

MATERIAL IDENTIFICATION USING RF SENSORS AND CONVOLUTIONAL NEURAL NETWORKS

Gianluca Agresti, Simone Milani

Dept. of Information Engineering - University of Padova, Padova, Italy
Email: {gianluca.agresti,simone.milani}@dei.unipd.it

ABSTRACT

Recent years have assisted a widespreading of RF-based tracking and mapping algorithms for a wide range of applications, ranging from environment surveillance to human-computer interface.

This work presents a material identification system based on a portable 3D imaging radar-based system, the Walabot sensor by Vayyar Technologies; the acquired three-dimensional radiance map of the analyzed object is processed by a Convolutional Neural Network in order to identify which material the object is made of. Experimental results show that processing the three-dimensional radiance volume proves to be more efficient than processing the raw signals from antennas. Moreover, the proposed solution presents a higher accuracy with respect to some previous state-of-the-art solutions.

Index Terms— material classification, radar, convolutional neural network, 3D RF sensors, reflectance

1. INTRODUCTION

Material and object identification have been extensively investigated in many practical applications, ranging from automotive to security and forensic analysis. To quote some of the proposed solutions, infrared imaging has been used to evaluate the ripeness of fruits or other food in industry and agriculture [1, 2], hyperspectral processing and classification have enabled a more accurate image segmentation [3], and muon tomography has permitted the detection of masked radioactive materials [4].

In this list, it is easy to notice that different types of sensors and technologies have been used, from infrared sensors to hyperspectral imaging. More recently, radar-based systems have been taken into consideration for this task. In [5], a synthetic aperture radar installed on board of an aerial vehicle is used to material classification and image segmentation. The approach in [6] employs a millimeter-wave portable radar system, which was designed for gesture recognition, and adapts it to the classification of multiple materials and objects.

This paper focuses on the problem of material identification using a portable 3D imaging radar-based system, which enables a three-dimensional mapping of the radiated environment. To this purpose, we employed the Walabot sensor by Vayyar Technologies [7], which consists in a radar-based device working in the [6.3 – 8.3] GHz bandwidth. Antennas emit a set of amplitude modulated signals and the device estimates a 3D Radio-Frequency (RF) mapping of the radiated space.

The samples are then processed by a Convolutional Neural Network (CNN) that permits to identify the material with an accuracy of 93 %. In this task, we focused on classifying building materials,

This work has been partially supported by the University of Padova project Phylo4n6 prot. BIRD165882/16.

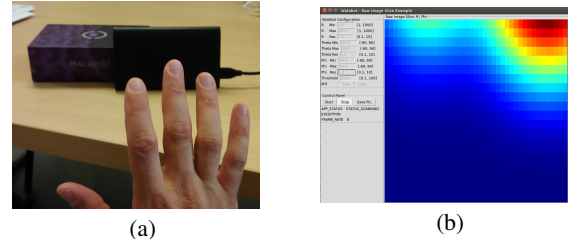


Fig. 1: The Vayyar Technologies Walabot device. a) external case, b) interface.

like different types of wood, bricks and mortars. This technology could be extremely useful in the evaluation of the material state and composition, which have direct consequences on its stability. It can also be used in detecting fraud materials or hidden objects.

The main contribution of the current work consists in proving the effectiveness of processing the reflectance map provided by the device firmware instead of the raw samples from the couples of transmitting/receiving antennas (like in [6]); this representation of the environment operates an effective dimensionality reduction which lighten the computational effort without losing accuracy. Moreover, the work evaluates the robustness of using a Convolutional Neural Network (CNN) in place of Random Forests for the classification even when the acquired datasets is limited. Experimental tests were run on different days and environments in order to verify the replicability of the obtained results; in all these instances, the obtained final accuracy was confirmed.

The structure of the paper can be summarized as follows. Section 2 overviews the different 3D sensors and the technology they adopt. Section 3 describes the Walabot sensor and its characteristics, while Section 4 presents the analyzed classification strategies. Experimental results (Section 5) and the final conclusions (Section 6) ends the work.

