IEEE ComSoc IEEE Communications Society | IEEE COMPUTER SOCIETY | IEEE Signal Processing Society | VTS Connecting the Mobile World | IEEE Transactions on Machine Learning in Communications and Networking

Received 06 October, 2022; revised 02 February, 2022; accepted 07 March, 2023; Date of publication XX Month, XXXX; date of current version 09 March, 2023.

Digital Object Identifier 10.1109/TMLCN.2023.1234567

Indoor Localization with Robust Global Channel Charting: A Time-Distance-Based Approach

Maximilian Stahlke[1,2], George Yammine[1], Tobias Feigl[1], Bjoern M. Eskofier[2], and Christopher Mutschler[1]

[1]Fraunhofer IIS, Fraunhofer Institute for Integrated Circuits IIS, Division Positioning and Networks, 90411 Nuremberg, Germany
[2]Department Artificial Intelligence in Biomedical Engineering (AIBE), Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), 91052 Erlangen, Germany

Corresponding author: Maximilian Stahlke (email: maximilian.stahlke@iis.fraunhofer.de).

This work was supported by the Fraunhofer Lighthouse project "6G SENTINEL" and by the Federal Ministry of Education and Research of Germany in the programme of "Souverän. Digital. Vernetzt." joint project 6G-RIC (16KISK020K).

ABSTRACT Fingerprinting (FP) significantly improves the indoor localization performance in non-line-of-sight-dominated areas. However, its deployment and maintenance is cost-intensive as it needs ground-truth reference systems for both the initial training and the adaption to environmental changes. Recently, channel charting (CC) has been investigated because it is an unsupervised method that does not need labeled measurement data. CC uses pairwise-distance metrics to estimate the physical distance between channel-state-information (CSI) measurements in the proximity. CC can then learn the underlying manifold of the channel measurements. While CC has shown promising results in modelling the local geometry of the radio environment, a deeper insight into CC for localization using multi-anchor large-bandwidth measurements is still pending. We therefore contribute a novel distance metric for CC that approaches a global linear correlation to the physical distance based on time-synchronized single-input/single-output (SISO) CSIs. This allows to learn the environment's global geometry in a channel chart without annotations. We leverage a Siamese network, which enables CC-assisted FP only using a linear transformation from the chart to the real-world coordinates. We compare our approach to the state-of-the-art of CC on two different real-world data sets recorded with a 5G and UWB radio setup. Unlike FP-based localization which needs a large number of labeled data points to achieve a superior localization accuracy, our approach outperforms FP when only a few labeled data samples are available.

INDEX TERMS Channel charting, machine learning, fingerprinting, localization, 5G, UWB, time-based measurements.

I. INTRODUCTION

INDOOR localization techniques are a key enabler for several downstream tasks in health care, industrial production or networking [1]. Although several localization techniques based on camera images [2], lidar sensors [3] or visible light communication [4] already exist, radio-based localization is still one of the most promising technologies [5]. Approaches such as angle-of-arrival (AoA) [6], [7] or time-of-arrival (ToA) [8] localization can achieve accuracies in the centimeter range given a line-of-sight (LoS) path between transmitters and receivers. However, this requirement can often not be met due to the complexity of indoor environments. While there are methods that can mitigate the non-line-of-sight (NLoS) effects [9] by identification and exclusion [10] or error estimation [11], a majority of radio units still need LoS for localization. To anyway achieve highly accurate localization in NLoS-dominated areas, FP methods can be used. FP exploits the CSI collected in a radio environment. The CSI contains location-specific radio information caused by reflections, scattering and absorptions [12]–[18]. To train a FP model, CSI measurements have to be labeled using a ground-truth reference system, which is often very expensive, while another common problem of FP is that environmental changes can alter the location-specific fingerprints, making

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

regular updates inevitable, which again requires labeled data [13], [16].

Recently, a new concept called CC has been proposed. CC generates a chart on collected CSI measurements to reflect the (local) geometry of the environment in an unsupervised manner, i.e., no ground-truth labels are needed [19]. CC exploits the fact that many sensor modalities lie on a manifold, which allows to model the physical geometry of the environment for localization [20], [21]. CC focuses on CSI measurements that have a clear spatial correlation as radio signals follow the electromagnetic wave propagation model, which is a smooth function over space. Environmental influences like absorption and reflection do not disrupt this smoothness. What is observed at the receiver is a sum of the different impinging waves which again results in a smooth function over space. This resulting function, which gradually changes in space, then lies on a differentiable manifold that can be exploited for channel charting.

