Micro hand gesture recognition system using ultrasonic active sensing method

Project Hug (Hand Ultrasonic Gesture) September 2015 -May 2016

Demonstration: https://www.youtube.com/watch?v=8FgdiIb9WqY

Authors:

Yu Sang sangy16@mails.tsinghua.edu.cn

Quan Wang wangquan.thu@aliyun.com

Mentor:

Yimin Liu yiminliu@tsinghua.edu.cn

Intelligent Sensing Lab, Dept. of Electronic Engineering,

Tsinghua University,

Beijing, China

Micro hand gesture recognition system using ultrasonic active sensing method

Yu Sang, Quan Wang, and Yimin Liu, *Member, IEEE*

***Abstract*—We define micro hand gesture recognition system as which uses micro dynamic hand gestures within a time interval for classification and recognition to achieve human-machine interaction. Our Project Hug (Hand-Ultrasonic-Gesture), with ultrasonic active sensing, pulsed radar signal processing, and time-sequence pattern recognition is presented in this paper for micro hand gesture recognition. We leverage one single channel to detect both range and velocity precisely, reducing the hardware complexity. And to make use of sequencial range-Doppler information, high dimensional features are symbolized, which significantly reduceds the computation complexity and power consumption. A real-time prototype was released and an average recognition accuracy of 90.5% for seven gestures was achieved .**

I. Introduction

There is a growing need for touchless ways to interact with the smart world. In scenarios such as wearable devices, driving assistance, smart home and virtual reality, interaction based on haptic controls and touch screens may be physically or mentally limited. Thus the development of touchless ways for interaction is becoming more and more important.

Considering of precision and agility, we'd like to address technologies which base on micro dynamic hand gestures involing the movements of multiple fingers are promising approches for this purpose. Rather than wide-range and large-shift actions such as waving or rolling hands, these micro hand gestures such as tapping, clicking, rotating, pressing, rubbing, etc (see Fig. 1), are exactly the ways how people using their hands in the real physical world. Directly using the hand and multi-fingers to control devices is a natural, flexible, and user-friendly way without extra learning efforts.

Numerous approaches based on RGB-D or infrared cameras for gesture recognition have been developed even commertialized such as Microsoft Kinect or Leap Motion. However, to our best knowledge, these camera-based approches with vision or depth perform well on relative wide-range gestures, but cannot be applied for micro hand gesture recognition due to the limitation of resolving ability and environment factors. These approaches are not computational efficient to achieve millimeter-level precision to track multi-fingers' movements. Besides, their reliability in harsh enviroments, such as in dark night, or under direct sunlight is another problem. In comparision, approaches based on ultrasonic or radio frequency (RF) signals can obtain information of range/depth, direction and instantaneous velocity, which is the core signature of micro dynamic hand gestures. These approaches can achieve millimeter-level precision to distinguish different fingers regardless of environment condition. The hardware can be integrated and miniaturized with MMIC or MEMS techniques, consuming much less energy (400μW at 30fps for 1m maximum range with ultrasonic MEMS technique) than CMOS image sensors (>250mW) [1]. We present our Project Hug (Hand-Ultrasonic-Gesture), to build a micro hand gesture recognition system with active ultrasonic sensors, radar signal processing technique, and time-sequence pattern recognition methods in this paper. The main contributions of our project are summarized as follows:

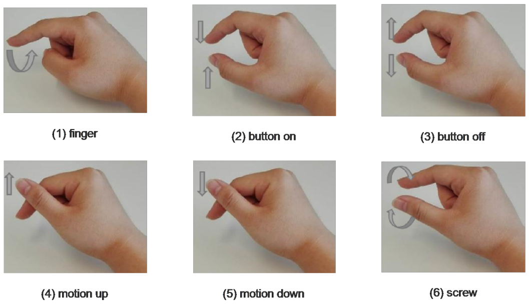

Fig. 1. Examples of micro hand gestures, named by finger, button on (BtnOn), button off (BtnOff), motion up (MtnUp), motion down (MtnDn), and screw, respectively.

