Through-Wall Object Recognition and Pose Estimation

Ruoyu Wang[a], Siyuan Xiang[a] Chen Feng[a]*, Pu Wang[b], Semiha Ergan[a], Yi Fang[a]

[a]Tandon School of Engineering, New York University, Brooklyn, NY 11201, USA
[b]Mitsubishi Electric Research Labs (MERL), Cambridge, MA 02139, USA
E-mail: {ruoyuwang, siyuan, cfeng, semiha, yfang}@nyu.edu, pwang@merl.com

Abstract -

Robots need to perceive beyond lines of sight, e.g., to avoid cutting water pipes or electric wires when drilling holes on a wall. Recent off-the-shelf radio frequency (RF) imaging sensors ease the process of 3D sensing inside or through walls. Yet unlike optical images, RF images are difficult to understand by a human. Meanwhile, in practice, RF components are often subject to hardware imperfections, resulting in distorted RF images, whose quality could be far from the claimed specifications. Thus, we introduce several challenging geometric and semantic perception tasks on such signals, including object and material recognition, fine-grained property classification and pose estimation. Since detailed forward modeling of such sensors is sometimes difficult, due to hidden or inaccessible system parameters, onboard processing procedures and limited access to raw RF waveform, we tackled the above tasks by supervised machine learning. We collected a large dataset of RF images of utility objects through a mock wall as the input of our algorithm, and the corresponding optical images were taken from the other side of the wall simultaneously as the ground truth. We designed three learning algorithms based on nearest neighbors or neural networks, and report their performances on the dataset. Our experiments showed reasonable results for semantic perception tasks yet unsatisfactory results for geometric ones, calling for more efforts in this research direction.

Keywords -

Through-Wall Imaging; Object Recognition; Pose Estimation; Deep Learning

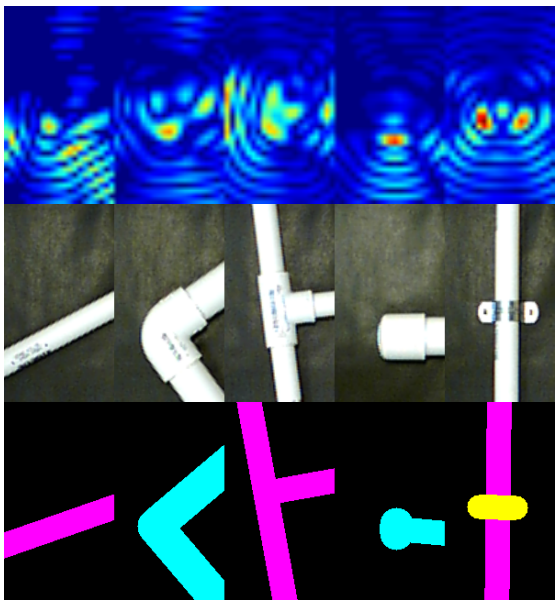


Figure 1. Overview of our problem. The first row are Walabot-returned 3D radar images (sliced at 1cm-depth for visualization) through a wooden board over the device. The second row are photos taken from the opposite side of the wooden board over Walabot. The third row are visualizations of geometric and semantic information of PVC pipe components estimated by our method. Different classification results are best viewed in color.

1 Introduction

It is often necessary to detect or even recognize the occluded objects on job sites for safety reasons. Ground penetrating radar (GPR) system has been applied for subsurface object detection. For example, GPR system can be used to survey the geological information on the construction site to predict potential construction safety hazards [1]. Signal detected by GPR can also be fused with other data. Li et al. [2] proposed a method to survey and visualize underground utilities by integrating the pre-processed GIS data with GPR signal, whose location is determined by RTK GPS.

Besides GPR, through-wall imaging techniques are widely used in fields such as fire protection, hostage rescue, flaw detection, and building construction. It is especially useful to estimate the content and structure beyond line-of-sight during maintenance and renovation. Abdel et al. [3] reviewed the state-of-the-art techniques of location and condition assessment for underground water pipelines. By employing radio frequency (RF) sensors, a through-wall imaging system can extract information through obstructions, which is impossible by using conventional line-of-

*Chen Feng is the corresponding author.

sight sensors, such as optical cameras or LIDAR.