Most channel-charting approaches only exploit the local distance in the manifold, which lead to charts that does not represent the global coordinates of the environment. Instead, it resembles only the relative geometry of the radio environment. To use CC for positioning, some semi-supervised approaches [22]–[25] use few labeled data points to unroll the manifold into the physical domain and enhance its consistency. It is mainly used to leverage several downstream tasks for multiple-input/multiple-output (MIMO) communications, which do not necessarily need accurate global coordinates. They improve tasks like pilot assignment [26], UE grouping [27], radio resource management [28] or beam forming [29]–[31].

However, all of the existing work on CC is based solely on MIMO systems with either single or multiple unsynchronized base stations with mostly less than 50 MHz bandwidth. Due to the low bandwidth and the lack of synchronization, the networks are less suited for localization and are rather optimized for communication. This renders current CC approaches inappropriate for high-precision positioning tasks.

Thus, this paper proposed the following contributions. First, we investigate the abilities of CC for high-precision indoor localization employing time-synchronized SISO radio systems with large bandwidths. Second, we contribute a novel geodesic CSI distance metric and its mathematical derivation that provides a linear correlation to the physical distance. This allows to unroll the manifold and represent the global structure of the CSI data in the physical space, which is necessary for localization. Third, we introduce a positioning framework, which utilizes a Siamese network to estimate a channel chart with global consistence, that only needs a linear transformation for positioning. Such parametric models allow to efficiently predict on unseen data, which is necessary for positioning. Fourth, we compare all the relevant methods for CC with ours on two real-world data sets using UWB and 5G radio systems. We also show that supervised FP needs 50 times more labeled data points to achieve similar results as CC-assisted FP, rendering CC-assisted FP a competitive candidate for indoor localization.

The remainder of this article is structured as follows. Section II discusses related work. Next, Section III provides details about local CIR distances and global geodesic distances. Section IV describes how we approximate them using a Siamese neural network, and how we preprocess the data. Section V describes our experimental setup. The numerical results are presented in Section VI and discussed in Section VII. Section VIII concludes.

II. RELATED WORK

Studer et al. [19] first described the concept of CC for multi-antenna systems, e.g., MIMO architectures, to embed the local radio geometry. As channel charts only yield poor geometries using a single base station (BS), Deng et al. [32] proposed a multi-point CC approach that fuses the channel information of multiple BSs to achieve a better spatial consistency of the channel charts. CC typically consists of two components, (i) a distance metric, which reflects the physical distance between two channel measurements and (ii) the dimensionality reduction, which encodes the high dimensional channel information into a 2D embedding reflecting the position.

As a distance metric, Studer et al. [19] propose to use the free-space path loss between two measurements, which is proportional to the physical distance if the translation is collinear and the path-loss coefficient is constant. Moreover, Magoarou et al. [33] enhanced the proposed metric by making it insensitive to fast fading effects by putting the channel measurements into phase. As the distance metric is only valid for collinear measurements, Agostini et al. [34] proposed a grouping of collinear measurements by exploiting the angular information encoded in the CSI measurements of fully digital MIMO setups. To achieve channel charts with higher quality, in fully digital MIMO setups, multipath components (MPCs) can be extracted using MUltiple Signal Identification Classification (MUSIC) [35], which allows to cluster MPCs and apply the path-loss model for every component providing an improvement of the original metric. However, MUSIC needs a known number of MPCs, which is often difficult to obtain especially under low signal-to-noise ratio (SNR) situations [36]. Moreover, a data association of MPCs after extraction is necessary, which is non-trivial especially in complex environments with dense multipath propagation with varying MPCs.

For indoor localization, multiple (synchronized) SISO radio setups are usually preferred in order to fully cover the environment. MIMO-capable base stations, in contrast, are more expensive than their SISO counterparts, due to their higher hardware complexity mostly useful for communication tasks. Compared to MIMO setups, SISO base stations have a simple hardware concept as they have only one receiving antenna but therefore cannot estimate any AoA information for positioning. Therefore, ToF or TDoA estimations are typ-

ically used for positioning [8]. However, this also means that for SISO setups, the angular-based distance metrics presented in the literature cannot be used. We therefore contribute a novel distance metric based on raw CSI measurements in the time-domain exploiting ToF and TDoA information dedicated for low cost SISO positioning systems.