2. RELATED WORKS

The last decade has assisted an accelerated development of active 3D sensors, which radiates the scene to be acquired at different frequencies of the electromagnetic spectrum. Some widely-adopted solutions processed infrared light signals in order to estimate a depth map of the scene [8, 9], localize hand poses [10], identify materials and characterize their states [11]. Such devices were widely employed in Human-computer Interfaces (HCI), robotic vision, 3D reconstruction, and many other control applications. Unfortunately, the efficiency of these sensors was deeply affected by the noise coming from external radiating source; as a matter of fact, their applica-

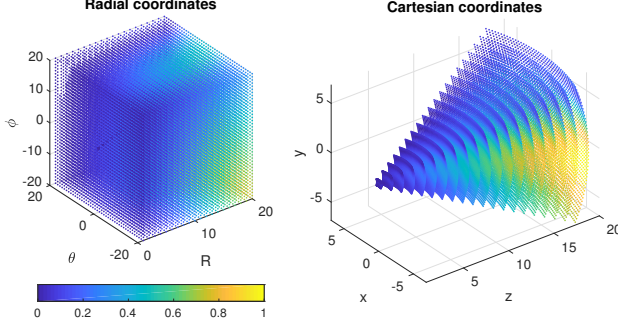


Fig. 2: Example of signal $I(R, \theta, \phi)$.

tions are limited to indoor environments with a controlled lighting.

Laser-based sensors, such as the LIDAR systems [12], can easily overcome such limitations, and therefore, their applications are nowadays investigated in many fields [13]. Unfortunately, such devices only enable a geometric modelization of the scene and are still available at an expensive cost.

Recent works have shown that RF sensors working in the SHF spectrum band can be used for similar applications. Wi-Fi antennas have been used to scan the 3D space, detect intruders, track people or objects with high precision [14], analyze volumes [15]. Such solutions have proved to be robust also in outdoor environment with respect to others using visible or near-visible radiations. Moreover, in many of the considered cases, occlusions do not prevent a correct reconstruction of the environment.

RF localization technology has been used for decades in aircraft tracking, security scanners and non-destructive testing and evaluation. They have also been employed in the monitoring of vibration in bridges or other large structures [16]. Similar systems have also been used in the diagnostics of working equipment [17] or detecting changes in the speed of the machinery [18]. Recently, the development of smaller and cheaper devices has extended the range of possible applications, such as presence detection, recognizing walking patterns [19], and detecting breathing and sleep patterns [20].

In addition, the technical development has allowed to build smaller and more accurate sensors which are able to monitor and recognize gestures [21, 22, 23]. Off-the-edge researches on Human-Computer Interaction (HCI) are nowadays exploring the use of millimeter-waves radar to detect and recognize small-sized motions to control or operate different devices [24]. Anyway, it is worth noticing that in RF systems the acquired signal depends on the shapes of objects, as well as on the material. This fact has suggested the idea of exploiting such sensors to characterize the composition of the objects. To the best of our knowledge, few works have been published on the subject. Among the most recent, the solution in [6] adopts a millimeter-wave radar to sense materials and body parts. The approach presented in this paper aims at improving such approach by employing a 3D spatialization of the signals and classifying it using a Deep Neural Network architecture.

3. THE WALABOT SENSOR

In our approach we adopted the radar-based RF sensor Walabot Pro [7], which senses the environment by transmitting, receiving and recording signals from an array of linearly polarized broadband antennas. This device is a wide band MIMO radar and it is equipped with 18 antennas, 4 of them are used to transmit amplitude modulated signals in the frequency range [6.3, 8.3] GHz (European

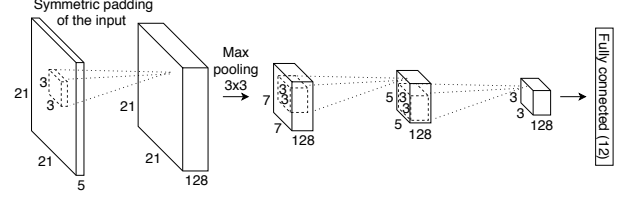


Fig. 3: Structure of the adopted CNN. Solid lines denote data tensors, dashed lines denote CNN kernels.

model), and the remaining 14 are used as receivers. The average transmit power is $25 \mu W$ and so it can be used safely in public places without restrictions.

