Article

Clustering-Driven DGS-Based Micro-Doppler Feature Extraction for Automatic Dynamic Hand Gesture Recognition

Chengjin Zhang 1,†, Zehao Wang 2,†, Qiang An 3,*, Shiyong Li 1, Ahmad Hoorfar 4 and Chenxiao Kou 1

1 Beijing Key Laboratory of Millimeter Wave and Terahertz Technology, Beijing Institute of Technology, Beijing 100081, China
2 Key Laboratory of Microwave Remote Sensing, National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China
3 Department of Biomedical Engineering, Fourth Military Medical University, Xi’an 710032, China
4 Antenna Research Laboratory, Center for Advanced Communications, Villanova University, Villanova, PA 19085, USA
* Correspondence: qiang.an.903@outlook.com; Tel.: +86-186-8291-7563
† These authors contributed equally to this work and should be considered as co-first authors.

Abstract: We propose in this work a dynamic group sparsity (DGS) based time-frequency feature extraction method for dynamic hand gesture recognition (HGR) using millimeter-wave radar sensors. Micro-Doppler signatures of hand gestures show both sparse and structured characteristics in time-frequency domain, but previous study only focus on sparsity. We firstly introduce the structured prior when modeling the micro-Doppler signatures in this work to further enhance the features of hand gestures. The time-frequency distributions of dynamic hand gestures are first modeled using a dynamic group sparse model. A DGS-Subspace Pursuit (DGS-SP) algorithm is then utilized to extract the corresponding features. Finally, the support vector machine (SVM) classifier is employed to realize the dynamic HGR based on the extracted group sparse micro-Doppler features. The experiment shows that the proposed method achieved 3.3% recognition accuracy improvement over the sparsity-based method and has a better recognition accuracy than CNN based method in small dataset.

Keywords: dynamic hand gesture recognition; sparse signal representation; dynamic group sparsity; micro-Doppler features

1. Introduction

Over the last decade, dynamic hand gesture recognition (HGR) has received an increasing research interest for human-machine interactions (HMI). It possesses great significance in a number of short-range contactless applications [1–6]. Getting rid of the physically contact sensors required in traditional electromyography (EMG)-based or glove-based HGR tasks [7–9] brings many benefits, including accessibility to all potential users (healthy persons, patients with limited mobility, those allergic to contact sensors), convenience of long-term monitoring (enabling automatic stop and go detection), mobility and flexibility of deployment (adaptation to all environmental and lighting conditions).

Various schemes have been proposed for non-contact dynamic hand gesture recognition, such as optical sensors [10–12], acoustic sensors [13], Wi-Fi [14,15], and radar-based methods [16]. Radar-based HGR has attracted a considerable attention and tremendous progress has been made since it can work in all lighting conditions, even in penetrating condition, and with a privacy preserving manner. The micro-Doppler features extracted from the spectrograms obtained using the short-time Fourier transform (STFT) analysis, are often utilized to characterize different hand gestures. In addition, some studies used wideband radar or multi-antenna radar systems to obtain distance or angle information [17–19]. Introducing more perspectives improves the ability of recognizing gestures in certain scenarios, such as using angle information to distinguish between flapping the hand to left and right or rotating the hand clockwise and counterclockwise.
Radar-based HGR task differs from the arm motion recognition [20,21], which has also grown rapidly in recent years. Arm motions are performed with the joint participation of the upper arm, the lower arm and the palm. With the involvement of the upper arm, a quite pronounced time-frequency distribution can be achieved due to the wide motion spreading range, rapid velocity changes, and large radar cross section (RCS) of the upper arm. Unlike arm motions, hand gestures involve the motions of the fingers, the palm and the lower arm. The motion expansion, the speed and the RCS of the hand gestures are much smaller than those of the arm motions. As a result, the motion features of hand gestures are strongly attenuated and degenerated, which make the recognition much more difficult than that of the arm motions. Thus, there is an urgent need to investigate more effective ways of conducting feature extraction and enhancement for hand gesture recognition.

The lower dimensional features are attractive for recognition due to only small amount of data needed for classifier in comparison with the methods based on neural network. In refs. [22–25], various handcrafted features are extracted from time-frequency maps and used for HGR. Eigenspace features are also commonly used for HGR [26]. In ref. [3], application-specific features extracted using principal component analysis (PCA) were utilized to recognize dynamic hand gestures. The sparse reconstruction-based feature extraction approach has also achieved good performance and proven to be effective to handle with the gesture recognition task [27–29]. However, the approach only considered the fully sparse property. In fact, the micro-Doppler signatures of hand gestures exhibit a more important feature, that is the local clustering.