The second component of CC is its dimensionality reduction, which can be accomplished using both parametric approaches and non-parametric approaches. Studer et al. [19] investigated non-parametric approaches such as principal component analysis (PCA) or Sammon mapping to CC. Ponnada et al. [37] proposed an approach based on Laplacian-Eigenmaps with an extension to prediction also on unseen data, which is the main intention of CC. This is also done implicitly by mapping the high-dimensional channel information to 2D embeddings using parametric deep learning (DL) approaches. Methods such as autoencoders [38], constrained autoencoders [39], Siamese networks [22], and triplets [23], [40]–[43] have shown very promising results in modelling the local geometry of the channel information. Also, a combined algorithm is proposed, which first compresses the CSI using a convolutional autoencoder to efficiently apply uniform manifold approximation and projection (UMAP) on the lower-dimensional representation of the CSI [44]. While CC only models the local geometry of the area, there are also semi-supervised approaches that learn the mapping from the local to the global coordinate frame. This can either be achieved by an affine transformation after optimizing the local map [45], or as a constraint in the optimization loss [22]–[25]. While the objective of the aforementioned approach is to minimize the local distance between adjacent points, Magoarou et al. [33] used Isomap to optimize the channel chart between all CSIs on global distances, which are created as sum of local distances on the shortest path between two points. This allows to generate a chart with global instead of only local similarity.

III. CSI DISTANCES

To estimate a globally consistent channel chart, we need both a locally and globally valid distance metric for CSI. In the following we derive a local CSI distance in the time-domain and our approximation of the global distance, which is necessary for CC-assisted localization.

A. LOCAL CIR DISTANCE

The idea of CC is that CSI measurements are spatially correlated as the CSIs are similar at the same position and become more and more dissimilar with their physical distance in space. Hence, the goal of CC is to find a metric which (for any measurement instance i and j) resembles the physical distance so that

$$d_{\text{phy}}(\boldsymbol{x}_i, \boldsymbol{x}_j) \propto d_{\text{csi}}(\tilde{\boldsymbol{H}}_i, \tilde{\boldsymbol{H}}_j) \ , \tag{1}$$

where $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are two real-world coordinates and $\tilde{\boldsymbol{H}}_i$ and $\tilde{\boldsymbol{H}}_j$ their approximated channel impulse responses (CIRs)

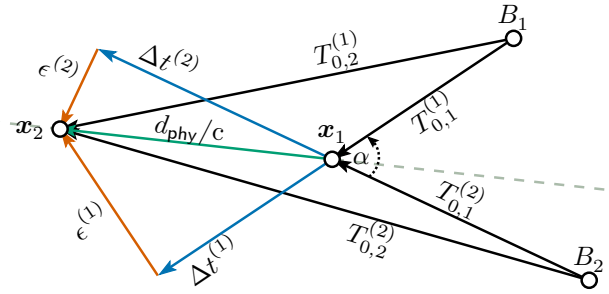


FIGURE 1. Schematic view of a radio environment with two base stations B_1 and B_2 showing the geometric context of ToF measurements w.r.t. the physical distance of a displacement of a radio unit from $\boldsymbol{x}_1$ to $\boldsymbol{x}_2$. The above-shown quantities are the (time-)lengths of the described vectors.

between all base base stations and the UE, measured at the receiver. The channel model of a radio signal can be defined as

$$h(t) = \sum_{n=0}^{N_{\text{p}}-1} a_n \delta(t - T_n) \ , \tag{2}$$

where N_{p} is the total number of arriving signal paths from the transmitter at the receiver, n is the index of the current path, a_n is the complex gain of the n^{th} MPC, $\delta(\cdot)$ is the Dirac delta function, and T_n the delay of the component. To measure the CIR, a bandwidth-limited measurement signal $s(t)$ is first transmitted and

$$y(t) = h(t) * s(t) \ , \tag{3}$$

defines the (noise-free) signal received at the base station, where $*$ is the convolution operator. The approximate CIR is then obtained from the autocorrelation:

$$\tilde{h}(\tau) = \int_{-\infty}^{\infty} y(t) s^*(t - \tau) \, \mathrm{d}t \ . \tag{4}$$

The CIR of length T for a measurement snapshot i for base station k is given in vector form as

$$\tilde{\boldsymbol{h}}_i^{(k)} \overset{\text{def}}{=} \left[\tilde{h}_i^{(k)}(0), \ldots, \tilde{h}_i^{(k)}(T-1)\right] \ . \tag{5}$$