The transmitting and receiving antennas are coupled generating 40 signals which are extracted by using the Wallabot SDK. Such system analyzes the received data via a beamforming algorithm and estimates reflected signal intensity at different points of the radiated space. The radiated space is sampled following a spherical coordinate reference system; each sample is rescaled and quantized into an 8-bits integer number.

As a matter of fact, different types of information can be acquired by the device. The implemented API permits acquiring the raw signals associated to the 40 couples of transmitting/receiving antennas, which are sampled at frequency of 100GHz and represented by arrays of 8192 values in double precision. In this work, we will denote such signals as $s_a(t)$, where $a = 0, \dots, 39$ is the index of the couple of antennas and $t = 0, \dots, 8191$ is the sampling instant. Alternatively, the firmware of the device can estimate an electromagnetic characterization of the radiated volume by processing the samples $s_a(t)$ and computing the amount of power reflected by a three-dimensional point. Such an information is formatted by the electromagnetic reflectance signal $I(R, \theta, \phi)$, which reports the reflectance (represented by an 8 bits integer) of a point localized at radial coordinates (R, θ, ϕ) . An example of the reflectance signal $I(R, \theta, \phi)$ is reported in Fig.2. In the adopted set-up, the signal $I(R, \theta, \phi)$ is sampled with ϕ, θ varying in the range $[-10^\circ, 10^\circ]$ and R varying in the range $[1, 5]$ cm with sampling step 1. By using these acquisition parameters, it is possible to extract just the local and superficial data from the target material, trying to reduce possible interferences coming from the environment. As a results, $I(R, \theta, \phi)$ is made of $21 \times 21 \times 5$ samples represented by 8-bits integers.

In the following sections, we will show that, differently from [6], it is possible to classify materials processing the reflectance $I(R, \theta, \phi)$ in place of $s_a(t)$. Experimental results show that $I(R, \theta, \phi)$ permits representing the acquired signal in a compact form without losing classification accuracy.

4. MATERIAL IDENTIFICATION STRATEGY

As it was anticipated, signals $s_a(t)$ and $I(R, \theta, \phi)$ can be used to identify the radiated material. To this purpose, two different machine learning strategies have been used and compared.

4.1. Random Forest approach

Following the approach in [6], we designed an *all-vs-all* material classifier using the *Random Forest* algorithm. Data are formatted into 1D arrays $\mathbf{s} = \text{vec}[s_0(t)|s_1(t)|\dots|s_{39}(t)]$ and $\mathbf{r} = \text{vec}I(R, \theta, \phi)$. The acquired arrays are partitioned into 30 different random subsets; for each subset, a classification tree is generate. Classification results are then combined using a majority

Table 1: Acquired materials and datasets

Materials				Datasets	
ID	material	ID	material	ID	labels
a	polystyrene	b	cement blocks	D1	h,i,j,k
c	leccese stone	d	extr. solava red	D2	all
e	desk Qe	f	desk Ae	D3	h,i,j,k,l
g	stab. cement	h	wall	D4	h,i,j,k,l
i	floor	j	wood1		
k	wood2	l	glass		

voting approach. Considering that the size of arrays $\mathbf{s}$ and $\mathbf{r}$ are quite large (approximately 2.6MB and 2.2kB, respectively), using some dimensionality reduction strategies can be useful. To this purpose, we considered two different types of projections Π_{PCA} and Π_{sNPE} , which were obtained from Principal Component Analysis (PCA) [25] and supervised Neighbourhood Preserving Embedding (sNPE) [26], respectively. The resulting arrays $\mathbf{r}' = \Pi_d^r \mathbf{r}$ and $\mathbf{s}' = \Pi_d^s \mathbf{s}$, where $d = \{PCA, sNPE\}$, are processed by the *Random Forest* classifier in the training and testing phases.

As for the tensor $I(R, \theta, \phi)$, it is possible to generate a more accurate classifier using CNN.

4.2. The proposed solution based on Deep Learning

Recent years have assisted to a widespread use of Deep neural networks (DNNs) in several classification problems. More specifically, whenever dealing with signals, Convolutional Neural Networks (CNNs) have proved to be significantly effective for their capability of processing large amounts of correlated inputs with a correspondingly-smaller amount of parameters to be optimize if compared to other DNN architectures. For these reasons, they have been largely employed in computer vision applications among which some examples are depth estimation from a single color image, flow motion estimation and semantic segmentation [27, 28, 29, 30].

