Advancing Hand Gesture Recognition with High Resolution Electrical Impedance Tomography

Yang Zhang               Robert Xiao           Chris Harrison
Carnegie Mellon University, Human-Computer Interaction Institute
5000 Forbes Avenue, Pittsburgh, PA 15213
{yang.zhang, brx, chris.harrison}@cs.cmu.edu

ABSTRACT
Electrical Impedance Tomography (EIT) was recently employed in the HCI domain to detect hand gestures using an instrumented smartwatch. This prior work demonstrated great promise for non-invasive, high accuracy recognition of gestures for interactive control. We introduce a new system that offers improved sampling speed and resolution. In turn, this enables superior interior reconstruction and gesture recognition. More importantly, we use our new system as a vehicle for experimentation – we compare two EIT sensing methods and three different electrode resolutions. Results from in-depth empirical evaluations and a user study shed light on the future feasibility of EIT for sensing human input.

Author Keywords
Electrical Impedance Tomography; EIT; hand gestures; smartwatch; bio-impedance; biometrics; input.

ACM Classification Keywords
H.5.2. [Information interfaces and presentation]: User interfaces – Input devices and strategies.

INTRODUCTION
Tomography is an imaging technique that estimates the cross-sectional interior structure of objects through the use of an external, penetrating signal [27]. In this work, we use Electrical Impedance Tomography (EIT) [10,22], which uses surface electrodes and high frequency alternating current (AC) to measure internal electrical impedance. By placing many electrodes around an object, it is possible to reconstruct the internal impedance distribution and infer the interior structure [10]. EIT is safe for long-term continuous operation (low voltage, no ionizing radiation), non-invasive to the wearer (rests on skin without the need for e.g., conductive gel) and can be made inexpensive (~$50). For these reasons, we recently adapted the technology for use in a gesture-sensing smartwatch called Tomo [43]. Tomo provided a glimpse into the applicability of EIT for input sensing. To further expand the feasibility of this promising technique, we sought to explore technical improvements to low-cost, worn, EIT sensing. The most obvious parameter that can be varied is the number of electrodes. Intuitively, more electrodes will produce a denser mesh of sensed paths (Figure 1), which should yield a superior reconstructed image.

In addition to varying the number of electrodes, EIT systems can also choose between two very different types of EIT sensing (see e.g., [1,11] for more details). The first method is “two-pole” sensing, in which impedance measurements are captured from each pair of skin electrodes – one acting as an emitter and the other as a receiver. This method is affected by skin impedance [1,8,20] and so larger electrodes are typically used for greater contact area with the skin, precluding dense electrode arrays. However, this approach is popular due to its technical simplicity.

Alternatively, EIT systems can use a more sophisticated “four-pole” scheme, which excites an adjacent pair of electrodes with an AC signal and measures the voltage between another pair of electrodes (Figure 5). This process is repeated for all possible emission and measurement pairs. This differential measurement approach makes four-pole sensing less sensitive to contact conditions at the skin [1,15].

We built our EIT system such that it could toggle between two- and four-pole sensing schemes, as well as 8-, 16- and 32-electrode configurations (i.e., six EIT configurations in

Figure 1. The number of sensed paths (grey lines) dramatically increases as electrode count grows (red dots). For reference, Tomo [43] uses a two-pole, 8-electrode scheme (upper left).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
RELATED WORK
Our work broadly intersects with two large research domains: 1) hand gesture sensing from a wrist- or arm-worn device, and 2) electrical impedance tomography.

Wrist Gesture Sensing
Humans naturally gesture with their hands, which can form many static poses and kinetic gestures. For this reason, they have long been studied for controlling interactive systems. For example, Digits [28] reconstructed a 3D model of the hand using a wrist worn camera, while Way et al. [40] used a time-of-flight camera to sense small free-hand gestures. Similarly, zSense [39] used infrared sensors deployed on a ring or smartwatch to detect both static and dynamic poses. Other systems have leveraged IMUs on the wrist to detect dynamic hand motions such as finger rubbing and hand waving [29,41]. Researchers have also utilized arm contour changes that occur when performing different hand gestures, including sensing capacitance [32], pressure [13,26], and proximity [17].

