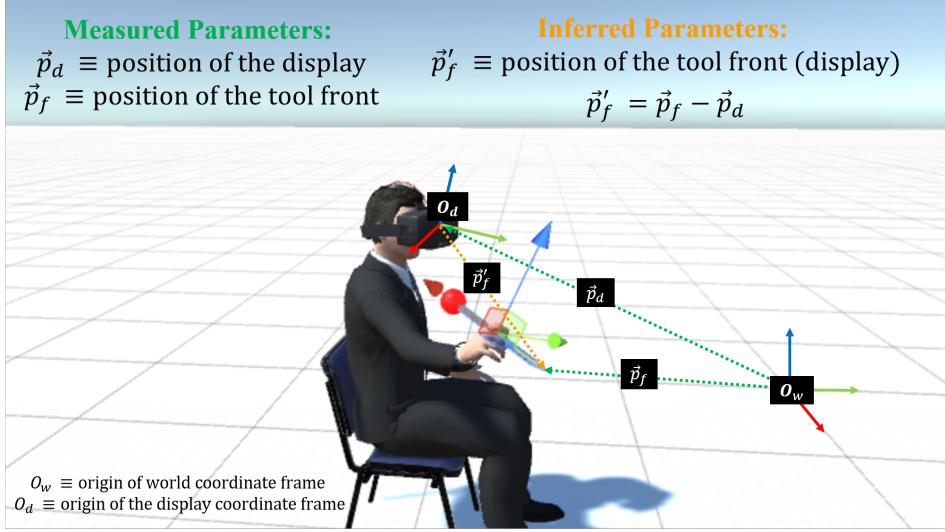


Figure 4: Position tracking with a world-referenced geometry

B.2 Option 2: Display-referenced geometry

The display-referenced tracking geometry is shown in Figure 5.

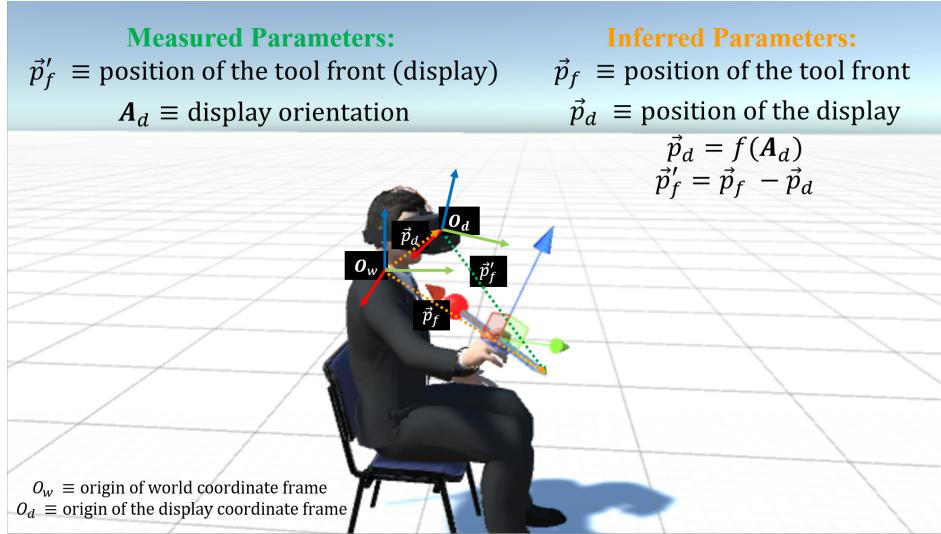


Figure 5: Position tracking with a display-referenced geometry

B.3 Geometry Selection: Comparing portability and tracking volume

Tracking geometries 1 and 2 have advantages and disadvantages when evaluated using the usability design objectives. Relative to the usability criteria, the two geometries differ greatly in their portability and tracking volume.

The world-referenced tracking geometry requires a fixed coordinate system in the user's physical environment, decreasing portability by increasing the configuration overhead for using the direct input device. However, once this coordinate system is established, the user can provide direct input to a VR application over a very large tracking volume, up to the size of the room in which the tracking system is located. The display-referenced geometry, on the other hand, has the advantage of enabling a user to provide direct input to a VR application without pre-calibrating the physical environment in which they are using the VR display. The tradeoff is that the tracking volume is small. Since the origin of the world coordinate system in the display-referenced geometry is located beneath the user's head, the user must remain seated while using a direct input device with a display-referenced tracking geometry.

C. Technology

A combination between optical and inertial sensors was chosen to track the 3D Interaction Tool's position and orientation. The chart in Figure 6 summarizes the comparison between various alternatives for the choice of the tracking technology.

Tracking Technology	Advantages	Disadvantages
Mechanical	Very high accuracy and precision . Good portability .	Small tracking volume .
Magnetic	Very high accuracy and precision . Low latency . Good robustness .	Very high cost .
Ultrasound	High accuracy and precision .	Poor portability . Incompatible with slender form factor . Poor robustness to acoustic noise.
Optical	High accuracy and precision . Low configuration overhead . Low cost .	Poor robustness to occlusions and tool velocity. High latency . Narrow field of view .
Inertial	High accuracy and precision of orientation. Very low latency . Low cost .	Low accuracy and precision of position over large time scales.

Figure 6: Comparison of five tracking technologies

D. Final Design

The final system layout is shown in figure 7.