In this paper, in the aim of further enhancing hand motion features and improving the recognition performance, we proposed a novel strategy to jointly consider the sparsity and clustering properties of the micro-Doppler signatures and used a dynamic group sparsity (DGS) model [30] to extract corresponding features. Firstly, the relationship between the radar echoes of hand gestures and their corresponding micro-Doppler signatures are established using a time-frequency dictionary. Secondly, the micro-Doppler features are modeled using structured priors and extracted using the DGS-Subspace Pursuit (DGS-SP) algorithm [30]. Then, the features are fed into an SVM classifier. At last, experiments with data collected by a 24 GHz continuous wave (CW) radar are carried out to verify the efficacy of the proposed method. The results demonstrate that the structured feature is beneficial to improve the accuracy of dynamic hand gesture recognition.

The remainder of this paper is organized as follows. In Section 2, the DGS-based structured sparse model and the DGS-SP-based feature extraction algorithm are detailed. In Section 3, the dynamic hand gesture experiments are implemented, and the recognition accuracy is presented to verify the effectiveness of the proposed method in comparison with the sparse only method and the convolutional neural network (CNN) method. Section 4 summarizes the paper.

2. Materials and Methods

2.1. Micro-Doppler Signatures of Dynamic Hand Gesture

Time-frequency analysis is the most common approach to conduct motion recognition task. Usually, the short-time Fourier transform (STFT) is applied to process the radar records so as to obtain the time-frequency representation,

\[ y(n, k) = \sum_{m=0}^{L-1} y(n + m)h(m)e^{-j2\pi km/N} \]  

(1)

where \( s(\cdot) \) represents the demodulated echo data, \( n = 0, \ldots, N - 1 \) denotes the time index, \( k = 0, \ldots, K - 1 \) is the discrete frequency index, \( h(\cdot) \) is a Hanning window with length \( L \). An example spectrogram of flipping fingers is illustrated in Figure 1. The length of Hanning window is set to 64 (0.064 s), the overlap of two consecutive windows is 63, and \( K \) is set to 256.
It can be clearly observed that most parts of the spectrogram are populated by background noises with relatively weak energy. Only in several localized and concentrated regions does the spectrogram possess stronger energy. The observation in fact states that the time-frequency distribution of dynamic hand gestures presents not merely completely sparseness features as discussed in ref. [27]. More precisely, it exhibits obvious local clustering characteristics, which can be more perfectly approximated by structured sparsity models.

It has been well studied and proven that sparsity based micro-Doppler features extraction methods can be of great benefit for HGR task in limited dataset scenarios [27,28]. However, no related work so far was reported to consider the clustering nature of the spectrograms of dynamic hand gestures.

### 2.2. Sparsity Model of Dynamic Hand Gesture

Firstly, we present the definition of a $K$-sparse signal. If a signal $x \in C^M$ can be approximated by $K \ll M$ non-zero coefficients under certain transformation, the signal is a called a $K$-sparse signal. Compressed sensing (CS) theory states that if the signal is sparse in a certain domain, the original signal can be accurately recovered using sparse reconstruction technique with a reduced observation [25].

Under complete sparsity hypothesis, by denoting the raw radar echoes of dynamic hand gestures as $y \in C^N$, the following sparse representation of $y$ in time-frequency domain holds,

$$y = \Phi x$$

where $\Phi \in C^{N \times M}$ represents a time-frequency dictionary, $x \in C^M$ is a sparse vector. The above model states that the radar echo of dynamic hand gesture can be approximated using linear superposition of a series of basis signals, which can have various forms. This paper adopts the Gaussian-windowed Fourier basis signal [31], which can be expressed as,

$$\Phi(n, m) \triangleq \Phi(n | t_m, f_m) = \frac{1}{2\pi} \exp \left[ -\frac{(n - t_m)^2}{\sigma^2} \right] \exp(jf_m n)$$

where $t_m$ and $f_m$ stand for the time and frequency shift of the basis signal, respectively, and $\sigma$ denotes the variance of the Gaussian window, namely, the scaling factor. And $n = 1, \ldots, N$ is the time shift index, while $m = 1, \ldots, M$ denotes the frequency shift index.

The parameter $\sigma$ is usually selected based on experience. Here, we set it to be 16. The values of $t_m$ and $f_m$ are empirically set to be $\{0.25\sigma, 0.5\sigma, 0.75\sigma, \ldots, 0.25\sigma \times N/(0.25\sigma)\}$ and $\{\frac{1}{4\pi}, \frac{2}{4\pi}, \frac{3}{4\pi}, \ldots, 2\pi\}$, respectively, where $\lfloor \cdot \rfloor$ denotes a rounding down operation [31].