In the following, we will base the derivation on the time of flight (ToF). For the time-difference-of-arrival (TDoA) case, the differences in the derivation can be found in Appendix A. For two measurements that share the same MPCs in the same order, we define the ToF as

$$\Delta t_{n,ij}^{(k)} \overset{\text{def}}{=} \left| T_{n,i}^{(k)} - T_{n,j}^{(k)} \right| \ , \tag{6}$$

for the positions $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ of the CSI measurements $\tilde{\boldsymbol{h}}_i$ and $\tilde{\boldsymbol{h}}_j$ obtained at base station B_k. $T_{n,i}^{(k)}$ and $T_{n,j}^{(k)}$ are the time delays of the n^{th} MPC from base station B_k at position i and j. If the displacement of the position $\boldsymbol{x}_j$ is collinear to both the position $\boldsymbol{x}_i$ and the position of the base station $\boldsymbol{x}_b$, then

$$d_{\text{phy}}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \text{c} \, \Delta t_{n,ij} \ , \tag{7}$$

where c is the speed of light. However, if they are not collinear the distance approximation is erroneous and can be defined

as

$$d_{\mathsf{phy}}(\boldsymbol{x}_i, \boldsymbol{x}_j) \overset{\text{def}}{=} \mathrm{c}\sqrt{\left(\Delta t_{n,ij}^{(k)}\right)^2 + \left(\epsilon_{n,ij}^{(k)}\right)^2}\,, \tag{8}$$

where the error $\epsilon_{n,ij}^{(k)}$ is assumed to be orthogonal to the time difference $\Delta t_{n,ij}^{(k)}$, as shown in Fig. 1.[1] Following the example case with two base stations shown in Fig. 1, base station B_1 has a delay of the first direct path $T_{0,1}^{(1)}$ to position x_1 and a delay of $T_{0,2}^{(1)}$ to position $\boldsymbol{x}_2$. The vector difference of the delay $\Delta t^{(1)}$ to the expected delay $d_{\mathsf{phy}}/\mathrm{c}$ can be expressed as the error $\epsilon^{(1)}$:

$$\left(\epsilon^{(1)}\right)^2 = \frac{d_{\mathsf{phy}}^2}{\mathrm{c}^2} - \left(\Delta t^{(1)}\right)^2\,. \tag{9}$$

As we do not know the direction of movement, we assume to have the similar likelihood for every direction, which leads to a arcsine distribution with a boundary of $[0, d_{\mathsf{phy}}/\mathrm{c}]$ for the error $\epsilon^{(1)}$. The standard deviation depends on the physical distance

$$\sigma_\epsilon(d_{\mathsf{phy}}) = \frac{1}{\sqrt{8}}\frac{d_{\mathsf{phy}}}{\mathrm{c}}\,, \tag{10}$$

which means that the standard deviation of the error monotonically increases with the distance between the two measurements. If we have multiple base stations, e.g., B_1 and B_2, the errors $\epsilon^{(1)}$ and $\epsilon^{(2)}$ are geometrically related

$$\epsilon^{(2)} = \frac{d_{\mathsf{phy}}}{\mathrm{c}}\cos\left(\pi - \alpha \mp \cos^{-1}\left(\frac{\mathrm{c}\,\epsilon^{(1)}}{d_{\mathsf{phy}}}\right)\right)\,, \tag{11}$$

where the sign in front of the $\cos(\cdot)^{-1}$ depends on whether B_1 and B_2 lie on opposite sides of the line of motion of the UE (i.e., the green line between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ in Fig. 1) or on the same side. This means that for every $\alpha > 0$, the sum of the errors $\epsilon^{(1)}$ and $\epsilon^{(2)}$ have a lower bound on the the sum of distances

$$2\,d_{\mathsf{phy}} > \mathrm{c}\,(\epsilon^{(1)} + \epsilon^{(2)})\,. \tag{12}$$

This relation also holds for every MPC arriving at the receiver and allows us therefore to define a distance metric

$$d_{\mathsf{s}}(\tilde{\boldsymbol{H}}_i, \tilde{\boldsymbol{H}}_j) \overset{\text{def}}{=} \sum_{k=0}^{N_{\mathsf{b}}-1}\sum_{n=0}^{N_{\mathsf{p}}-1} \Delta t_{n,ij}^{(k)}\,, \tag{13}$$

which has a monotonically increasing mean and standard deviation, and a lower bound of the error. Here, the matrices $\tilde{\boldsymbol{H}}_i$ and $\tilde{\boldsymbol{H}}_j$ collect all vector-form CIRs, i.e.,

$$\tilde{\boldsymbol{H}}_i \overset{\text{def}}{=} \begin{bmatrix} \tilde{\boldsymbol{h}}_i^{(0)} \\ \vdots \\ \tilde{\boldsymbol{h}}_i^{(N_{\mathsf{b}}-1)} \end{bmatrix} \quad \text{and} \quad \tilde{\boldsymbol{H}}_j \overset{\text{def}}{=} \begin{bmatrix} \tilde{\boldsymbol{h}}_j^{(0)} \\ \vdots \\ \tilde{\boldsymbol{h}}_j^{(N_{\mathsf{b}}-1)} \end{bmatrix}\,. \tag{14}$$