With the development of RF sensors and signal processing algorithms, many through-wall imaging techniques have been proposed. Ahmad et al. [4] provided a method of digital beam-forming to capture 3D images behind a single uniform wall. Zhuge et al. [5] have created a real-time 3D near-field imaging algorithm with the application of Fast Fourier Transform (FFT) for 2-D multiple-input-multiple-output (MIMO) sensor array. By implementing their methods, a 3D image of object reflectivity can be formed with high accuracy and computational efficiency. In 2017, Karanam et al. [6] used WiFi router and WLAN card mounted on two UAVs to get the 3D image of an area surrounded by brick walls. These techniques are well designed in the phase of signal processing and imaging, with clearly defined forward sensor models and wave propagation model. Besides, to get higher resolution images, their antenna arrays are relatively large. In Zhuge's method, the self-designed antenna array is 0.54 m wide along both azimuth and elevation directions [5], and in Ahmad's method, their simulated antenna array can reach 2.4m [4]. Such large antenna arrays are less portable for mobile robotics applications. Also, as far as we are aware of, not much research has been done for multi-class object recognition from through-wall images.

As we mentioned before, through-wall imaging techniques benefit from increasingly accessible inexpensive off-the-shelf RF sensors. Walabot, a portable UWB-based (ultra-wideband) sensor developed by Vayyar Imaging Ltd, can provide API for users to capture 3D RF images. It has about the same size as a cell phone, drawing power and sending data through a single USB cable. Thus it can be easily incorporated in mobile robotic applications. However, due to commercial reasons, some key parameters of the Walabot sensor model are unknown to users, so detailed forward modeling of the sensor's physical process and its calibration is either impractical or cost-inefficient. Besides, the small aperture size causes low ambiguity resolution, resulting in ripple-like artifacts in the signal, as shown in the top of Figure 1. In practice, RF components are often subject to hardware imperfections (e.g., the phase noise of oscillators and mixers, nonlinear distortion from RF amplifiers, mutual coupling between antennas [7, 8]), resulting in distorted RF images, whose quality is far from the claimed specifications in user manuals. This distortion may be more severe for low-cost RF imaging systems [9]. Therefore, objects are difficult to be recognized intuitively by humans from these 3D images.

Fortunately, with the help of machine learning, some categorical and geometric information of objects could still be restored from raw 3D images, such as material, size, shape, and pose, etc. Moreover, these utility objects, such as PVC (polyvinyl chloride) pipe components, have important fine-grain attributes including thickness and contents (full of water or empty). They are all crucial for the aforementioned construction applications. The bottom of Figure 1 illustrates several results of our method based on the Convolutional Neural Network (CNN).

We highlight the contribution of our paper as follows:

- We introduce several challenging geometric and semantic tasks for better understanding and utilizing RF-based through-wall images, including materials recognition, fine-grained property classification and pose estimation.
- We propose to use machine learning methods to address the above tasks while bypassing detailed RF sensor modeling and electromagnetic wave propagation modeling that is sometimes inaccessible to end users, which showed promising results.
- We develop methods for efficiently collecting a large number and variety of through-wall images using an off-the-shelf RF sensor, with ground truth automatically generated from machine vision.
- We will make our dataset publicly available to stimulate more research in this direction.

2 Related Work

As previously mentioned, there are many research works aimed at image formation from RF signals. However, another branch of research is focused on extracting information from radar images or even raw signals. Among those, there are some works on RF object recognition and pose estimation, which is closely related to this paper.

Object Recognition. Yeo et al. [10] trained a random forest classifier for object recognition using handcrafted statistical features from raw signals. Their method has been tested for object classification, transparent material classification, and body part classification. Their method reaches over 90% accuracy on all three tasks. However, their method was not implemented with obstructions between the sensor and targets. Avrahami et al. [11] developed an activity recognition method through a cashier's counter, using a Walabot sensor, which is the same sensor used in our paper. Their method projects the 3D images sensed by Walabot to 2D plane, then extracts handcrafted statistical features for classification. By using an SVM classifier, their method reaches 90.5% accuracy. Zhao et al. [12] used RF signals reflected from a person's body to recognize one's emotion by extracting the person's heartbeats. Their method reaches comparable accuracy to on-body ECG monitors. All these above methods need handcrafted feature extractions. Wang et al. [13] proposed a method for gesture recognition using Google Soli [14], based on deep learning. However, Soli is a dedicated and customized millimeter wave radar system, which operates at a much higher frequency than Walabot, thus has a