More similar to our work are systems that use bio-sensing. For example, electromyography (EMG) senses the electrical signals produced by muscle activation. Saponas et al. [33,34] built an EMG-based system that supported 4 pinch gestures at 79% accuracy. Another popular approach is bioacoustics, which use vibrations that propagate through the body upon performing hand gestures. For example, Hambone [14] used contact microphones to detect 4 gestures with ~90% accuracy. The Sound of One Hand [3] used a similar setup to detect finger gestures such as rub, tap, and flick. Skinput [21] also leveraged bioacoustics to detect 8 flicking/pinchning gestures at 87.3% accuracy.

EIT Image Reconstruction
Tomographic image reconstruction is well studied in the signal processing literature, and a number of popular algorithms exist. A good overview of current approaches can be found in [38]. The basic goal of EIT image reconstruction is to obtain the “conductivity image” of the interior of an object. The interior is discretized using a finite element method to generate a mesh, and the conductivity at each mesh element is computed. Generally, finer meshes provide higher resolution output images but increase the computational requirements.

Early algorithms were based on linear back-projection (e.g., [9]), which is commonly used for PET and CT image reconstruction (using gamma and X-rays respectively). This technique assumes that the electrical current travels approximately along certain fixed equipotential lines. However, as this does not accurately capture the complete 3D movement of the electric signals, accuracy can be limited [42]. Tomo [43] employed this method to obtain computationally inexpensive and straightforward image reconstruction.

More recent methods use least-squares optimization to find the “best-fit” image [38]. To solve this nonlinear problem, systems often employ Newton-Raphson or Gauss-Newton iteration, which are computationally expensive. For real-time use, sophisticated single-iteration methods, such as the maximum a posteriori estimator [2], are able to produce acceptable images rapidly after performing significant pre-computation. We use the latter approach.

Applications of Electrical Impedance Tomography
Electrical Impedance Tomography has been applied to many application domains. For example, due to its safe and non-invasive nature, it has been widely used in clinical applications such as sensing lung ventilation [35], brain function [25] and blood flow [19]. EIT has also seen use in geophysics (e.g., monitoring mountain permafrost [23]), environmental science (e.g., underground pollutant detection [12]), biology (e.g., tree inspection [36] and industrial monitoring (e.g., measuring liquid flow [16]).

Our previous Tomo system [43] showed that two-pole EIT could be made compact and low cost, making it more amenable for use in consumer electronics. Tomo used an off-the-shelf AD5933 impedance converter chip for measuring bio-impedance, which takes 3.6 milliseconds to perform one measurement. Sensing all combinations of its 8 electrodes (28 pairs) to generate a single image frame takes roughly 100 ms, yielding a frame rate of 10 Hz. Unfortunately, this hardware cannot easily scale to 16 and 32 electrode configurations, which have 120 and 496 electrode pairings respectively (see Table 1).

As we will describe in greater detail in the next section, our new EIT system uses both custom hardware and software, granting us tight control over the entire sensing pipeline. In contrast to Tomo, our system can complete a single measurement in just 0.33 ms, which allows us to achieve 100 frames per second in a two-pole, 8-electrode configuration. More interestingly, it enables us to scale to greater numbers of electrodes (Table 1). This, in turn, offers superior sensing accuracy, as we discuss in our evaluation.

<table>
<thead>
<tr>
<th>Table 1. Performance characteristics of Tomo and our new setup. We extrapolate hypothetical performance (grey region) for 16 and 32 electrode versions of Tomo.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of Electrodes</strong></td>
</tr>
<tr>
<td><strong>(Two-pole)</strong></td>
</tr>
<tr>
<td>-------------------------</td>
</tr>
<tr>
<td>Number of Measurements</td>
</tr>
<tr>
<td>per Frame</td>
</tr>
<tr>
<td>Frame Rate (Hz)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
IMPLEMENTATION
Our setup has two main components: a wristband (Figure 2), which features an array of electrodes to be worn by a user, and a custom-designed data capture and processing board. Our board is built around a Cortex-M4-based Teensy 3.2 system-on-chip (SoC), augmented with a voltage controlled current source (VCCS), direct digital synthesis (DDS) IC, and ADC preamp (Figure 3). The board also features multiplexers that allow for dynamic electrode selection, enabling different EIT configurations. A schematic view of our system is shown in Figure 4. The total cost of our hardware is $80, which could be made both smaller and less expensive in a high volume commercial application.

