Biying Fu, Fraunhoferstrasse 5, 64283 Darmstadt, Germany
Telephone: +49 6151 155-214, E-Mail: biying.fu@igd.fraunhofer.de

Master Thesis Summary - Dinesh Vaithyalingam Gangatharan: Activity Recognition On Unmodified Consumer Smartphones Via Active Ultrasonic Sensing

Introduction

Sensor miniaturisation and streaming classification techniques can be used to recognize human behaviours and contexts. This is extremely valuable to realize smart environments, e.g. to support healthy and independent living. The most important parameters to sense include indoor location, gestures, or emergencies like falls. Up to now, activity recognition systems face a number of sensitive drawbacks. For example, camera-based systems induce privacy issues and are costly to deploy. Body-worn systems are inconvenient to wear over long periods of time. Highly visible systems may introduce social stigma and modify the well-known living environment. In this project, we explore the possibility for the use of a new, unobtrusive, physical principle to sense and recognize human activities using off-the-shelf smartphone. A person's smartphone is a cornucopia of information. The huge variety of sensors in today's mobile phones makes these devices a prime target for human activity recognition.

Our novel approach is to develop a novel activity recognizing system using an unmodified smartphone. We profit from integrated microphones and loudspeakers without additional hardware components needed. The advantage of this system is therefore that it can be easily installed on a smartphone and put into action. An android application has already been developed in former master thesis which is able to send a high frequency sound in the near ultrasound range, e.g. 20 kHz. Using the received echo from the microphone, the information caused by movement in midair around the device will be extracted. In this thesis we intend to improve the performance of the existing system with respect to noise cancellation and other modified classification schemes.

Related Work

TODO: exchange some of the related works...

Acoustic Sensing

In the last decade, acoustic sensing has proven to be a very unobtrusive technology for sensing interactions within a human's environment. In contrast to ultrasound, acoustic sensing does not rely on actively transmitted signals. Although speech recognition is a well established research field, acoustic sensing has not yet pervaded other areas on a wide scale. The sensing method is based on the fact that almost every interaction with the environment leaves acoustic traces. This ranges from tapping onto a button, moving along a surface, or interacting with devices such as water taps.

Solid materials are very well-suited for transporting acoustic noise along their surface, making them an ideal candidate for passive acoustic sensing of humans and their environment. Choosing a suitable technology for sound-pickups is not straight-forward, developers face challenges in frequency responses and coupling to materials that strongly depend on the use case. Techniques range from using microphones, microphones combined with stethoscopes, piezo-electric vibration elements, to accelerometers (MEMS) [7].

Interactions on Surfaces

Sensing interactions with a wide variety of materials has been applied for sensing hand movements and touches. Even when using walls as a material for conducting sounds, ranges up to 8 m from touching the surface to the microphone can be achieved [6]. In order to achieve these reasonable high resolutions, a common technique is to use stethoscopes with an attached microphone. When scratching along a surface with a finger nail, sound frequencies up to 3 KHz are generated and conducted within the material [6]. Using this approach, finger and hand movements along walls, tables, screens, and even fabric can be recognized.

Detecting knocks, scratches and taps can be implemented very reliably using acoustic sensing [13]. Attaching four microphones to the corners of a screen, having a length of 1 m, enables to detect the point of a tapping finger with a resolution of 2-4 cm. Measuring the time-of-arrival can act as a reliable method for determining the finger position. When capturing interactions with devices, such as a screen,

Biying Fu, Fraunhoferstrasse 5, 64283 Darmstadt, Germany
Telephone: +49 6151 155-214, E-Mail: biying.fu@igd.fraunhofer.de

a lot of information is lost when touches are just analyzed by means of binary decisions [11, 14]. Using acoustic methods, different force levels of finger taps can be distinguished [14]. When distinguishing two different force levels, the authors reached accuracies ranging from 99 % for two force intervals to 58 % for six force intervals.

The human body is also a well-suited medium for conducting sound. The authors of [7] use an array of piezo-electric vibration elements to an upper arm. By measuring the acoustic waves propagating on the human skin it is possible to recognize different tap locations on the forearm. The resulting signals differ in amplitude and frequency that enable a user-dependent classification. Here, picking up sound with frequencies of up to 100 Hz was utilized to infer up to 10 discrete positions with an accuracy of 81.5 % on the user's skin. Reducing the number to only four locations increases the accuracy to 91.2 %.