- We use pulsed radar signal processing technique to obtain time-sequencial range-Doppler features. We measure objects' distances and velocities precisely and simultaneously through a single channel, which can reduce the hardware complexity.
- A hidden markov model (HMM) approach is used to classify the time sequencial range-Doppler features. We propose a state transition mechanism to symbolize the features which significantly compresses the data and extract the most intrinsic signatures. In addition, instantaneous velocity deduced from the Doppler effect is directly involved in motion detection and segmentation in realtime application. These approaches improve the computation and energy efficiency of the system.
- We released a realtime micro hand gesture recognition prototype and demonstrate the control of a music player with our new system.

Related Work: In 2012, Microsoft Research introduced a work that leveraged the speakers and microphones already embedded in commodity devices to sense in-air gestures[2]. In 2013, N. Patel et al developed WiSee, a novel gesture recognition system that leveraged wireless signals to enable whole-home sensing and recognition under complex conditions[3]. These projects arouse researches based on ultrasound or RF signals to recognize wide-range and large-shift human gestures.

Similar to our ideas, Google announced Project Soli at 2015, aiming at controlling smart devices and wearables through human micro hand gestures[4]. Our project differs from Google Soli in several aspects. Ultrasound transmits at a much slower speed than light, enable us to obtain millimeter soundwave at much lower frequencies (less than 1MHz), which can reduce the front-end circuit complexity. Soli uses a frequency modulated continuous wave radar and a direct-sequence spread spectrum radar to acquire the range and Doppler, while we develop the pulsed radar based on ultrasonic sensors and simultaneously measure the range and Doppler with high resolution.

II. A brief description of Project Hug

A. Pulse radar Signal Processing

Pulse radar measures the target range via round trip delay and uses the Doppler effect to determine the velocity by emitting pulses and processing the returned signals at a specific pulse repetition frequency. Two important procedures in this method are fast-time sampling and slow-time sampling, while the former refers to the sampling inside the pulse repetition interval (PRI), and determines the range; the latter refers to the sampling across multiple pulses to determine the Doppler shift.

In our project, we use the pulsed radar signal processing method to detect the palm and the fingers' movement. The Chirp pulses are applied to improve the SNR as well as maintain the range resolution. Thus the baseband waveform can be expressed as:

$$x(t) = \sum_{m=0}^{M-1} a(t - mT), \tag{1}$$

where

$$a(t) = \cos(\pi \frac{B}{\tau} t^2 - \pi B t) \qquad 0 \leq t \leq \tau. \tag{2}$$

In the expressions, $a(t)$ is the single pulse with time duration τ, B is the bandwidth, T stands for the PRI, and M is the total number of pulses which is carefully chosen to meet the stop-and-hop assumption. In terms of range or velocity resolution, micro hand gesture recognition requires millimeter-level range resolution and centimeter-per-second-level velocity resolution to discriminate actions of different fingers. System parameters can be chosen according to the required resolution R_d and v_d (3), where c is the propagation speed and λ is the wavelength of the ultrasound:

$$R_d = \frac{c}{2B}, \qquad v_d = \frac{\lambda}{2MT}. \tag{3}$$

Brief implemention of the ultrasonic signal processing is given as follows. Leveraging one single channel, received reflections of the ultrasonic waves is processed via the I-Q quadrature demodulator and the low-pass filter to get baseband signals. With fast-time sampling, the matched filters are applied to detect ranges of the palm and fingers. With slow-time sampling, FFTs are applied to the samples of different pulses which correspond to the same time delay to detect objects' velocities. This implemention generates a 3-dimentional (range-Doppler-time) data cube as the output. Fig. 2 shows an example "button off" gesture, with four range-Doppler planes sampled at different time from the data cube.

