People Counting using Multistatic Passive WiFi Radar with a Multi-Input Deep Convolutional Neural Network

Chong Tang[a], Wenda Li[a], Shelly Vishwakarma[a], Fangzhan Shi[a], Simon Julier[b], and Kevin Chetty[a]

[a]Department of Security and Crime Science, University College London, London, UK
[b]Department of Computer Science, University College London, London, UK

ABSTRACT

Accurately counting numbers people is useful in many applications ranging from intelligent environments and security/law enforcement, to management of areas that experience high levels of footfall such as transport hubs and shopping malls. Currently, camera-based systems assisted by computer vision and machine learning algorithms represent the state-of-the-art. However, they have limited coverage areas and are prone to blind spots, obscuration by walls, shadowing of individuals in crowds, and rely on optimal positioning and lighting conditions. Moreover, their ability to image people raises ethical and privacy concerns. WiFi-based RF sensing systems are attracting growing attention above owing to their ability to uncooperatively detect people through solid opaque barriers, provide 360 degrees of surveillance coverage, derive target velocity/Doppler characteristics, and operate in all light levels and environmental conditions that inhibit optical sensors.

In this paper we propose a distributed multistatic passive WiFi radar (PWR) consisting of 1 reference and 3 surveillance receivers, that can accurately count up to six test subjects using Doppler frequency shifts and intensity data from measured micro-Doppler (µ-Doppler) spectrograms. To build the person-counting processing model, we employ a multi-input convolutional neural network (MI-CNN) that takes into account features such as the movement speed of the targets, and their aspect angle to the receiver. The results demonstrate a 96% classification accuracy for six subjects when data from all three surveillance channels are utilised, and significant improvements when compared with the classification accuracy from one- or two-surveillance channels.

Keywords: Multi-channel Passive WiFi Sensing, People Counting, Convolutional Neural Network, Wireless Sensing

1. INTRODUCTION

Counting the numbers of people in both in- and outdoor environments, and in a privacy-conscious manner has a wide range of possible applications. For example, to assist with the management of people flows, provide situational awareness in security & policing scenarios, acquiring visitor data in retail and entertainment venues, and ensuring room occupancy numbers are not exceeded when social distancing measures are in place. Currently, the most frequently deployed solutions for counting numbers of people are camera-based technologies that make use of computer vision and machine learning algorithms.[1–3] However, their coverage areas are limited to the cameras' field-of-view, and other factors such as blind spots, obscuration by walls and shadowing by other individuals, which all have a detrimental effect on the systems performance. Additionally, they rely on being optimally positioned and having sufficient lighting and contrast conditions. Moreover, their ability to image people raises ethical and privacy concerns. These limitations on both the performance and ability to deploy such systems have driven the emergence of wireless-based RF sensing technologies, particularly systems that

Further author information: (Send correspondence to Chong Tang, Kevin Chetty and Simon Julier)
Chong Tang: E-mail: chong.tang.18@ucl.ac.uk
Wenda Li: E-mail: wenda.li@ucl.ac.uk
Shelly Vishwakarma: E-mail: s.vishwakarma@ucl.ac.uk
Fangzhan Shi: E-mail: fangzhan.shi.17@ucl.ac.uk
Simon Julier: s.julier@ucl.ac.uk
Kevin Chetty: E-mail: k.chetty@ucl.ac.uk

exploit WiFi communication networks which are now a ubiquitous in urban environments. These RF systems offer robust and reliable solutions to the issues described above. Furthermore, they do not produce images of people which could raise ethical questions and privacy concerns.

One area of WiFi based sensing that has attracted significant attention concerns systems that make use of the WiFi channel state information (CSI). These have been examined extensively for numerous sensing tasks, including counting people.[4,5] Although studies have shown CSI based WiFi sensing can perform with relatively high accuracy, the phase information from the WiFi subcarriers is generally unusable due to unsynchronized local oscillators in a WiFi network.[6,7] This can act as a barrier to further development for example to increase the sensitivity and specificity of detecting large numbers /crowds of people.