According to the theory of CS [32,33], for a $K$-sparse signal $x$, if $K \ll N < M$ holds, the sparse time-frequency distribution $x$ in Equation (2) can be recovered by,

$$\hat{x} = \arg \min_x \|y - \Phi x\|_2^2 \text{ s.t. } \|x\|_0 \leq K,$$
where $\| \cdot \|_0$ and $\| \cdot \|_2$ denote the $L_0$ and $L_2$ norms, respectively. Equation (4) can be effectively solved using many algorithms [31,34,35]. Once the sparse coefficient is obtained, the raw radar echo of dynamic hand gestures can be approximated as follows,

$$
\hat{\mathbf{y}} = \mathbf{\Phi} \hat{\mathbf{x}} = \sum_{k=1}^{K} \hat{x}_{ik} \mathbf{\Phi}(n|t_{ik}, f_{ik})
$$

(5)

The time-frequency distribution of the reconstructed radar echo $\hat{\mathbf{y}}$ is shown in Figure 2, with $K$ set to be 24. Compared with the raw spectrogram in Figure 1, it is obvious that the position and strength of the dominant time-frequency components are well preserved and highlighted. However, the noises are not removed perfectly. The reason is that the complete sparsity assumption is valid not only for time-frequency signatures of the hand gesture, but also for the noise components. Therefore, it is unavoidable that part of the noise components would be recovered as well in the reconstructed spectrogram.

Figure 2. Spectrogram of reconstructed signal using the OMP algorithm.

2.3. Dynamic Group Sparsity Model of Dynamic Hand Gesture

In fact, the time-frequency distribution of dynamic hand gesture shows obvious clustering property while the noises tend to arbitrarily spread throughout the spectrogram. Meanwhile, the pattern of the cluster is not limited to any specific structure. Thus, we remodel the time-frequency distribution of dynamic hand gesture using a dynamic group sparsity model with more flexibility, in which an element surrounded by non-zero elements has a higher probability of being non-zero, and vice versa.

The dynamic group sparsity signal can be defined as follows: if a signal $\mathbf{x} \in \mathbb{C}^M$ can be approximated by $K \ll M$ non-zero coefficients under some linear transforms and these $K$ nonzero coefficients are clustered into $q \in \{1, 2, \cdots, K\}$ groups, the signal is called a dynamic $G_{K,q}$-sparse signal [30]. In this work, the group sparse representation of $\mathbf{y}$ in time-frequency domain is expressed as follows,

$$
\mathbf{y} = \mathbf{\Phi} x_{K,q}
$$

(6)

where $x_{K,q} \in \mathbb{C}^M$ is a group sparse vector. An effective algorithm, called Dynamic Group Sparsity-Subspace Pursuit (DGS-SP) [30], can be used to recover the above $G_{K,q}$-sparse signal, which is expressed as,

$$
\hat{x}_{K,q} = \text{DGS-SP}(\mathbf{y}, \mathbf{\Phi}, K, \beta)
$$

(7)

where $\beta$ denotes the weights of the neighbors. The most important feature of the DGS-SP algorithm is the introduction of a unique pruning process in the iteration, as described in Algorithm 1 below. Firstly, for the vector $\mathbf{v}$ needed to be pruned, the neighbor indices of each element are calculated. Then the neighbors are weighted and summed according to the weight coefficient $\beta$, and the result is recorded as $z$. The first $K$ maximum values in $z$ are taken as the pruned vector. After that, embedding the DGS pruning into the Subspace Pursuit (SP) algorithm results in the so-called DGS-SP algorithm. The pseudocodes of the
DGS pruning and the DGS-SP algorithm are described in Algorithms 1 and 2, respectively. Different neighboring structures, namely the structured priors, can be adopted when conducting the pruning, as shown in Figure 3 [36,37], which comes with different recognition performance. It will be detailed in Section 4.

![Group structures used in the DGS Pruning algorithm: (a) cross-shape structure, (b) vertical structure, (c) horizontal structure, (d) eight neighbors structure.](image)

**Algorithm 1.** DGS pruning.

**Input:** Signal \( v \in \mathbb{R}^M \), sparsity \( K \), weights for neighbors \( \beta \)  
**Output:** solution support \( \text{supp}\{ v, K \} \)

**Step:**
1. compute the index \( N_v \in \mathbb{R}^{M \times l} \) of the corresponding neighbors, \( l \) is equal to the number of non-zero elements in \( \beta \);
2. compute the weights \( w = [\beta_1, \beta_2, \ldots, \beta_I], \beta_i = [\beta_i, \beta_i, \ldots, \beta_i]_I^T \) for \( m = 1 \) to \( M \) do
3. compute \( z(m) = v^2(m) + \sum_{i=1}^I w^2(m,i) r^2(N_v(m,i)) \), end for
4. let \( \text{supp}\{ v, K \} \) be the index corresponding to the first \( K \) maximum values in \( z \).

