Demo: High-Precision Acoustic Motion Tracking

Wenguang Mao, Jian He, Huihuang Zheng, Zaiwei Zhang, Lili Qiu
{wmao,jianhe,huihuang,zaiwei,lili}@cs.utexas.edu
The University of Texas at Austin

ABSTRACT
Video games, virtual reality, augmented reality, and smart appliances all call for a new way for users to interact and control them. This paper develops high-preCision Acoustic Tracker (CAT), which aims to replace a traditional mouse and let a user control various devices by moving a smartphone in the air. At its heart lies a distributed Frequency Modulated Continuous Waveform (FMCW) that can accurately estimate the distance between a transmitter and a receiver that are separate and unsynchronized. We further develop an optimization framework to combine FMCW estimation with Doppler shifts to enhance the accuracy. We implement CAT on a mobile phone. The performance evaluation and user study show that our system achieves high tracking accuracy and ease of use using existing hardware.

CCS Concepts
• Human-centered computing → Ubiquitous and mobile computing systems and tools;

Keywords
Tracking; Acoustic signals; Doppler shift; FMCW.

1. INTRODUCTION
Motivation: Video games, Virtual Reality (VR), Augmented Reality (AR), and Smart appliances (e.g., smart TVs) all call for a new way for users to interact with them. This paper develops CAT, which aims to replace a traditional mouse and let a user play games, interact with VR/AR headsets, and control smart appliances by simply moving a smartphone in the air. Achieving high tracking accuracy is essential to provide enjoyable user experience.

Overview: We develop high-preCision Acoustic motion Tracking (CAT) system, which turns a mobile phone into a motion controller. CAT can be potentially used to control game consoles, VR/AR, and smart appliances. Refer to [2] for further details.

There has been significant work on motion tracking and localization. Recent works reduce the tracking error significantly by using many antennas and new spectrum (e.g., 60 GHz). Despite significant work on localization and tracking, achieving mm-level tracking on commodity devices remains an open challenge. Therefore, we aim to achieve high tracking accuracy, minimize error accumulation, and remove the need of special hardware. A unique feature of our approach is that it uses existing hardware already available, while achieving high accuracy and ease of use.

We use audio signals for tracking because (i) it propagates slowly, which makes it possible to achieve high accuracy, (ii) it can be supported by commodity devices thanks to widely available speakers and microphones, and (iii) its processing cost is low due to its low sampling rate.

In our system, the speakers serve as anchor points and play specially designed audio signals. A smartphone analyzes the received signals to accurately estimate the distances and velocities with respect to the speakers using a distributed Frequency Modulated Continuous Waveform (FMCW) and Doppler effect. It then fuses the distance and velocity in an optimization framework for accurate motion tracking.

Doppler shift is a well known phenomenon where the signal frequency changes as a sender or receiver moves. By tracking the amount of frequency shift, we can estimate the mobile’s velocity with respect to each speaker. To determine its position, the speed needs to be integrated over time, which incurs error accumulation.

To minimize error accumulation, we develop a novel FMCW-based approach to directly estimate the distance between the mobile and speakers. Our FMCW differs from existing approaches due to the distributed nature of our system: the speakers (i.e., senders) and the microphone on the mobile (i.e., receiver) are separate and unsynchronized. In this case, the transmission time, which is required by traditional FMCW approaches, is not known by the receiver. Our distributed FMCW addresses the issue using the following steps: i) find a reference point and determine its absolute position, ii) estimate the distance change with respect to the reference point when a mobile moves, 3) derive the absolute distance between the current point and speakers. In this way, we no longer need the transmission time. Moreover, the separate sender and receiver have different sampling frequencies. To address the issue, we develop a simple procedure to calibrate and compensate for the frequency offset.

Furthermore, we develop an optimization framework to incorporate FMCW measurements and Doppler shifts over time for accurate tracking. These two types of measurements are complementary: the former gives distance estimation, which does not have error accumulation, while the latter provides more accurate distance change in a short term. The framework can further incorporate IMU sensors (e.g., accelerometers and gyroscopes) to improve accuracy.

Our major contributions include: (i) a distributed FMCW ap-
proach that achieves accurate distance estimation without requiring synchronized and co-located sender and receiver; (ii) an optimization framework to combine distance and velocity estimation over multiple time intervals to accurately track motion, and an efficient algorithm to solve it on a mobile device; (iii) prototype systems that achieve 5-7mm tracking error in 2D and 8-9mm error in 3D.

2. OUR APPROACH

Velocity estimation: The Doppler effect is a well known phenomenon where the frequency of a signal changes as a sender or receiver moves. Without loss of generality, we consider only the receiver moves while the sender remains static. The frequency changes with the velocity as \( F' = F \frac{v}{c} \), where \( F \) is the original frequency of the signal, \( F' \) is the amount of frequency shift, \( v \) is the receiver’s speed towards the sender, and \( c \) is the propagation speed of sound waves. Therefore, by measuring the amount of frequency shift \( F' \), we can estimate the receiver’s velocity with respect to the sender, which can further be used to get distance and location. We use the approach described in [3] to estimate the Doppler shift.