In the proposed *all-vs-all* material classification system, we decided to use a CNN which processes the 3D map $I(R, \theta, \phi)$ captured by the Walabot device and defined in Section 3. CNNs are a good choice for this task since they are able to internally extract features related to the data spatial neighbourhood without requiring the huge number of parameters of *fully-connected* networks. Due the relatively small amount of labelled data, we used *k-fold cross-validation* in order to find a structure for the network that was nor to complex (incurring in over-fitting), nor to simple (incurring in under-fitting).

Different CNN structures were tested ending up with the configuration reported in Figure 3. The selected network contains a stack of 3 convolutional layers, each one followed by a *ReLU* non-linearity, and a fully-connected layer that is used as the output layer. The convolutional layers have 128 filters each of window size 3×3 pixels. Only the first layer uses a symmetric padding of 1 pixel on the input data and it is followed by a pooling layer with window 3×3 and stride 3. The last layer is a fully-connected layer composed by 12 nodes. The output of this layer is fed to a softmax which produces a probability distribution over the 12 class labels.

Made exception for the first layer, padding was not used for the other layers in order to reduce the overall dimension step by step. This structure of the CNN was sufficient to reliably classify the tested material as we will show in Section 5.

The network has been trained in order to minimize the softmax cross entropy between the output of the *CNN* and the class labels. We used the Stochastic Gradient Descent optimization with momentum 0.9 and minibatches of size 32. The convolutional kernels were initialized with Xavier's procedure [31]. The learning rate was se-

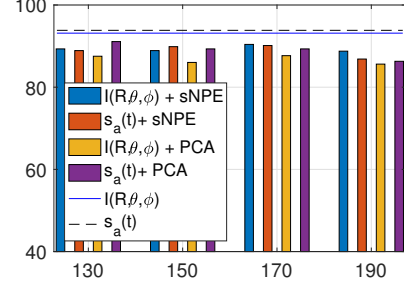


Fig. 4: Classification accuracy using $I(R, \theta, \phi)$ and $s_a(t)$. Vertical axis reports the accuracy on D1, horizontal axis reports the number of features extracted with PCA or sNPE.

lected equal to $5 \cdot 10^{-3}$ for the first 10 epochs and then it was reduced by the 20% every 10 epochs. The network was trained for a total of 200 epochs. These hyperparameters of the *CNN* were empirically selected maximizing the validation accuracy on the *k-fold cross-validation* procedure applied on the training set. When the system hyper-parameters have been chosen, the complete training set has been used for the training. The network has been implemented using the MatConvNet framework [32].

5. EXPERIMENTAL RESULTS

In our experimental tests, we evaluated the classification accuracy in identifying different building materials. To this purpose, we acquired different datasets which will be described in the following section.

5.1. Data collection

In order to verify the effectiveness of the described classifiers, we collected four different datasets (D1, D2, D3, D4) with different materials, which are reported in Table 1. Dataset D1 consists in 4 material classes, namely *wall*, *glass*, *wood1*, *wood2*,¹ which are made of about 100 acquisitions of signals $s_a(t)$ and $I(R, \theta, \phi)$. The acquisitions D1 were used to evaluate the performances using $\mathbf{s}$ or $\mathbf{r}$.

These classes were extended in dataset D2, including two different desks, different types of bricks and mortars, plastic material, and composed materials (*floor*). The final datasets consists of 12 different material types, which were acquired on the same day.

Dataset D3 and D4 consists in 5 of the materials contained in D2 (*wall*, *floor*, *wood1*, *wood2* and *glass*) but captured on different days and instances. This ensured the reproducibility of the reported results and permitted evaluating the robustness of the approach. For each acquisition the standard device calibration routine was disabled. The datasets are publicly available at http://lstm.dei.unipd.it/paper_data/Walabot_material_classification.