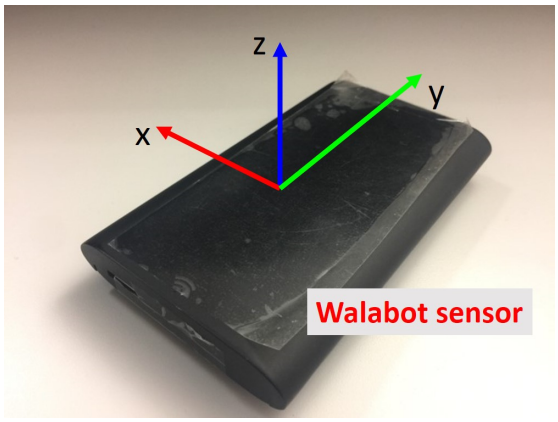


Figure 2. Walabot Coordinate System. The z axis is towards the object being detected.

higher spatial resolution. Also, there were no obstructions between the sensor and the target in their work.

Pose Estimation. Pose estimation from optical sensors has been studied for a long time [15–17]. However, pose estimation from RF sensors currently has not attracted a lot of research attention. Adib et al. has done a lot of work on pose estimation using RF sensors. In 2014, Adib et al. [18] proposed a 3D motion tracking system, WiTrack, which can track the 3D motion of a human, and estimate the direction of a pointing hand. Then they [19] presented the RF-Capture system, which can track the limbs of a human body. Most recent research on pose estimation from RF signal is from Zhao et al. [20]. In this paper, they create a deep neural network approach that estimates human 2D poses through the wall. Our work is inspired by this line of research, yet our focus is on construction robotics applications that require a different set of semantic and geometric perception tasks.

3 Method

3.1 Sensor Introduction

Walabot is an ultra-wideband MIMO array that operates at frequency 3.3-10 GHz. The model used in this research has 18 antennas. The coordinate system is defined as illustrated in Figure 2. The sensing arena used in our experiment is -5 to 5 cm in x direction, -9 to 9 cm in y direction and 1 to 10 cm in z direction. The resolution is set to 0.5 cm. In this case, the raw 3D image sensed by Walabot is a tensor of dimension $37 \times 21 \times 19$.

3.2 Dataset Design

Our dataset is designed for two tasks. The first task is material-based classification and pose estimation for linear objects. In this task, we set 5 classes of objects made from different materials. They are all common in-wall structures, including cable, wooden stud, steel pipe, PVC pipe, and background. The pose estimation task is designed only for 3 classes of straight objects: cable, wooden stud, and steel pipe. For PVC pipes, multiple shapes of pipe components are considered, and their pose estimation is designed in the second task.

The second task is more fine-grained compared to the first one, which focuses only on PVC pipes. This task contains multi-label classification and pose estimation. The PVC pipe components are classified from three aspects:

- **Thickness.** We use two types of PVC pipes in our experiment, one is 1/2" thick, the other is 1" thick. We use binary labels on these two types of pipe.
- **Content.** The PVC pipes are either full of water or empty, which is another binary classification problem.
- **Shape.** The PVC pipe components in the sensing arena are classified as 5 shapes: straight, elbow, tee, end and strap. Different shapes of objects have different ways of pose parameterization, as Figure 3 shows. It is worth to mention that, for straight components, two translational parameters x and y are enough to characterize the pose. For other types of components, two translational parameters x, y and one rotational parameter θ are used to characterize the pose. The specifications of our dataset are listed in the experimental part. For straight components, the pose is described by the coordinates of the foot of the perpendicular from the original point to the line. For elbow components, the pose is described by the coordinates of the intersecting point of the two pipes, and the angle of the angular bisector. For tee components, the pose is described by the coordinates of the intersecting point, and the angle of the pipe which constitute the vertical part of letter T. For end component, the pose is described by the coordinates of the center of the cap, and the angle of the ray away from the cap. For the strap component, the pose is described by the coordinates of the strap and the angle of the perpendicular line. The pose of the straight components also applies in task 1.

Note that we currently ignore situations when multiple objects appear simultaneously in the sensing arena, and only focus on RF images of a clean background, which is not very uncommon in practices.

