Gesture Recognition for Smart Home Applications using Portable Radar Sensors

Qian Wan, *Student Member, IEEE*, Yiran Li, *Student Member, IEEE,* Changzhi Li, *Senior Member, IEEE*, and Ranadip Pal, *Senior Member, IEEE*

Abstract—**In this article, we consider the design of a human gesture recognition system based on pattern recognition of signatures from a portable smart radar sensor. Powered by AAA batteries, the smart radar sensor operates in the 2.4 GHz industrial, scientific and medical (ISM) band. We analyzed the feature space using principle components and application-specific time and frequency domain features extracted from radar signals for two different sets of gestures. We illustrate that a nearest neighbor based classifier can achieve greater than 95% accuracy for multi class classification using 10 fold cross validation when features are extracted based on magnitude differences and Doppler shifts as compared to features extracted through orthogonal transformations. The reported results illustrate the potential of intelligent radars integrated with a pattern recognition system for high accuracy smart home and health monitoring purposes.**

I. Introduction

Research on human gesture classification for smart home applications using Wi-Fi [1] or radar signals [2] has been on the rise in recent years. On one hand, smart gesture recognition without requiring any tag attached to human body offers many advantages compared to contact-based solutions such as accelerometers, and thus is attractive for next generation elderly care (e.g., home utility control) and full body gaming. On the other hand, the radio-based approach is more robust than optic-based approaches because radio frequency (RF) signals can easily penetrate multiple indoor obstacles and is insensitive to ambient light.

Some of the key requirements for smart home gesture recognition systems are that the hardware used to map the gestures to signals is relatively small and cost effective, and the software can produce highly accurate classification in real time. On the hardware side, majority of the reported noncontact RF gesture recognition systems either use expensive commercial transceiver modules such as the NI USRP, or rely on bulky bench top systems. On the algorithm side, robust gesture recognition algorithms that can reliably differentiate multiple gestures with subtle differences are highly desirable.

In this article, we consider the design of a new system based on a cost-effective and lightweight smart radar sensor unit and a high accuracy gesture pattern recognition system, which is based on time and frequency domain features extracted from the smart radar signals. A K-Nearest Neighbor classification approach is applied to process the radar-received data. Multiple experiments are performed to demonstrate the effectiveness of the proposed system.

Research partially supported by the NSF under Grant ECCS- 1254838

All authors are with the Electrical and Computer Engineering Department at Texas Tech University, Lubbock, TX, USA, 79409. (corresponding author phone: 806-834-8631; e-mail: ranadip.pal@ttu.edu).

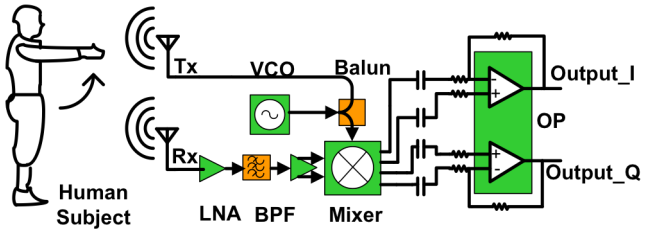


Fig. 1. *Simplified block diagram of the AC-coupled CW mode portable radar sensor.*

II. Radar System and Gesture Recognition Methods

A. Portable Smart Radars

A 2.4 GHz AC-coupled CW mode portable radar sensor is used to in this work to sense different gestures. Fig.1 shows the simplified block diagram of this radar sensor. The radar sensor design adopts homodyne direct direct-conversion architecture [3]. The transmitter and receiver chains are implemented on the same board. As shown in the block diagram, the receiver chain consists of a low noise amplifier (LNA), a band pass filter, a gain block, a mixer and a baseband operational amplifier (OP). A voltage-controlled oscillator (VCO) and a balun form the transmitter chain of the radar sensor to generate a single-tone -3 dBm carrier signal at a frequency of 2.4 GHz and also provide a local oscillator (LO) signal for the mixer in the receiver chain. It should be noted that the transmitted power is more than 1000 times weaker than the peak power of a GSM cell phone, and is thus safe for practical applications.