5.2. Input data evaluation

A first set of experiments was performed to evaluate the effectiveness of using $I(R, \theta, \phi)$ instead of $s_a(t)$. To this purpose, we evaluated the classification accuracy of *Random Forest* classifier on dataset D1: 70 % of acquired samples are used for training, while the remaining 30 % of left samples are used in validation. Classification performance was evaluated using either $I(R, \theta, \phi)$ or $s_a(t)$, which

¹Classes *wood1* and *wood2* are two different types of plywood, which present a different level of density

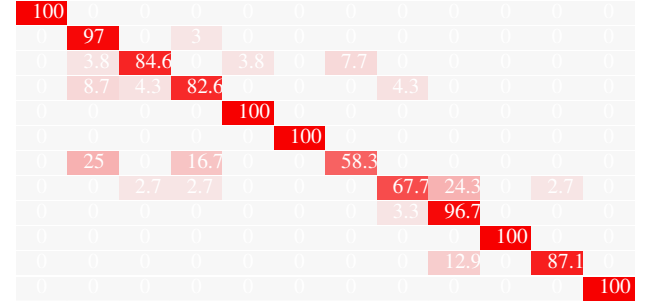
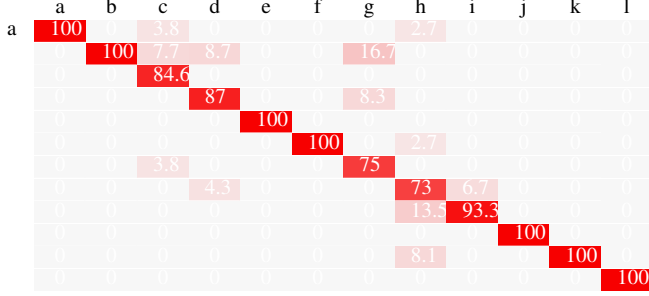


Fig. 5: Confusion matrix for the classification with CNN (a) and Random Forest (b) on D2 dataset. Column labels report the actual class, row label the predicted one.

were preprocessed by the dimensionality reduction strategies PCA and sNPE in order to reduce the size of feature arrays to N_F . The obtained results (for different N_F) are reported in Fig. 4, together with the classification results on the original data. It is possible to notice that using $I(R, \theta, \phi)$ does not generate significant differences in accuracy, although the entailed computational complexity and amount of memory decrease significantly. For such reasons, in the following we will employ $I(R, \theta, \phi)$ only.

5.3. Evaluation of the classification strategy

The second set of experiments were performed to evaluate the accuracies of the *Random Forest* and CNN classifiers on the task of discriminating the 12 materials contained in D2. This dataset is particularly challenging since it contains materials that are similar to each other, e.g. different types of stones and wooden plates. D2 has been split in a training (70%) and a test set (30%). The CNN hyperparameters (described in Section 4.2) have been selected by *k-fold cross-validation* on the extracted training set (both CNN and the *Random Forest* have been trained on this).

Fig. 5(a) and 5(b) report the confusion matrix respectively for the CNN and *Random Forest* classifications on the test set. The CNN is able to outperform with a mean accuracy of 93.3% the *Random Forest* classifier that has a mean accuracy of 89.5%. The *Random Forest* confuses some classes as the couple (*stabilized cement*, *cement block*) and (*wall*, *floor*), instead the CNN is more robust to the problem of class mismatch.

5.4. Evaluation of the performance reproducibility

In order to test the reliability of the proposed material classification system, we tested the *Random Forest* and CNN trained on the training set extracted from D2 on the datasets D3 and D4. These datasets contain 5 materials also present in D2 but acquired on different days.

As it is possible to observe from Fig. 6, both the *Random Forest* and CNN classifiers share stable performances when they are tested on new samples, in particular the CNN outperforms the *Random Forest* classifier obtaining a mean accuracy of 94% and 92.6% respectively on D3 and D4. The *Random Forest* classifier has mean accuracy of 92.3% and 90.9% respectively on D3 and D4.