Distance estimation: As shown in [3], the Doppler shift alone can be used to provide reasonable tracking for a short time. However, since the Doppler shift gives a velocity estimate, it has to be integrated over time to get a distance estimate. Therefore, the error grows over time. For tracking over a relative long time interval, the accuracy of the Doppler shift based tracking degrades.

To avoid error accumulation, we estimate the distances between the speakers and the mobile based on the propagation delay, which can be used to determine the location directly. Instead of sending a sharp pulse signal or pseudo-random sequence using large bandwidth [1, 4], we use FMCW-based approach to estimate the propagation delay, which can achieve high estimation accuracy with moderate bandwidth usage [1].

FMCW approach lets each speaker transmit periodic chirp signal, whose frequency sweeps linearly from \( f_{\min} \) to \( f_{\max} \) in each period. The frequency within each sweep is given by \( f = f_{\min} + Bt/T \), where \( B \) is the signal bandwidth, \( T \) is the sweep time.

The receiver mixer (i.e., multiplies) the received signal with the transmitted signal, and applies FFT on the mixed signal. It can be shown that the peak frequency \( f_p \) in the spectrum relates with the distance as \( R = f_p cT / B \).

Traditional FMCW assumes that the transmitter and receiver are co-located and share the same clock. However, in our system, the speakers and microphone are separate and unsynchronized. Thus, we develop a new distributed FMCW approach to support this situation. In our approach, we apply FMCW technique to derive the change of distance to the speaker when the mobile moves from one position to another. Moreover, we propose a scheme to find a reference point and leverage this point to convert the distance change to the absolute distance to the speaker. Also, we explicitly take into account the impact of movement on FMCW to improve its accuracy. We further account for the impact of sampling frequency offset between the speaker and microphone to get more accurate distance estimation.

Combining distance and velocity estimations: We propose the following optimization framework that combines the Doppler shift and FMCW measurements for accurate motion tracking. Specifically, we minimize the following function:

\[
\sum_{i \in [k-n+1 .. k]} \sum_{j} \alpha (|z_i - c_j| - |z_0 - c_j| - d_{FMCW}^{i,j})^2 + \\
\sum_{i \in [k-n+2 .. k]} \beta (|z_i - c_j| - |z_{i-1} - c_j| - v_{i,j}^{doppler}T)^2
\]

where \( k \) is the current processing interval, \( n \) is the number of intervals used in the optimization, \( z_i \) denotes the mobile’s position at the beginning of the \( i \)-th interval, \( z_0 \) denotes the reference position, \( c_j \) denotes the \( j \)-th speaker’s position, \( d_{FMCW}^{i,j} \) denotes the distance change from the reference location with respect to the \( j \)-th speaker at the \( i \)-th interval, \( v_{i,j}^{doppler} \) denotes the velocity with respect to the \( j \)-th speaker during the \( i \)-th interval, \( T \) is the interval duration, and \( \alpha \) and \( \beta \) are the relative weights of the measurement from FMCW and Doppler shifts, respectively.

The objective reflects the goal of finding a solution \( z_i \) that best fits the FMCW and Doppler measurements. The first term captures the distance calculated from the coordinates should match the distance estimated from the FMCW, and the second term captures the distance traveled over an interval should match with the distance derived from the Doppler shift. Our objective consists of terms from multiple intervals to improve the accuracy. The formulation is general, and can support 2-D or 3-D coordinates.

The only unknowns in the optimization are the mobile’s positions over time (i.e., \( z_i \)). The speakers’ coordinates \( c_j \) can be determined using the method proposed by [3]. \( d_{FMCW}^{i,j} \) and \( v_{i,j}^{doppler} \) are derived from FMCW and Doppler shift measurements, respectively. We develop a novel algorithm to efficiently solve this non-convex optimization problem.

3. PROPOSED DEMO AND EQUIPMENT

The proposed demo consists of two parts. First, we will demonstrate the tracking accuracy by comparing the distance measurements with readings from a ruler. We let a speaker continuously transmit specific acoustic signals. We invite a user to place a smartphone at an arbitrary distance away from the speaker. The smartphone will analyze the acoustic signal to compute and display the distance from the speaker. The user can easily evaluate the accuracy of our tracking system by comparing the readings displayed by the smartphone with the readings measured by the ruler.

Second, we will apply our tracking to motion-based computer gaming. We make the smartphone running CAT serve as a motion controller for video games by mapping the phone’s movement into a cursor movement in games using Windows API mouse_event. We invite a user to play a motion-based game using the smartphone. The smartphone acts as a game controller and analyzes the audio signal transmitted by several speakers to determine its location in real-time and feeds back to the computer, which will update the cursor position in the game. Users can assess the accuracy of CAT based on their gaming experience (e.g., whether the cursor in the game moves as they intend).

We will bring a laptop, smartphone, and speakers required for the demo. We need a space at least 3 m × 3 m, and a power outlet. We would like to have access to our demo booth at least 2 hours in advance of the session to test our tracking system.

Acknowledgements: This work is supported in part by NSF Grants IIP-1546831, CNS-1343383, and Cisco gift funding.

4. REFERENCES