The received signal is first captured and amplified by the receiver chain, then down-converted to baseband. The baseband OP amplifies the *I*/*Q* signals. An NI-USB6009 data acquisition module digitizes the baseband *I*/*Q* signals with 200-Hz sampling frequency and transmits them to a laptop through an USB port.

In order to capture different gestures for this work, the radar sensor was placed on a table 1m above ground. A person sat in front of the radar and performed different gestures. Each gesture was repeated several times. The distance between the radar and the person was in the range of 2 meters.

B. Gesture Recognition System

For classifying the diverse categories of motions based on acquired radar signals, we map our problem to a traditional pattern recognition framework consisting of the following steps: feature selection, classification and error estimation. Since radar *I* and *Q* signals are time domain signals with multiple variations (such as time shift or speed) for same category of gestures, treating the I and Q signals at each time instant as a feature is not advisable. Thus, we first considered feature extraction based on Principal Component Analysis (PCA) [4] that maps the signals into orthogonal components with highest variance.

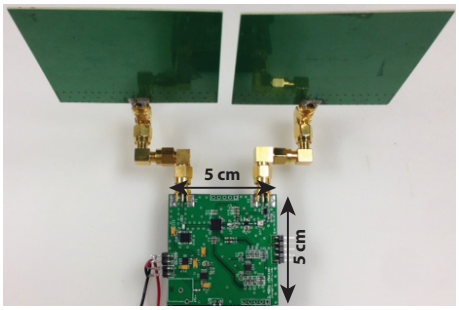
We also considered feature extraction based on capturing physical attributes such as relative direction and speed of motion during the duration of gesture. The details of these feature extractions are included in the Results section. The classifier used for all the pattern recognition analysis in this article was K-Nearest Neighbors [5] with k=3. In this supervised learning *k*NN model, each sample can be assigned to the class most common among its *k* nearest neighbors based on Euclidean distance between the feature vectors. We have normalized each feature individually before applying *k*NN. We didn't try other classifiers since KNN with features extracted based on physical attributes was providing us a very high classification rate. The classification accuracy was calculated using 10 fold cross validation [6].

III. Experimental Results

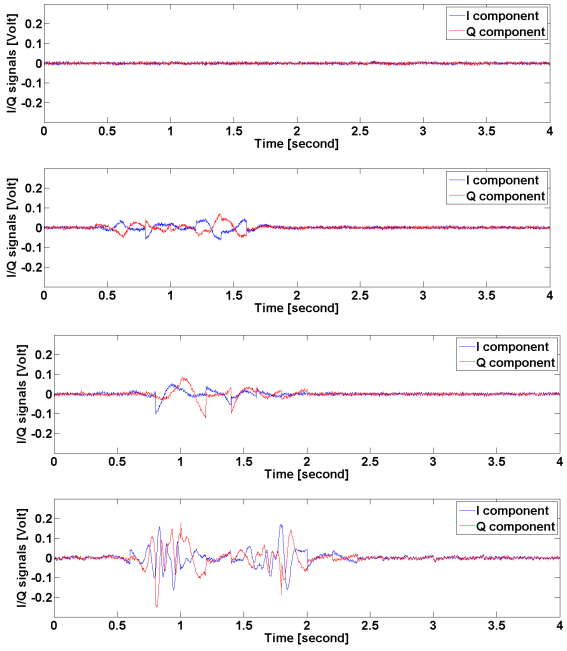
A. Experiments with Basic Motions

We initially sampled four familiar motions (*no movement*, *shaking head*, *nodding* and *hand lifting*) from the same person with 20 repetitions for each motion resulting in a sample set of 80 samples with four categories. Table 1 shows the classification accuracy for 11 different motion classification combinations. For instance, M_2M_4 denotes the classification problem of categorizing shake-head or hand-lifting movements. The second column denotes 3NN 10-fold cross validation results using top 20 features extracted based on PCA. We observe that accuracy is highest (90%) for categorizing *shaking head* and *nodding* movements but average classification accuracy stays low at 64%. We also applied 3NN by changing the number of top features with similar results (results not included). These results seem to indicate that PCA may not be suitable to extract the most discriminative features from input radar signals. Accordingly, we considered involving physical features in our classification algorithm for the purpose of achieving a higher recognition rate. Since the relative distance between the radar and the user is related to the magnitudes of captured radar signals, we next consider the signal magnitudes in feature generation.

