RF-Interfaces For User Input & Enhanced Spatial Perception

BEN PEARMAN, ANAND KUMAR, WESLEY WILLETT, ADITYA NITALLA, University of Calgary, Canada

1 INTRODUCTION

In this work we present a series of design concepts which we refer to as RF-Interfaces (Radio Frequency Interfaces). These concepts explore several interaction opportunities which may become possible thanks innovations in the field of RF-Vision. While much of the existing work in RF-Vision focuses on technical evaluations and model benchmarks, we attempt to take a user-centered approach by first considering the design goals of such a system and the future uses it may afford. We explain how RF-Interfaces can overcome the limitations of visual tracking systems, such as occlusion and lighting variability, and enable spatial rendering through obstructive environments. We then provide a set of design concepts for four future interactions that we're working towards. We will conclude with a summary of our technical implementation, its performance thus far, and our plans for future work.

2 RF-VISION

Visible light occupies the 400 - 700 (nm) section of the electromagnetic (EM) spectrum. This is a tiny fraction of the total range of EM signals which can be observed by manufactured sensing modalities. We're used to using these signals for things such as communications technology (Wifi, radio) and medical imaging (X-Ray, CT), however recent advances in deep learning have shown the potential of other sections of the EM spectrum for things like device input and sensory augmentation.

RF-Vision can be conceptualized as a subset of Computer Vision, with some additional roots in signal processing and radar applications. RF-Vision leverages radio frequencies to extract information about the physical world, translating this data into formats usable by software applications or human operators. One notable example of RF-Vision is Project Soli from Google, which enabled gesture recognition on the Pixel 4 smartphone. Soli detected hand gestures within a hemispherical range of 1–2 feet from the screen, showcasing the potential of RF-based systems for close-range interaction [9]. Since then, many researchers have continued to push the boundaries of RF-Vision in various application spaces such as seeing through walls [1, 6, 7], occlusion robust gesture & hand detection [5, 8, 13, 14], environment and infrastructure monitoring [3, 10, 11], and human presence / biosignal detection [4, 12] .

Traditional visual tracking methods for hand gesture input in devices like AR and VR headsets are often hindered by environmental factors such as poor lighting conditions and physical occlusions (a user's hand behind a desk). RF-Vision circumvents these issues by providing continuous gesture recognition regardless of visibility. Additionally, RF sensing's ability to penetrate physical barriers opens possibilities for augmented human perception, such as detecting hidden obstacles, locating individuals in compromised scenarios, and navigating hazardous environments safely.

2.1 RF-Interfaces - Design Futures

We present 4 design concepts that imagine future applications of RF-Interfaces. Figures 1 & 2 demonstrate *ubiquitous user input*, while figure 3 & 4 demonstrate *enhanced spatial perception.* In all four designs, we imagine an RF sensing device embedded in a head mounted display (HMD). Also of note is that designs 1 & 2 would be utilizing an isotropic

Author's address: Ben Pearman, Anand Kumar, Wesley Willett, Aditya Nitalla, University of Calgary, Department of Computer Science, Calgary, Canada.

antenna, while designs 3 and 4 use a directional one. Isotropic antennas create a spherical emission pattern of the RF signal, which enables multi-directional sensing for gesture recognition. Directional antennas are used when the user only cares about a specific known area, like searching avalanche debris (shown in figure 4).

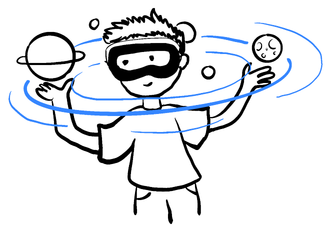


Fig. 1. A user manipulates 3D objects in XR, even when their left hand is not visible to the HMD's cameras.

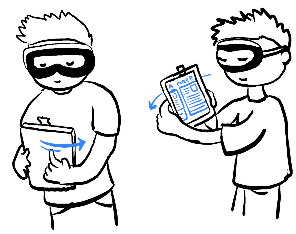


Fig. 2. Occluded / busy hand gesture recognition. User can swipe the back of a physical clipboard to move AR overlays on it.

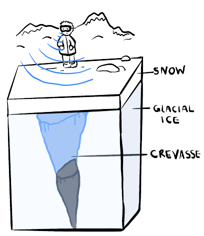


Fig. 3. Ground penetrating radar on a user's HMD highlights snow bridges on a glacier.