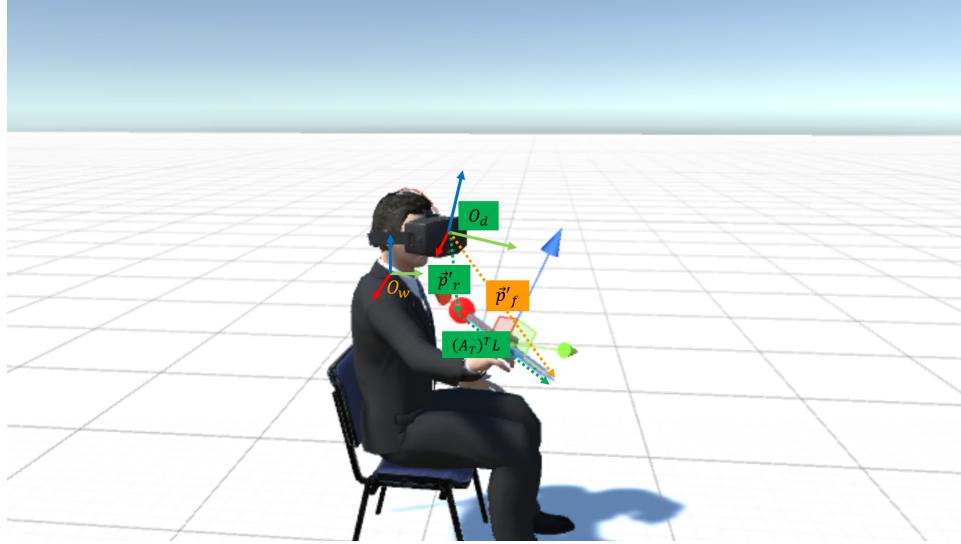


Figure 7: Final tracking system design: a display-referenced tracking topology

D.1 Tracking measurements

Not all of these three tracking targets are measured directly. The tracking targets are inferred from the measurements of the position of the rear of the interaction tool relative to the camera, the attitude of the camera, and the attitude of the tool. The position of the tool relative to the camera is denoted by \vec{p}'_s , the attitude of the camera relative to the inertial coordinate system by \mathbf{A}_c and the attitude of the tool relative to the inertial coordinate system by \mathbf{A}_t . The vector \vec{p}'_s is measured by finding the .

D.2 Inferring the tool position

The position of the front of the tool in world coordinates is given by \vec{p}_f . This position is inferred through measurements of the position of the rear of the tool and the tool's orientation. The equation for the position of the front of the tool is

$$\vec{p}_f = \vec{p}_r + \mathbf{A}_t^{-1} \vec{p}_l$$

where the position of the rear of the tool, \vec{p}_r is given by

$$\vec{p}_r = \mathbf{A}_c^{-1} \vec{p}'_s - \vec{p}_c$$

The attitude matrix, \mathbf{A}_t^{-1} , is the rotation from the body coordinates of the tool to the world coordinate system.

In order to render the tool on the display for the user, the position of the tool must be known to the VR application in the coordinate system of the VR display. The position of the front of the tool in display coordinates is \vec{p}'_f and it is found by multiplying the tool's position in the world coordinate system by the attitude of the display. Additionally, the position of the display must be added to the tool position before multiplying by the display attitude matrix:

$$\vec{p}'_f = \mathbf{A}_d(\vec{p}_f + \vec{p}_d)$$

The tool attitude in the coordinate system of the display is denoted by \mathbf{A}'_t and it is found by composing the rotation from the tool's body coordinates to world coordinates with the display attitude matrix:

$$\mathbf{A}'_t = \mathbf{A}_d \mathbf{A}_t^{-1} \vec{p}_l$$

This is the tool orientation visible to the user.

IV. Implementation

The system's physical layout is shown in ???. For simplicity, the tracking camera is assumed to be fixed. The model for this system is geometrically equivalent to the one presented in Figure 7. The webcam of the computer is used as the tracking camera. The inertial measurement unit, the Adafruit BNO055, is attached to the breadboard. A colored foam sphere is attached to the rear of the interaction tool to serve as the tracking target.

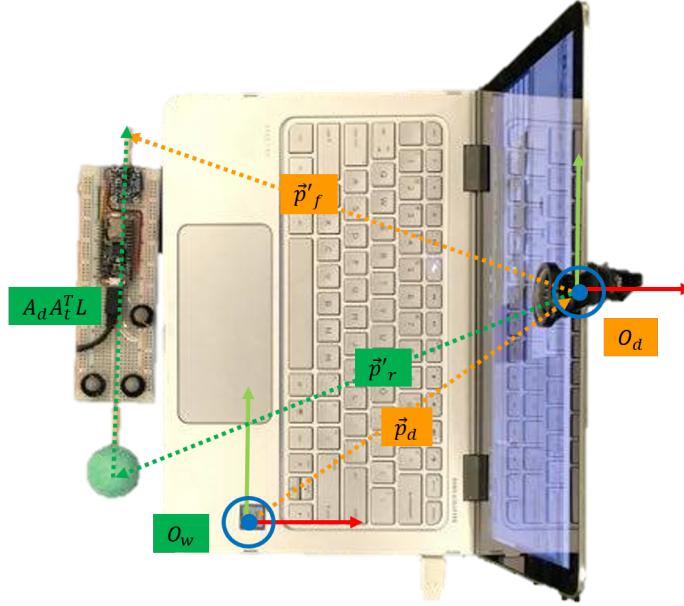


Figure 8: Geometrically equivalent tracking system

The virtual reality application runs on the computer shown in Figure ???. While this virtual reality application is not immersive, it creates a virtual 3D environment displayed to the user on the computer screen, where they can interact with virtual objects. Thus, this geometry is sufficient to prove the usefulness of the 3D Interaction Tool for virtual reality applications.

The spherical tracking target is chosen because its size in each image does not depend on its orientation. Additionally, if the angle subtended by the tracking target in the field of view is small, then the midpoint of the tracking target in world coordinates is on the same plane as the vanishing points on the edges of the target.

The block diagram of the implemented tracking system is shown in Figure 9.

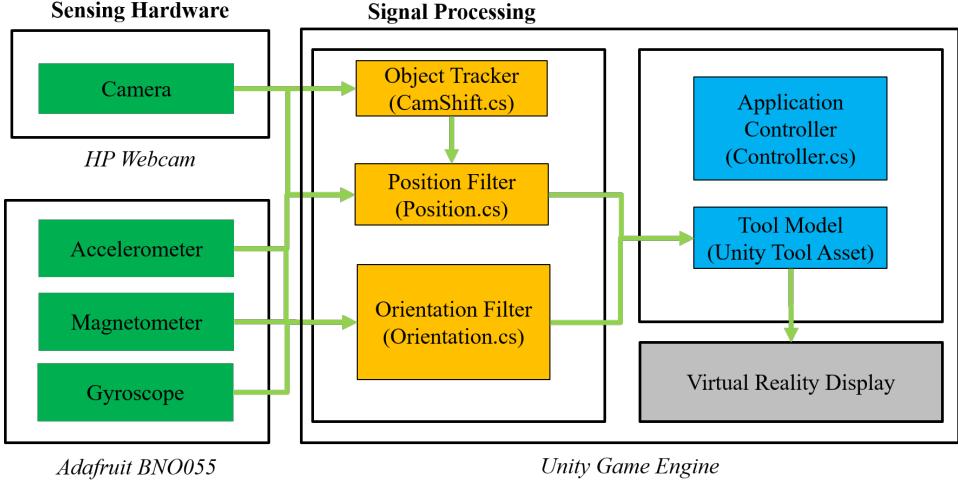


Figure 9: Full system diagram of implementation

V. Orientation Tracking

A. Mathematical Representation of Orientation

The orientation of the interaction tool relative to the world coordinate frame can be represented in a number of ways. The two representations that will be important here are the attitude matrix and the attitude quaternion. First, the attitude of the orientation tool can be described by the attitude matrix A . The attitude matrix typically maps a vector in the reference frame to a vector in the body frame [2]. This mapping is given by the relation,