**Algorithm 2.** DGS-SP.

**Input:** Sparsity \( K \), observation matrix \( \Phi \), original signal \( y \), weight for neighbors \( \beta \)  
**Output:** Sparse approximation \( \hat{x} \)

**Initialization:**
1. the residual \( r^0 = y \);
2. the solution support \( S^0 = \emptyset \);
3. the atom set \( \Phi^0 = \emptyset \);
4. the iteration index \( l = 0 \);

**Iteration:** At the \( l \)th iteration, go through the following steps:
1. \( l = l + 1 \);
2. compute \( v' = r^T \Phi \);
3. \( \Omega = \text{DGSPruning}\left( v', K, \beta \right) \);
4. \( S^l = S^l-1 \cup \Omega \);
5. \( \Phi^l = \Phi_{S^l} \);
6. compute \( b = \arg\min_{\hat{x}} \| y - \Phi^l \hat{x} \|_2 \);
7. \( \Omega = \text{DGSPruning}\left( b, K, \beta \right) \);
8. \( S^l = S^l_\Omega \);
9. \( \Phi^l = \Phi_{S^l} \);
10. compute \( \hat{x}^l = \arg\min_{\hat{x}} \| y - \Phi^l \hat{x} \|_2 \);
11. update \( r^l = y - \Phi^l \hat{x}^l \);
12. if \( \| r^l \|_2 \geq \| r^{l-1} \|_2 \), quit the iteration.
The time-frequency distribution of the reconstructed radar echo \( \hat{y} \) using DGS-SP algorithm is shown in Figure 4. The group sparsity level is set to 24, which is consistent with the level in Section 2.2. The third structure as shown in Figure 3c is selected as the neighboring structure. As compared to the results in Figure 2, it is obvious that the micro-Doppler signatures of flipping fingers are well recovered while the noise components being significantly suppressed, which indicate that the proposed approach possesses better noise-isolation performance than the OMP algorithm.

![Figure 4. Spectrogram of reconstructed signal using the DGS-SP algorithm.](image)

### 2.4. Feature Extraction of Dynamic Hand Gesture

It was detailed in Section 2.2 that the time-frequency distribution of the hand gesture echo \( y \) can be approximated by a group of basis signals with parameter sets \([|x_{i_k}|, t_{i_k}, f_{i_k}]\), where \(|x_{i_k}|\) denotes the intensity of the specific time-frequency cell \((t_{i_k}, f_{i_k})\). Thus, the sets in effect serve to be representative features directly related to the data content of different hand gestures. By extracting such the discrete parameter sets of the pre-designated K-sparsity signal, the feature vectors can be formulated as below and utilized for subsequent hand gestures recognition.

\[
f(y) = (t_{i_1}, \cdots, t_{i_K}, f_{i_1}, \cdots f_{i_K}, \cdots, |x_{i_1}|, \cdots |x_{i_K}|)
\]

(8)

Then, we turn to the recovered spectrograms for a visually intuitive comparison of the distribution of the extracted features using the proposed method and the OMP-based method. In Figure 5, the white triangles in the spectrogram represents the time-frequency locations \((t_{i_k}, f_{i_k})\) of the extracted feature vector. As can be observed from the results, the features selected by DGS-SP algorithm are more focused around the major micro-Doppler signatures than that of the OMP method. Since the major micro-Doppler signatures contribute to the discrimination of different hand gestures it implies that the extracted feature vectors using the proposed approach can achieve better hand gesture recognition performance.

![Figure 5. Demonstration of the feature vectors and the corresponding time-frequency distribution, yielded by: (a) DGS-SP; (b) OMP.](image)
3. Results

3.1. Data Collection and Feature Extraction

The dataset of dynamic hand gestures utilized in this paper was collected using a software-defined radar, SDR-KIT 2400T2R4, developed by Ancortek Inc., USA [38]. The platform is composed of two transmitters and four receivers. It can work either in frequency modulated continuous wave (FMCW) mode with the operating frequency ranging from 24 GHz to 26 GHz or in a single-tone CW mode with the frequency fixed at any intermediate value. We in this work used a laptop connected to the Ancortek millimeter wave radar to record and process the radar echoes of hand gestures. And the Python programming language is employed to implement all related signal processing method.

