ENHANCING BEAMFORMED FINGERPRINT OUTDOOR POSITIONING WITH HIERARCHICAL CONVOLUTIONAL NEURAL NETWORKS

João Gante[⋆] *Gabriel Falcão*[†] *Leonel Sousa*[⋆]

[⋆] INESC-ID, IST, Universidade de Lisboa, Portugal
[†] Instituto de Telecomunicações, University of Coimbra, Portugal

ABSTRACT

With 5G millimeter wave communications, the resulting radiation reflects on most visible objects, creating rich multipath environments. The radiation is thus significantly shaped by the obstacles it interacts with, carrying latent information regarding the relative positions of the transmitter, the obstacles, and the mobile receiver. Through a pre-estabilhed codebook of beamforming patterns transmitted by a base station, the concept of beamformed fingerprints for mobile devices' outdoor positioning has been previously proposed. In this paper, a tailored hierarchical convolutional neural network is proposed to further leverage the structure in the aforementioned hidden information. Average errors of down to 3.3 meters are obtained on a simulation environment based on realistic outdoor scenarios, containing mostly non-line-of-sight positions, making it a very competitive and promising alternative for outdoor positioning.

Index Terms— 5G, Beamforming, Deep Learning, mmWaves, Outdoor Positioning.

1. INTRODUCTION AND BACKGROUND

With mmWaves, the propagation changes dramatically: the resulting radiation has severe path loss properties and reflects on most visible obstacles [1]. To counteract the aforementioned characteristics, beamforming (BF) is usually employed in systems containing multiple-input and multiple-output (MIMO) antennas, enabling steerable and focused radiation patterns.

With that recent focus, new mmWave positioning systems were proposed [2]. The achievable accuracy in certain conditions is remarkable, having sub-meter precision in indoor [3] and ultra-dense line-of-sight (LOS) outdoor scenarios [4]. Nevertheless, in order to be broadly applicable for outdoor localization, a mmWave positioning system must also be able to accurately locate with devices in non-line-of-sight (NLOS) locations. This requirement, allied to non-linear propagation phenomena such as reflections and diffractions, poses serious challenges to any mmWave positioning method directly based on the measured time and/or distance.

1.1. Related Work

The works developed in [5–9] address this concern, attempting to locate devices in both LOS and NLOS situations. The method in [5] applies compressed sensing on information gathered from static listeners, while in [6] multiple access points are used to create a location fingerprint database of received powers and angles-of-arrival (AoA). In [7], the authors use multiple BF transmissions and an iterative algorithm to estimate the position and orientation of the device. The same parameters are obtained in [8], through the estimation of the AoA, time of arrival (ToA), and angle of departure (AoD), making simultaneous use of LOS and NLOS transmissions.

However, the methods referred so far have a hard time complying with typical outdoor situations: [5] and [6] assume that each device is always in range of multiple static transceivers, while the other two methods struggle with NLOS locations, requiring multiple transmission paths reflecting in at least three different surfaces [7] or preferring to not disclose the performance results for those locations [8]. Finally, the method proposed in [9] overcomes the aforementioned restrictions by creating a fingerprint database of uplink pilots transmitted to a distributed massive MIMO base station (BS), and then resolving the position using a Gaussian process regression, obtaining a large root-mean-square-error (RMSE) value, around 35m.

For the 5G BSs, which are expected to be positioned in elevated positions of urban scenarios, the majority of the obstacles will be buildings, and thus mainly static for a significant amount of time. As result, successive measurements of the received power delay profile (PDP) at a given position are expected to remain comparable until a significant change in the area occurs. Therefore, if a BS transmits short pulses employing a sequence of directive BF patterns, so as to cover all possible angles of transmission, each position covered by the system will likely have a unique set of PDPs. In [10], we proposed the use of that pattern, which we coined as beamformed fingerprint, as a foundation for an accurate mmWave outdoor positioning method. In this paper, we design a new hierarchical convolutional neural network (CNN [11]), tailored to the spatial properties of the positioning problem, which significantly improves the ability to decode the incoming data into a geolocation.