Wristband
We made a leather wristband with 32 evenly spaced stainless steel electrodes (Figure 2). Each electrode measures $3.4 \times 15.3$ mm. The average human forearm diameter is roughly 190 mm in circumference, suggesting an upper-bound electrode width of ~6 mm in a 32-electrode configuration. The band is secured to the user with a Velcro strap.

Excitation Signal
We use an AD5930 [7] DDS IC and an AD8220-based VCCS [5] to generate the EIT excitation signal. The AD5930 is configured to output 40 kHz sinusoidal waves (the same frequency used in Tomo [43]). This signal is then fed into the VCCS to output a constant 300 µA AC current (0–6 Vpp depending on the load impedance).

Multiplexing
Two 32-to-1 multiplexers (ADG732 [6]) connect the VCCS terminals to any two electrodes, forming the signal-projection pair. Two more multiplexers connect the preamp buffer terminals to two electrodes to form the voltage-measuring pair. In two-pole EIT sensing, we measure the voltage between the VCCS output and the receiver electrode. In this case, the voltage measurement pair is the same as the signal projection pair. In the four-pole scheme, the voltage measurement pairs are different from the current projecting pair. Note the system captures the voltage difference between electrodes (i.e., we do not compare to ground).

Analog Sampling
We first amplify our signal with a preamp. The gain value is adjusted to maintain a consistent dynamic range when switching between four-pole and two-pole measurement schemes. We also use a high pass filter with a 15.6 kHz cutoff frequency to remove ambient EMI (e.g., from power-line noise). The input signal is then biased by AVDD/2 (1.65 V) and sampled by our microprocessor’s ADC at 2 MHz with 12-bit resolution.

Data Acquisition
Once the multiplexer has selected the appropriate electrodes, we wait 100 µs for the DC bias on AC coupling capacitor to stabilize. We then collect 250 samples, or roughly five periods of the 40 kHz excitation signal (collecting multiple periods to reduce noise). The root-mean-square (RMS) of the samples is calculated to form a single measurement. The sensor then moves to the next measurement, reconfiguring the multiplexers accordingly. After it collects all values for the current frame, it sends the RMS measurements to a laptop over Bluetooth. The number of measurements and our system’s frame rate can be found in Table 1.

Two-Pole and Four-Pole Measurement Schemes
As noted in the introduction, the simplest EIT setup uses a two-pole scheme. In each measurement, one pair of electrodes is used for both signal emission and voltage measurement, and all pairs are tested for a total of $N_e \choose 2 = N_e(N_e - 1)/2$ measurements.
In four-pole sensing, separate pairs are used for signal projection and voltage measurement. The Adjacent Drive method [11,24] is the most common projection pattern – the signal is applied through one pair of adjacent electrodes and the voltage difference is measured between the $N_e - 3$ other adjacent electrode pairs. This is repeated for all emitter pairs, resulting in a total of $N_e(N_e - 3)$ measurements.

Figure 5A illustrates the first signal projection round for the four-pole scheme. The signal is emitted using electrodes 1 and 2, and the voltage differences $V_1, V_2, \ldots V_4$ are measured sequentially with five electrode pairs: 3-4, 4-5, \ldots 7-8. In the second projection round (Figure 5B), the next pair (2-3) is used for emission, and the voltage differences are measured sequentially with another five electrode pairs. This process is repeated sequentially until a full loop as been completed ($8 \times (8 - 3) = 40$ measurements).

Baseline Calibration
Although we endeavored to make our sensing board and wristband as consistent as possible, there are nonetheless small variations in the pathways to different electrodes. To prevent this from impacting our tomography, we must obtain a baseline measurement of the impedance between all pairs of electrodes. With 928 electrode pairings in our four-pole configuration, it was impractical to perform these calibrations manually (as was done for Tomo’s [43] 28 pairs). Instead, we calibrated our setup by wrapping the wristband around a homogenous, electrically conductive material and capturing one frame of data using our standard sensing pipeline to serve as a baseline. We experimented with a variety of materials: ground beef in Saran wrap, lean pork chop, cylinder of ice, saltwater bath and Jell-O, with the latter producing the best results.

Interior Image Reconstruction
We performed all interior image reconstruction on a 13" 2015 MacBook Pro with a 2.7 GHz Intel Core i5 processor. The reconstruction algorithm was derived from the EIDORS toolkit [37], which provides a large library of different solvers. After preliminary experimentation, we chose the nodal one-step Gauss-Newton iterative solver, which produces a maximum a posteriori (MAP) estimate [2] of the conductivity at each node of the finite element mesh.