6. CONCLUSIONS

The paper presented a material classification strategy based on the Walabot RF sensor. The acquired three-dimensional reflectance volume is processed by a CNN in order to identify the material of

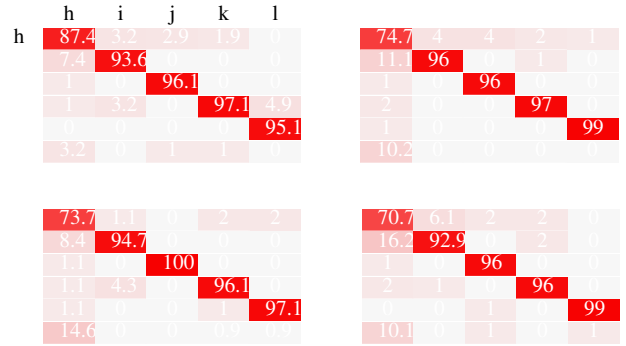


Fig. 6: Results on datasets D3 (first row) and D4 (second row) for CNN (first column) and Random Forest (second column). Label ω stands for other materials not used as labels. Column labels report the actual class, row label the predicted one.

the analyzed object. Experimental tests were run on different object sets in different days and environments in order to verify the reproducibility of the approach, and the obtained results show a good detection accuracy even on materials whose characteristics are very close. Future works will be oriented to the reconstruction of accurate object shapes.

7. REFERENCES

- [1] Shuihua Wang, Yudong Zhang, Genlin Ji, Jiquan Yang, Jianguo Wu, and Ling Wei, "Fruit classification by wavelet-entropy and feedforward neural network trained by fitness-scaled chaotic abc and biogeography-based optimization," *Entropy*, vol. 17, no. 8, pp. 5711–5728, 2015.
- [2] Douglas Fernandes Barbin, Ana Lucia de Souza Madureira Felicio, Da-Wen Sun, Suzana Lucy Nixdorf, and Elisa Yoko Hirooka, "Application of infrared spectral techniques on quality and compositional attributes of coffee: An overview," *Food Research International*, vol. 61, pp. 23 – 32, 2014.
- [3] N. Keshava, "Distance metrics and band selection in hyperspectral processing with applications to material identification and spectral libraries," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 7, pp. 1552–1565, July 2004.
- [4] S. Pesente, S. Vanini, M. Benettoni, G. Bonomi, P. Calvini, P. Checchia, E. Conti, F. Gonella, G. Nebbia, S. Squarcia,