3.3 Data Collection

Figure 4 shows our data collection system. The wooden board here serves as the mock wall's shell. Walabot is

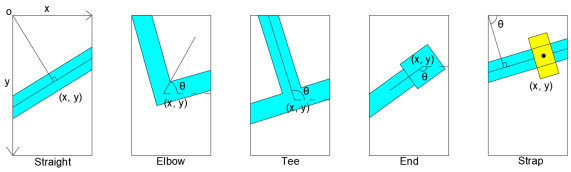


Figure 3. The definition of pose for 5 shapes of PVC pipe components.

placed right under the wooden board. Objects are placed on the upper surface of the wooden board with different poses against a black background. The camera takes photos of the objects, and synchronously, Walabot records the data it sensed from below the board. The photos are warped to the cross-section of the upper surface and the sensing arena using a homography prior to any processing. This allows us to also employ the cross-modal data annotation strategy like in [20]. But we do not need a pre-trained deep neural network to get the object poses from a camera. Instead, we developed a simple marker-based approach to track object poses from a camera, which is very robust under our controlled lab environment. This method is based on estimating the skeletons of color tapes. Figure 5 shows the procedure of pose estimation from color tapes.

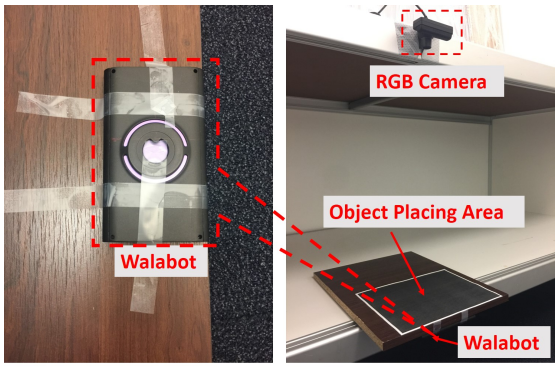


Figure 4. Cross-modal Data Collection System. RGB camera is placed above the object placing area with a black background while Walabot is placed under the object placing area (on the other side of the wooden board). Therefore, RF signals and images of the object can be collected simultaneously. The left image shows the bottom view of the Walabot in the right image.

Color Statistics. The tape is placed against the background, then the mean RGB value of pixels on the color tape and standard deviation of the distances of pixels on the color tape to the mean RGB value is calculated respectively.

Find Color Tape. For each pixel on the image taken by the camera, calculate the distance to the mean RGB value. If the distance is less than a certain threshold times the standard deviation, this pixel is classified as that color. In this case, the mask of color tape is created. Small holes on the mask are closed by morphological operations.

Figure 5. Ground truth pose generation. From left to right: Warped image, detected red mask, detected green mask, green and red tape skeletons estimation. The pose of corresponding objects can be calculated according to the skeleton of color tapes.

Skeleton Extraction. For the color tape mask image, detect line segments using LSD [21]. Merge the two longest line segments by averaging their closest end points. The merged new line segment is regarded as the skeleton of the color tape. The pose of the object is then calculated according to the new segment.

3.4 Baseline and Proposed Methods

To show the superiority of our proposed CNN-based method, we compared it with two baseline methods based on K-Nearest Neighbors (KNN) and Multi-Layer Perceptron (MLP) in all classification tasks respectively. The two baseline methods need hand-craft features, which come from Principal Component Analysis (PCA). Unlike the two baseline methods, our proposed CNN-based method does not need hand-craft features. Instead, the features are learned by the neural network.

PCA+KNN: We stretch the 37×21×19 tensor to a 14763 dimension vector. We use PCA to reduce its dimension to 5. Then a KNN classifier is applied to these 5-dimensional vectors.

PCA+MLP: We use PCA to reduce the tensor to 10 dimensions. There are 3 hidden layers in our MLP classifier with 20, 40 and 20 neurons respectively. The number of neurons in the output layer depends on the classification task.

CNN: Our CNN-based method has a share-weight structure. The input tensor is treated as a 19-channel image. After 5 interleaved 3×3 convolution layers and 2×2 max pooling layers, the tensor is reduced to a 512-dimension vector. Then this vector is sent to different sub-networks for different purposes. In task 1, the vector

is sent to a classification sub-net which predict the materials and a regression sub-net which estimate the poses of straight objects. In task 2, the vector is sent to three classification sub-nets which predict the thickness, content, and shape of the object, and a regression sub-net which estimates the poses of different PVC pipe components. The total loss function is the weighted sum of the losses of each sub-net. The total loss functions for the two tasks are formulated as following (**w** is the set of learnable network parameters):

Task 1:

$$J(\mathbf{w}) = -\frac{1}{N}\sum_{i=1}^{N}\Big(\sum_{c=1}^{5}\underbrace{y_c^{(i)}\log\hat{p}_c^{(i)}}_{\textbf{classification loss}} + C_i\lambda_0\ \underbrace{\|\mathbf{x}_i-\hat{\mathbf{x}}_i\|^2}_{\textbf{translational loss}}\Big) \tag{1}$$