Body Movements

In order to detect whole-body movements, one can distinguish between body-mounted and stationary acoustic sensors. In the domain of body-mounted sensors, multiple physical parameters can be captured using sound. For example, using the sound produced by muscle fibers enables Yamakawa et al. to detect different finger movements [24]. In contrast to muscles, conducting sounds through bone also allows for recognizing bending of a human elbow [21]. Chewing, eating and drinking activities can be captured with microphones mounted inside a human ear [1] or a stethoscope placed near the throat [26].

Popescu et al. employ acoustic sensing to detect fall situations [16]. Based on multiple microphones, the height of the sound-source is estimated. If the height is below a threshold of approximately 60 cm, an emergency situation can be derived. However, the authors state that the number of false positives is still too high for real-world deployments. Sound-source estimation can also be used for person localization by microphone arrays [3]. Such systems provide an accuracy of approximately 1 m, while being able to detect multiple sources of sound.

Object Usage

One of the first approaches to sense objects with acoustic sensing is the intelligent ping pong table at MIT, developed in 1999 [9, 23]. The impact of a ball on a table is captured by four microphones and the impact position is calculated. An enhanced game experience is achieved by interactive visualizations that are projected onto the table's surface. This method approaches its limits when multiple objects shall be recognized. SurfaceLink connects multiple devices that are placed on the same surface by gestures [4]. For instance, carrying out a swipe gesture on a table triggers a data transfer from one device to the other.

StickEar [27] aims for equipping objects and parts of the environment with acoustic sensing. The authors developed small stickable wireless sensor nodes. Attached to objects, they can passively sense interactions, e.g. with microwaves or books. Deployed on the wall, they can act very similar as Scratch Input [6], presented previously. The sensed data is then augmented with data from an accelerometer and a touch button. PANDAA picks up ambient sounds and uses them for ranging and localization among multiple devices. Its iterative approach enables to achieve accuracies for localization of approximately 17 cm.

Ultrasound Sensing

Not all interactions leave such nicely analyzable traces as in acoustic sensing. When it comes to in-the-air interactions or detecting stationary states, passively picking up acoustic noise is not sufficient anymore. In contrast to acoustic sensing, actively emitting non-hearable sounds can overcome the problem of recognizing stationary objects. Here, sounds are emitted and the reflected signal is measured by one or multiple microphones. In ultrasound sensing, one has to distinguish between active messaging nodes and passive backscattering techniques. The latter ones analyze backscattered signals to infer the position of body parts or the location of a person [5]. Signals from other devices received and transmitted by nodes allows for implementing an active messaging system.

Body Movements

Detecting finger and hand movements in free air is often achieved by analyzing a backscattered ultrasound signal. Here, worn nodes that use ultrasound as a messaging system can rarely be found. [10] apply ultrasonic waves to unobtrusively recognize one-handed gestures. The authors employ the Doppler effect which introduces frequency shifts in the backscattered signal of a moving human body part. A single transmitter and three receiver microphones are sufficient to determine a 3D movement. Due to the availability of ultrasound capabilities in consumer hardware, ultrasound approaches have also been ported to consumer smartphones and laptops. For example, SoundWave makes it possible to recognize gestures in front of the screen in ordinary laptops [5].

Biying Fu, Fraunhoferstrasse 5, 64283 Darmstadt, Germany
Telephone: +49 6151 155-214, E-Mail: biying.fu@igd.fraunhofer.de

In order to detect whole-body movements, worn nodes as well as passively backscattered signals can be used. Multiple microphones allow for ranging and localization based on the angle or time-difference of arrival. Moreover, exploiting physical effects like the Doppler effect helps to detect changes within the environment. This enables to recognize body parts moving away or towards a sound pickup [5]. Tarzia et al. [22] use a similar technique to determine user presence near a consumer laptop. Using a measurement window of only 10 s results in an accuracy of approximately 96 %, discriminating the two classes of absence and presence. Extending the windows to 25 s almost exceeds perfect accuracy.

Attaching ultrasound nodes to a human body represents a very convenient way to identify and track a person. Techniques which use simple backscattering are not able to identify objects and thus leave a certain amount of ambiguity in their results. Randell et al. present an active ultrasonic positioning system that combines RF and ultrasound for object and person localization [18]. Relying on a secondary modality with faster wave propagation enables to reliably calculate the time-of-flight of the ultrasound signal [8, 20]. This technique has found wide applications in people localization and the corresponding tags became very small or integrated into devices [2].