2. BEAMFORMED FINGERPRINT POSITIONING

A critical component of learnable fingerprint data is its consistency, as it then allows the system to extract helpful information from expected patterns. Thus, an immutable BF codebook for the transmitter must be selected, resulting in a fixed set of transmitted radiation patterns. Furthermore, to obtain equivalent fingerprints, different target devices must employ comparable detection schemes. To comply with both requirements, the system depicted in fig. 1 was originally suggested in [10]. It operates in four distinct phases, as labeled in the diagram, whose details are further described below.

Phase A: During this phase, a fixed codebook ($\mathbf{C}_{Tx}$) is considered, containing B_{Tx} directive BF patterns. Before a position estimate becomes possible, the BS must transmit a sequence of B_{Tx} pulses. Assuming a BS with N_S antennas, the frequency-domain signal for the i-th transmitter BF at a mobile device with N_R antennas, $y \in \mathbb{C}$, can be written as

Acknowledgments: This work was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with references UID/CEC/50021/2013, UID/EEA/50008/2013, and FCT Grant No. FRH/BD/103960/2014.

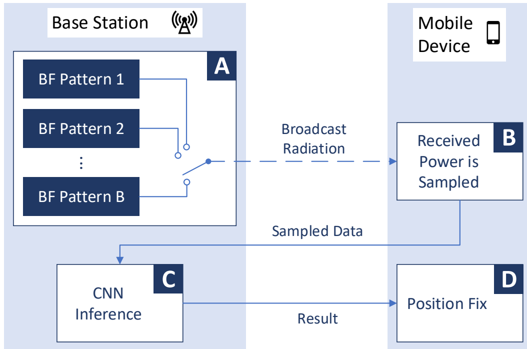

Fig. 1. Overall scheme of the system proposed in [10]. The mobile device samples the received PDPs from radiation transmitted through a fixed set of beamforming patterns, resulting in a unique beamformed fingerprint that can then be translated into its position.

$$y = \mathbf{w}^T \mathbf{H}\mathbf{f}_i x + \mathbf{w}^T \mathbf{z}, \tag{1}$$

where the superscript T denotes a matrix transpose, $\mathbf{w} \in \mathbb{C}^{N_R \times 1}$ corresponds to the (optional) beamforming at the receiver, $\mathbf{H} \in \mathbb{C}^{N_R \times N_S}$ is the channel matrix, $\mathbf{f}_i \in \mathbb{C}^{N_S \times 1}$ denotes the currently selected transmitter beamforming, $x \in \mathbb{C}$ is the waveform to be detected, and $\mathbf{z} \in \mathbb{C}^{N_R \times 1}$ represents noise. Since the transmitter beamforming is codebook-based, it is important to state that $\mathbf{f}_i \in \mathbf{C}_{Tx}$ ($\mathbf{C}_{Tx} = \{\mathbf{f}_1, \ldots, \mathbf{f}_{B_{Tx}}\}$).

As the system transmits the sequence of beamformed pulses, it is important to avoid losing information due to destructive interference. As such, after the time allocated to the measurement of the desired fingerprint data for a given pulse, a minor time interval (T_{guard}) should be considered before the transmission of the following pulse, to account for longer paths with multiple reflections.

Phase B: Given the consistent data requirement, a sampling rate must be defined, and the receivers should be synchronized with the BS transmissions. If the system is expecting beamforming at the receiver, a target array gain must also be established for all receivers. In that case, the receivers would have to define their own BF codebook, $\mathbf{C}_{Rx}$, containing B_{Rx} elements ($\mathbf{C}_{Rx} = \{\mathbf{w}_1, \ldots, \mathbf{w}_{B_{Rx}}\}$), so as to search over all angles-of-arrival (AoA) with similar gain[1]. With BF at the receiver, the device would have to sample each original transmission B_{Rx} times, storing the maximum measured value for each data point. The acquired data from the i-th transmitter BF, $\mathbf{d}_i$, can thus be written as

$$d_i[n] = \max_{j=1,\ldots,B_{Rx}} y_j(nT), \quad n = 0, 1, \ldots, N-1, \tag{2}$$

where y_j is the time-domain sampled data using the receiver beamforming $\mathbf{w}_j$, T is the sampling period, and N is the number of samples to gather per transmitter BF. It should be noted that the obtained fingerprint data ($\mathbf{d}$) has a negligible dependency on the mobile device orientation if the receiver BF codebook does cover all AoAs, since it considers the maximum value among all used receiver BFs.