In the experiment, we use only one tranceiving antenna pair to collect the scattered data. The data acquisition setup and the schematic diagrams of four dynamic hand gestures are illustrated in Figure 6. The radar system operates at 24 GHz with a sampling frequency of 1 kHz. The separation between the antenna front and the human hand is about 0.3 m. In total, four types of hand gestures are considered, including (a) snapping fingers, (b) flipping fingers, (c) clenching hand, and (d) clicking fingers. Four human subjects were recruited to conduct the experiment. Each of them repeated the four gestures for 25 times. The duration of radar recording lasts for 15 s. Thus, a complete dynamic hand gesture cycle is about 0.6 s. In this way, we can obtain 400 records of four hand gestures, 100 for each. Then we use the same parameters as introduced in Section 2.1 to calculate the spectrograms of the four types of dynamic hand gestures, and the results are shown in Figure 7.

![Figure 6. Illustrations of four different dynamic hand gestures: (a) snapping fingers, (b) flipping fingers, (c) clenching hand, (d) clicking fingers.](image)

![Figure 7. Spectrograms of received signals corresponding to four dynamic hand gestures: (a) snapping fingers, (b) flipping fingers, (c) clenching hand, (d) clicking fingers.](image)
The reconstructed spectrograms and the feature vectors are shown in Figure 8, by using the OMP method and the proposed method, respectively, with the same parameters described in Section 2. The results show that the DGS-SP method, by modeling clustering characteristics of spectrograms of hand gestures using structured prior, has better performance in extracting key hand motion information and suppressing noises in spectrogram.

![Figure 8](image_url)

**Figure 8.** Demonstration of the feature vectors and the corresponding time-frequency distributions of the four hand gestures by DGS-SP: (a–d), and by OMP: (e–h), with \( K = 24 \).

### 3.2. Recognition Accuracy Using Different Classifiers

Next, we consider the recognition accuracy. Each record of dynamic hand gesture was processed using the proposed method for each sparsity level. By repeating the above process for all the collected data samples, a dataset of feature vectors of different dynamic hand gestures of different sparsity levels is constructed. Then, several traditional machine learning classifiers are separately employed for recognition, including the Decision Trees, Naïve Bayes Classifiers, support vector machine (SVM) and K-nearest neighbor (KNN). Eighty percent of the dataset is selected as the training set, and the rest is used as the testing set. One hundred Monte Carlo trials are performed to produce an average recognition accuracy for each sparsity level. The mean recognition accuracy for four types of hand gestures obtained using different classifiers under various sparsity levels are depicted in Figure 9.

![Figure 9](image_url)

**Figure 9.** Recognition accuracy of the proposed method using different classifiers versus different sparsity levels.
Clearly, SVM performs best for all the four types of hand gestures and thus is chosen as our classifier. Such a selection means that the maximum inter-class distance is obtained using the specific classifier, and thus it is more competent for the hand gesture recognition task when compared to other classifiers. The highest recognition accuracy rate reached 91.3% with sparsity level being set to 48.

3.3. Recognition Accuracy under Different Dynamic Group Structures

The neighboring structure plays a key role in the DGS-SP method, which is also the key to distinguish it from the OMP method. Different neighboring structures can result in different reconstruction results, which will ultimately affect the classification performance. The results corresponding to different group structures (see Figure 5) are illustrated in Figure 10.

![Figure 10](image)

**Figure 10.** Recognition accuracy of the proposed method under different neighboring structures versus different sparsity levels.

Among the four different neighboring structures, the type (c) structure yields an overall highest recognition accuracy. The underlying rationality lies in the fact that the time-frequency distributions of hand gestures exhibit a more evident vertical expansion rather than lateral expansion due to the high sensing frequency and the short time duration when conducting hand gestures. Therefore, it is more desirable to apply such a neighboring structure with prominent lateral expansion to preserve the information related to the time dimension. Thus, the group structure (c) is suggested to be exploited to model the clustering nature of the spectrograms of the dynamic hand gestures.

3.4. Comparison with the OMP Method

The recognition performance of the proposed approach is given in this subsection, in comparison with the OMP-based method [27]. To give a more intuitively illustration, the reconstructed spectrograms of the snapping fingers under different sparsity levels using the DGS-SP approach and the OMP approach are first shown in Figure 11 with the selected feature vectors’ coordinates highlighted using white triangles in the spectrograms.