$$\mathbf{b} = \mathbf{A}\mathbf{r}$$

Second, the attitude of the interaction tool can be represented as a quaternion. A rotation in three-dimensions is equivalent to an axis of rotation and angle of rotation. If \hat{n} is the axis of a rotation and θ is the angle of the rotation, then the quaternion \bar{q} represents the attitude:

$$\bar{q} = \begin{bmatrix} \vec{q} \\ q_4 \end{bmatrix}$$

with $\vec{q} = \hat{n} \sin \frac{\theta}{2}$ and $q_4 = \cos \frac{\theta}{2}$. The quaternion must have length 1: $\bar{q}^T \bar{q} = 1$.

The attitude matrix and the quaternion representing attitude are related by the following equation, where A is a function of \bar{q} [3]:

$$A(\bar{q}) = (q_4^2 - \|\vec{q}\|^2)I_{3x3} + 2\vec{q}\vec{q}^T - 2q_4[\vec{q}\times]$$

where $[\vec{q}\times]$ is the skew-symmetric cross product matrix,

$$[\vec{q}\times] = \begin{bmatrix} 0 & -q_3 & q_2 \\ q_3 & 0 & -q_1 \\ -q_2 & q_1 & 0 \end{bmatrix}$$

The relationship between A and \bar{q} can also be expressed as the product of two matrices [2]:

$$A(\bar{q}) = \Xi(\bar{q})^T \Psi(\bar{q})$$

with

$$\Xi(\vec{q}) = \begin{bmatrix} q_4 \mathbf{I}_{3 \times 3} + [\vec{\mathbf{q}} \times] \\ -\vec{q}^T \end{bmatrix}$$

$$\Psi(\vec{q}) = \begin{bmatrix} q_4 \mathbf{I}_{3 \times 3} - [\vec{\mathbf{q}} \times] \\ -\vec{q}^T \end{bmatrix}$$

B. Naive Measurement Model

The TRIAD estimation method finds the attitude matrix A that maps three reference frame vectors \vec{r}_k to their corresponding body frame vectors \vec{b}_k for $k = 1, 2, 3$ [4]. Two observations of reference frame vectors provide sufficient information to compute a right-handed orthonormal triad of vectors in the reference frame. If \vec{v}_1 and \vec{v}_2 are reference frame vectors for the accelerometer and the magnetometer, respectively, then

$$\vec{v}_1 = \begin{bmatrix} 0 \\ 0 \\ 9.81 \end{bmatrix} \quad \text{and} \quad \vec{v}_2 = \begin{bmatrix} 0.3 \\ -1 \\ 0 \end{bmatrix}$$

so that the gravity vector points down and the magnetometer vector points in the direction of magnetic North in Boston, relative to the world reference frame. The orthonormal vector triad representing the reference frame is given by,

$$\vec{r}_1 = \frac{\vec{v}_1}{\|\vec{v}_1\|}, \quad \vec{r}_2 = \frac{\vec{v}_1 \times \vec{v}_2}{\|\vec{v}_1 \times \vec{v}_2\|}, \quad \vec{r}_3 = \vec{r}_1 \times \vec{r}_2$$

Let \vec{w}_1 be the measurement of the normalized unit vector from the accelerometer and \vec{w}_2 be the measurement of the normalized unit vector from the magnetometer, then the orthonormal vector triads for the reference vectors are given by,

$$\vec{b}_1 = \frac{\vec{w}_1}{\|\vec{w}_1\|}, \quad \vec{b}_2 = \frac{\vec{w}_1 \times \vec{w}_2}{\|\vec{w}_1 \times \vec{w}_2\|}, \quad \vec{b}_3 = \vec{w}_1 \times \vec{w}_2$$

If there was no noise in the system, then it would be the case that $\vec{b}_k = A \vec{r}_k$ for $k = 1, 2, 3$. An equivalent way of writing this in matrix form is the following:

$$M_b = A M_r$$

where

$$M_b = \begin{bmatrix} \vec{b}_1 & \vec{b}_2 & \vec{b}_3 \end{bmatrix}$$

$$M_r = \begin{bmatrix} \vec{r}_1 & \vec{r}_2 & \vec{r}_3 \end{bmatrix}$$

Since M_r is an orthogonal matrix, its inverse is equal to its transpose: $M_r^{-1} = M_r^T$. Thus, the solution for the attitude matrix is,

$$A = M_b M_r^T$$

C. Dynamic Measurement Model

The rate of change of the attitude can be written in the form of a general state space model:

$$\dot{\vec{x}} = A(\vec{x}, t)\vec{x} + B(\vec{x}, t)\vec{u}$$

The kinematic equation for the change in attitude as a function of the angular rate can be written as a rate of change of the attitude matrix or a rate of change in the quaternion [3]. The angular velocity vector is given by $\vec{\omega}$. The rate of change of the attitude matrix is,

$$\dot{A}(t) = [\vec{\omega} \times] A(t)$$

In this equation, $[\vec{\omega} \times]$ is an instance of the state transition matrix $A(\vec{x}, t)$ from the general schema for a state space model. The kinematic relation for the rate of change of the quaternion can be written in the following two ways:

$$\dot{\bar{q}} = \frac{1}{2} \Omega(\vec{\omega}) \bar{q}$$

$$\dot{\bar{q}} = \frac{1}{2} \Xi(\bar{q}) \vec{\omega}$$

where the first relation is a state space model with state transition matrix $A(\vec{\omega}) = \Omega(\omega)$, where the state transition matrix is a function of the angular rate vector $\vec{\omega}$:

$$\Omega(\vec{\omega}) = \begin{bmatrix} -[\vec{\omega} \times] & \vec{w} \\ -\vec{\omega}^T & 0 \end{bmatrix}$$

The second kinematic equation for the rate of change of the quaternion is an instance of the schema for the state space model where the angular rate $\vec{\omega}$ is treated as the input \vec{u} to the system and the input matrix $B(\bar{q}) = \Xi(\bar{q})$ is a function of the current attitude \bar{q} .