High quality data can be obtained with a small T and pulse duration (< 100 ns). In such conditions, the received radiation is detected in clusters, containing voids that are large enough to be reliably detected [13]. The ability to distinguish these voids provides a meaningful shape to the resulting data, enhancing the learning capabilities of the system. As it was shown in [10], the position of the acquired non-zero samples in the data contains nearly all the extractable information. Therefore, we proposed a binary detection of the signal's existence when acquiring the data, further reducing the hardware requirements of the proposed system.

Phases C and D: After the required fingerprint data $\mathbf{d}$ is obtained, the previously trained CNN can finally infer the device position, which will the be relayed back to the mobile device. It should be noted that each deployed BS will have their own dataset and, therefore, their own CNN inference model. It should be noted that while gathering a labeled dataset for each BS might seem like a considerable effort, practical systems can always fall back to GPS in order to effortlessly label their outdoor positioning samples.

3. HIERARCHICAL CNNS

The neural network is a circuit analogous to a biological brain, comprised of a number of basic elements called neurons that are stacked in multiple fully connected layers. The vector containing the output of the l-th layer of neurons $\mathbf{n}_l$ can be written as

$$\mathbf{n}_l = a\left(\mathbf{C}_l\, \mathbf{n}_{l-1} + \mathbf{b}_l\right), \tag{3}$$

where $\mathbf{C}_l$ depicts the weight matrix, $\mathbf{b}_l$ is the bias, and a is an activation function, a non-linear differentiable function. The first layer ($\mathbf{n}_0$), also known as input layer, is fed in with the input data $\mathbf{d}$, which is a beamformed fingerprint in the context of this paper.

Due to the non-linear activation functions, a neural network is a good candidate to learn the non-linear phenomena commonly encountered in a mmWave transmission, such as reflections and diffractions. To map the input fingerprint data to the target label, the network is trained using a gradient-based algorithm which iteratively updates the neurons' learnable parameters. For the proposed system, the neural network is trained to perform a regression, minimizing the mean square error (MSE) to the data's labeled position $\mathbf{p_d}$, *i.e.*,

$$\hat{\mathbf{p}}^* = \arg\min_{\hat{\mathbf{p}}} \mathbf{E}\Big\{\left(\hat{\mathbf{p}} - \mathbf{p_d}\right)^{\mathbf{T}}\left(\hat{\mathbf{p}} - \mathbf{p_d}\right)\Big\}, \tag{4}$$

where $\hat{\mathbf{p}}^*$, which is the output of the neural network's last layer, denotes the estimated position given $\mathbf{d}$. The usage of this loss function can be interpreted as a minimization of the euclidean distance between the labeled position and its estimate. After being trained, the network is able to provide estimates for new, unseen data.

Consider now the two indexing dimensions of the beamformed fingerprint data samples, the time-domain sample number and the transmitter BF index. If the sequence of BF indices corresponds to a continuous sweep over the azimuth, it is possible to extract information not only from the individual data points, but also from their *sequence* along those two dimensions. With CNNs, the convolutional layer is introduced, where the neural network can learn the most effective set of short filters to apply on the received data, and thus also extracting information from their sequence. With convolutional layers, more than one feature can be learned from the previous layer's output, which can be seen as a higher-order abstraction. For the l-th convolutional layer of neurons $\mathbf{N}$, which is now a matrix, the output of the f-th feature can be written as

$$\mathbf{N}_l^f = a\left(\sum_{d=1}^{D}\left(\mathbf{C}_l^{f,d}\, \mathbf{N}_{l-1}^d\right) + \mathbf{1} \times b_l^f\right), \tag{5}$$

where D is the number of features in the previous layer, $\mathbf{1}$ is a bi-dimensional matrix of ones, the bias b_l^f is now a single scalar, and each $\mathbf{C}_l^{f,d}$, now denoting a bi-dimensional filter, is a doubly block

[1]It should be noted that even though different mobile devices would likely have different implementations of the receiver BF codebook, due to dissimilar antenna arrays, a CNN trained with perturbations is usually resilient to them [12]. If the ecosystem ends up containing particularly diverse devices, it would be possible to train a separate CNN for each type of devices.

