This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LSENS.2018.2810093, IEEE Sensors Letters

Gesture Recognition Using mm-Wave Sensor for Human-Car Interface

Karly Smith[1], Clément Csech[2], David Murdoch[3], and George Shaker [3,4*]

[1] Biomedical Engineering, Department of Systems Design Engineering, University of Waterloo, Waterloo, Ontario, Canada
[2] Department of Biomechanics and Bioengineering, Universite de Technologie de Compiegne, Compiegne, France
[3] Department of Electrical and Computer Engineering, University of Waterloo, Ontario, Canada
[4] Department of Mechanical and Mechatronics Engineering, University of Waterloo, Waterloo, Ontario, Canada
* Senior Member, IEEE

Received xx, revised xx, accepted xx, published xx, current version xx. (Dates will be inserted by IEEE; "published" is the date the accepted preprint is posted on IEEE Xplore®; "current version" is the date the typeset version is posted on Xplore®).

Abstract—This paper details the development of a gesture recognition technique using a mm-wave radar sensor for in-car infotainment control. Gesture recognition is becoming a more prominent form of human-computer interaction, and can be used in the automotive industry to provide a safe and intuitive control interface that will limit driver distraction. We use a 60 GHz mm-wave radar sensor to detect precise features of fine motion. Specific gesture features are extracted and used to build a machine learning engine that can perform real-time gesture recognition. This paper discusses the user requirements and in-car environmental constraints that influenced design decisions. Accuracy results of the technique are presented, and recommendations for further research and improvements are made.

Index Terms—60 GHz mm-wave radar, gesture sensing, human-car interface, random forest classifier, machine learning

I. INTRODUCTION

In the automotive industry, vehicular infotainment systems have grown in popularity and complexity over the past several years. Mainstream car manufacturers now offer up to 700 infotainment and environmental controls for the driver and passengers to manipulate [1]. However, increased functionality within the vehicle has increased potential causes for driver distraction. The main causes of driver distraction are categorized in [2] as visual, cognitive, manual, and auditory. Studies have shown that visual and manual distractions when combined have the most impact on driving performance [2]. This paper presents a gesture detection system using a mm-wave radar sensor for intuitive human-vehicular interaction (HVI). Many different gesture sensing and processing techniques have been developed in recent years. Previous gesture detection systems have used camera based sensors (IR, colour, etc.), depth based sensors, and wearable sensors such as gloves embedded with 3D tracking technology [2-11]. However, these systems all have significant drawbacks that affect their usability. Camera based sensors are susceptible to changes in light, colour, background, and have high computational costs due to extensive image processing [3,4]. Depth based sensors are very good at detecting changes in position, however they cannot detect orientation or specific hand shapes [5]. Wearable technology may interfere with other tasks the user does in daily life, and limits system input to whoever is wearing the input device.

Alternatively, we believe that radar sensors present a viable system solution. Radars are not affected by variable lighting changes inside a car, and are able to detect specific hand and finger orientations with precision. The radar system described in this paper provides real-time visionless infotainment control to driver and passengers without wearable components, decreasing risk of driver distraction and allowing multiple user input. Previous work has been done on in-car gesture sensing that combines short-range radar, time of flight depth sensors, and colour cameras for gesture detection [6]. That system used a FMCW monopulse 25GHz radar in conjunction with camera based data for detection and a convolutional neural network for near real-time recognition [6]. In comparison, the system presented in this paper uses 60GHz radar for finer spatial resolution, and a random forest classifier algorithm for real time recognition.

II. SYSTEM DESIGN

Using a wireless radar sensor for detection and recognition of gestures offers several advantages over other systems currently in use. Automobile manufacturers currently offer touchscreens, voice control, Bluetooth phone connection, and other methods of infotainment control. Interfaces that require tactile manual input such as a touchscreen also require small amounts of visual attention to navigate, taking the driver's eyes off the road. Voice control does not require manual or visual control, however if the in-car environment is noisy with music or conversation voice control is not a viable option. As highlighted earlier, there are several other systems for gesture and posture detection in development that use IR and depth cameras for sensing [2-11], however cameras are affected by light conditions and obstacles in the field of view. As well, issues of privacy and user compliance arise when cameras are in use. Radar is advantageous because it is not affected by light or sound in the environment, it can be embedded in devices, has very precise

Corresponding author: K. A. Smith (k62smith@edu.uwaterloo.ca). IEEE Sensors Letters discourages courtesy authorship; please use the Acknowledgment section to thank your colleagues for routine contributions.