There are two good reasons to use the quaternion representation of attitude: first, the rate of change of the quaternion is linearly related to the state of the quaternion; and second, successive rotations are computed by taking the product of quaternions [2]. The following identity holds of the composition of quaternions [3]:

$$A(\bar{q}_1) A(\bar{q}_2) = A(\bar{q}_1 \otimes \bar{q}_2)$$

with the operation $A(\bar{q})$ finding the attitude matrix A that corresponds with the quaternion \bar{q} . Following the notation in [2], the product of the quaternion is linear in each of the quaternions \bar{q}_1 and \bar{q}_2 :

$$\bar{q}_1 \otimes \bar{q}_2 = \begin{bmatrix} \Psi(\bar{q}_1) & \bar{q}_1 \end{bmatrix} \bar{q}_2 = \begin{bmatrix} \Xi(\bar{q}_1) & \bar{q}_1 \end{bmatrix} \bar{q}_2$$

The inverse of the quaternion is found by negating the imaginary component of the quaternion:

$$\bar{q}^{-1} = \begin{bmatrix} -\vec{q} \\ q_4 \end{bmatrix}$$

C.1 Filter Initialization

The initial state of the orientation is estimated using the TRIAD approach. The rotation matrix output of the TRIAD approach is converted to a quaternion to initialize the dynamic model. In order to initialize the covariance, an initial uncertainty in the orientation of 10 degrees is assumed.

The Kalman filter assumes process noise and measurement noise in the dynamics. The process noise for the dynamic model of rotations assumes that the body is fixed at each time step, so the noise in the dynamics must come from noise on the gyroscope input. The root-mean squared noise density of gyroscope on the Adafruit BNO055 given in the data sheet as $0.014 \frac{\text{deg}}{\sqrt{\text{Hz}}}$ [?]. To find the mean squared noise, this value is squared and multiplied by bandwidth of the signal, taken here to be the sample rate, 47 Hz. This yields a value of approximately $(0.1)^2 \frac{\text{deg}}{\text{s}}$. Converting to radians, the mean squared noise of the gyroscope is $2.8e-6 \frac{\text{rad}}{\text{s}}$. Thus, the process noise Q is initialized as

$$Q = \begin{bmatrix} 2.8e-6 & 0 & 0 \\ 0 & 2.8e-6 & 0 \\ 0 & 0 & 2.8e-6 \end{bmatrix}$$

Similarly, the measurement noise is initialized with noise values for the accelerometer and magnetometer listed in the datasheet [?]. Again, squaring the noise densities and multiplying by the bandwidths yields mean squared errors of $1.1e - 6g^2$ and $17\mu T^2$ for the accelerometer and the magnetometer, respectively. The six dimensional measurement noise matrix, R , is initialized as

$$R = \begin{bmatrix} 1.1e - 6 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.1e - 6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.1e - 6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 17 & 0 & 0 \\ 0 & 0 & 0 & 0 & 17 & 0 \\ 0 & 0 & 0 & 0 & 0 & 17 \end{bmatrix}$$

C.2 Filter Tuning

The Kalman filter assumes that the measurement noise and process noise are Gaussian. However, this assumption rarely holds, so the noise parameters of the filter must be tuned to achieve better performance of the filter. The measurement noise parameters is particularly important. Increasing the measurement noise parameter, represented by R in the Kalman filter equations, decreases the Kalman gain.

Intuitively, a larger value for the measurement noise means that the variability of the measurements is high, so the filter is less responsive to measurements and more responsive to the model. Conversely, if the measurement noise parameter decreases, the Kalman gain increases. The graph in figure ?? shows how the Kalman filter responds to a step input in the roll for different values of R , the measurement gain.

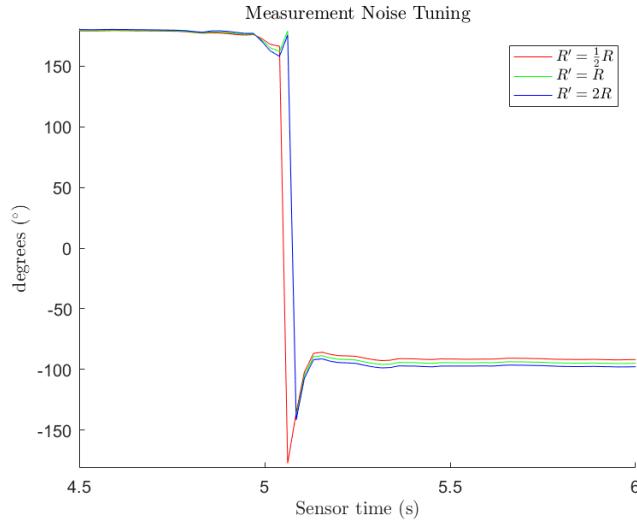


Figure 10: Tuning the measurement noise parameter of the Kalman filter

Lower values of R lead to a faster step response but more overshoot because the Kalman gain is high. Higher values of R have a slower step response but less overshoot because the Kalman gain is lower.

VI. Position Tracking

A. Measurement Model

The following is a model of a single frame from the camera. In this frame, shown in image Figure 11 the red circular object represents the rear of the interaction tool. The spherical object of the rear of

the interaction tool is called the tracking target. The width of the target along the x dimension of the image is s_u .