Passive WiFi radar (PWR) offers a good alternative to the above issues. PWR has itself also been studied for various applications from through-the-wall sensing[8] and vehicle tracking[9] to gesture recognition[10] and localisation of UAV's.[11] Moreover, unlike CSI-based methods, the use of time-frequency spectrograms in the PWR not only considers the received signal strength, but also uses the phase information to reveal the micro-Doppler (µ-Doppler signatures), providing richer information relating to micro-motions associated with the target. Previously, our work[12] demonstrated the potential of using a single-channel PWR for people counting. To develop the system for deployment in more complex and challenging scenarios, and address issues around shielding as well as having only a single look-angle on the target, we propose a multistatic PWR system for counting people. Experiments were limited to small numbers of test subjects due to Covid-19 restrictions but will be increased in future work. We design a multi-input deep convolutional neural network (MI-CNN) to access data from different receivers under various conditions, which can efficiently help manage various information and extracted features. The experimental results show that the system can achieve up to 96.33% accuracy in identifying up to six people, providing significant improvement compared with a single- or dual-channel PWR system.

2. SIGNAL PROCESSING

In the IEEE 802.11 standard, WiFi transmission signals are OFDM modulated and the source signal $S_{ref}(t)$ can be modelled as:

$$S_{ref}(t) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} a_n e^{j2\pi nt} \tag{1}$$

where N is the number of OFDM symbos for each carrier a_n, and n represents the n^{th} OFDM. The signal will be reflected by objects from multiple paths, resulting in delayed and phase-shifted copies of $S_{ref}(t)$. So, the received signal $S_{sur(t)}$ can be described as the summation of these copies:

$$S_{sur}(t) = \sum_{p} A_p e^{j2\pi f_d t} x(t-\tau) + n(t) \tag{2}$$

where p is the number of reflected paths, A_p is the attenuation factor, τ and f_d represent time delay and Doppler shift for the p^{th} path, respectively. Additionally, $n(t)$ is used to describe the noise. PWR employs separately located reference and surveillance receiver channels to respectively collect $S_{ref}(t)$ and $S_{sur(t)}$, and then applies cross ambiguity function (CAF) processing to measure variations in received signal strength as well as Doppler information.

The CAF processing takes Fast Fourier Transform (FFT) of cross-correlated signals from $S_{sur(t)}$ and $S_{ref}(t)$ to extract range (from τ) and Doppler (from f_d) information, which can be described as:

$$CAF(\tau, f_d) = \int_0^T S_{sur}(t) S^*_{ref}(t-\tau) e^{j2\pi f_d t} dt \tag{3}$$

where $*$ is the complex conjugate for cross-correlating complex numbers, and T_i is the integration time that determines the Doppler resolution as: $\delta f_d = \frac{1}{T_i}$.

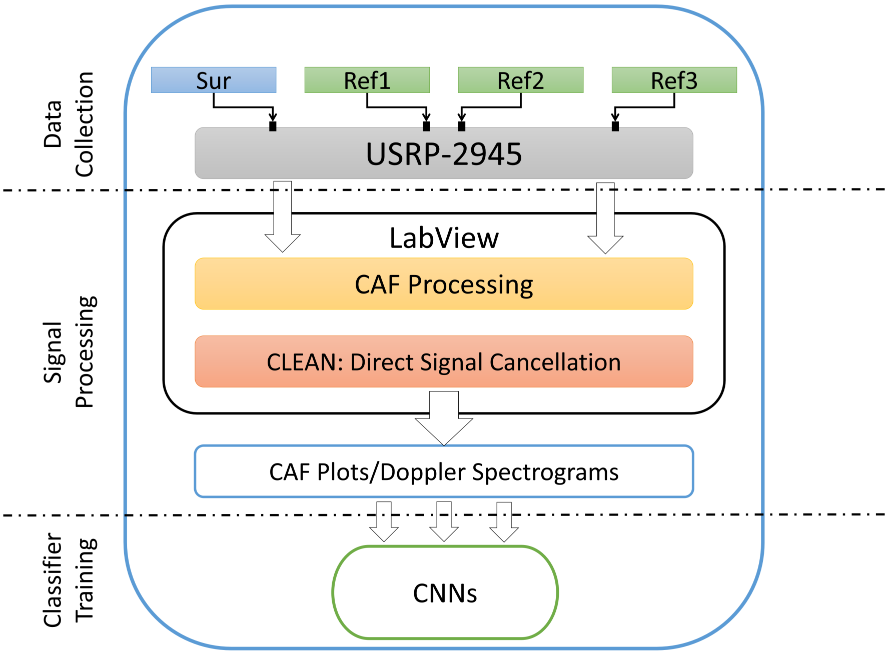

Figure 1. Block diagram of the multistatic PWR system

Furthermore, to avoid the impacts caused by the component from the direct path between the WiFi access point and the receiver. We can utilise the CLEAN algorithm to suppress the direct signal interference (DSI). The CLEAN algorithm is based on a self-cancellation mechanism, and can be described as:

$$CAF^{'}(\tau, f_d) = CAF(\tau, f_d) - \alpha CAF_{self}(\tau - T_k, f_d) \tag{4}$$

where $CAF_{self}(\tau, f_d)$ is the CAF over the reference channel and α is the maximum absolute value of $CAF(\tau, f_d)$. Finally, the cleaned CAF values can be used to generate μ-Doppler spectrograms from which features can be extracted to provide estimates of the number of people present in the area being monitored.Furthermore, our processing incorporates two additional strategies; the first is a batch processing techniques to lower the computational overhead, facilitating real-time output; the second relates to maintaining high-performance, even when only low bandwidth WiFi beacon signals are being transmitted.[10]

3. MULTI-CHANNEL PWR SYSTEM AND NEURAL NETWORK

Within a passive bistatic radar geometry, a single surveillance channel comprising of a directional antenna has a limited field of view that depends on its beamwidth. Additionally, it only has a single viewpoint of the target which may or may not be favourable for measuring µ-Doppler fluctuations i.e. a target may be side-on to the antenna shielding reflections arising from arm and leg motions on the far side of their body. To circumvent these issues we have advanced our bistatic PWR system described in[12] to a multistatic architecture (consisting of one reference and three surveillance channels) permitting the spatial distribution of the receivers around areas being monitored to provide three aspect angles on the target. Additionally, careful placement of the surveillance antennas can provide a wider coverage area permitting measurements of reflected WiFi signals from numerous people in an area. The increase in the signal processing overhead of our updated multistatic system in minimal as each surveillance channel will perform CAF processing with the reference channel independently. In this section, we briefly introduce the new multistatic PWR system and explain how our MI-CNN is employed to extract and classify features from the measured spectrograms generated from our experimental work.

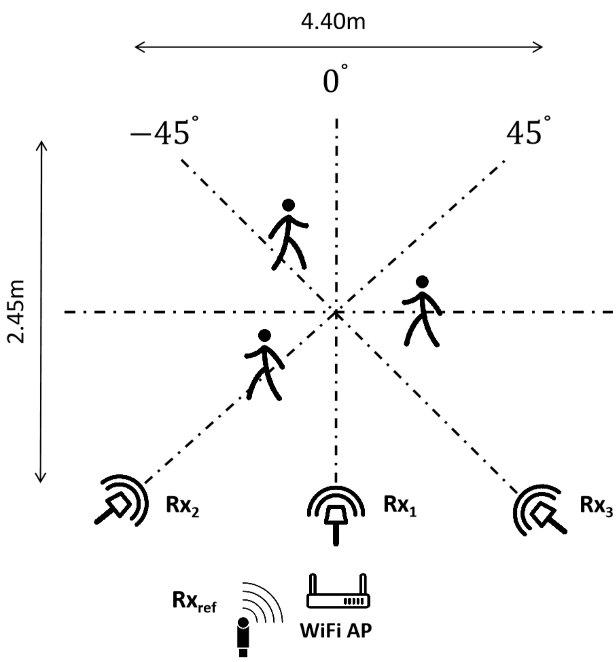

Figure 2. Multi-channel PWR people counting system

Networks	Type of Networks	Network Architecture
C1	CNN	ResNet50[13]
C2	CNN	ResNet50[13]
C3	CNN	ResNet50[13]
F	FCN	Linear(3*2048, 128), BatchNorm1d(), ReLU(), Dropout(), Linear(128, 64), BatchNorm1d(), ReLU(), Dropout(), Linear(64, 4), Softmax()

Table 1. The architecture details of the neural network

3.1 System Overview

The multistatic PWR is built around the Universal Software Radio Peripheral (USRP-2945). The block diagram of the system is illustrated in Fig. 1.