Task 2:

$$\begin{aligned} J(\mathbf{w}) = -\frac{1}{N}\sum_{i=1}^{N}\Big(\sum_{l=1}^{3}\sum_{c=1}^{K_l}\underbrace{{}^{l}y_c^{(i)}\log{}^{l}\hat{p}_c^{(i)}}_{\textbf{classification loss}} \\ + \lambda_0\ \underbrace{\|\mathbf{x}_i-\hat{\mathbf{x}}_i\|^2}_{\textbf{translational loss}} + C_i\lambda_1\ \underbrace{D(\theta_i,\hat{\theta}_i)^2}_{\textbf{rotational loss}}\Big) \end{aligned} \tag{2}$$

Here, c represents different classes; N is the number of training samples; y, $\mathbf{x}$, and θ are the ground truth values for label, translational vector, and rotational angle respectively; $\hat{p}$, $\hat{\mathbf{x}}$, and $\hat{\theta}$ are the predicted values for class probability, translational vector, and rotational angle respectively, which are functions of $\mathbf{w}$. λ_0 and λ_1 are weight coefficients of the translational and rotational loss, and both of them are hyperparameters. In task 1, $C_i = 0$ when the ith sample is "no object" or PVC pipe, in other cases, $C_i = 1$. In task 2, $C_i = 0$ when the shape of the ith sample is straight, in other cases, $C_i = 1$. We use circular distance for angular loss, since θ is in [-π, π), the loss should be their shortest distance on a circle. $D(\theta,\hat{\theta}) = |\theta-\hat{\theta}|$, when $|\theta-\hat{\theta}| < \pi$. $D(\theta,\hat{\theta}) = 2\pi - |\theta-\hat{\theta}|$, when $|\theta-\hat{\theta}| >= \pi$.

4 Experiments

For both of the two tasks, our CNN based method is trained 500 epochs with batch size of 10. λ_0 and λ_1 are both set to 10. We choose Adam optimizer with learning rate 0.0001. The confusion matrix and accuracy are calculated to evaluate the performance of classification. Root Mean Square Error (RMSE) is used to characterize the accuracy of pose estimation. Compared to the two baseline methods, PCA+KNN and PCA+MLP, our CNN based method shows better performance in overall accuracy for all experiments. Furthermore, the CNN based method is more efficient than the PCA+KNN method, without storing training data for prediction.

4.1 Material Classification and Pose Estimation

Figure 6 shows the classification result of task 1. Bkg., Cb., S.P., W.S., P.P. represent background, cable, steel pipe, wooden stud and PVC pipe respectively. The overall accuracy of our CNN based method far exceeds the other two methods, especially for the classification result of Cb.. Classification accuracy on all 5 classes exceeds 95%. For pose estimation, Figure 10 shows that our CNN-based method can achieve the overall error around 1.3 cm in both x and y directions on relatively large experimental objects. The translational error of pose estimation for straight objects in our experiment is 1.8 cm. For all types of objects, steel pipes have the highest pose estimation accuracy, due to their high reflectivity to RF waves.

4.2 Thickness, Content, Shape Classification and Pose Estimation

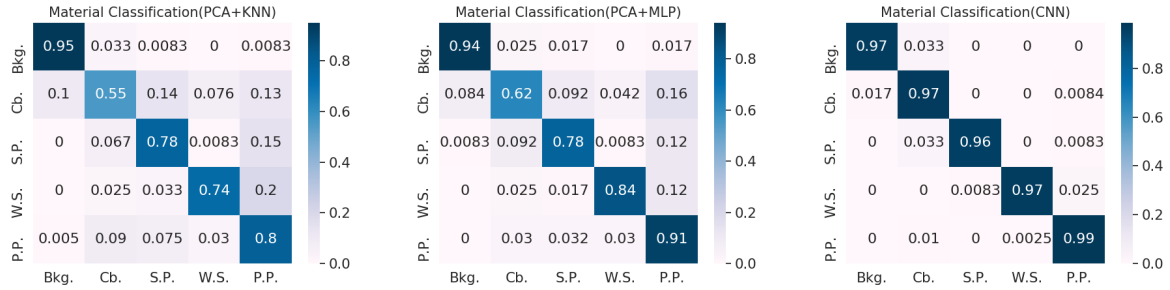
Figure 7, 8, 9 show our thickness, content and shape classification results for task 2 respectively.