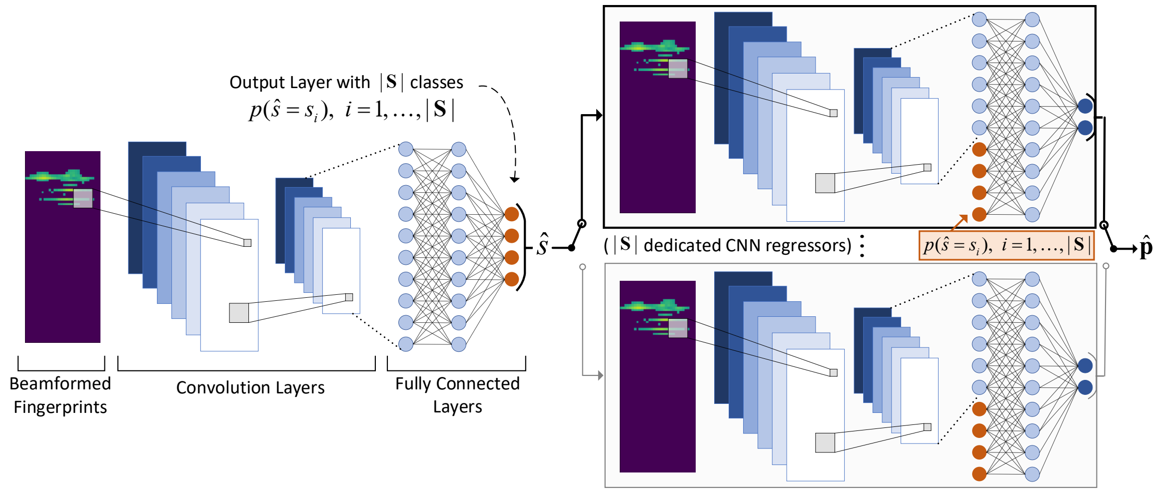

Fig. 2. Overview of the proposed hierarchical model, where the total area can be divided into $|\mathbf{S}|$ sub-areas, each with a dedicated CNN regressor. A CNN classifier is used to select the most suitable regressor $\hat{s}$, which in turn yields the position estimate $\hat{\mathbf{p}}$. To enhance the regressor precision, it is also fed with the output layer of the CNN classifier, which can be seen as a coarse estimate.

circulant matrix (which is a special case of a Toeplitz matrix). In this case, the input layer ($\mathbf{N}_0$) is fed in with the beamformed fingerprint data $\mathbf{d}$. Due to its new structure, if $\mathbf{C}_l^{f,d}$ represents a filter with $L1$ rows and $L2$ columns, it only contains $L1 \times L2$ learnable parameters. Although there is a different learnable filter for each pair of features on two subsequent convolutional layers, the number of learnable parameters in a convolutional layer is significantly lower than in a fully connected layer, for equally performing neural networks [11].

To further refine the learning mechanism, a hierarchical model inspired by the work in [14] is proposed, as depicted in fig. 2. For each BS, the covered area can be seen as a set of sub-areas $\mathbf{S}$ ($\mathbf{S} = \{s_1, \ldots, s_{|\mathbf{S}|}\}$). If each sub-area contains a dedicated CNN, those $|\mathbf{S}|$ CNNs can specialize on their own data partition. As adjacent positions are very likely to contain similar data patterns, each dedicated CNN will have fewer types of patterns to learn, thus facilitating the learning process.

The sub-areas can be seen as coarse positions and, as result, identifying the sub-area s of a new data sample should be an accessible task. Thus, a separate CNN classifier is used to predict the most likely $\hat{s}$, indicating which dedicated CNN should be used to estimate the device location. As mentioned, the predicted $\hat{s}$ can be seen as a coarse position estimate and, therefore, the selected CNN regressor is also fed in with the output layer of the CNN classifier, so as to enhance its precision. To train the classifier, the cross-entropy between prediction and ground truth is minimized, such as

$$p(\hat{\mathbf{s}}) = \underset{p(\hat{s}_i),\ i=1,\ldots,|\mathbf{S}|}{\arg\min} \mathbf{E}\Big\{ - \sum_i p(s_i|\mathbf{d}) \log(p(\hat{s}_i)) \Big\}, \quad (6)$$

where $p(\hat{\mathbf{s}})$ denotes the output vector of the classifier neural network, containing the predicted probabilities $p(\hat{s} = s_i)$ for a certain input data $\mathbf{d}$. After obtaining the classifier's output, the most suitable dedicated CNN $\hat{s}$ is selected by determining

$$\hat{s} = \underset{i=1,\ldots,|\mathbf{S}|}{\arg\max}\ p(\hat{s} = s_i). \quad (7)$$

