12.1 3D Ultrasonic Gesture Recognition

Richard J. Przybyla1, Hao-Yen Tang2, Stefan E. Shelton2, David A. Horsley3, Bernhard E. Boser4

1University of California, Berkeley, CA, 2University of California, Davis, CA

Optical 3D imagers for gesture recognition suffer from large size and high power consumption. Their performance depends on ambient illumination and they generally cannot operate in sunlight. These factors have prevented widespread adoption of gesture interfaces in energy- and volume-limited environments such as tablets and smartphones. Wearable mobile devices, too small to incorporate a touchscreen more than a few fingers wide, would benefit from a small, low-power gestural interface. Gesture recognition using sound is an attractive alternative to overcome these difficulties due to the potential for chip-scale size, low power consumption, and ambient light insensitivity. Using pulse-echo time-of-flight, MEMS ultrasonic rangers work over distances of up to a meter and achieve sub-mm ranging accuracy [1,2]. Using a 2-dimensional array of transducers, objects can be localized in 5 dimensions.

This paper presents an ultrasonic 3D gesture-recognition system that uses a custom transducer chip and an ASIC to sense the location of targets such as hands, down to sub-mm scales in reality. The system block diagram is shown in Fig. 12.1.1. Targets are located using pulse-echo time-of-flight methods. Each of the 10 transceiver channels interfaces with a MEMS transducer, and each includes a transmitter and a readout circuit. Echoes from off-axis targets arrive with different phase shifts for each element in the array. The off-chip digital beamformer realigns the signal phase to maximize the SNR and determine target location.

The 450μm diameter piezoelectric micromachined ultrasound transducers (pMUTs) used in this work are made up of a 2.2μm thick AlN/Mo/AIn/AIn stack deposited on a Si wafer and released with a back-side through-wafer etch. The bottom electrode is continuous, while each pMUT has a top electrode lithographically defined to actuate the trampoline mode. Each pMUT can transmit and receive sound waves, and is operated at its resonance of 217kHz ± 2kHz with a bandwidth of 12kHz. The impedance of the transducers is defined to actuate the trampoline mode. Each pMUT can be bi-polar switched, setting the transducer’s bottom electrode to 16V to permit bi-polar actuation of the transducer. The transmitter then excites the transducer with a 32Vp-p transmit switches, setting the transducer’s bottom electrode to 16V to permit bi-polar actuation of the transducer. The transmitter then excites the transducer with a 32Vp-p square wave for 30 cycles at the transmit frequency fTX which is locked to 1/16th of the sampling frequency fs. At the end of the transmit phase, the frequency autotuning loop as it is enabled. An initial 57kHz offset frequency is turned on, and a resistor converts the ringdown current to a voltage that is quantized by a comparator. The high in-band gain provided by the 4-th order ΔΣ ADC is I/Q demodulated, filtered, and downsampled off-chip. A digital beamformer [3] processes the received signals to maximize the receive SNR and determine the x-angle location of the target. This process can be repeated in the orthogonal angle axis to implement 3D beamforming; in this work we forgo 3D beamforming since the tiny y-axis aperture does not provide any y-axis resolution.

The thermal noise in the front-end amplifier and the thermal motion of air limit the minimum detectable echo. The input-referred noise of the amplifier is 11nV/√Hz, and the noise voltage of the transducer is 6nV/√Hz at resonance. Figure 12.1.4 shows the measured signal-to-noise ratio vs. range for a 127mm×181mm flat rectangular target. Figure 12.1.4 also shows the rms error in the range and direction measurement. Amplitude noise in the received signal limits the accuracy of the time-of-flight estimate. Figure 12.1.5 shows the output of the digital beamformer from a single measurement, which captures the echoes from a user’s hands and head as he poses as shown. The system tracks objects between 45mm to 1m away and over an angular range of ±45°. Echoes from targets at a range of 1m return after 5.8ms, and this sets the maximum measurement rate of the system at 172 frames per second (fps).

The output of each ΔΣ ADC is I/Q demodulated, filtered, and downsampled off-chip. A digital beamformer [3] processes the received signals to maximize the receive SNR and determine the x-angle location of the target. This process can be repeated in the orthogonal angle axis to implement 3D beamforming; in this work we forgo 3D beamforming since the tiny y-axis aperture does not provide any y-axis resolution.

References:


Figure 12.1.1: System block diagram.

Figure 12.1.2: Readout circuit with mixed CT/SC architecture for inherent antialiasing. All structures are implemented differentially.

Figure 12.1.3: Ringdown frequency offset measurement and tuning loop settling behavior.

Figure 12.1.4: Signal-to-noise ratio and target localization accuracy vs. range for 127mm×181mm flat rectangular target.

Figure 12.1.5: Echo from user’s hands and head when posing as shown. Color axis shows y-angle position of the targets. Beamformed data is thresholded at 12dB SNR.

Figure 12.1.6: Comparison table.
Figure 12.1.7: CMOS die photo and MEMS ultrasound die photo.