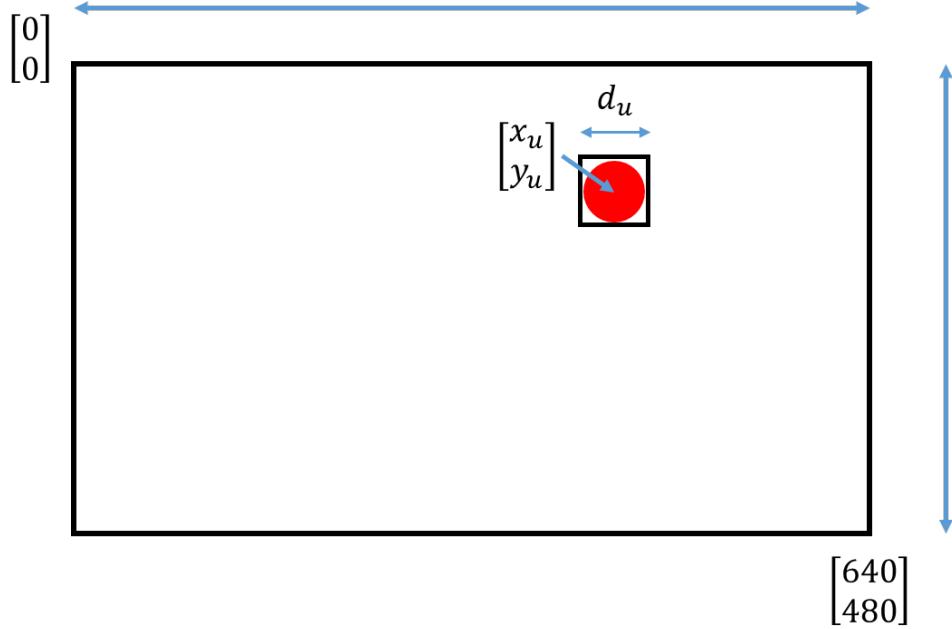


Figure 11: Computer vision tracking model

The midpoint of the target in the ideal undistorted image is $\vec{p}_u = \begin{bmatrix} x_u & y_u \end{bmatrix}^T$. The center of the undistorted image is c_x in the x direction and c_y in the y direction. The focal lengths in the x and y directions are f_x and f_y , respectively. These parameters are found through the camera calibration routine, explained in detail below. The measurements of the center of the target and its width are related to the hidden state $\vec{p}'_s = \begin{bmatrix} x_s & y_s & z_s \end{bmatrix}^T$ in the following way:

$$\vec{y} = h(\vec{x}) = \begin{bmatrix} x_u \\ y_u \\ s_u \end{bmatrix} = \begin{bmatrix} \frac{f_x x_s}{z_s} + c_x \\ \frac{f_y y_s}{z_s} + c_y \\ \frac{s_c}{z_s \sqrt{\frac{1}{f_x^2} + \frac{1}{f_y^2}}} \end{bmatrix}$$

These relationships are found through using a simple pin-hole camera model [5].

A.1 Camera Calibration

The measurement model requires knowledge of the intrinsic camera parameters, f_x, f_y, c_x, c_y , and the radial and tangential distortion factors. There are two steps to translating a point from camera coordinates to image coordinates [5]. First, the pin-hole camera model shows how to find the ideal image coordinates of an object, $(\mathbf{x}_u, \mathbf{y}_u)$ based on the position of the object in camera coordinates, $\vec{p}'_s = (x, y, z)$:

$$\begin{aligned} \mathbf{x}_u &= f_x \frac{x}{z} \\ \mathbf{y}_u &= f_y \frac{y}{z} \end{aligned}$$

Second, the radial and tangential distortion relations show how the actual image coordinates, $(\mathbf{x}_u, \mathbf{y}_u)$, relate to the image coordinates on the actual or distorted image plane, $(\mathbf{x}_d, \mathbf{y}_d)$. The radial distortion correction equation is as follows:

$$\begin{aligned}\mathbf{x}_u &= x_d(1 + k_1r^2 + k_2r^4 + k_3r^6) \\ \mathbf{y}_u &= y_d(1 + k_1r^2 + k_2r^4 + k_3r^6)\end{aligned}$$

and the tangential distortion equation is given by,

$$\begin{aligned}\mathbf{x}_u &= x_d + (2p_1xy + p_2(r^2 + 2x^2)) \\ \mathbf{y}_u &= y_d + (p_1(r^2 + 2y^2) + 2p_2xy)\end{aligned}$$

Thus, the five distortion parameters are summarized in a vector D :

$$D = \begin{bmatrix} k_1 & k_2 & p_1p_2 & k_3 \end{bmatrix}$$

The calibration parameters are discovered by using an open source implementation provided by OpenCV. The camera collects a series of images that contain a checkerboard pattern. The calibration computes the transformation from the image coordinates of the checkerboard to its real world coordinates. The calibration images look like the ones in figure 1 below.

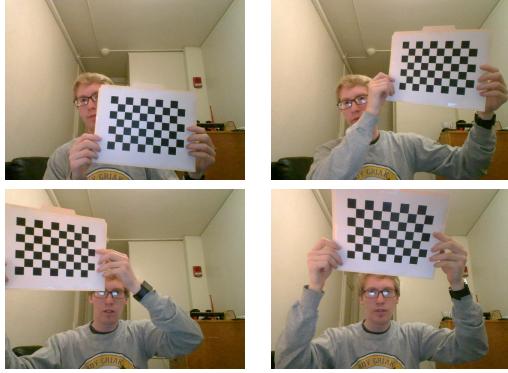


Figure 12: Camera calibration routine

A.2 Object Tracking

The signal processing layer contains an object tracker to find the rear of the interaction tool in each frame. Since a robust estimate of the object size in each image is required, its edges must be estimated as exactly as possible. One simple way to estimate the edges of the spherical tracking target is to segment the tracking target from the background by using color. The color of the object is used to determine which pixels belong to it and which do not. Additionally, the color-based object tracking must happen in real time. To meet the demands of real-time, color-based object tracking, the CamShift algorithm was chosen [6].

A.3 Tool Maneuvers

The possible maneuvers of the interaction tool are classified into two categories: maneuvers in the image plane and those in the depth dimension. For a user of the 3D Interaction Tool, side-to-side or up-and-down movements of the interaction tool map to motions in the image plane of the tracking camera, while movements of the tool towards and away from their body map to motions in the depth dimension. More precisely, the maneuvers in the image plane are movements of the interaction tool that occur perpendicular to the viewing angle of the camera.

The convention established for the purposes of testing is that the image plane is the yz -plane of the coordinate system and the depth dimension maps to the negative x -dimension. Maneuvers in the x -dimension encompass the movements of the tool towards and away from the tracking camera. Figure 13 compares the actual motion of the interaction tool to the estimates made by the tracking system.