In the thickness classification experiment, our CNN based method outperforms the other two methods, with an overall 0.90 accuracy, while the accuracy of the PCA+KNN method and PCA+MLP method is 0.66 and 0.75 respectively.

For content classification, our PCA+MLP baseline method has overall 0.92 accuracy, slightly lower than the CNN based method, whose overall accuracy is 0.98. The relatively high accuracy of the CNN based method showcases that our algorithm is quite suitable for content classification. Besides, the high reflectivity of water makes the signals very distinguishable between full and empty PVC pipes.

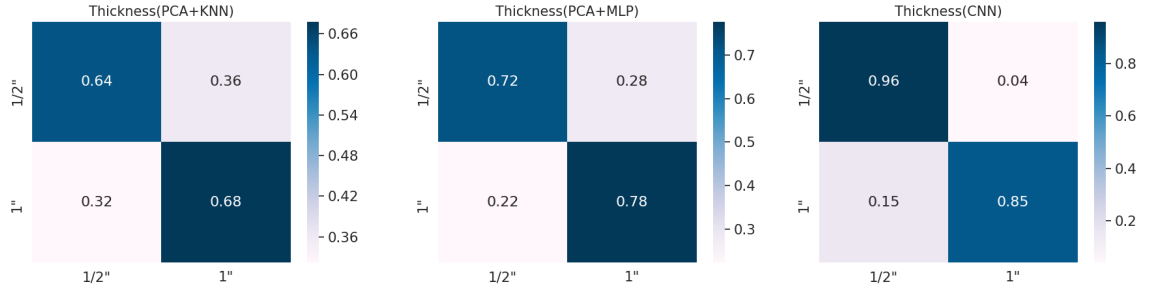
However, even using the CNN-based method, our experimental result of the overall accuracy of the shape classification is not satisfying, compared to material, content and thickness classification, with only 0.82 of accuracy overall. We think it might be caused by noise generated during the scanning process. Either pre-processing the RF signals to filter out noise or using larger training dataset might improve our experimental result.

Figure 11 shows the result of the pose estimation for the 5 shapes of PVC pipe components. Notice that straight shaped object does not need any rotational parameter to represent its pose, thus the rotational error ϵ_θ is not estimated in our experiment result. The overall error in x axis and y axis is 1.0 cm and 1.5 cm respectively. The translational error is 1.8 cm and the rotational error is 47.6 degree. Note that for all shapes of objects, error in x-axis is always smaller than the error in y-axis, which could be due to the arrangement of antennas of the sensor. Compared to translational error, rotational error is very large. The source of error could come from the noise of



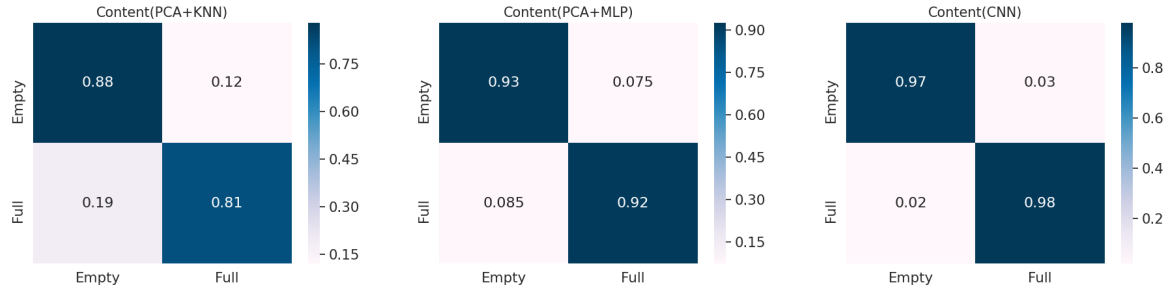
Overall Accuracy: 0.78 Overall Accuracy: 0.85 Overall Accuracy: 0.98

Figure 6. Material Classification Confusion Matrix. 120, 119, 120, 120, 402 training data for Bkg., Cb., S.P., W.S., P.P. (background, cable, steel pipe, wooden stud ,and PVC pipe), respectively.



Overall Accuracy: 0.66 Overall Accuracy: 0.75 Overall Accuracy: 0.90

Figure 7. Thickness Classification Confusion Matrix. 200 and 202 training data for 1/2" thick and 1" thick PVC pipe respectively.