For people counting, the multistatic PWR system employs use one reference channel (Rx_{ref}) and three surveillance channels (Rx_1, Rx_2 and Rx_3) for monitoring the transmitted WiFi and reflected target signals respectively. Fig. 2 shows the experimental topology: Rx_{ref} and Rx1 are configured in a monostatic geometry while Rx_{ref} has bistatic geometry with Rx_2 and Rx_3. We define the aspect angle of Rx_{ref} as 0°, so that the aspect angles of three surveillance antennas are 0°, 45°and -45°. Combing data from these receivers can provide more detailed target characteristics from various perspectives to assist target detection and counting.

3.2 Multi-input Neural Network

To handle the recorded data from multiple receivers, we design a MI-CNN for extracting and classifying features. μ-Doppler spectrograms were generated through standard time-frequency processing and treated as one-channel

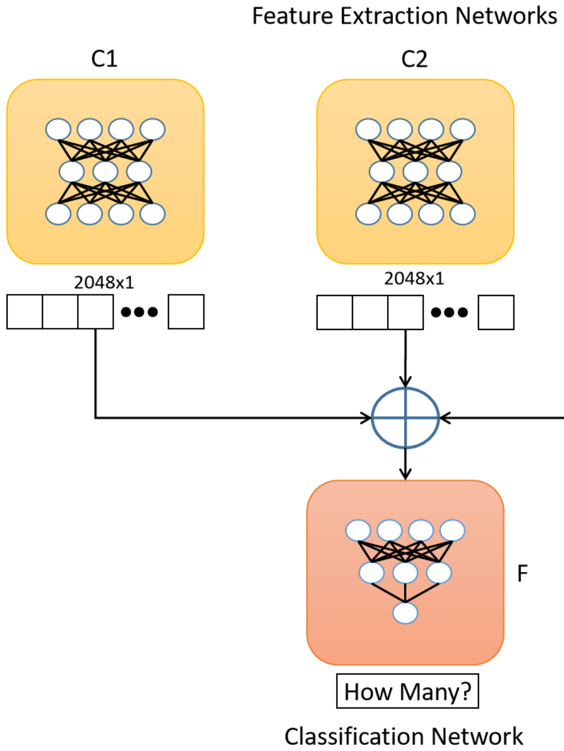

Figure 3. The architecture of the multi-input CNN

images, permitting the CNN to learn spatial and intensity patterns within the images themselves. The architecture of the network is presented in Fig. 3. We first have several individual feature extraction networks based on ResNet50[13] to receive data from different channels, where the number of the networks is same as the number of channels. Each feature extraction network outputs 2048 by 1 feature vectors, and then these vectors are concatenated together and passed into a classification fully-connected network (FCN). In Table 1, we list the detailed architecture and parameters of the CNNs. During the training phase, we use the stochastic gradient descent (SGD) optimization technique and set its learning rate to 0.001.

4. EXPERIMENTS AND EXPERIMENTAL RESULTS

The aim of the experiments was to explore the improvements in counting accuracy when increasing the number of aspects on the target, afforded through multistatic PWR. If any improvements are realised, future work will then be to carry out a more detailed quantitative analysis on the effects of the look angles themselves on the performance on the system. This section outlines the experiments designed to initially assess the people counting ability of the multistatic PWR. To that end we have performed a qualitative analysis of the results output, and report classification accuracy to validate the expected improved performance of the system.

4.1 Data Collection

The three surveillance receivers were positioned to cover the entire space of a 2.45 x 4.40m room and record continuously. Six participants were present in the room initially and were instructed to walk around continuously at a typical walking pace in a random manner, utilising the whole space. The experiment lasted 30 minutes in total, and at 5-minute intervals one participants was instructed to leave the room being monitored. Ground-truth data was recorded from a camera system and an ultrawideband (UWB) tracking system for labelling purposes. Though target localisation is not withing the scope of this work, the positional information recorded by the UWB system permitted in-depth interpretation and validation of the results. Moreover, it forms part of our larger OPERAnet study to enable future work.[14]