This algorithm is parameterized by a single hyperparameter $\mu$, which controls the smoothing of the output. In all of our experiments, we fix the hyperparameter value at $\mu=0.03$ for the two-pole configuration and $\mu=0.001$ for the four pole configuration. The solver requires a precomputation step, which takes about three seconds on our laptop; subsequent image reconstruction is carried out as a single matrix multiplication, taking 2.4 ms per image (numbers provided for 32 electrodes, the most computationally expensive case).

Hand Gesture Classification
Since interior image reconstruction is sufficiently fast to be used in real-time system, we chose to derive machine learning features purely from the reconstructed images. Specifically, we downsampled the reconstructed images to $16 \times 16$ pixels, and used the resulting 256 data points as raw features. We used a support vector machine (SVM) implementation provided by the Weka Toolkit [18] for classification (SMO; polynomial kernel with default parameters).

FIDELITY EXPERIMENTS
We ran a series of experiments to evaluate the fidelity of our six different EIT configurations. These were designed to elicit fidelity characteristics of the tomographic output.

Apparatus
In order to analyze the quality of our tomographic reconstruction, we needed a stable and known reference input. For this, we used a 20 cm diameter acrylic cylinder filled with 3 cm of saline water (9000 ppm NaCl, approximating the conductivity of human tissue). We used acrylic shapes to make different EIT “phantoms” [15,19,30] for testing; the impedance of these shapes is approximately $10^{17}$ times that of saline water at 25°C.

Number of Interior Features
We varied the number of the cylinders (2.25” diameter) in the bath from one to five to see how well different EIT configurations can resolve distinct objects (Figure 6). As one would expect, as resolution increases, the ability to discern dense constellations of objects improves. For example, in both two- and four-pole sensing schemes, five objects are not discernable when using 8 electrodes. Even at 16 electrodes, it is hard to separate them. However, with 32 electrodes, all five objects can be readily segmented. Overall, at each electrode level, four-pole offers improved fidelity (particularly with respect to background uniformity) compared to two-pole.

Size of Interior Features
Resolving the size of features is also important, as the human body has bones and muscles of varying cross-sectional size. Additionally, we wished to see if small features were detectable. To explore this, we tested our EIT configuration.
tions with circular reference shapes of three different sizes (1.5”, 2.25” and 3”). We used all combinations of these three objects as reference inputs. Output is shown in Figure 7, showing the significant fidelity improvement offered by four-pole sensing. Specifically, both relative and absolute object sizes are clearly distinguishable in the 16 and 32 electrode versions of our four-pole setup, while the two-pole setup has difficulty even separating the two objects.

Shape of Interior Features
Unlike directional waves (e.g., X-rays), electric current is not confined to a path ray or a plane, and thus a change in impedance anywhere in the domain (not just along the path ray) will affect the measurement. Furthermore, this “soft field problem” [30] causes the electrical path to curve away from objects along the path, resulting in incomplete occlusion (unlike the perfect occlusion from e.g. an optical mesh system [31]). Thus, shape recovery is particularly challenging for EIT systems.

To test this property in our EIT setup, we used paired small and large squares, triangles, cylinders and rectangles (small size 1.5”; large 3”), as seen in Figure 8. As with our size experiment, two-pole sensing only discerns that two objects are present, but little other detail is visible. Four-pole performs substantially better, with steady improvements in fidelity as electrode resolution grows. However, even four-pole reconstruction with 32 electrodes does not resolve sharp edges (e.g., our triangle reference shape). We are very likely approaching the upper limit of what can be reconstructed using low-cost EIT.

Biological Reference
While the various reference shapes used above are useful for controlled testing, they do not approximate biological tissue. Unfortunately, no one at our institution was willing to give us a cross-sectional slice of their arm to use as a ground truth. Instead, as a biological reference, we used a cross-cut lamb shoulder (Figure 9, left). This closely emulates human tissue, and further offers reference features with more subtle impedance differences (unlike salt water and plastic, which are polar opposites with respect to conductivity). As can be seen in Figure 9, two-pole sensing struggles with reconstruction at any electrode resolution, in line with our previous results. In particular, two-pole has trouble localizing the high impedance bone when it is roughly in the middle, yielding a bulls-eye-like impedance distribution. Four-pole does better, and with 32 electrodes, the bone and fat are both visible.