B. Time sequencial range-Doppler feature extraction and classification

A pattern recognition module is cascaded after the pulse signal processing to classify the time sequence range-Doppler signitures. However, it is ineffective to feed these data cubes directly into some frequently-used classifiers such as support vector machine or k-nearest neighbors. Two main challenges are in front of us. First, we need to deal with uncertain feature size because people may perform the gesture at different velocity. Second, the raw range-Doppler-time data cubes lead to unbearable computation costs for realtime and energy constrained applications.

We aim at finding a way to extract and compress features containing more useful information and less noise as well as using effective models to deal with time sequencial features. Concerning approaches to recognize micro hand gestures, information of the whole range-Doppler plane is not needed. Instead, the detected objects should be the focus. Furthermore, precise ranges and velocities of objects are not necessary, while the moving tendencies and relations of objects are core features for recognition. Based on the above analysis, we adopt a state transition mechanism to summarize each detected object's state (moving direction, discretized velocity, whether merge with or separate from other objects, reflection intensity, etc.) by tracking the objects moving.

When detecting a gesture, the previous signal processing module generates a range-Doppler-time data cube which is the stack of range-Doppler planes among the time dimension. With total N frames of range-Doppler planes in the data cube, we compress the data to a feature sequence $\boldsymbol{S} = \{\boldsymbol{s}_t\}_{t=1}^{N}$, where range-Doppler plane in frame t is extracted as a symbol $\boldsymbol{s}_t$, which can be described as:

$$\boldsymbol{s}_t = \sigma(n_t, \boldsymbol{v}_t, \boldsymbol{r}_t, \boldsymbol{L}_t), \tag{4}$$

where n_t is the number of detected objects, $\boldsymbol{v}_t$ and $\boldsymbol{r}_t$ are $n_t \times 1$ componets vector containing each object's velocities or ranges, and $\boldsymbol{L}_t$ is the tracking result matrix indicating whether each object is merged or split from previous objects or should be labeled as noise. A mapping function σ discretizes and maps these information to the summary symbol $\boldsymbol{s}_t$ which can be stored using just several bits.

We use the HMM to digging out the patterns of the symbolized states and Fig. 3 shows the work flow. As described in equation (5), the HMM used in our module is composed of a hidden state transition model and a multi-nomial emitting

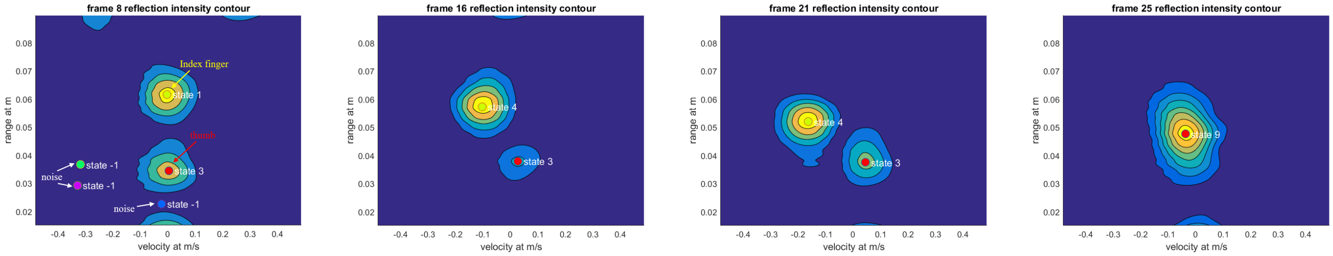

Fig. 2. Range-Doppler feature frames sampled in the "button off" (see Fig. 1) gesture. The first subfigure shows the index finger and thumb separate at a 3cm distance. Noises will be removed using time smoothing to increase robustness as the 3 marked "noise" objects in the first subfigure. The second and third subfigures show the index finger is moving down with an acceleration while the thumb almost keeps static with a tiny velocity moving up. The last subfigure shows two fingers get touched. Noting that an object's trajectory will always be a curve in the range-Doppler plane. The symbolized states are labeled at the center of the detected objects.