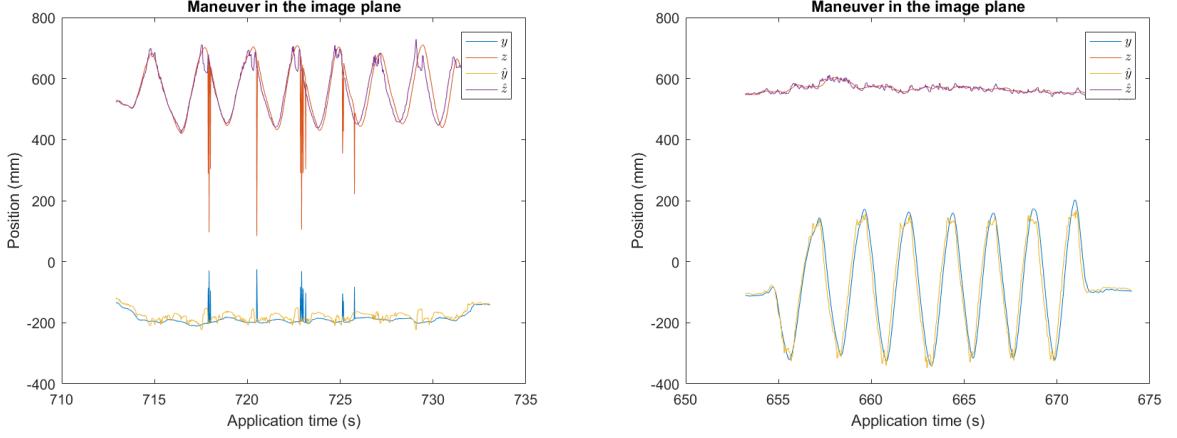


Figure 13: Vertical and horizontal motion in the image plane

In 13, the traces on the left correspond with vertical maneuvers of the tool and the traces on the right correspond to horizontal movements of the tool. In both graphs it is apparent that the tracking system is very accurate with low bias. The estimates of \hat{y} and \hat{z} map almost exactly to the true values.

The second set of maneuvers are those in the depth dimension. Figure 14 shows a trace of the actual depth x and the estimated depth \hat{x} .

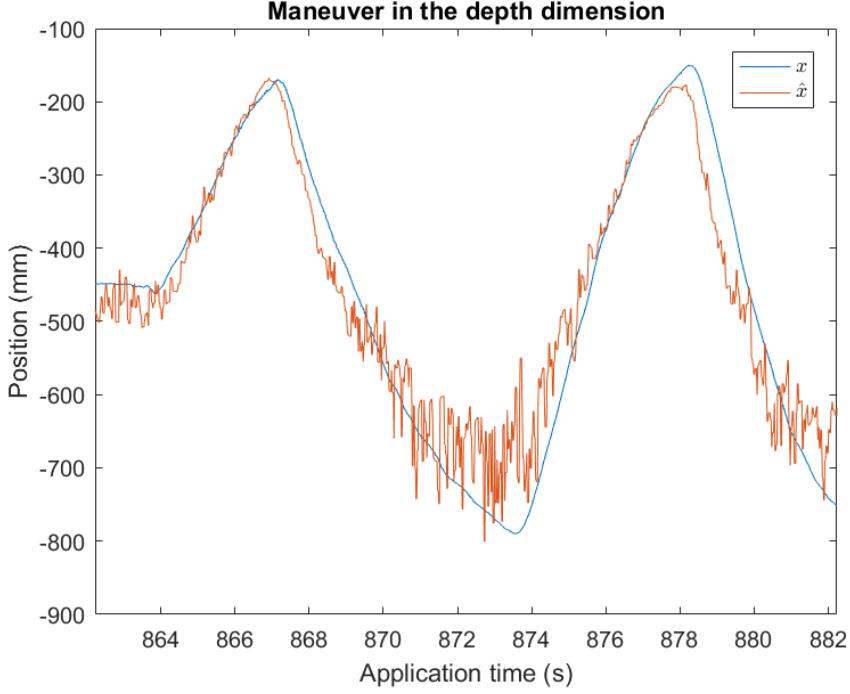


Figure 14: Motion of the interaction tool in the depth dimension

The natural conclusion from the traces in 14 is that the depth tracking accuracy and precision depend on how far away the interaction tool is from the tracking camera.

A.4 Error Analysis of Motion in the Image Plane

The accuracy and precision of the tracking system depends on the direction of the tool maneuvers. This is evident from analyzing the covariance of the error in the y and z directions for vertical and

horizontal maneuvers. Figure 15 shows the error ellipse associated with motion in the image plane. On the left is the error ellipsee associated with vertical motion and on the right is the error associated with horizontal motion.

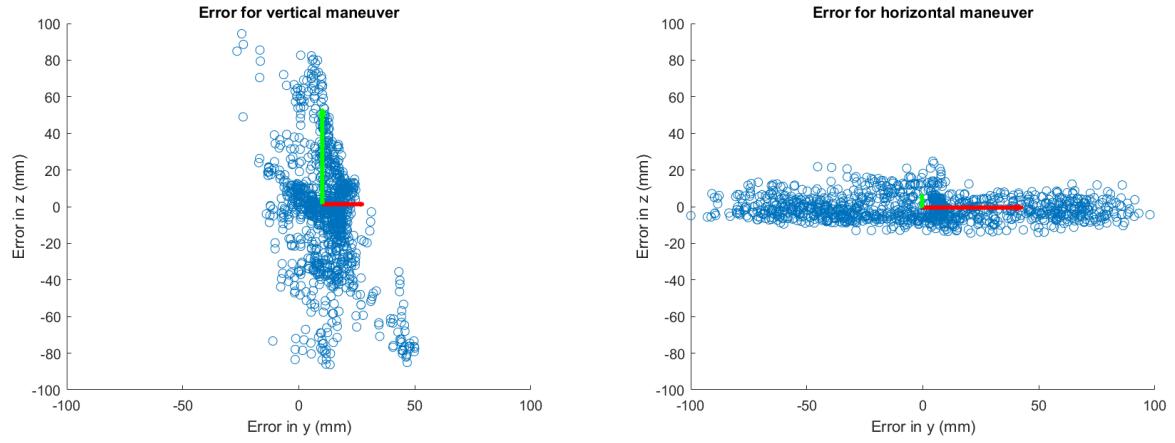


Figure 15: Tool tracking errors in the yz-plane

In 15 two arrows are placed with their origins at the mean of the yz-error distribution and their length corresponding to the square root of the variance along the direction they point. It is clear from the magnitudes of these arrows that their is a breakdown of precision of the tracking system along the direction of motion for the 3D Interaction Tool.

What explains the higher variance along the direction of motion for the interaction tool? There are essentially two sources of error for tracking the target relative to the camera. First, there is error that comes from the camera calibration routine in the form of error of intrinsic parameters of the camera – errors in fx , fy , for example. Second, there is error that comes from finding the centroid of the object in each image frame. Broadly, we could classify the calibration errors as systemic and the image processing errors as dynamic. Since the variance of the tool tracking increases in the direction of motion of the object, it is not likely that this error comes from a systemic problem with the calibration routine. Therefore, we can attribute the decrease in precision to error in the image segmentation routine.

Why does motion along a particular direction increase the error in the image segmentation? When the object is moving, its boundaries become blurred along the direction of motion in each frame. These blurred edges increase the uncertainty about the location of the object along the direction of motion. Therefore, when the object is moving along a particular direction, the uncertainty of its location in the direction of motion increases, but quickly drops again when the motion of the object halts.