$\Delta t_{n,ij}^{(k)}$ in (13) denotes the time-difference observed at the k^{th} BS. To calculate the distance metric, the estimation of the time delays of every arriving path is necessary. However, as the extraction of the delay of the MPCs from measured CIRs is very challenging due to the bandwidth limited signal $s(t)$ [46], we consider a simple approximation by subtracting the time-aligned CIRs:

$$d'_{\mathsf{s}}(\tilde{\boldsymbol{H}}_i, \tilde{\boldsymbol{H}}_j) \overset{\text{def}}{=} \sum_{k=0}^{N_{\mathsf{b}}-1}\sum_{t=0}^{T-1} \left|\,|\tilde{h}_i^{(k)}(t)| - |\tilde{h}_j^{(k)}(t)|\,\right|\,, \tag{15}$$

where k is the index of the N_{b} base stations and t is the time-index of the CIR of total length T.

Proposition 1: The limited window of every CIR includes the absolute position of the paths w.r.t. their ToF or TDoA, where the window is aligned by the first direct path of arrival (FDPoA). If the CIRs are recorded at the same position, the distribution of power within the CIRs is equal, which leads to $d'_{\mathsf{s}}(\tilde{\boldsymbol{H}}_i, \tilde{\boldsymbol{H}}_i) = 0$, whereas measurements at different positions have a time-shift of the paths within the CIR and therefore also a shift of the power. This leads to less overlap within the CIRs and hence to a higher absolute difference that is proportional to the sum of time differences of the arriving paths.

Proposition 2: As long as the main lobes of the bandwidth-limited signals are overlapping in the CIRs, the physical distance and the CIR distance have a non-linear relationship, and have a linear one if the displacements between two CIR measurements is small. (The detailed derivation of this linear relationship can be found in Appendix B.) We therefore have

$$d_{\mathsf{phy}}(\boldsymbol{x}_i, \boldsymbol{x}_j) \propto d_{\mathsf{s}}(\tilde{\boldsymbol{H}}_i, \tilde{\boldsymbol{H}}_j) \propto d'_{\mathsf{s}}(\tilde{\boldsymbol{H}}_i, \tilde{\boldsymbol{H}}_j) \tag{16}$$

for small distances between CSI measurements, which means that our metric is locally restricted to its spatial neighborhood.

B. GLOBAL GEODESIC DISTANCES

The distance metric $d'_{\mathsf{s}}(\tilde{\boldsymbol{H}}_i, \tilde{\boldsymbol{H}}_j)$ between CSIs is restricted to short distances between two CSI measurements. Hence, our optimization is restricted to each sample's neighborhood. This only yields a local similarity and consistency of the channel chart. However, to use CC for localization we have to learn a channel chart reflecting the global geometry of the environment. To achieve this, we use the idea of Isomap [47] which creates global distances as the sum of local distances on the shortest paths on the manifold, i.e., *geodesic distances*. Here, we estimate a matrix of pair-wise distances

$$\boldsymbol{D}_{\mathsf{pw}} \overset{\text{def}}{=} \begin{bmatrix} d'_{\mathsf{s}}(\tilde{\boldsymbol{H}}_0, \tilde{\boldsymbol{H}}_0) & \cdots & d'_{\mathsf{s}}(\tilde{\boldsymbol{H}}_0, \tilde{\boldsymbol{H}}_{N-1}) \\ \vdots & \ddots & \vdots \\ d'_{\mathsf{s}}(\tilde{\boldsymbol{H}}_{N-1}, \tilde{\boldsymbol{H}}_0) & \cdots & d'_{\mathsf{s}}(\tilde{\boldsymbol{H}}_{N-1}, \tilde{\boldsymbol{H}}_{N-1}) \end{bmatrix}\,, \tag{17}$$

$\boldsymbol{D}_{\mathsf{pw}} \in \mathbb{R}^{N\times N}$, to create a neighborhood graph. The shortest paths between each coordinates of the channel chart $d'_{\mathsf{s}}(\tilde{\boldsymbol{H}}_i, \tilde{\boldsymbol{H}}_j)$ are then estimated using a shortest-path estimator, e.g., via the Dijkstra algorithm [48].[2] As the computational complexity of shortest-path algorithms grows with the number of neighbors, the number of edges has to

[1] In the figure and derivations, $n = 0$, $i = 2$ and $j = 1$ are assumed whenever not explicitly given, e.g., $\epsilon_{0,21}^{(k)} \equiv \epsilon^{(k)}$ and $\Delta t_{0,21}^{(k)} \equiv \Delta t^{(k)}$.