Fig. 4. Full-body pose estimation reveals a skier's orientation after being buried in an avalanche.

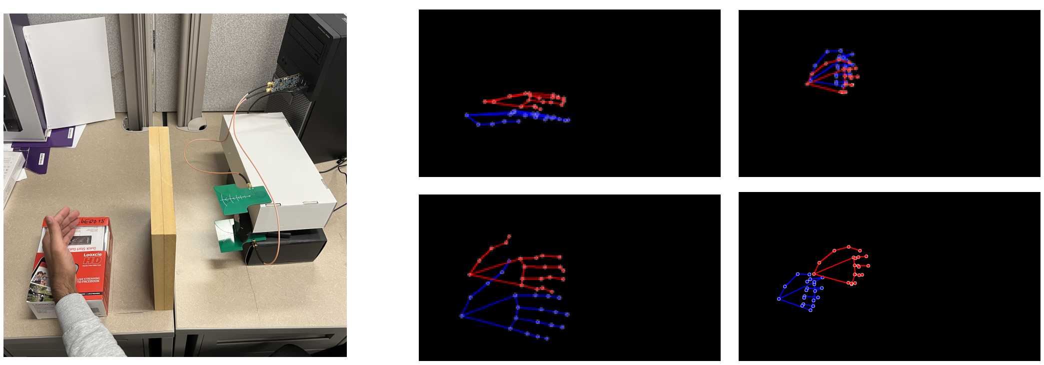
Figures 1 & 2 illustrate how RF-Vision addresses a common limitation of AR/VR HMDs: the reliance of cameras on a direct line of sight for gesture recognition. In these designs, RF transceivers complement rather than replace cameras, as camera-based gesture recognition currently exceeds state-of-the-art RF-Vision accuracy (our system achieves approximately 85% accuracy under clear line-of-sight conditions). This hybrid approach leverages the strengths of both camera and RF technologies, resulting in a robust and versatile input method unaffected by darkness or visual obstructions.

Figures 3 & 4 demonstrate RF signals' inherent ability to penetrate materials, enabling advanced spatial awareness applications. A typical use case would be any situation involving objective hazards which are not immediately visible, as well as search and rescue. Mountaineers are frequently exposed to this type of hazard, for example, when traversing a glacier, the mountaineer relies on snow bridges to cross crevasses in the glacial ice, however they cannot usually distinguish where these bridges are. The current method used by most mountaineers is to use a telescopic probe to poke into the snow to feel if there's solid ice under it, this is a slow process which is prone to failure.

Figure 4 illustrates an avalanche burial scenario. The current industry standard for recovering buried victims is to use an avalanche transceiver, which can locate the victim within about a 2 meter radius. The rescuer then has to guess the victims orientation. This is problematic since the victim has a finite amount of oxygen under the snow, and a full extraction can take several hours [2]. In ideal situations, the rescuers get lucky and expose the victims head first. Our design circumvents this problem by implementing full body human pose detection (HPE) using RF signals, allowing the rescuer to make the best judgment about where to start digging. This scenario can be extrapolated to many other cases involving search and rescue operations such as collapsed buildings, fires, and any situation where knowing a person's position would be valuable while they aren't within a clear line of sight.

3 PROTOTYPE

As a stepping stone towards the design futures described above, we have created a prototype using a consumer grade software defined radio (SDR) called the LimeSDR Mini 2.0, which features a full-duplex transceiver. This minimal complexity hardware platform has allowed us to begin demonstrating the feasibility of our designs. The SDR has been equipped with 2 log-periodic antennas (1 Rx, 1 Tx) to create a simple monostatic radar system. So far, we have used this system to demonstrate 2 key capabilities in our design futures: occlusion robust *gesture recognition*, and *spatial localization.* We achieve this by transceiving a frequency modulated continuous wave signal. The raw signal is composed of complex numbers which we transform into phase and amplitude components $\text{Phase} = \arctan 2(Q, I), \quad \text{Amplitude} = I^2 + Q^2$, which are then turned into input tensors for two 1-Dimensional Convolutional Neural Networks (1D-CNN). These networks have shown strong classification accuracy for gesture input (87% across three levels of occlusion), as well as accurate spatial localization.



(a) Prototype setup.

(b) Hand pose estimation from RF signals (red) with ground truth (blue).

Fig. 5. Experimental setup and model predictions. The user's hand is placed behind a 4.5cm thick wood block, the SDR is plugged into the computer in the top right of fig (a), and the antennas are the green card-like objects on the right side of the block.