A.5 Error Analysis of Motion in the Depth Dimension

The distribution of errors is larger for depth-tracking than tracking in the image plane. This result is displayed graphically in Figure 16.

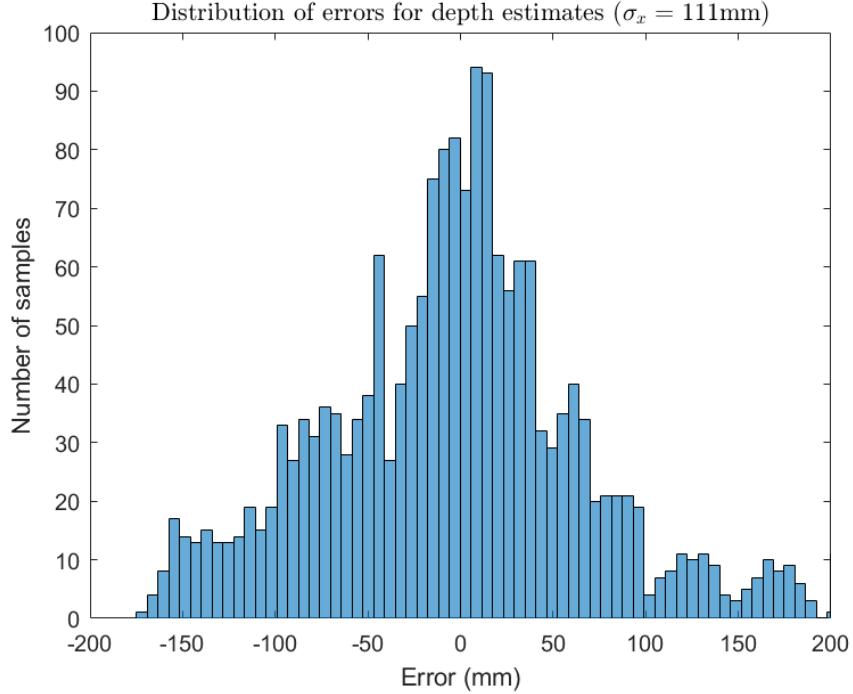


Figure 16: Distribution of depth-tracking errors

While the bias of the tracking errors tends to zero, the variance is much larger than for the depth-tracking in the image plane.

B. Dynamic Measurement Model

Since the user's movement of the interaction tool is voluntary, no assumptions can be made about its overall trajectory. However, human-motion obeys minimum-jerk constraints [7]. Since the jerk of the tool's motion is small, it can be modelled as a white-noise process. The dynamical model for the tool's motion along the x axis is estimated by the discrete time state-space model

$$\vec{x}_{k+1} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ \dot{x} \end{bmatrix} + \begin{bmatrix} 0 \\ \Delta t \end{bmatrix} \ddot{x} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w_k$$

where w_k is the jerk. Additionally, the motion of the tool is assumed to be coordinate uncoupled, since the user is free to move the tool in any direction. Therefore this same model can be extended to motion along the y and z axes as well.

VII. Performance Analysis

A. Estimation Error

Given the tool model $\hat{\vec{p}}_f = \hat{\vec{p}}_r + \hat{\vec{p}}_l$, the overall error covariance Σ_{pf} is a function of the error due to the position estimation and the error to the orientation estimation. Since the estimation of \vec{p}_r and \vec{p}_l occur independently, it is reasonable to assume that the errors in the estimation of these vectors is not correlated. Thus, the covariances of the errors for these two vectors simply add to yield the covariance of the error for the tool front:

$$\Sigma_{pf} = \Sigma_{pr} + \Sigma_{pl} \quad (1)$$

B. Covariance Transformations

Here are three properties of covariance transformations, useful for analysing the performance of the tracking system as a whole:

(1) Given two independent random variables X, Y , and their sum $Z = X + Y$ how does the covariance of $Z \Sigma_Z$ relate to the covariances Σ_X and Σ_Y of X and Y , respectively? In this case the covariances simply sum so that

$$\Sigma_Z = \Sigma_X + \Sigma_Y$$

(2) Given two random variables X and Y related by the linear operator T so that $Y = TX$ how does the covariance of Y relate to the covariance of X ? If Σ_X is the covariance of X and Σ_Y is the covariance of Y , then

$$\Sigma_Y = T\Sigma_XT^T$$

(3) Given two random variables X and Y , and the non-linear relationship that $Y = f(X)$, how does the covariance of Y change with the covariance of X ? As a first-order approximation, the randomness of Y can be approximated as a linear function of X [8]:

$$Y - \mu_Y = T(X - \mu_X) \quad (2)$$

Now, given the linear relationship, $Y - \mu_Y = T(X - \mu_X)$, the covariance of Y is related to the covariance of X by

$$\Sigma_Y = T\Sigma_XT^T$$

using property (2) of the covariance transformation under a linear operator.

C. Estimation Error for $\hat{\vec{p}}_l$: Finding position error due to orientation error

How do errors in the orientation estimation lead to errors the tool position estimation? In the tool model, the position of the front of the tool is estimated from three pieces of information: the position of the rear of the tool \vec{p}_r , the orientation of the tool \mathbf{A}_t , and the tool length in body coordinates \vec{p}'_l . This relationship is described by the equation

$$\vec{p}_f = \vec{p}_r + \vec{p}_l = \vec{p}_r + \mathbf{A}_t^T \vec{p}'_l$$

There is no error in the measurement of \vec{p}'_l ; that is a well known parameter of the interaction tool. The error in \vec{p}_l comes from the estimation of the rotation matrix from body coordinates to world coordinates.

The attitude matrix \mathbf{A}_t is parameterized by the quaternion $\bar{\mathbf{q}}_t$ estimated from the EKF so that $\vec{p}_l = \mathbf{A}_t(\bar{\mathbf{q}}_t)^T \vec{p}'_l$. This expression is helpful because it describes the tool span vector \vec{p}_l as a function of $\bar{\mathbf{q}}_t$. The relationship between \vec{p}_l and $\bar{\mathbf{q}}_t$ must be linearized in order to understand how errors in the estimation of $\bar{\mathbf{q}}_t$ translate to errors in \vec{p}_l .