Then, quantitative comparison via evaluating the recognition accuracy by using 40% and 80% of the data as the training datasets are given in Figure 12. It can be found that the proposed method outperforms the OMP-based method when choosing the sparsity level larger than about 25, meaning that the proposed method is more robust with respect to the sparsity levels. The proposed method yields the highest recognition accuracy of 91.1% (mean value for the four gestures) when the sparsity level is set to be 48, while that value for OMP is 87.8% with the sparsity level set to be 8.
Figure 10. Recognition accuracy of the proposed method under different neighboring structures for each one, as shown in Figure 12. The recognition performance of the proposed approach is given in this subsection, in comparison with the CNN method and OMP. The corresponding confusion matrix in this scenario for the DGS-SP approach is depicted in Table 1.

Table 1. Confusion matrix of the proposed method.

|                  | Snapping Fingers | Flipping Fingers | Clenching Hand | Clicking Fingers |
|------------------|------------------|------------------|----------------|------------------|
| Snapping fingers | 81%              | 4%               | 7%             | 8%               |
| Flipping fingers | 1%               | 98%              | 0%             | 1%               |
| Clenching hand   | 2%               | 0%               | 97%            | 1%               |
| Clicking fingers | 7%               | 4%               | 0%             | 89%              |
3.5. Comparison with the CNN Method

Finally, the recognition performance of the proposed method, OMP-based method and the convolutional neural network (CNN) method, another widely adopted approach [39] as illustrated in Figure 13, is analyzed with different sizes of training dataset. The size of the training data varies from 10% to 90% with a step of 10%. The sparsity level is set to be 48 for the proposed method, and the value for OMP is chosen as 8 (the best parameter for each one, as shown in Figure 12). The results are shown in Figure 14.

Note that the recognition performance of the proposed method is better than that of OMP. Meanwhile, in the case of small dataset, the proposed method achieves higher recognition accuracy than the CNN method.

4. Conclusions

In this paper, we investigated and exploited four structured prior of the time-frequency distributions of dynamic hand gestures in order to enhance the hand motion features and further improve the recognition performance. A dynamic group sparsity model and DGS-Subspace Pursuit algorithm was utilized to model the spectrograms of the hand gestures. Such a modeling can well isolate the features of dynamic hand gestures and the noise components in the spectrograms. By experiment, we choose SVM as the classifier and sparsity level is set to be 48. Experimental result shows that the proposed method improves recognition accuracy about 3.3% over the OMP-based method and has a better performance than CNN based method in small dataset. The experiment demonstrated the effectiveness of the proposed method. The method proposed in this work is not only suitable for the recognition of simple hand gestures, but also for the recognition of more complex sign language gesture using phased array radar or smart metasurface. And for the future work, we would like to exploit dual band millimeter wave radar and configurable beam radiation, etc.
complex sign language gestures, which can enhance the feature extraction process. For future work, we would like to exploit dual-hands’ gesture recognition and more subtle and complex sign language gesture using phased array radar or smart metasurface. And for the limitation of hardwares, we feel that the commercial millimeter wave radar still has many places to improve, such as the operating distance improvement under the low power illumination, the antenna array configuration that enables user and environment sensitive and configurable beam radiation, etc.

**Author Contributions:** Conceptualization, C.Z., Z.W., Q.A. and S.L.; formal analysis, C.Z., Z.W. and S.L.; methodology, C.Z., Z.W., Q.A. and S.L.; validation, C.Z., Z.W., C.K. and A.H.; writing—original draft, C.Z., Z.W., Q.A. and S.L.; writing—review & editing, Q.A., S.L. and A.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Data sharing is not applicable to this article.

**Conflicts of Interest:** The authors declare no conflict of interest.

**References**

1. Wu, Q.; Zhao, D. Dynamic hand gesture recognition using FMCW radar sensor for driving assistance. In Proceedings of the 10th International Conference on Wireless Communications and Signal Processing, Hangzhou, China, 18–20 October 2018.

2. Ma, Y.; Liu, Y.; Jin, R.; Yuan, X.; Sekha, R.; Wilson, S.; Vaidyanathan, R. Hand gesture recognition with convolutional neural networks for the multimodal UAV control. In Proceedings of the Workshop on Research, Education and Development of Unmanned Aerial Systems, Linköping, Sweden, 3–5 October 2017.

3. Wan, Q.; Li, Y.; Li, C.; Pal, R. Gesture recognition for smart home applications using portable radar sensors. In Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA, 26–30 August 2014.