Digital Object Identifier: 10.1109/LSEN.XXXX.XXXXXXX (inserted by IEEE).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LSENS.2018.2810093, IEEE Sensors Letters

resolution, offers real time recognition, and does not require recording of an image of the user [7].

A. Gesture Detection and Recognition

In this work, we utilized a 60GHz frequency modulated continuous wave (FMCW) mm-wavelength radar sensor. The sensor hardware consists of an 8mmx11mm radar chip with two Tx and four Rx antennas shown in Fig. 1.

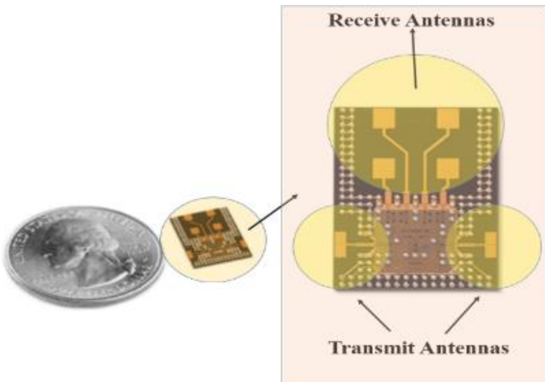
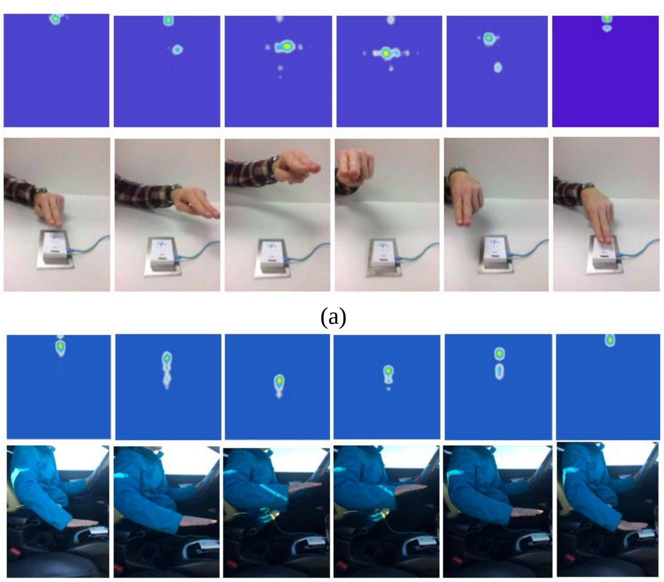


Fig.1. Radar chip photo showing size comparable to a nickel

The radar board was fitted inside an after-market car. Machine learning was used to record the radar signature of each set of hand gestures (as shown in Fig. 2 a and b), train a model using a random forest classifier algorithm, and then perform recognition using that model. The random forest classifier was used because of the higher success rate earlier as highlighted in [15]. Features of the received signal were processed and made easily accessible for manipulation in C language. The features used in this project were range, acceleration, energy total, energy moving, velocity, velocity dispersion, spatial dispersion, energy strongest component, and movement index [12].



(a)

(b)

Fig. 2. Timestamped images of gesture progression with corresponding range doppler signature progressions

Each gesture was assigned a classifier number, then twenty of each were recorded and timestamped for data collection. Twenty samples of the background were also recorded and assigned a classifier, so the system would accurately recognize the absence of a gesture. The collected data was then used to create a random forest classifier, allowing real time recognition of future gestures. The parameters of the random forest classifier were set as follows:

1. Forest Size = 10
2. Forest Max Depth = 10
3. Forest Min Samples = 50
4. Classifier Min Count = 30
5. Classifier Buffer Size = 50

Increasing the number of trees within the classifier increased robustness, and altering the ratio of classifier min count to classifier buffer size changed the precision of recognition. A ratio of 3:5 between classifier min count and classifier buffer size ensures that for classification to occur 30 out of the 50 minimum samples required must be classified in the same category. Forest size and forest depth were both set to 10 to ensure adequate forest size and average tree depth for classification, while minimizing computational cost on the system.