As $|\mathbf{S}|$ grows, a trade-off is expected: the specialized CNNs will have a smaller area to cover, while the accuracy for the CNN classifier decreases. Since the dedicated CNNs map their predictions to the complete area, they might be able to recover from previous classification errors, as long as it is a recurrent (and thus learnable) mistake. On the other hand, non-recurrent misclassifications have a significant penalty on the system, especially when training, where a misclassified sample is tied to the training set for $\hat{s}$ (with $\hat{s} \neq s_i$). This can be seen as simultaneously adding noise to the training set for $\hat{s}$, while depriving s_i of meaningful samples. The results in [14] also reflect this trade-off, with hierarchical models outperforming traditional CNNs unless there are too many data partitions.

The application of the hierarchical model is completely transparent to the mobile device. For the BS, while it might require some additional computing resources, it is scalable in respect to the number of area partitions: the predictions always require the same number of operations (one coarse classification and one fine-grained regression), and the training time barely changes (it depends mainly on the total number of training samples).

4. SIMULATIONS AND EXPERIMENTAL RESULTS

4.1. Evaluation Apparatus

To evaluate the proposed system accuracy, a dataset using mmWave ray-tracing simulations in the New York University (NYU) area is used, containing fingerprint data from 160801 different positions. The propagation specifications in Table 1 were inherited from the experimental measurements in [13] and, in [16], it was shown that these ray-tracing simulations matched the aforementioned experimental measurements.

Table 1. Simulation Parameters

Parameter Name	Value
Carrier Frequency	28 GHz
Transmit Power	45 dBm
Codebook Size	32 (155° arc with 5° between entries)
Receiver Grid Size	160801 (400 × 400 m, 1 m between Rx, 1 m above the ground)
Convolutional Layers	1 layer (8 features with 3 × 3 filters, 2 × 1 max-pooling)
Hidden Layers	12 (256 neurons each)
Class. Output	Softmax with $|\mathbf{S}|$ classes
Regression Ouput	2 Linear Neurons (2D position)
Samples per Tx. BF	82 (4.1 μs @ 20 MHz)
Assumed Rx. Gain	10 dBi (as in [15])
Detection Threshold	−100 dBm

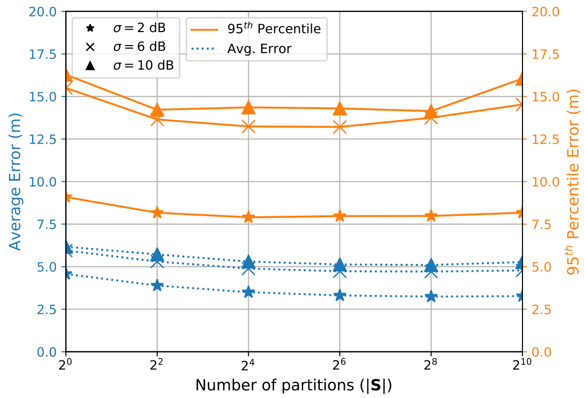

Fig. 3. Average and 95^{th} percentile prediction errors for multiple number of partitions and noise levels (σ). While it is a tool to extract additional accuracy, an excessive number of partitions has adverse consequences.

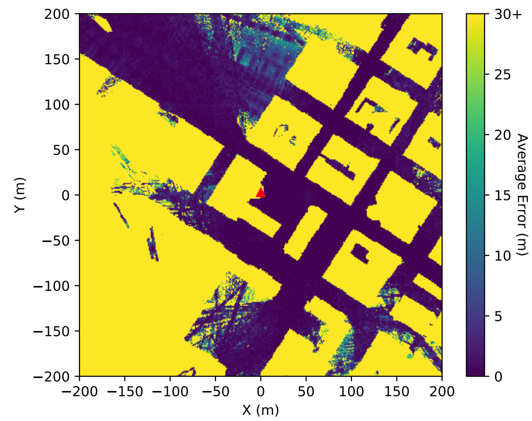

Fig. 4. Average error per covered position, assuming $|\mathbf{S}| = 64$ and a noise σ of 6 dB. Given that the transmitter is at the center of the image (red triangle), it is possible to confirm that being in a NLOS position is not a constraint for the proposed system.