Object Usage

Transmitting dedicated information about the object and its manipulations requires an active messaging system. One of the first approaches combine RF with ultrasound to achieve a reference for time-of-flight measurements [17]. BeepBeep [15] solely relies on ultrasound generated by smartphones, that then supports ranging among objects. When more than three objects are involved, it is possible to determine a relative position based on the information about multiple devices. [12] follows a very similar approach: It enables users to wear a virtual uniform which emits sounds at specified frequencies. The system is intended to communicate the social state to other people. Sensing device location in a car was investigated by [25]: By classifying the mobile phone's position it is possible to differentiate between the driver using the phone and a passenger. Reynolds et al. [19] introduce an ultrasound position sensing system for tangible objects above LCD-screens. The authors present interactive 'pucks' that communicate by emitting and receiving ultrasound and reconstruct their position.

Scientific Contribution

The scientific challenge is to try to improveclear the back-scattered signal from the vicinity of the smartphone and hence to improve the performance of the existing system with respect to activity recognition. Based on the cleared echo signal, signal processing and machine learning techniques can be investigated to infer a user's activities. The overall system will be evaluated for its recognition rate with the use-case of recognition certain predefined users activities. The project will be conducted over a six months period within a master thesis.

Detailed Work Plan

... TODO:

The following points are expected to be completed in the scope of this work and thesis:

- Checkout the Noise Suppressor function in android sdk
- Checkout the performance of using two microphones instead of one to get more 2D information
- Find other features to increase the recognition rate
- Explore other classification schemes

References

[1] Oliver Amft, Holger Junker, and Gerhard Tröster. Detection of eating and drinking arm gestures using inertial body-worn sensors. In *Proceedings of the 9th IEEE International Symposium on Wearable Computers (ISWC '05)*, pages 160–163. IEEE, 2005.

[2] Gaetano Borriello, Alan Liu, Tony Offer, Christopher Palistrant, and Richard Sharp. Walrus: Wireless acoustic location with room-level resolution using ultrasound. In *Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services*, MobiSys '05, pages 191–203, New York, NY, USA, 2005. ACM.

[3] Ionut Constandache, Sharad Agarwal, Ivan Tashev, and Romit Roy Choudhury. Daredevil: Indoor location using sound. *SIGMOBILE Mob. Comput. Commun. Rev.*, 18(2):9–19, June 2014.

[4] Mayank Goel, Brendan Lee, Md. Tanvir Islam Aumi, Shwetak Patel, Gaetano Borriello, Stacie Hibino, and Bo Begole. Surfacelink: Using inertial and acoustic sensing to enable multi-device interaction on a surface. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, pages 1387–1396, New York, NY, USA, 2014. ACM.

[5] Sidhant Gupta, Daniel Morris, Shwetak Patel, and Desney Tan. Soundwave: Using the doppler effect to sense gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1911–1914, New York, NY, USA, 2012. ACM.

[6] Chris Harrison and Scott E. Hudson. Scratch input: Creating large, inexpensive, unpowered and mobile finger input surfaces. In *Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology*, UIST '08, pages 205–208, New York, NY, USA, 2008. ACM.

[7] Chris Harrison, Desney Tan, and Dan Morris. Skinput: Appropriating the body as an input surface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 453–462, New York, NY, USA, 2010. ACM.

[8] S. Helal, B. Winkler, Choonhwa Lee, Y. Kaddoura, L. Ran, C. Giraldo, S. Kuchibhotla, and W. Mann. Enabling location-aware pervasive computing applications for the elderly. In *Pervasive Computing and Communications, 2003. (PerCom 2003). Proceedings of the First IEEE International Conference on*, pages 531–536, March 2003.

[9] Hiroshi Ishii, Craig Wisneski, Julian Orbanes, Ben Chun, and Joe Paradiso. Pingpongplus: Design of an athletic-tangible interface for computer-supported cooperative play. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '99, pages 394–401, New York, NY, USA, 1999. ACM.

[10] K. Kalgaonkar and B. Raj. One-handed gesture recognition using ultrasonic doppler sonar. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pages 1889–1892, April 2009.

[11] Pedro Lopes, Ricardo Jota, and Joaquim A. Jorge. Augmenting touch interaction through acoustic sensing. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '11, pages 53–56, New York, NY, USA, 2011. ACM.