USER STUDY
The ultimate objective for our EIT system is interactive control. To evaluate the feasibility of the different EIT configurations we support, we ran a hand gesture recognition user study. In order to compare our results with Tomo [43], we adopted its experimental procedure and gesture set (which were themselves chosen to be comparable to prior work). Specifically, we used a finger pinch gesture set: Index Pinch, Middle Pinch, Ring Pinch, and Little Pinch, and a hand gesture set, designed around coarse hand motions: Fist, Stretch, Right, Left, Thumbs Up, Spider-Man and Index Pinch. A relax gesture was included in both sets as the neutral state, resulting in a total of five pinch gestures and eight hand gestures.

Participants
We recruited 10 participants (four female), with a mean age of 23. The study took approximately thirty minutes to complete and paid $20. All of our participants were right-handed, so the sensor wristband was fitted to participants’ left arm (where a watch would usually be). Participants were seated for the duration of the study with their left elbow resting on a table. After participants felt comfortable, data collection began.
Procedure

We first collected data using four-pole sensing. Participants were asked to perform all 11 gestures, one gesture at a time, in a random order. While holding a gesture pose, the system automatically switched between 8, 16 and 32 electrode configurations, with ten trials recorded for each (for a total of 30 trials). This process took 15 seconds and our laptop beeped to indicate the participant could relax. We then switched to two-pole sensing mode, and repeated the same procedure. Thus, a round of data collection captured 10 trials × 11 gestures × 3 electrode resolutions × 2 EIT schemes = 660 gesture trials.

In total, we ran five rounds of data collection, yielding 33,000 gesture trials (660 trials per round × 5 rounds × 10 participants). To estimate the gesture recognition accuracy of our EIT system, we performed a post hoc, leave-one-out train/test experiment. Specifically, we trained our classifier on four rounds of a user’s data, and tested on the fifth (all combinations). This ensured that any two data points close in time (which naturally tend to be more similar) were either in the test set or the training set, but never both.

RESULTS AND DISCUSSION

There are several high level conclusions to be drawn from our results. First, increasing electrode count strictly improves recognition accuracy (roughly +3.8% when moving from 8 to 16, and +1.5% when moving from 16 to 32). Secondly, four-pole handily out-performs two-pole sensing (on average, 6.8% better across all electrode resolutions and gesture sets). Interestingly, four-pole recognition accuracies do not dip when combining hand and pinch gesture sets, which suggest they are complementary. Unsurprisingly, our best performing configuration is four-pole sensing with 32 electrodes; on the 11-gesture hand+pinch set, it achieves a mean accuracy of 94.3% across our ten participants – a 67.8% reduction in gesture recognition error vs. Tomo [43]. Figure 10 provides an overview of these results. Overall, our results suggest EIT is competitive with more mature bio-sensing input methods, including bioacoustics (e.g., smartwatches that auto-authenticate when worn) and bio-factors (e.g., bone density, pulse, and blood pressure). Medical-grade EIT systems have shown this to be possible [8,19].

Currently, we only recognize discrete hand gestures. However, we have analog data of muscle movement inside the arm, and so we believe it may be possible to build a continuous kinetic model using regression. In addition, with our improved sample rate and tomographic resolution, it may be possible to capture biometric data (e.g., smartwatches that auto-authenticate when worn) and bio-factors (e.g., bone density, pulse, and blood pressure). Medical-grade EIT systems have shown this to be possible [8,19].

To facilitate comparison to prior work, we deliberately adopted the gesture set used in Tomo [43]. However, our system is not limited to these gestures, and it is likely that other gestures can be detected. Alternate gesture sets may even provide better recognition accuracies, and future work could explore a wider variety of gestures to find those with distinctive EIT signatures. Further extending Tomo, future iterations could explore other sensing locations, including the face, throat, legs and feet, which would require new electrode form factors.

Finally, there are also technical improvements we wish to explore. For example, using (or combining) different excitation frequencies could yield better reconstructions [35]. Swept frequency excitations may also be possible. We also want to try the “auto rotation” approach proposed by Amma et al. [4], which demonstrated a high-density EMG sensor band that automatically detected the orientation with respect to the arm. This orientation data was used to digitally rotate the signal in order to provide robust classification no matter how the user wore the device.

ACKNOWLEDGMENTS

This research was generously supported by the David and Lucile Packard Foundation and a Google Faculty Research Award.
REFERENCES