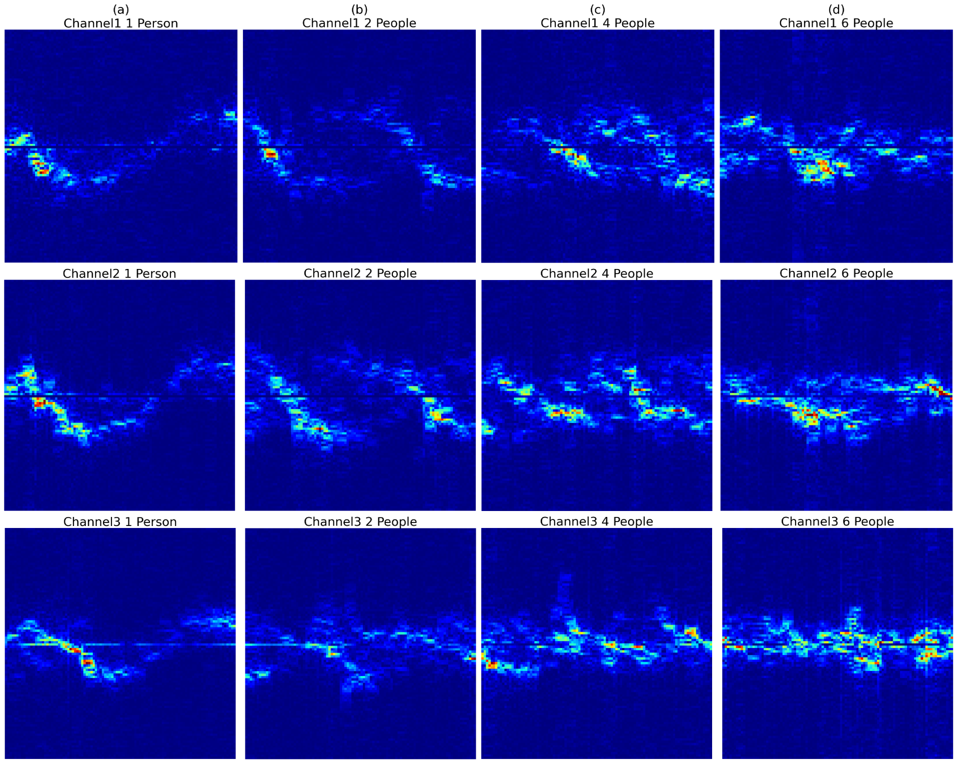

Figure 4. The qualitative results

4.2 Experimental Results

4.2.1 Qualitative Presentation

Figure 4 presents the spectrograms of four occupancy states from the three surveillance receiver channels. For all receiver channels, as the number of people present increases, it can be observed that both the complexity of the Doppler patterns and signal intensity also increase. Closer inspection of channels 1 and 2 reveal that as the room is occupied firstly by a single person, followed by another, a single sinusoidal trace is superimposed with another. Comparing spectrograms of different channels, but for the same number of people highlights marked differences in the recorded spectrograms which are attributed to the aspect angle of the receiver on the target, as well as shielding and multipath effects. This suggests that the combination of all three-receiver channel through the MI-CNN permits more comprehensive perception of the dynamic nature of people within the surveillance area to enable improved counting performance.

4.2.2 Classification Results

Post-processing analysis focused on examining the effect of using different numbers and combinations of receiver channels on classification performance. Table 2 shows the classification results achieved when processing data from the surveillance channels individually, in different pairing combinations and when using recorded data from all three. Table 2 also compares the impact of using different volumes of training data on the accuracy of the model. It clear that the using more surveillance channels provides enhanced classification accuracies of over 90%, and 96.3% in the best-case scenario where 80% of the data is used for training. The results indicate that aggregating Doppler information from different perspectives is beneficial for people counting.

5. CONCLUSION

In this paper, we propose a multistatic PWR system for the people counting. Using both qualitative and quantitative approaches to assess the Doppler spectrograms generated when people are present and in-motion

	80%	70%	60%	50%
ch1	86.24%	84.05%	80.47%	75.66%
ch2	92.66%	86.50%	81.58%	78.28%
ch3	91.74%	88.95%	83.25%	77.90%
ch1 and ch2	93.25%	91.41%	87.44%	82.39%
ch1 and ch3	94.49%	92.02%	89.30%	81.52%
ch2 and ch3	95.41%	90.80%	90.69%	86.64%
ch1, ch2 and ch3	**96.33%**	**93.25%**	**92.09%**	**90.52%**