- G. Viesti, A. Zenoni, and G. Zumerle, "First results on material identification and imaging with a large-volume muon tomography prototype," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 604, no. 3, pp. 738–746, 2009.
- [5] Richard A Albanese and Richard L Medina, "Materials identification synthetic aperture radar: progress toward a realized capability," *Inverse Problems*, vol. 29, no. 5, pp. 054001, 2013.
- [6] Hui-Shyong Yeo, Gergely Flamich, Patrick Schrempf, David Harris-Birtill, and Aaron Quigley, "Radarcats: Radar categorization for input & interaction," in *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, 2016, UIST '16, pp. 833–841.
- [7] Vayyar Imaging, "Walabot - technical brief," Apr. 24 2016.
- [8] Simone Milani and Giancarlo Calvagno, "Correction and interpolation of depth maps from structured light infrared sensors," *Image Commun.*, vol. 41, no. C, pp. 28–39, Feb. 2016.
- [9] S. Zennaro, M. Munaro, S. Milani, P. Zanuttigh, A. Bernardi, S. Ghidoni, and E. Menegatti, "Performance evaluation of the 1st and 2nd generation kinect for multimedia applications," in *Proc. of IEEE ICME*, June 2015, pp. 1–6.
- [10] Frank Weichert, Daniel Bachmann, Bartholomäus Rudak, and Denis Fisseler, "Analysis of the accuracy and robustness of the leap motion controller," *Sensors*, vol. 13, no. 5, pp. 6380–6393, 2013.
- [11] "N. Salamati, C. Fredembach, and S. Susstrunk", "Material classification using color and nir images," in *Color and Imaging Conference, 17th Color and Imaging Conference Final Program and Proceedings*, Jan. 2009, pp. 216–222.
- [12] B. Schwarz, "Lidar: Mapping the world in 3D," *Nature Photonics*, vol. 4, pp. 429–430, July 2010.
- [13] R. H. Rasshofer and K. Gresser, "Automotive radar and lidar systems for next generation driver assistance functions," *Advances in Radio Science*, vol. 3, pp. 205–209, 2005.
- [14] C. Chen, Y. Chen, Y. Han, H. Q. Lai, and K. J. R. Liu, "Achieving centimeter-accuracy indoor localization on wifi platforms: A frequency hopping approach," *IEEE Internet of Things Journal*, vol. 4, no. 1, pp. 111–121, Feb 2017.
- [15] Chitra R. Karanam and Yasamin Mostofi, "3d through-wall imaging with unmanned aerial vehicles using wifi," in *Proc. of ACM/IEEE ICIPSN*, New York, NY, USA, 2017, IPSN '17, pp. 131–142, ACM.
- [16] M. Pieraccini, M. Fratini, F. Parrini, and C. Atzeni, "Dynamic monitoring of bridges using a high-speed coherent radar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 11, pp. 3284–3288, Nov 2006.
- [17] Chia Chen Ciang, Jung-Ryul Lee, and Hyung-Joon Bang, "Structural health monitoring for a wind turbine system: a review of damage detection methods," *Measurement Science and Technology*, vol. 19, no. 12, pp. 122001, 2008.
- [18] V. C. Chen, F. Li, S. S. Ho, and H. Wechsler, "Micro-doppler effect in radar: phenomenon, model, and simulation study," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 42, no. 1, pp. 2–21, Jan 2006.
- [19] Michael Otero, "Application of a continuous wave radar for human gait recognition," *Proc. SPIE*, vol. 5809, pp. 538–548, 2005.
- [20] Tauhidur Rahman, Alexander T. Adams, Ruth Vinisha Ravichandran, Mi Zhang, Shwetak N. Patel, Julie A. Kientz, and Tanzeem Choudhury, "Dopplesleep: A contactless unobtrusive sleep sensing system using short-range doppler radar," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2015, UbiComp '15, pp. 39–50.
- [21] P. Molchanov, S. Gupta, K. Kim, and K. Pulli, "Short-range fmcw monopulse radar for hand-gesture sensing," in *2015 IEEE Radar Conference*, May 2015, pp. 1491–1496.
- [22] Qifan Pu, Sidhant Gupta, Shyamnath Gollakota, and Shwetak Patel, "Whole-home gesture recognition using wireless signals," in *Proc. of Annual ICCMCN*, New York, NY, USA, 2013, MobiCom '13, pp. 27–38, ACM.
- [23] Bryce Kellogg, Vamsi Talla, and Shyamnath Gollakota, "Bringing gesture recognition to all devices," in *Proceedings of the 11th USENIX Conference on Networked Systems Design and Implementation*, 2014, NSDI'14, pp. 303–316.
- [24] Jaime Lien, Nicholas Gillian, M. Emre Karagozler, Patrick Amihoud, Carsten Schwesig, Erik Olson, Hakim Raja, and Ivan Poupyrev, "Soli: Ubiquitous gesture sensing with millimeter wave radar," *ACM Trans. Graph.*, vol. 35, no. 4, pp. 142:1–142:19, July 2016.
- [25] R.O. Duda, P.E. Hart, and Stork D.G., *Pattern Classification, 2nd Edition*, Wiley, Nov. 2000.
- [26] Xianhua Zeng and Siwei Luo, "A supervised subspace learning algorithm: Supervised neighborhood preserving embedding," in *Advanced Data Mining and Applications*, Reda Al-hajj, Hong Gao, Jianzhong Li, Xue Li, and Osmar R. Zaiane, Eds., Berlin, Heidelberg, 2007, pp. 81–88, Springer Berlin Heidelberg.
- [27] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
- [28] David Eigen and Rob Fergus, "Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture," in *Proc. of IEEE ICCV*, 2015, pp. 2650–2658.
- [29] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *arXiv preprint arXiv:1606.00915*, 2016.
- [30] Alexey Dosovitskiy, Philipp Fischer, Eddy Ilg, Philip Hausser, Caner Hazirbas, Vladimir Golkov, Patrick van der Smagt, Daniel Cremers, and Thomas Brox, "Flownet: Learning optical flow with convolutional networks," in *Proc. of IEEE ICCV*, 2015, pp. 2758–2766.
- [31] Xavier Glorot and Yoshua Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 2010, pp. 249–256.
- [32] Andrea Vedaldi and Karel Lenc, "Matconvnet: Convolutional neural networks for matlab," in *Proceedings of the 23rd ACM international conference on Multimedia*, ACM, 2015, pp. 689–692.