Fig. 2 shows the radar hardware used for the experiment and the typical *I*/*Q* channel signals for the four considered movements. We extracted both the *I* and *Q* magnitude differences between the highest crest and the lowest trough of each signal. Subsequently, utilizing these two magnitude difference features, we trained our *k*NN model and the results illustrate that the 10-fold CV classification rates are significantly increased as shown in Table 1 (column 4 denoted as *MagDiff*). The accuracy rates are all > 92% and 100% for some of the classification combinations. Fig. 3 illustrates the distinguishability among the four motions based on the magnitude difference of I and Q signals. Note that *k*NN using a combination of top 20 PCA features and Magnitude Difference features was also implemented as shown in column 3 of Table 1 where the performance actually dropped since the PCA features still remained significant in the model formation process.



(a)



(b)

Fig. 2. *(a) The radar used for experiment. (b) Sampled radar signals for four motions: no movement, shaking head, nodding and lifting hand (ordered from top to bottom).*

Fig. 3 shows that the four movements of *shake head*, *head nod*, *hand lifting* and *no movement* can be accurately classified by the simple classifier designed from only two magnitude difference features. However, if the user performs more complicated gestures, the accuracy of our recognition system will be limited. We next consider a set of experiments with complicated gestures.

B. Experiments with three gestures

In this experiment, we added *hand pushing* motion for comparison with *hand lifting*, here the user differs from the one in Fig. 2 and we have three categories composed of 20 samples each. The gestures are shown in Fig. 4 and their

magnitude differences are shown in Fig. 5. The differentiability between *shaking head* and *hand lifting* is still quite distinct, but there are overlapping points between *hand pushing* and *hand lifting*. Consequently, besides the magnitude difference features (related to the relative distance from the radar), we further investigated the feasibility of extraction of orientation features.

Table1. 10-fold cross validation classification rates among four sets of motions. M_1: *no movement*, M_2: *shaking head*, M_3: *nodding*, M_4: *hand lifting*. PCA represents features extracted from Principal Component Analysis while MagDiff denotes features based on Magnitude Differences.

	PCA	PCA + MagDiff	MagDiff
M_1M_2	70.0%	70.0%	100.0%
M_1M_3	82.5%	82.5%	100.0%
M_1M_4	55.0%	55.0%	100.0%
M_2M_3	90.0%	90.0%	92.5%
M_2M_4	55.0%	55.0%	100.0%
M_3M_4	52.5%	52.5%	97.5%
$M_1M_2M_3$	71.7%	71.7%	93.3%
$M_1M_2M_4$	55.0%	56.7%	100.0%
$M_1M_3M_4$	56.7%	56.7%	98.3%
$M_2M_3M_3$	56.7%	56.7%	93.3%
$M_1M_2M_3M_4$	57.5%	58.8%	96.3%

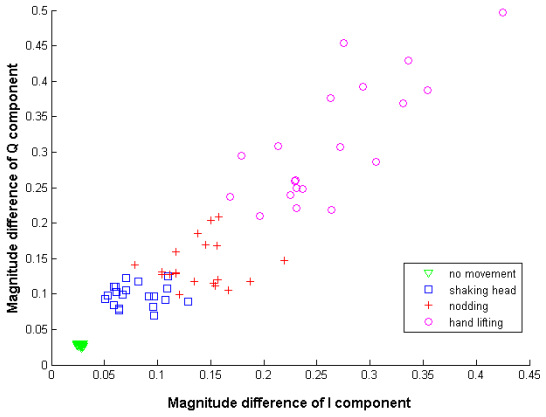


Fig. 3. *Plot of four common human motions in terms of magnitude differences between I and Q radar signals.*