[12] Aleksandar Matic, Alban Maxhuni, Venet Osmani, and Oscar Mayora. Virtual uniforms: Using sound frequencies for grouping individuals. In *Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication*, UbiComp '13 Adjunct, pages 159–162, New York, NY, USA, 2013. ACM.

[13] J.A Paradiso, Che King Leo, N. Checka, and Kaijen Hsiao. Passive acoustic sensing for tracking knocks atop large interactive displays. In *Sensors, 2002. Proceedings of IEEE*, volume 1, pages 521–527 vol.1, 2002.

[14] Esben Warming Pedersen and Kasper Hornbæk. Expressive touch: Studying tapping force on tabletops. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 421–430, New York, NY, USA, 2014. ACM.

[15] Chunyi Peng, Guobin Shen, and Yongguang Zhang. Beepbeep: A high-accuracy acoustic-based system for ranging and localization using cots devices. *ACM Trans. Embed. Comput. Syst.*, 11(1):4:1–4:29, April 2012.

[16] M. Popescu, Yun Li, M. Skubic, and M. Rantz. An acoustic fall detector system that uses sound height information to reduce the false alarm rate. In *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pages 4628–4631, Aug 2008.

[17] Nissanka B. Priyantha, Anit Chakraborty, and Hari Balakrishnan. The cricket location-support system. In *Proceedings of the 6th Annual International Conference on Mobile Computing and Networking*, MobiCom '00, pages 32–43, New York, NY, USA, 2000. ACM.

[18] Cliff Randell and Henk L. Muller. Low cost indoor positioning system. In *Proceedings of the 3rd International Conference on Ubiquitous Computing*, UbiComp '01, pages 42–48, London, UK, UK, 2001. Springer-Verlag.

[19] Matthew Reynolds, Alexandra Mazalek, and Glorianna Davenport. An acoustic position sensing system for large scale interactive displays. In *Sensors, 2007 IEEE*, pages 1193–1196, Oct 2007.

[20] Andreas Savvides, Chih-Chieh Han, and Mani B. Strivastava. Dynamic fine-grained localization in ad-hoc networks of sensors. In *Proceedings of the 7th Annual International Conference on Mobile Computing and Networking*, MobiCom '01, pages 166–179, New York, NY, USA, 2001. ACM.

[21] Kentaro Takemura, Akihiro Ito, Jun Takamatsu, and Tsukasa Ogasawara. Active bone-conducted sound sensing for wearable interfaces. In *Proceedings of the 24th Annual ACM Symposium Adjunct on User Interface Software and Technology*, UIST '11 Adjunct, pages 53–54, New York, NY, USA, 2011. ACM.

[22] Stephen P. Tarzia, Robert P. Dick, Peter A. Dinda, and Gokhan Memik. Sonar-based measurement of user presence and attention. In *Proceedings of the 11th International Conference on Ubiquitous Computing*, Ubicomp '09, pages 89–92, New York, NY, USA, 2009. ACM.

[23] Xiao Xiao, Michael S. Bernstein, Lining Yao, David Lakatos, Lauren Gust, Kojo Acquah, and Hiroshi Ishii. Pingpong++: Community customization in games and entertainment. In *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology*, ACE '11, pages 24:1–24:6, New York, NY, USA, 2011. ACM.

[24] Shumpei Yamakawa and Takuya Nojima. A proposal for a mmg-based hand gesture recognition method. In *Adjunct Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*, UIST Adjunct Proceedings '12, pages 89–90, New York, NY, USA, 2012. ACM.

Biying Fu, Fraunhoferstrasse 5, 64283 Darmstadt, Germany
Telephone: +49 6151 155-214, E-Mail: biying.fu@igd.fraunhofer.de

[25] Jie Yang, Simon Sidhom, Gayathri Chandrasekaran, Tam Vu, Hongbo Liu, Nicolae Cecan, Yingying Chen, Marco Gruteser, and Richard P. Martin. Detecting driver phone use leveraging car speakers. In *Proceedings of the 17th Annual International Conference on Mobile Computing and Networking*, MobiCom '11, pages 97–108, New York, NY, USA, 2011. ACM.

[26] Koji Yatani and Khai N. Truong. Bodyscope: A wearable acoustic sensor for activity recognition. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, UbiComp '12, pages 341–350, New York, NY, USA, 2012. ACM.

[27] Kian Peen Yeo, Suranga Nanayakkara, and Shanaka Ransiri. Stickear: Making everyday objects respond to sound. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, pages 221–226, New York, NY, USA, 2013. ACM.