4. Jani, A.B.; Kotak, N.A.; Roy, A.K. Sensor Based Hand Gesture Recognition System for English Alphabets Used in Sign Language of Deaf-Mute People. In Proceedings of the 2018 IEEE SENSORS, New Delhi, India, 28–31 October 2018.

5. Sawasdee, S.; Pumrin, S. Elderly care notification system using hand posture recognition. In Proceedings of the 2014 Fourth International Conference on Digital Information and Communication Technology and its Applications, Bangkok, Thailand, 6–8 May 2014.

6. Kuno, Y.; Murashina, T.; Shimada, N.; Shirai, Y. Intelligent wheelchair remotely controlled by interactive gestures. In Proceedings of the 15th International Conference on Pattern Recognition, Barcelona, Spain, 3–7 September 2000.

7. Zou, M.F.; Tafreshi, R.; Langari, R. A Multi-Window Majority Voting Strategy to Improve Hand Gesture Recognition Accuracies Using Electromyography Signal. *IEEE Trans. Neural Syst. Rehabil. Eng.* 2020, 28, 427–436. [CrossRef] [PubMed]

8. Jaramillo, A.G.; Benalcazar, M.E. Real-time hand gesture recognition with EMG using machine learning. In Proceedings of the 2017 IEEE Second Ecuador Technical Chapters Meeting, Salinas, Ecuador, 16–20 October 2017.

9. Rahimian, E.; Zabibi, S.; Asif, A.; Farina, D.; Atashzar, S.F.; Mohammadi, A. Mohammad. FS-HPG: Few-Shot Learning for Hand Gesture Recognition via Electromyography. *IEEE Trans. Neural Syst. Rehabil. Eng.* 2021, 29, 1004–1015. [CrossRef] [PubMed]

10. Hussain, S.; Saxena, R.; Han, X.; Khan, J.A.; Shin, H. Hand gesture recognition using deep learning. In Proceedings of the 2017 International SoC Design Conference, Seoul, Korea, 5–8 November 2017.

11. Benitez-Garcia, G.; Prudente-Tixteco, L.; Castro-Madrid, L.C.; Toscano-Medina, R.; Olivares-Mercado, J.; Sanchez-Perez, G.; Villalta, L.J.G. Improving Real-Time Hand Gesture Recognition with Semantic Segmentation. *Sensors* 2021, 21, 356. [CrossRef] [PubMed]

12. Yu, C.; Wang, X.; Huang, H.; Shen, J.; Wu, K. Vision-Based Hand Gesture Recognition Using Combinational Features. In Proceedings of the Intelligent Information Hiding and Multimedia Signal Processing, Darmstadt, Germany, 15–17 October 2010.

13. Siddiqui, N.; Chan, R.H. A wearable hand gesture recognition device based on acoustic measurements at wrist. In Proceedings of the International Conference of the IEEE Engineering in Medicine & Biology Society, Seogwipo, Korea, 11–15 July 2017.

14. Zhou, G.; Jiang, T.; Liu, Y.; Liu, W. Dynamic gesture recognition with Wi-Fi based on signal processing and machine learning. In Proceedings of the 2015 IEEE Global Conference on Signal and Information Processing, Orlando, FL, USA, 14–16 December 2015.

15. Zhang, Y.; Dong, S.; Zhu, C.; Balle, M.; Zhang, B.; Ran, L. Hand Gesture Recognition for Smart Devices by Classifying Deterministic Doppler Signals. *IEEE Trans. Microw. Theory Tech.* 2021, 69, 365–377. [CrossRef]

16. Park, J.; Cho, S.H. IR-UWB Radar Sensor for Human Gesture Recognition by Using Machine Learning. In Proceedings of the 2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS), Sydney, NSW, Australia, 12–14 December 2016.
17. Wang, Y.; Ren, A.; Zhou, M.; Wang, W.; Yang, X. A Novel Detection and Recognition Method for Continuous Hand Gesture Using FMCW Radar. *IEEE Access* 2020, 8, 167264–167275. [CrossRef]

18. Yu, J.-T.; Yen, L.; Tseng, P.H. mmWave Radar-based Hand Gesture Recognition using Range-Angle Image. In Proceedings of the 2020 IEEE 91st Vehicular Technology Conference, Antwerp, Belgium, 25–28 May 2020.

19. Skaria, S.; Al-Hourani, A.; Lech, M.; Evans, R.J. Evans. Hand-Gesture Recognition Using Two-Antenna Doppler Radar with Deep Convolutional Neural Networks. *IEEE Sens. J.* 2019, 19, 3041–3048. [CrossRef]

20. Amin, M.G.; Zeng, Z.; Shan, T.; Guendel, R.G. Automatic Arm Motion Recognition Using Radar for Smart Home Technologies. In Proceedings of the 2019 International Radar Conference, Toulon, France, 23–27 September 2019.