Doppler shift [7] is the change in the observed frequency as the user moves relative to the radar receiver. For instance, when a user moves his hand towards the receiver, we see positive Doppler frequencies, and negative Doppler frequencies will show up when he pulls his hand backwards from the receiver. In order to see the micro-Doppler information of each gesture, Short-Time Fourier Transform (STFT) was applied to the output *I*/*Q* signals with a sliding window size of 2.56 second (512 samples) [8]. Hamming window was used in this work. Note that no extra filter was used to fully preserve all the motion information. Fig. 6 demonstrates three kinds of Doppler spectrogram corresponding to the motions in Fig. 4 for three repetitions. The Doppler profiles show that each individual movement has similar Doppler shifts (similarity across each column), that are distinct from each other (differences among each row).

Fig. 4. *The three gestures used for experimental set B. Each row represents a gesture from left to right: Hand Pushing (top row), Hand Lifting (middle row), Head Shake (bottom row). The orientation of the radar was to the left for the first two rows and in front for the third row.*

We next present the feature extraction algorithm based on Doppler shift images similar to the ones shown in Fig. 6. Let *P* denote the pixel values in a Doppler shift image. The rationale and steps for feature extraction are as follows:

- We observed in the Doppler shift images that the location of the brightest area changes across different gestures, while remaining stable for the same category of motion. Meanwhile, the light blue shape of the Doppler shift appears to change with each category of gesture. Thus, two Doppler pattern features can be generated based on the above observations.
- Note that the highest energy in Fig. 6 concentrates around zero frequency, thus we set $P_{i,j} = min(P)$ if $P_{i,j} > 0.8max(P)$ to eliminate the high energy around zero frequency. Subsequently, an averaging window (here it is 7×7) was applied over each image to get the coordinates of the brightest area. Following that, we can obtain the height of this area (i.e. frequency corresponding to highest magnitude) to be used as an input feature.
- In order to get the Doppler shape information, we divided each image to upper and lower parts separated by the zero frequency line. Before that, we set $P_{i,j} = min(P)$ if $P_{i,j} > 0.6max(P)$ to get rid of high energy frequencies. Let *AvgP*

represents the averaged pixel values. If it is true that $max(AvgP_{upper}) > 0.3\max(AvgP)$ and $max(AvgP_{lower}) > 0.3\max(AvgP)$, that means there exists light blue shapes, otherwise we define this kind of image as a representation of *shaking head* motion, giving the feature a numerical value of *0*. For other kinds of motions, we denote X_{upper} and X_{lower} as the time coordinate of the brightest energy in upper and lower parts, respectively. If $X_{upper} < X_{lower}$, we consider that this image represents *hand pushing*, giving the feature a numerical value of *0.5*. Otherwise, if $X_{upper} >= X_{lower}$, this corresponds to *hand lifting* and we provide the feature a numerical value of *1*.

Finally, we are capable of extracting two features (magnitude differences) from the time domain and two features (Doppler patterns) from the spectrogram. For the purpose of evaluating the performance of our recognition system based on these 4 features, we compared it with PCA and magnitude difference features on the same dataset. As shown in Table 2, our final recognition system has been substantially improved by adding Doppler features. It outperforms both PCA based *k*NN and magnitude difference based *k*NN and produces an average 10 fold cross validation classification rate of 98%. It should be noted that it is feasible to modify the technique for extracting the fourth feature (Doppler shape information) by adding further shape extraction approaches if more categories are being differentiated.