[2] Note that we do not need to run the estimator for each of the pairs in the matrix separately as due to the *principle of optimality* [49], any pairs of points on a sub-path of a shortest path themselves constitute an optimal path.

be minimized while keeping the accuracy high. We found that in our case, choosing 5 neighbors already leads to valid geodesic distances.

This allows us to define valid distances for all points, also for far points, on the manifold

$$\boldsymbol{D}_{\text{geo}} \overset{\text{def}}{=} \begin{bmatrix} d'_{\text{geo}}(\tilde{\boldsymbol{H}}_0, \tilde{\boldsymbol{H}}_0) & \cdots & d'_{\text{geo}}(\tilde{\boldsymbol{H}}_0, \tilde{\boldsymbol{H}}_{N-1}) \\ \vdots & \ddots & \vdots \\ d'_{\text{geo}}(\tilde{\boldsymbol{H}}_{N-1}, \tilde{\boldsymbol{H}}_0) & \cdots & d'_{\text{geo}}(\tilde{\boldsymbol{H}}_{N-1}, \tilde{\boldsymbol{H}}_{N-1}) \end{bmatrix}, \tag{18}$$

$\boldsymbol{D}_{\text{geo}} \in \mathbb{R}^{N \times N}$, by summing up the local distances on the shortest path between two points

$$d'_{\text{geo}}(\tilde{\boldsymbol{H}}_i, \tilde{\boldsymbol{H}}_j) \overset{\text{def}}{=} \sum_{p \in \mathcal{P}} d'_{\text{s}}(\tilde{\boldsymbol{H}}_p, \tilde{\boldsymbol{H}}_{p+1}) \,, \tag{19}$$

where p and $p+1$ are the indices of all neighboring pairs of points on the shortest path $\mathcal{P}$. As our CIR distance is linear with a constant slope for small displacements, we also have a linear geodesic distance if the spatial density of neighbors is high, as the sum of linear elements with the same slope will also be linear.

IV. CHANNEL-CHARTING-ASSISTED LOCALIZATION

Intentionally, Isomap performs dimensionality reduction on an enclosed data set, which therefore restricts the method to predict the embedding on unseen data. The full geodesic distance matrix has to be calculated including the previous and new data to predict on unseen data, which is both time-consuming and impractical for real-time locating systems. Hence, at first glance, Isomap is inappropriate for CC, that predicts on unseen data. While there are extensions of the original Isomap algorithm, such as the landmark Isomap [50], which can interpolate the embedding on unseen data, Dimal et al. [51] have shown that Siamese networks outperform the classical approaches in terms of generalization. In this approach the multidimensional scaling (MDS) is replaced by a Siamese network [52], which encodes the input data into the 2D space and then learns the distance between two inputs in the embedding given a distance metric. The goal of the Siamese network is to learn the proposed geodesic distance $d'_{\text{geo}}(\tilde{\boldsymbol{H}}_i, \tilde{\boldsymbol{H}}_j)$ with the objective

$$\mathcal{L} = \left| d'_{\text{geo}}(\tilde{\boldsymbol{H}}_i, \tilde{\boldsymbol{H}}_j) - \|\boldsymbol{z}_i - \boldsymbol{z}_j\|_2 \right| , \tag{20}$$

where $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$ are the two-dimensional outputs of the neural network for the inputs $\tilde{\boldsymbol{H}}_i$ and $\tilde{\boldsymbol{H}}_j$. Instead of utilizing all data like in the Isomap algorithm, we sample randomly from the available geodesic data until the Siamese network has converged. After optimization, the neural network has learned the global geometry of the radio environment. Since the neural network uses the CSI measurements as input data, the network learns a transformation from the manifold of the input data to a 2D embedding. This allows the network to interpolate and therefore also to efficiently process unseen data without recalculating the geodesic distance matrix.

TABLE 1. Parameters of the Siamese network architectures for the UWB and 5G radio setups.