Let the small deviations in the orientation be represented by the error quaternion $\delta\bar{\mathbf{q}}_t$ so that $\vec{p}_l = \mathbf{A}(\delta\bar{\mathbf{q}}_t)\mathbf{A}_t(\bar{\mathbf{q}}_t)^T \vec{p}'_l$, where $\mathbf{A}(\delta\bar{\mathbf{q}}_t)$ is the attitude matrix parameterized by the orientation error. Then, the goal here is to linearize the relationship between \vec{p}_l and $\delta\bar{\mathbf{q}}_t$, and to accomplish we find the matrix first-order derivative of \vec{p}_l with respect to the three dimensional representation of the error quaternion $\delta\vec{q}$:

$$\frac{\partial \vec{p}_l}{\partial \delta\vec{q}} = \frac{\partial}{\partial \delta\vec{q}} \mathbf{A}(\delta\bar{\mathbf{q}}_t)\mathbf{A}_t(\bar{\mathbf{q}}_t)^T \vec{p}'_l$$

Now simplify $\mathbf{A}(\delta\bar{\mathbf{q}}_t)$ using the assumptions that $\delta q_4 \approx 1$ and $|\delta\vec{q}| \approx 0$, both of which will be true for small errors:

$$\mathbf{A}(\delta\bar{\mathbf{q}}_t) = (|\delta q_4|^2 - |\delta\vec{q}|^2)I_{3x3} + 2\delta\vec{q}\delta\vec{q}^T - 2\delta q_4 \begin{bmatrix} \delta\vec{q}^\times \\ \end{bmatrix} = I_{3x3} - 2 \begin{bmatrix} \delta\vec{q}^\times \\ \end{bmatrix}$$

Pplugging in this expression for $\mathbf{A}(\delta\bar{\mathbf{q}}_t)$:

$$\frac{\partial \vec{p}_l}{\partial \delta \vec{q}} = \frac{\partial}{\partial \delta \vec{q}} (I_{3x3} - 2 \begin{bmatrix} \delta \vec{q}^\times \end{bmatrix}) \mathbf{A}_t (\bar{\mathbf{q}}_t)^T \vec{p}'_l$$

Note that $\mathbf{A}_t (\bar{\mathbf{q}}_t)^T \vec{p}'_l$ is a constant vector – in fact it is simply the estimate for \vec{p}_l !

$$\frac{\partial \vec{p}_l}{\partial \delta \vec{q}} = \frac{\partial}{\partial \delta \vec{q}} (I_{3x3} \vec{p}_l - 2 \begin{bmatrix} \delta \vec{q}^\times \end{bmatrix} \vec{p}_l) = -2 \frac{\partial}{\partial \delta \vec{q}} \begin{bmatrix} \delta \vec{q}^\times \end{bmatrix} \vec{p}_l$$

The quantity $\begin{bmatrix} \delta \vec{q}^\times \end{bmatrix} \vec{p}_l$ is simply $\delta \vec{q} \times \vec{p}_l = -\vec{p}_l \times \delta \vec{q}$. Therefore,

$$\frac{\partial \vec{p}_l}{\partial \delta \vec{q}} = -2 \frac{\partial}{\partial \delta \vec{q}} (-\vec{p}_l \times \delta \vec{q}) = 2 \begin{bmatrix} \vec{p}_l^\times \end{bmatrix} \frac{\partial}{\partial \delta \vec{q}} \delta \vec{q} = 2 \begin{bmatrix} \vec{p}_l^\times \end{bmatrix}$$

Now we can write deviations in \vec{p}_l as a linear function of the quaternion error or angular error ($\delta \vec{q} = \frac{1}{2} \delta \vec{\theta}$):

$$\delta \vec{p}_l = 2 \begin{bmatrix} \mathbf{A}_t (\bar{\mathbf{q}}_t)^T \vec{p}'_l^\times \end{bmatrix} \delta \vec{q} = \begin{bmatrix} \mathbf{A}_t (\bar{\mathbf{q}}_t)^T \vec{p}'_l^\times \end{bmatrix} \delta \vec{\theta}$$

This is written in the form of Equation 2. Thus, if $\Sigma_{\theta t}$ is the covariance of the estimation of the tool's orientation, then the covariance Σ_l of \vec{p}_l is

$$\Sigma_l = \left[\mathbf{A}_t (\bar{\mathbf{q}}_t)^T \vec{p}'_l^\times \right] \Sigma_{\theta t} \left[\mathbf{A}_t (\bar{\mathbf{q}}_t)^T \vec{p}'_l^\times \right]^T = \mathbf{A}_t (\bar{\mathbf{q}}_t)^T \left[\vec{p}'_l^\times \right] \Sigma_{\theta t} \left[\vec{p}'_l^\times \right]^T \mathbf{A}_t (\bar{\mathbf{q}}_t) \quad (3)$$

VIII. Conclusion

IX. Future Work

References

- [1] J. Verhage, “Goldman Sachs Has Four Charts Showing the Huge Potential in Virtual and Augmented Reality.” <http://www.bloomberg.com/news/articles/2016-01-13/goldman-sachs-has-four-charts-showing-the-huge-potential-in-virtual-and-augmented-reality>, 2016. Accessed: 2016-03-19.
- [2] J. L. Crassidis, F. L. Markley, and Y. Cheng, “Survey of nonlinear attitude estimation methods,” *Journal of guidance, control, and dynamics*, vol. 30, no. 1, pp. 12–28, 2007.
- [3] E. J. Lefferts, F. L. Markley, and M. D. Shuster, “Kalman filtering for spacecraft attitude estimation,” *Journal of Guidance, Control, and Dynamics*, vol. 5, no. 5, pp. 417–429, 1982.
- [4] M. D. Shuster, “Deterministic three-axis attitude determination,” *Journal of Astronautical Sciences*, vol. 52, no. 3, pp. 405–419, 2004.
- [5] R. Y. Tsai, “A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses,” *Robotics and Automation, IEEE Journal of*, vol. 3, no. 4, pp. 323–344, 1987.
- [6] G. R. Bradski, “Real time face and object tracking as a component of a perceptual user interface,” in *Applications of Computer Vision, 1998. WACV'98. Proceedings., Fourth IEEE Workshop on*, pp. 214–219, IEEE, 1998.
- [7] T. Flash and N. Hogan, “The coordination of arm movements: an experimentally confirmed mathematical model,” *The journal of Neuroscience*, vol. 5, no. 7, pp. 1688–1703, 1985.
- [8] R. M. Haralick, “Propagating covariance in computer vision,” *International journal of pattern recognition and artificial intelligence*, vol. 10, no. 05, pp. 561–572, 1996.