While the used ray-tracing software (Wireless InSite 3.0.0.1 [17]) was unable to control BF patterns, a physically rotating horn antenna was used, producing similar directive radiation patterns. The received power data was sampled at 20 MHz over a spawn of 4.1 μs, for each of the 32 elements in $\mathbf{C}_{Tx}$. Regarding BF at the receiver, a 10 dBi gain was considered (akin to [15]). When the area is split for the hierarchical model, only powers of 4 partitions are considered, where each physical dimension is subsequently bisected (e.g. when 64 partitions are considered, each dimension is bisected 8 times, resulting in partitions with 50×50 m^2).

In the proposed system, noise is added to the obtained ray-tracing data following a log-normal distribution (also known as *slow fading*). The noise was introduced before applying a detection threshold of -100 dBm, and the data is binarized after adding the noise and applying the detection threshold. For each CNN training epoch, a new training set is generated, consisting of the original ray-tracing dataset entries with added random noise. Since the system is expected to be used to predict physical positions for which it already has training samples, the test set is also generated from noisy samples of the ray-tracing data. The higher the noise level is, the more different the training and test sets are expected to be. When evaluating the average results, a total of 10 test sets are used.

The selected CNNs follow a typical architecture, whose hyperparameters (depicted in Table 1) were selected after the empirical testing[2] of a random hyperparameter search [18]. The classification and the $|\mathbf{S}|$ regression CNNs share the same configuration and hyperparameters, except for the input of the first fully connected layer and the output layer (see fig. 2 and Table 1). While sub-optimal, they are all trying to extract similar information given similar data, and thus using a single hyperparameter set yields satisfying results while alleviating the search complexity.

4.2. Simulation Results and Discussion

The first assessed parameter is the number of data partitions ($|\mathbf{S}|$), as shown in fig. 3. It is interesting to notice that the predictions for $|\mathbf{S}| > 64$ keep roughly the same average error, at the expense of an increased 95^{th} percentile error. This means that although more specialized regressors yield improved predictions on correctly classified samples, the higher number of misclassified samples reverts those gains, as discussed in Section 3. Considering a partion-less dataset (*i.e.*, $|\mathbf{S}| = 1$), the average error ranges from 4.57 m to 6.17 m, for low and high noise values, respectively, with a 95^{th} percentile error never exceeding 16.3 m. The best results were obtained when $|\mathbf{S}| = 64$, with an average error ranging from 3.31 m to 5.13 m and a 95^{th} percentile error never exceeding 14.3 m. For a moderate noise level of 6 dB, when those results are compared to the previous work in [10], the average and 95^{th} percentile errors were reduced in 55% and 58%, respectively. It is important to clarify that the selected partitions (subsequent bisections of the considered area) are very likely to be sub-optimal. Nevertheless, they demonstrate the applicability of hierarchical data to the considered problem, achieving performance gains with minimal effort.

In fig. 4, the average prediction error per position is show, considering $\sigma = 6$ dB and $|\mathbf{S}| = 64$. Given that most of the lower left corner has no mmWave coverage in the ray-tracing simulations, and that the solid yellow figures throughout the figure are buildings, it is possible to conclude that the system is always able to return a positioning prediction as long as the mobile user is covered by the mmWave network, with or without line-of-sight. For this configuration, the predictions have an RMSE of 19.7 m, which denotes quite superior performance in all aspects when compared to [9], whose simulations obtained an RMSE of 35 m. Moreover, it is important to point out that [9] considers a lower noise level, with $\sigma = 5$ dB (we used 6 dB in our experiments), and its numerical simulations do not contain NLOS positions, as we do.

5. CONCLUSIONS

Throughout this paper, a system that is able to provide outdoor positioning through the transmission of mmWave beamformed fingerprints is further enhanced with hierarchical CNNs. With the inclusion of hierarchical CNNs, the received PDP similarity between adjacent positions is exploited, reducing the average error by 55% when compared to our previous work, while bringing no changes to the device-side part of the system. By providing accurate estimates even for NLOS positions, where other sub-meter-accuracy mmWave positioning algorithms struggle, the proposed system can be seen not only as an enabling component of the mmWave positioning techniques ecosystem, but also as a potential alternative to GPS systems when 5G networks become available.