B. Sensor Placement

Specific environmental and user constraints were considered when designing the system and gesture sets to ensure usability and robustness. The interior of a car is spatially complex which created detection challenges for the sensor. If placed too close to detectable objects, such as a gear shift, it often recognized false positives, or did not recognize gestures within the field. Placing the sensor where majority of the radar beam spread into free space and recording a robust background case mitigated detection of objects and identified those objects that did intrude into the field as non-targets. This spatial constraint, combined with the spatial constraints of the user position, lead to placement of the sensor on the center console (for use by front seat passenger and the driver) and between the backs of the front seats (for use by the back-seat passengers). The beam of the field spread upward into free space, where little movement from passengers naturally occurs. With these two placements (as shown in Fig. 3), the radar is within a comfortable arm's length reach for all passengers.

(a) (b)

Fig. 3. Shows the two placements of the radar sensor, (a) for the driver and front-seat passenger, (b) for the backseat passengers

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LSENS.2018.2810093, IEEE Sensors Letters

C. Connections to In-Car Infotainment System

When a gesture was detected, its classifier number was added to the body of a POST request. This POST request was then sent to an Android phone (paired to a car in older car models) or Android auto system (available in newer car models), which, acting as a server, parsed the POST request for the classifier number. The Android platform would then parse the request to receive the classifier number. Each classifier number was associated with some action, which the system could then undertake, using either Android intents for phone actions, or the Spotify API for media actions. This allowed for both the streaming of audio over the car speakers and the use of the physical car infotainment system to display relevant information. A visual representation of the entire system can be seen in Fig. 4.

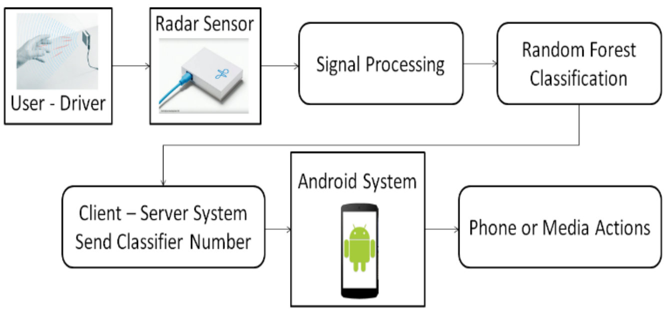


Fig. 4. Shows a flowchart of the system from gesture recognition to infotainment output

D. Gesture Set Design

The needs of the driver and characteristics of the system were the main considerations when designing the gesture sets. The 60 GHz sensor in use was designed to detect fine motion, with spatial resolution as fine as 0.3mm, which is quite an improvement from the 3.75cm range resolution of the 25GHz radar system presented in [13]; this enables the 60GHz radar to have greater precision and accuracy, smaller footprint allowing for spatial and temporal differences to be used to make gestures distinguishable from one another. Practical system implementation dictated the use of gestures with tolerance of large margins of error to create less driver distraction. By creating larger spatial zones for each gesture, and having gestures performed at varying speeds the driver would have more room for non-exact gestures and thus would need to devote less visual resources to the motion. Designing more robust gestures also allows multiple users to operate the system, as it will be more tolerant to the natural spatial and temporal variations of each user and the various gesture motions within different vehicle designs.

Two demonstration videos [14] of the system at work were recorded for the proof of concept system, the details of which are shown in Table 1. Only the phone function and playlist music function of the car infotainment system were included in the demonstration videos, however the system can control much more. As much as possible intuitive gestures that related to their corresponding functions were used to limit cognitive load of the user. Using universal signs such as a telephone gesture and numbers 1, 2, and 3 will make gestures easier to recall and natural to perform. Vehicular infotainment systems often use menu functions to navigate and organize all the options of control, which allows reuse of several gestures. The same gestures can be used in all menus for back, forward, select, etc., which will provide extensive functionality to the user while limiting the number of gestures they need to remember. Only main gestures that will be used to select which menu to access (i.e. contact list, map, music, etc.) will not be reusable. Reuse of gestures also increases the accuracy of the system, as fewer classification options has been shown to increase accuracy of recognition when using a random forest classifier [15].