4 FUTURE WORK

Future directions include expanding the dataset and RF Bandwidth of the system. An Ultrawide-Band radar system trained on a diverse dataset would create a more generalizable and robust model beyond what has been demonstrated here. The future system will also be implemented as an input method on an HMD for real-time performance.

REFERENCES

[1] Fadel Adib, Chen-Yu Hsu, Hongzi Mao, Dina Katabi, and Frédo Durand. 2015. Capturing the human figure through a wall. *ACM Transactions on Graphics* 34, 6 (2015), 219:1–219:13. https://doi.org/10.1145/2816795.2818072

[2] Tim Banfield. 2018. Skier Saved from Deep Under Rockies Avalanche. *Gripped Magazine* (April 2018). https://gripped.com/news/skier-saved-from-deep-under-rockies-avalanche/

[3] V. Cai. 2020. Designing a Narrowband Radar using GNU Radio and Software Defined Radio for Tomography and Indoor Sensing. *Proceedings Of The GNU Radio Conference* 5, 1 (2020). https://pubs.gnuradio.org/index.php/grcon/article/view/67

[4] G. Vinci et al. 2013. Six-Port Radar Sensor for Remote Respiration Rate and Heartbeat Vital-Sign Monitoring. *IEEE Transactions on Microwave Theory and Techniques* 61, 5 (2013), 2093–2100. https://doi.org/10.1109/TMTT.2013.2247055

[5] H. Kong et al. 2024. mmHand: 3D Hand Pose Estimation Leveraging mmWave Signals. In *2024 IEEE 44th International Conference on Distributed Computing Systems (ICDCS)*. https://doi.org/10.1109/ICDCS60910.2024.00102

[6] J. Geng, D. Huang, and F. De la Torre. 2022. DensePose from WiFi. *arXiv* (2022). https://arxiv.org/abs/2301.00250

[7] R. A. Güler, N. Neverova, and I. Kokkinos. 2018. DensePose: Dense human pose estimation in the wild. *arXiv* (2018). https://arxiv.org/abs/1802.00434

[8] Sijie Ji, Xuanye Zhang, Yuanqing Zheng, and Mo Li. 2024. Construct 3D Hand Skeleton with Commercial WiFi. In *Proceedings of the 21st ACM Conference on Embedded Networked Sensor Systems (SenSys '23)*. Association for Computing Machinery, New York, NY, USA, 322–334. https://doi.org/10.1145/3625687.3625812

[9] Jaime Lien, Nicholas Gillian, M. Emre Karagozler, Patrick Amihood, Carsten Schwesig, Erik Olson, Hakim Raja, and Ivan Poupyrev. 2016. Soli: ubiquitous gesture sensing with millimeter wave radar. *ACM Transactions on Graphics* 35, 4 (2016), 142:1–142:19. https://doi.org/10.1145/2897824.2925953

[10] A. Monk. 2020. *Exploration into the Use of a Software Defined Radio as a Low-Cost Radar Front-End*. Ph. D. Dissertation. Brigham Young University. http://hdl.lib.byu.edu/1877/etd11486

[11] I. Rodriguez, H. C. Nguyen, N. T. K. Jorgensen, T. B. Sorensen, and P. Mogensen. 2014. Radio Propagation into Modern Buildings: Attenuation Measurements in the Range from 800 MHz to 18 GHz. In *2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall)*. https://doi.org/10.1109/VTCFall.2014.6966147

[12] J. Sadhwani and M. S. Manikandan. 2021. Non-Collaborative Human Presence Detection Using Channel State Information of Wi-Fi Signal and Long-Short Term Memory Neural Network. In *2021 13th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*. Pitesti, Romania, 1–6. https://doi.org/10.1109/ECAI52376.2021.9515148

[13] J. Xu, B. Yang, Z. Chen, P. Lv, S. Zhang, and J. Luo. 2024. Ske-Fi: Estimating Hand Poses via RF Vision Under Low Contrast and Occlusion. *IEEE Internet of Things Journal* 11, 4 (2024), 6412–6425. https://doi.org/10.1109/JIOT.2023.3312316

[14] Shujie Zhang, Tianyue Zheng, Zhe Chen, Jingzhi Hu, Abdelwahed Khamis, Jiajun Liu, and Jun Luo. 2023. OCHID-Fi: Occlusion-Robust Hand Pose Estimation in 3D via RF-Vision. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 15112–15121.