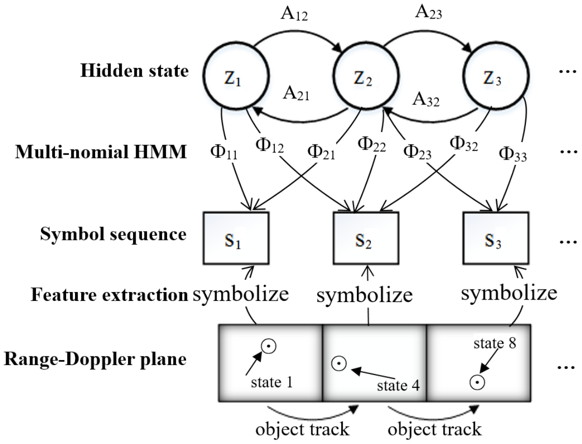

Fig. 3. feature extraction and classification work flow

model, where $\boldsymbol{z}$ is the hidden state, π is the initial hidden state distribution and $\boldsymbol{A}$ is the state transition matrix, as well as ϕ being the multi-nomial possibility distribution. For each pre-defined micro hand gesture, a HMM is trained through Baum-Welch algorithm. When determing the class C of a input gesture, likelihood of each model is calculated through the forward algorithm and compared with each other from a Bayesian posterior perspective as described by equation (6).

$$p(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta}) = p(\mathbf{z}_1|\boldsymbol{\pi})\Big[\prod_{n=2}^{N} p(\mathbf{z}_n|\mathbf{z}_{n-1},\mathbf{A})\Big]\prod_{m=1}^{N} p(\mathbf{x}_m|\mathbf{z}_m,\boldsymbol{\phi}), \tag{5}$$

$$p(C=n|\boldsymbol{X}) = \frac{p(\boldsymbol{X}|\mathrm{C}=n)p(\mathrm{C}=n)}{\sum_n p(\boldsymbol{X}|\mathrm{C}=n)p(\mathrm{C}=n)}. \tag{6}$$

Instantaneous velocity deduced from Doppler effect can also be used to directly detect and segment moving objects, which can further simplifies the classification algorithm in realtime applications.

III. Results

We developed a micro hand gesture recognition system by using ultrasonic active sensing. The time sequential range-Doppler signatures of multiple fingers' actions are determined by pulsed radar signal processing method. We propose a symbolization approch to compress and extract core features from the generated range-Doppler-time data cube and adopt a multi-nomial HMM to classify the sequential features. As shown in Table I, an average classification accuracy of **90.5%** is achieved, for six pre-defined gestures plus one "others" (including null or unconscious actions). We released a realtime micro hand gesture recognition system prototype and demonstrate the control of a music player with our new system. With low power comsumption and computation complexity, our work presents a promissing approch to interact with the smart world touchlessly.

TABLE I
Classification Confusion Matrix and Accuracy

	Others	**Finger**	**BtnOff**	**BtnOn**	**MtnUp**	**MtnDn**	**Screw**
Others	5616	75	94	70	91	82	69
Finger	1	734	0	0	0	0	17
BtnOff	11	7	656	1	36	1	23
BtnOn	23	6	0	651	2	20	32
MtnUp	11	0	49	1	636	21	10
MtnDn	9	0	3	48	5	685	6
Screw	3	48	8	27	4	8	722
accuracy	0.921	0.976	0.893	0.887	0.874	0.902	0.880

References

[1] R. J. Przybyla, H. Tang, A. Guedes, S. E. Shelton, D. A. Horsley, and B. E. Boser, "3d ultrasonic rangefinder on a chip," *IEEE Journal of Solid-State Circuits*, vol. 50, no. 1, pp. 320–334, 2015.

[2] S. Gupta, D. Morris, S. Patel, and D. Tan, "Soundwave: using the doppler effect to sense gestures," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2012, pp. 1911–1914.

[3] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *Proceedings of the 19th annual international conference on Mobile computing & networking*. ACM, 2013, pp. 27–38.

[4] Google ATAP, "Project soli," https://atap.google.com/soli/, 2015, [Online; accessed 29-May-2016].