Table 1. Summary of gesture demonstration characteristics

Demo	Gesture	Function
Phone (Driver)	Wiggle phone sign	Call/Hang up
Music (Backseat Passenger)	Wiggle fingers	Pause/Play
	Hold out one finger	Select Playlist 1
	Wiggle 2 fingers	Select Playlist 2
	Wave 3 fingers	Select Playlist 3

III. TESTING THE SYSTEM

For each demonstration, gesture sets were recorded by one individual to film the demonstration. It took significantly less time to record with one individual, however gesture sets created by one individual may not work for other users. Recording with multiple individuals will capture natural temporal and spatial variations introduced with each new user to create a more robust system. Another two gesture sets comprised of three gestures each were recorded by multiple individuals to test performance with more human variation. Five individuals were used to record each gesture set, each individual recording 20 of each gesture. The gesture set was later tested by the five recorders as well as three individuals who did not record, the results of which are shown in Table 2. To test, everyone performed each gesture 30 times and the accuracy percentage was recorded. Before testing the system, all participants had opportunity to practice and find the correct range and speed of each gesture. Instructions were provided during this practice to ensure each gesture was performed correctly; these results reflect user accuracy with a strong understanding of each movement. The participants' ability to learn the gestures alone was not evaluated and should be explored further to evaluate learnability of each gesture when only provided with written or video instruction. The system performed above 90% accuracy for all gestures on average. The gestures with the lowest accuracy were gestures 1 and 4 which may be attributed to human error; both had several incorrectly timed recordings. Gesture set 2 showed a slight decrease in accuracy going from the recorders to the users, which may be attributed to the spatial

Table 2. Summary of recognition accuracy results for two gesture sets

Number	Set 1	Recorder 1	Recorder 2	Recorder 3	Recorder 4	Recorder 5	User 1	User 2	User 3	Average
1	Low wiggle	100%	100%	87%	90%	93%	93%	80%	83%	91%
2	Turn over	100%	100%	100%	100%	100%	93%	100%	100%	99%
3	High grab	93%	100%	90%	100%	87%	100%	90%	100%	95%
-	Set 2	-	-	-	-	-	-	-	-	-
4	Swipe	100%	93%	87%	97%	93%	87%	87%	83%	91%
5	Large circle	100%	100%	100%	100%	97%	93%	93%	90%	97%
6	Small circle	100%	100%	93%	100%	100%	93%	93%	87%	96%

design of the gesture set. Gesture set 1 had more distinct spatial zones for each gesture, whereas the spatial zones for gesture set 2 had some overlap; those who recorded the gesture set had more familiarity with the spatial zones than the users who did not record. To improve the accuracy of the system more participants should be used to record, with each participant recording more gestures. It also may be valuable to distribute the gestures evenly when testing and recording rather than complete all samples for a gesture at once; it was observed that as participants fatigued variations were introduced that were not present when the gesture was done naturally only once.

IV. CONCLUSION

We have presented a gesture detection system using mm-wave radar for vehicular infotainment control. The gestures were designed to be distinguishable by the radar, be intuitive and memorable for the user, and fit the constraints of the environment. Specific decisions were made to maximize ease of use by drivers; further testing should be done to validate these usability design decisions. Demonstrations of the system in use were filmed and presented to showcase the use of intuitive gestures [14], ease of use by both driver and passengers, and the use of large spatial zones for more robust recognition. The accuracy of the system was tested with multiple users; it was found that involving more participants when recording a gesture set increased accuracy and robustness. In the future, studies are needed to define and optimize the required user input for the system training stages.

ACKNOWLEDGMENT

This work was supported by NSERC. The radar transceiver used in this work was provided by Google's Advanced Technology and Projects group (ATAP) through the Project Soli Alpha DevKit Early Access Grant.

REFERENCES

[1] Pickering C, Burnham K, and Richardson M (2007), "A research study of hand gesture recognition technologies and applications for human vehicle interaction," in 2007 IET 3rd Institution of Engineering and Technology Conference on Automotive Electronics, Warwick, UK, UK.