Overall Accuracy: 0.85 Overall Accuracy: 0.92 Overall Accuracy: 0.98

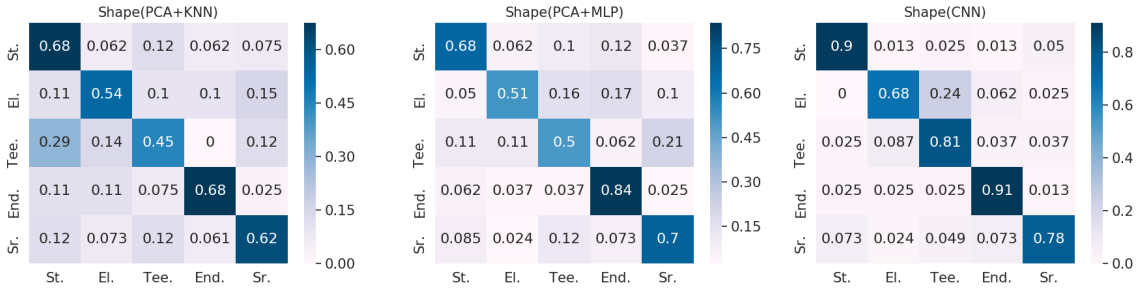
Figure 8. Content Classification Confusion Matrix. 200 and 201 training data for PVC pipes empty of water and filled with water respectively.

the signal, or the sensor's non-sensitivity to rotation.

5 Conclusions

In this paper, we explored the feasibility of using a CNN-based method for through-wall object recognition and pose estimation. With the CNN-based method, we could achieve very high performance on material, thickness, content classification for common in-wall structures. Our proposed method shows a large potential of application in construction industry.

Limitations and Discussions. However, there are some limitations in our proposed method. For the shape classification of five different PVC pipes, the overall accuracy is not satisfying, which calls for more efforts in this research direction. As discussed in the experimental part, we would



Overall Accuracy: 0.59 Overall Accuracy: 0.64 Overall Accuracy: 0.82

Figure 9. Shape Classification Confusion Matrix. 80 training data for shape Sr., End., Tee., El. respectively, and 82 training data for St..

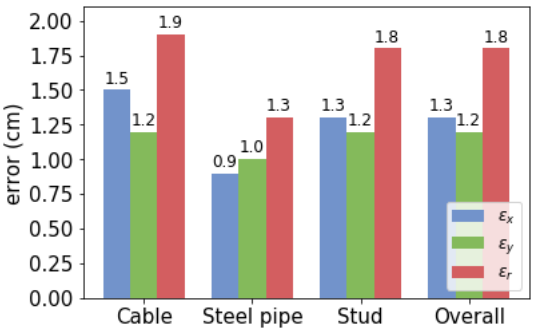


Figure 10. Pose estimation accuracy for Task 1. ϵ_x is the RMSE in x-axis, and ϵ_y is the RMSE in y-axis. Translational error $\epsilon_r = \sqrt{\epsilon_x^2 + \epsilon_y^2}$.

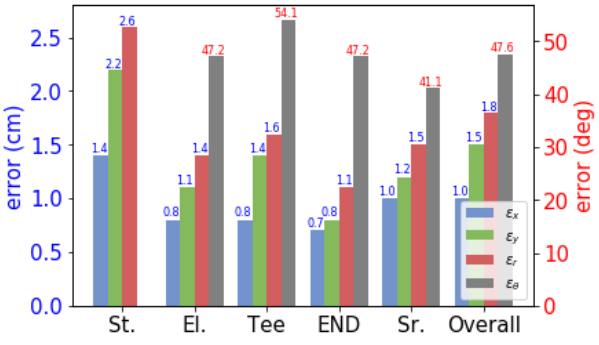


Figure 11. Pose estimation accuracy for Task 2. Translational error is defined the same as in task 1, and rotational error ϵ_θ is RMSE of θ. St., El., Tee, END, and Sr. represents straight, elbow, tee, end, and strap shape respectively.

test two possible solutions, filtering out noise or using a larger dataset, to improve the shape classification result in the future.

Future Work. One future work we are aiming towards is a useful application that can generate a detailed plumbing map behind a wall. We believe by generating such plumbing map, the designers and workers can retrieve the object information in the wall more easily.

References

[1] Xinjian Tang, Weizhong Ren, Tao Sun, and Zhicheng Zhang. Rock structures analysis from gpr radargram based on music method. In *Advanced Ground Penetrating Radar (IWAGPR), 2017 9th International Workshop on*, pages 1–4. IEEE, 2017.