Table 2. Classification accuracy: each row presents the results of using different channels (ch), each column presents the results of using different amounts of training data

within a room, we observe marked differences in the presence of Doppler traces and the intensity of the Doppler bins for different room occupancy conditions. Features associated with various room occupancy levels were then learnt by our MI-CNN to provide accurate estimates of the number of people present when tested in unseen conditions. The study was however limited by the number of people allowed to participate in the experiments due to Covid-19 restrictions. We hypothesize that the increased surveillance coverage and perspectives on targets when spatially distributing numerous receivers around room will address key issues associated with shielding and aspect angles on targets. Future work will therefore focus on experimentation involving large crowds of 10+ people in a room, and further evolution of our multistatic PWR to increase the number of surveillance channels available. To conclude, multistatic PWR could prove to be a valuable tool for people counting, and particularly useful if it is able to accurately estimate the presence of large number of people.

ACKNOWLEDGMENTS

This work is part of the OPERA project funded by the UK Engineering and Physical Sciences Research Council (EPSRC), Grant No: EP/R018677/1.

REFERENCES

[1] Jüngling, K., Bodensteiner, C., and Arens, M., "Person re-identification in multi-camera networks," in [*CVPR 2011 WORKSHOPS*], 55–61, IEEE (2011).

[2] Ma, L., Yang, X., and Tao, D., "Person re-identification over camera networks using multi-task distance metric learning," *IEEE transactions on image processing* **23**(8), 3656–3670 (2014).

[3] Lin, J., Ren, L., Lu, J., Feng, J., and Zhou, J., "Consistent-aware deep learning for person re-identification in a camera network," in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 5771–5780 (2017).

[4] Sobron, I., Del Ser, J., Eizmendi, I., and Vélez, M., "Device-free people counting in iot environments: New insights, results, and open challenges," *IEEE Internet of Things Journal* **5**(6), 4396–4408 (2018).

[5] An, H.-s. and Kim, S., "A deep learning based device-free indoor people counting using csi," *Journal of the Korea Institute of Information and Communication Engineering* **24**(7), 935–941 (2020).

[6] Tewes, S. and Sezgin, A., "Ws-wifi: Wired synchronization for csi extraction on cots-wifi-transceivers," *IEEE Internet of Things Journal* **8**(11), 9099–9108 (2021).

[7] Shi, F., Li, W., Amiri, A., Vishwakarma, S., Tang, C., Brennan, P., and Chetty, K., "Pi-nic: Indoor sensing using synchronized off-the-shelf wireless network interface cards and raspberry pis," IEEE (2022).

[8] Chetty, K., Smith, G. E., and Woodbridge, K., "Through-the-wall sensing of personnel using passive bistatic wifi radar at standoff distances," *IEEE Transactions on Geoscience and Remote Sensing* **50**(4), 1218–1226 (2011).

[9] Falcone, P., Colone, F., Macera, A., and Lombardo, P., "Two-dimensional location of moving targets within local areas using wifi-based multistatic passive radar," *IET Radar, Sonar & Navigation* **8**(2), 123–131 (2014).

[10] Li, W., Piechocki, R. J., Woodbridge, K., Tang, C., and Chetty, K., "Passive wifi radar for human sensing using a stand-alone access point," *IEEE Transactions on Geoscience and Remote Sensing* **59**(3), 1986–1998 (2020).

[11] Martelli, T., Murgia, F., Colone, F., Bongioanni, C., and Lombardo, P., "Detection and 3d localization of ultralight aircrafts and drones with a wifi-based passive radar," (2017).

[12] Tang, C., Li, W., Vishwakarma, S., Chetty, K., Julier, S., and Woodbridge, K., "Occupancy detection and people counting using wifi passive radar," in [*2020 IEEE Radar Conference (RadarConf20)*], 1–6, IEEE (2020).

[13] He, K., Zhang, X., Ren, S., and Sun, J., "Deep residual learning for image recognition," in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 770–778 (2016).

[14] Bocus, M. J., Li, W., Vishwakarma, S., Kou, R., Tang, C., Woodbridge, K., Craddock, I., McConville, R., Santos-Rodriguez, R., Chetty, K., et al., "Operanet: A multimodal activity recognition dataset acquired from radio frequency and vision-based sensors," *arXiv preprint arXiv:2110.04239* (2021).