Layer type	Output dimensions		Kernel size		Activation
	UWB	5G	UWB	5G	
Conv	8×6×200	8×6×49	3×3	3×3	ReLU
Conv	8×6×200	8×6×49	5×5	5×5	ReLU
Conv	8×6×200	8×7×50	15×15	8×8	ReLU
Conv	16×7×200	16×8×49	30×30	10×10	ReLU
Avg. Pooling	1×6×200	1×6×49	—	—	—
Fully Con.	300	200	—	—	ReLU
Fully Con.	2	2	—	—	—

A. ARCHITECTURE

As CC is very similar to FP, we employ a similar efficient architecture as proposed in [13]. The network consists of 4 convolutional layers and 2 dense layers. Batch normalization stabilizes the training and we apply Rectified Linear Units (ReLU) as activation functions except for the last layer, where we applied no activation function. We have no local pooling layers between the convolutional layers to keep the dimension of time, which has shown good results in time series downstream tasks [53]. Instead, we use a global average pooling of the channels before the dense layers. The kernel sizes of the convolutional layers are increased with the depth of the model to enhance the receptive field of the network. The kernel sizes being used depends on the effective input resolution of the CIRs. As we have two different radio systems with different resolutions of the CIRs, we have different kernel sizes of the neural network adapted for the respective bandwidth:

(i) for the UWB data, we have kernel sizes of $[3 \times 3]$ for the first, $[5 \times 5]$ for the second, $[15 \times 15]$ for the third, and $[30 \times 30]$ for the last layer;
(ii) for the 5G data, we have $[3 \times 3]$ for the first, $[5 \times 5]$ for the second, $[8 \times 8]$ for the third, and $[10 \times 10]$ for the last layer.

The parameters of both architectures are summarized in Table 1.

B. PREPROCESSING

For data preprocessing, we follow the idea from [13] to exploit both the raw CIR in the time domain and the corresponding ToF/TDoA using a CNN. We generate a 2D tensor of dimensions $[N_{\text{A}} \times L_{\text{w}}]$, with N_{A} anchors and L_{w} time-steps in the temporal resolution of the CIR. In our case, the CIRs are discrete-valued and of a fixed length. The CIRs are padded by the corresponding ToF/TDoA within the tensor to model the unique relative shift of the FDPoA for every position in the area.

An example of this process for a tensor with $N_{\text{A}} = 6$ anchors is shown in Fig. 2. Each CIR is shifted by its

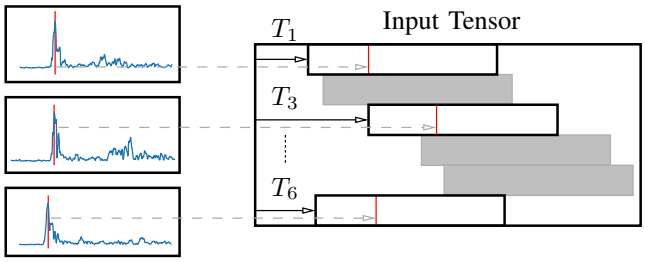


FIGURE 2. Exemplary illustration of the data preprocessing for the CIRs: they are padded with their respective ToFs within the tensor to model relative shift.

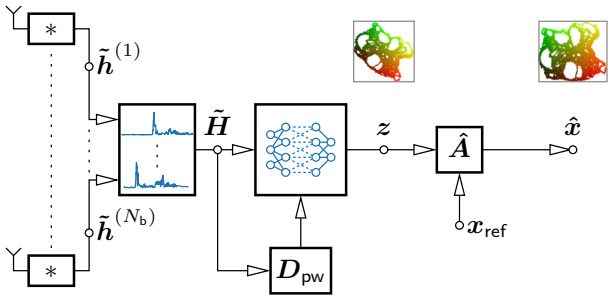


FIGURE 3. End-to-end model of our positioning pipeline. After the CIRs are estimated, they are first preprocessed and collected to form the input to the neural network for training and to estimate the geodesic-distances matrix. Once the chart is learned, a transformation is applied to obtain position estimates using the real-world frame.

corresponding ToF T_k, which not only preserves the MPC information but also its relative time-shift, which is unique for the environment. Note that L_w must be selected large enough to ensure that even after padding with the largest ToF/TDoA the CIRs MPC information is contained in the input tensor.