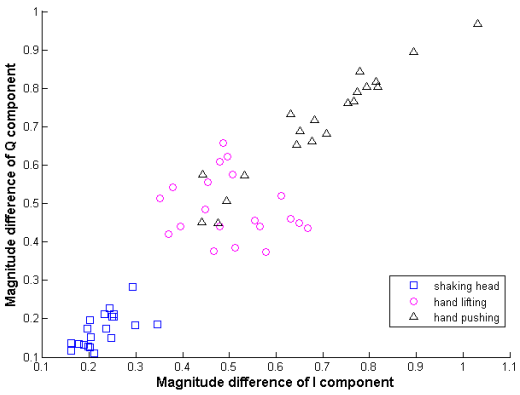


Fig. 5. *Plot of three common human movements in terms of magnitude differences of radar signals.*

IV. Conclusion

The article presents a proof-of-concept gesture recognition system based on portable smart radar sensors that achieves high accuracy for differentiating different types of hand movement or different types of head movement. The results for differentiating hand lifting and hand pushing based on four features suggest that further complicated set of gestures can potentially be recognized if we utilize signals from multiple radars placed at different locations in the home environment. For future research, we will consider significantly higher number of gestures and optimal placement of radars to improve the robustness and accuracy of the recognition system along with detailed comparison with state of the art gesture recognition systems.

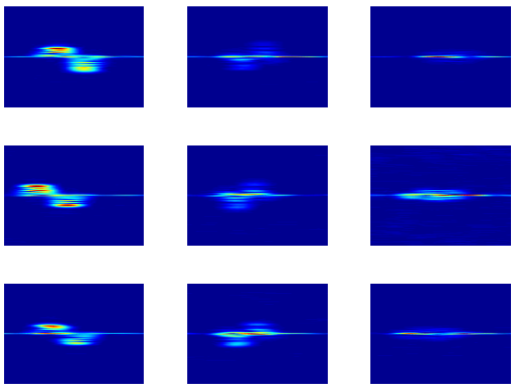


Fig. 6. *Frequency-time Doppler shift profiles of three common human motions for three samples.* Each column represents identical movement, motions are (from left to right): *hand pushing and pulling*, *hand lifting and lowering*, *shaking head*.

Table 2. 10-fold cross validation classification rates among sets of three motions. M_1: *shaking head*, M_2: *hand lifting*, M_3: *hand pushing*. PCA represents Principal Component Analysis, MagDiff denotes Magnitude Difference while Doppler denotes Doppler shift features.

	PCA	MagDiff	MagDiff + Doppler
M_1M_2	75.0%	100.0%	97.5%
M_1M_3	50.0%	100.0%	100.0%
M_2M_3	65.0%	77.5%	97.5%
$M_1M_2M_3$	51.7%	83.3%	96.7%

References

[1] P. Qifan, G. Sidhant, G. Shyamnath, P. Shwetak, "Whole-home gesture recognition usingwireless signals," presented at the Proceedings of the 19th annual international conference on Mobile computing & networking, Miami, Florida, USA, 2013.

[2] F. Adib, Z. Kabelac, D. Katabi, R. C. Miller., "3D Tracking via Body Radio Reflections", *Usenix NSDI'14*, Seattle, WA, April 2014

[3] C. Gu, R. Li, H. Zhang, A. Y. C. Fung, C. Torres, S. B. Jiang, C. Li, "Accurate respiration measurement using DC-coupled continuous-wave radar sensor for motion-adaptive cancer radiotherapy," *IEEE Trans. Biomed Eng.*, vol. 59. No. 11, pp. 3117-3123, November 2012.

[4] Jolliffe, Ian. Principal component analysis. John Wiley & Sons, Ltd, 2005.

[5] T. Cover, P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Information Theory*, vol. 13, no.1, pp. 21-27, 1967.

[6] Kohavi, Ron. "A study of cross-validation and bootstrap for accuracy estimation and model selection." IJCAI. Vol. 14. No. 2. 1995.

[7] V. Chen, Q. Shie, "Joint time-frequency transform for radar range-Doppler imaging," *IEEE Trans. Aerospace and Electronic Systems*, vol. 34, no. 2, pp. 486-499, 1998.

[8] Y. Kim, H. Ling, "Human activity classification based on micro-Doppler signatures using a support vector machine", *IEEE Trans. Geosci. Remote Sensing*, vol. 47, No. 5, pp. 1328-1337, May 2009.