21. Zhang, Z.; Tian, Z.; Zhou, M. Latern: Dynamic Continuous Hand Gesture Recognition Using FMCW Radar Sensor. *IEEE Sens. J.* 2018, 18, 3278–3289. [CrossRef]

22. Li, B.; Yang, J.; Yang, Y.; Li, C.; Zhang, Y. Sign Language/Gesture Recognition Based on Cumulative Distribution Density Features Using UWB Radar. *IEEE Trans. Instrum. Meas.* 2021, 70, 1–13. [CrossRef]

23. Sun, Y.; Fei, T.; Schliep, F.; Pohl, N. Gesture Classification with Handcrafted Micro-Doppler Features using a FMCW Radar. In Proceedings of the 2018 IEEE MTT-S International Conference on Microwaves for Intelligent Mobility, Munich, Germany, 15–17 April 2018.

24. Amin, M.G.; Zeng, Z.; Shan, T. Hand Gesture Recognition based on Radar Micro-Doppler Signature Envelopes. In Proceedings of the 2019 IEEE Radar Conference, Boston, MA, USA, 22–26 April 2019.

25. Ritchie, M.; Jones, A.M. Micro-Doppler Gesture Recognition using Doppler, Time and Range Based Features. In Proceedings of the 2019 IEEE Radar Conference, Boston, MA, USA, 22–26 April 2019.

26. Ritchie, M.; Jones, A.; Brown, J.; Griffiths, H. Hand gesture classification using 24 GHz FMCW dual polarised radar. In Proceedings of the International Conference on Radar Systems (Radar 2017), Belfast, UK, 23–26 October 2017.

27. Li, G.; Zhang, R.; Ritchie, M.; Griffiths, H. Sparsity-based dynamic hand gesture recognition using micro-Doppler signatures. In Proceedings of the 2017 IEEE Radar Conference, Seattle, WA, USA, 8–12 May 2017.

28. Li, G.; Zhang, R.; Ritchie, M.; Griffiths, H. Sparsity-Driven Micro-Doppler Feature Extraction for Dynamic Hand Gesture Recognition. *IEEE Trans. Aerosp. Electron. Syst.* 2018, 54, 655–665. [CrossRef]

29. Ding, C.; Yan, J.; Hong, H.; Zhu, X. Sparsity-based Feature Extraction in Fall Detection with a Portable FMCW Radar. In Proceedings of the 2021 IEEE International Workshop on Electromagnetics: Applications and Student Innovation Competition (iWEM), Guangzhou, China, 7–9 November 2021.

30. Huang, J.; Huang, X.; Metaxas, D. Learning with dynamic group sparsity. In Proceedings of the IEEE International Conference on Computer Vision, Kyoto, Japan, 27 September–4 October 2009.

31. Mallat, S.G.; Zhang, Z. Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Process.* 1993, 41, 3397–3415. [CrossRef]

32. Donoho, D.L. Compressed sensing. *IEEE Trans. Inf. Theory* 2006, 52, 1289–1306. [CrossRef]

33. Candes, E.J.; Tao, T. Near-Optimal Signal Recovery from Random Projections: Universal Encoding Strategies? *IEEE Trans. Inf. Theory* 2006, 52, 5406–5425. [CrossRef]

34. Dai, W.; Milenkovic, O. Subspace Pursuit for Compressive Sensing Signal Reconstruction. *IEEE Trans. Inf. Theory* 2009, 55, 2230–2249. [CrossRef]

35. Tropp, J.A.; Gilbert, A.C. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Inf. Theory* 2007, 53, 4655–4666. [CrossRef]

36. Wang, L.; Zhao, L.; Bi, G.; Wan, C. Sparse Representation-Based ISAR Imaging Using Markov Random Fields. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2015, 8, 3941–3953. [CrossRef]

37. Zhao, L.; Wang, L.; Yang, L.; Zoubir, A.M.; Bi, G. The Race to Improve Radar Imagery: An overview of recent progress in statistical sparsity-based techniques. *IEEE Signal Process. Mag.* 2016, 33, 85–102. [CrossRef]

38. Ancortek, Inc. SDR Evaluation Kit, Read Write. Available online: http://ancortek.com/sdr-kit-2400ad2 (accessed on 12 October 2020).

39. Kim, Y.; Toomajian, B. Hand gesture recognition using micro-Doppler signatures with convolutional neural network. *IEEE Access* 2016, 4, 7125–7130. [CrossRef]