[2] Young K, Regan M (2007), "Driver distraction: a review of the literature," Distracted Driving, Sydney, Australia: Australasian College of Road Safety, 2007, pp. 379-405.

[3] Wachs J P, Kolsch M, Stern H, and Edan Y (2011), "Vision based hand gesture applications," in Communications of the ACM, vol. 54, no. 2, pp. 60-71, doi: http://dx.doi.org/10.1145/1897816.1897838

[4] Breuer P, Eckes C, and Müller S (2007), "Hand gesture recognition with a novel IR time-of-flight range camera–a pilot study", in MIRAGE 2007, LNCS 4418, pp. 247–260, doi: http://dx.doi.org/10.1007/978-3-540-71457-6_23

[5] Keskin C, Kırac F, Kara Y E, and Akarun L (2011), "Real time hand pose estimation using depth sensors", in IEEE International Conference on Computer Vision Workshops, pp 1228-1234, doi: http://dx.doi.org/10.1109/ICCVW.2011.6130391

[6] Molchanov P, Gupta S, Kim K, and Pulli K (2015), "Multi-sensor system for driver's hand-gesture recognition," IEEE 11th International Conference and Workshops on Automatic Face and Gesture Recognition, Ljubljana, Slovenia, pp. 1-8, doi: http://dx.doi.org/10.1109/FG.2015.7163132

[7] Riener A, Rossbory M, and Ferscha M (2011), "Natural DVI based on intuitive hand gestures," INTERACT Workshop User Experience in Cars, pp. 62-66.

[8] Molchanov P, Gupta S, Kim K, and Kautz J (2015), "Hand gesture recognition with 3D convolutional networks," IEEE Conference on Computer Vision and pattern Recognition Workshops, Boston, MA, USA, pp. 1-7, doi: http://dx.doi.org/10.1109/CVPRW.2015.7301342

[9] Ohn-Bar E, and Manubhai-Trivedi M (2014), "Hand gesture recognition in real time for automotive interfaces: a multimodal vision based approach and evaluations," IEEE Transactions on Intelligent Transportation Systems, vol. 15, no. 6, doi: http://dx.doi.org/10.1109/TITS.2014.2337331

[10] Jacob Y, Manitsaris F, Lele G, Pradere L (2015), "Hand gesture recognition for driver vehicle interaction," IEEE Computer Society Workshop on Observing and Understanding Hands in Action of 28th IEEE conference on Computer Vision and Pattern Recognition, Boston, United States.

[11] Reissner U (2007), "Gestures and speech in cars," Chair for Computer Aided Medical Procedures and Augmented Reality, Department of Informatics, Technische Universitat Munchen, unpublished.

[12] Gillian N. (2017, Oct.). Gesture Recognition Toolkit, ver. 1.0. [Online]. Available: http://www.nickgillian.com/grt/.

[13] Molchanov P, Gupta S, Kim K, Pulli K (2015), "Short-range FMCW monopulse radar for hand-gesture sensing," IEEE Radar Conference, Johannesburg, South Africa, pp. 1491-1496, doi: http://dx.doi.org/10.1109/RADAR.2015.7131232

[14] Shaker G (2017, Nov 1), "Gesture recognition using mm-waves," Waterloo Artificial Intelligence Institute [Online]. Available: https://goo.gl/iRqkJC

[15] Lien J, Gillian N, Karagozler M E, Amihood P, Schwesig C, Olson E, Raja H, and Poupyrev I (2016), "Soli: ubiquitous gesture sensing with mm-wave radar," in ACM Transactions of Graphics, vol. 35, no. 4, article 142, doi: http://dx.doi.org/10.1145/2897824.2925953

[16] Naik S, Abhishek H R, Ashwal K N, Balasubramanya S P (2012), "A study on automotive human vehicle interaction using gesture recognition technology," International Journal of Multidisciplinary in Cryptology and Information Security, vol. 1, no. 2, pp. 6-12.

[17] Alpern M, Minardo K (2003), "Developing a car gesture interface for use as a secondary task," ACM Extended Abstracts on Human Factors in Computing Systems, Fort Lauderdale, United States, pp. 932-933, doi: http://dx.doi.org/10.1145/765891.766078