[2]The simulation code and used data are available at https://github.com/gante/mmWave-localization-learning

6. REFERENCES

[1] T. S. Rappaport, R. W. Heath, R. C. Daniels, and J. N. Murdock. *Millimeter Wave Wireless Communications*. Prentice Hall, 2014.

[2] J. A. del Peral-Rosado, R. Raulefs, J. A. López-Salcedo, and G. Seco-Granados. Survey of cellular mobile radio localization methods: from 1G to 5G. *IEEE Communications Surveys Tutorials*, 2017.

[3] K. Witrisal, P. Meissner, E. Leitinger, Y. Shen, C. Gustafson, F. Tufvesson, K. Haneda, D. Dardari, A. F. Molisch, A. Conti, and M. Z. Win. High-accuracy localization for assisted living: 5G systems will turn multipath channels from foe to friend. *IEEE Signal Processing Magazine*, 33(2), March 2016.

[4] M. Koivisto, A. Hakkarainen, M. Costa, P. Kela, K. Leppanen, and M. Valkama. High-efficiency device positioning and location-aware communications in dense 5G networks. *IEEE Communications Magazine*, 55(8), 2017.

[5] Xueyan Han, Jingjing Wang, Wei Shi, Qiuna Niu, and Lingwei Xu. Indoor precise positioning algorithm using 60GHz pulse based on compressive sensing. *Journal of Mathematics and Computer Science*, 2016.

[6] Z. Wei, Y. Zhao, X. Liu, and Z. Feng. DoA-LF: A location fingerprint positioning algorithm with millimeter-wave. *IEEE Access*, 5, 2017.

[7] A. Shahmansoori, G. E. Garcia, G. Destino, G. Seco-Granados, and H. Wymeersch. Position and orientation estimation through millimeter-wave MIMO in 5G systems. *IEEE Transactions on Wireless Communications*, 17(3):1822–1835, March 2018.

[8] Z. Abu-Shaban, X. Zhou, T. Abhayapala, G. Seco-Granados, and H. Wymeersch. Error bounds for uplink and downlink 3d localization in 5g mmwave systems. *IEEE Transactions on Wireless Communications*, pages 1–1, 2018.

[9] V. Savic and E. G. Larsson. Fingerprinting-based positioning in distributed massive MIMO systems. In *2015 IEEE 82nd Vehicular Technology Conference (VTC2015-Fall)*, Sept 2015.

[10] J. Gante, G. Falcao, and L. Sousa. Beamformed Fingerprint Learning for Accurate Millimeter Wave Positioning. *IEEE 88th Vehicular Technology Conference (VTC Fall)*, August 2018.

[11] Yoshua Bengio, Yann LeCun, and Geoffrey Hinton. Deep learning. *Nature*, 521:436–444, 2015.

[12] Kartik Audhkhasi, Osonde Osoba, and Bart Kosko. Noise-enhanced convolutional neural networks. *Neural Networks*, 78:15 – 23, 2016. Special Issue on Neural Network Learning in Big Data.

[13] Y. Azar, G. N. Wong, K. Wang, R. Mayzus, J. K. Schulz, H. Zhao, F. Gutierrez, D. Hwang, and T. S. Rappaport. 28 GHz propagation measurements for outdoor cellular communications using steerable beam antennas in New York city. In *2013 IEEE International Conference on Communications (ICC)*, pages 5143–5147, June 2013.

[14] Z. Yan, H. Zhang, R. Piramuthu, V. Jagadeesh, D. DeCoste, W. Di, and Y. Yu. HD-CNN: Hierarchical deep convolutional neural networks for large scale visual recognition. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 2740–2748, Dec 2015.

[15] T. Obara et al. Experiment of 28 GHz band 5G super wideband transmission using beamforming and beam tracking in high mobility environment. In *2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Sept 2016.

[16] J. Gante, G. Falcao, and L. Sousa. Data-aided fast beamforming selection for 5G. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2018.

[17] Wireless InSite web-page. `https://www.remcom.com/wireless-insite-em-propagation-software/`. Accessed: 9th of April, 2018.

[18] James Bergstra and Yoshua Bengio. Random search for hyper-parameter optimization. *J. Mach. Learn. Res.*, 13:281–305, February 2012.