[2] Shuai Li, Hubo Cai, and Vineet R Kamat. Uncertainty-aware geospatial system for mapping and visualizing underground utilities. *Automation in Construction*, 53:105–119, 2015.

[3] Mostafa Abdel-Aleem, Claude C Chibelushi, and Mansour Moniri. Multisensor data fusion for the simultaneous location and condition assessment of underground water pipelines. In *2011 International Conference on Networking, Sensing and Control*, pages 416–421. IEEE, 2011.

[4] Fauzia Ahmad, Yimin Zhang, and Moeness G Amin. Three-dimensional wideband beamforming for imaging through a single wall. *IEEE Geoscience and remote sensing letters*, 5(2):176–179, 2008.

[5] Xiaodong Zhuge and Alexander G Yarovoy. Three-dimensional near-field mimo array imaging using range migration techniques. *IEEE Transactions on Image Processing*, 21(6):3026–3033, 2012.

[6] Chitra R Karanam and Yasamin Mostofi. 3d through-wall imaging with unmanned aerial vehicles using wifi. In *Proceedings of the 16th ACM/IEEE International Conference on Information Processing in Sensor Networks*, pages 131–142. ACM, 2017.

[7] Cheng-Po Liang, Je-hong Jong, Wayne E Stark, and Jack R East. Nonlinear amplifier effects in communications systems. *IEEE Transactions on Microwave Theory and Techniques*, 47(8):1461–1466, 1999.

[8] Thomas H Lee and Ali Hajimiri. Oscillator phase noise: A tutorial. *IEEE journal of solid-state circuits*, 35(3):326–336, 2000.

[9] Emil Bjornson, Michail Matthaiou, and Mérouane Debbah. Massive mimo systems with hardware-constrained base stations. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3142–3146. IEEE, 2014.

[10] Hui-Shyong Yeo, Gergely Flamich, Patrick Schrempf, David Harris-Birtill, and Aaron Quigley. Radarcat: Radar categorization for input & interaction. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pages 833–841. ACM, 2016.

[11] Daniel Avrahami, Mitesh Patel, Yusuke Yamaura, and Sven Kratz. Below the surface: Unobtrusive activity recognition for work surfaces using rf-radar sensing. In *23rd International Conference on Intelligent User Interfaces*, pages 439–451. ACM, 2018.

[12] Mingmin Zhao, Fadel Adib, and Dina Katabi. Emotion recognition using wireless signals. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, pages 95–108. ACM, 2016.

[13] Saiwen Wang, Jie Song, Jaime Lien, Ivan Poupyrev, and Otmar Hilliges. Interacting with soli: Exploring fine-grained dynamic gesture recognition in the radio-frequency spectrum. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pages 851–860. ACM, 2016.

[14] Jaime Lien, Nicholas Gillian, M Emre Karagozler, Patrick Amihood, Carsten Schwesig, Erik Olson, Hakim Raja, and Ivan Poupyrev. Soli: Ubiquitous gesture sensing with millimeter wave radar. *ACM Transactions on Graphics (TOG)*, 35(4):142, 2016.

[15] Alexander Toshev and Christian Szegedy. Deeppose: Human pose estimation via deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1653–1660, 2014.

[16] Alejandro Newell, Kaiyu Yang, and Jia Deng. Stacked hourglass networks for human pose estimation. In *European Conference on Computer Vision*, pages 483–499. Springer, 2016.

[17] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. *arXiv preprint arXiv:1611.08050*, 2016.

[18] Fadel Adib, Zachary Kabelac, Dina Katabi, and Robert C Miller. 3d tracking via body radio reflections. In *NSDI*, volume 14, pages 317–329, 2014.

[19] Fadel Adib, Chen-Yu Hsu, Hongzi Mao, Dina Katabi, and Frédo Durand. Capturing the human figure through a wall. *ACM Transactions on Graphics (TOG)*, 34(6):219, 2015.

[20] Mingmin Zhao, Tianhong Li, Mohammad Abu Alsheikh, Yonglong Tian, Hang Zhao, Antonio Torralba, and Dina Katabi. Through-wall human pose estimation using radio signals. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[21] Rafael Grompone Von Gioi, Jeremie Jakubowicz, Jean-Michel Morel, and Gregory Randall. Lsd: A fast line segment detector with a false detection control. *IEEE transactions on pattern analysis and machine intelligence*, 32(4):722–732, 2010.