C. COORDINATE TRANSFORMATION

After training, the channel chart reflects only the local geometry of the environment, which means that the physical distances are preserved, but the channel chart might be rotated, translated or scaled (as the geodesic distance is only proportional to the physical distance). To exploit the estimated channel chart for localization, we therefore have to estimate an affine transformation from the local channel chart coordinate frame to the global real-world coordinate frame. Theoretically, we only need 3 points to define an affine transformation between two coordinate systems in the 2D space. However, as a chart will always have imperfections, we cannot select the optimal points in it to do the transformation. We therefore use several points to estimate a transformation matrix with a least-squares solution

$$\hat{\boldsymbol{A}} \overset{\text{def}}{=} \underset{\boldsymbol{A}}{\operatorname{argmin}} \|\boldsymbol{A}\boldsymbol{z} - \boldsymbol{x}_\mathrm{ref}\|_2^2 \,, \tag{21}$$

where $\hat{\boldsymbol{A}}$ is the estimated transformation matrix, $\boldsymbol{z}$ are the points of the channel chart, and $\boldsymbol{x}_\mathrm{ref}$ are the coordinates in the real-world frame. Once the matrix is at hand, the estimated positions in the real-world frame $\hat{\boldsymbol{x}}$ are obtained as

$$\hat{\boldsymbol{x}} \overset{\text{def}}{=} \hat{\boldsymbol{A}}\boldsymbol{z} \,. \tag{22}$$

An evaluation on the effect of the number of selected reference points $\boldsymbol{x}_\mathrm{ref}$ for the transformation on the performance is shown in Section D.

D. END-TO-END POSITIONING PIPELINE

Our end-to-end positioning pipeline is shown in Fig. 3. First, the CIRs $\tilde{\boldsymbol{h}}$ are estimated at the N_b base stations. Next, they are preprocessed as described in Section B and are collected in $\tilde{\boldsymbol{H}}$. We use $\tilde{\boldsymbol{H}}$ to estimate a distance matrix $\boldsymbol{D}_\mathrm{pw}$ of geodesic distances as described in Section B. The input tensor is also fed to the neural network to learn a unique mapping to the channel-chart position. During the training, the distance matrix $\boldsymbol{D}_\mathrm{pw}$ is used to enforce the Siamese network to preserve the pair-wise distances between corresponding input tensors in the 2D embedding. After training, a few labeled reference data samples $\boldsymbol{x}_\mathrm{ref}$ are used estimate an affine transformation from the local channel-chart coordinate frame to the global real-world frame. Finally, the Siamese network can be used to estimate local channel-chart positions $\boldsymbol{z}$ on unseen data and with a linear transformation, i.e., $\hat{\boldsymbol{A}}$, to transform them to the real-wold coordinate frame (i.e., to calculate $\hat{\boldsymbol{x}}$) for localization.

V. EXPERIMENTAL SETUP

A. MEASUREMENT DATA

To show the capabilities of CC for positioning, we compare the performance with two different radio setups. The first setup is a 5G downlink TDoA setup and the second is a UWB ToF radio setup.

5G Data. For our first experiment, we use a 5G downlink TDoA setup with six commercial off-the-shelf software-defined-radio BS. The radio system has a bandwidth of 100 MHz and a center frequency of 3.7 GHz. The BS transmit power is set to 20 dBm. All BSs are highly synchronized by means of a common signal generator. The recording frequency is 6.6 Hz. Fig. 4 shows a schematic view and the real-world environment. The transceivers (green) are placed at the edges of the recording area at a height of 6-7 m. We created a typical industrial setup with reflective walls (red), an industrial truck (orange), a forklift (gray), and small (blue) and large shelves (purple) to block the LoS to the receiver and to create dense multipath propagation. The receiver is placed at a height of 1.95 m on a handcart, which is moved by a person. Also for the 5G setup, we record a training data set with 18,722 bursts and a test data set with 15,721 bursts on different trajectories within the same environment.

UWB Data. For our second experiment, we use a UWB radio setup with 6 transceivers acting as base station (BS) and one recording robot platform with a maximum velocity of $v_\mathrm{max} = 0.2\,\mathrm{ms}^{-1}$. We configured the system for ToF acquisition as a two-way ranging setup with a bandwidth of

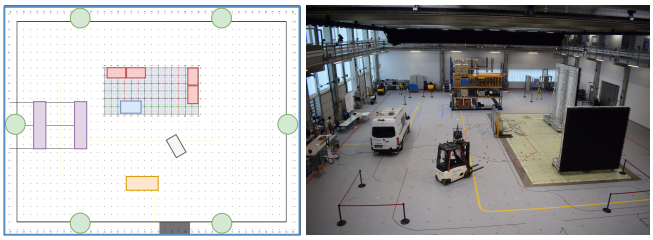


FIGURE 4. Schematic top view (left) of the environment (right). The environment consists of reflective walls (red), an industrial truck (orange), a fork lift (gray), large shelves (purple), and a small shelf (blue). The BSs are indicated as green dots and are placed at a height of 6–7 m. The recording area has a size of $25\,\mathrm{m} \times 16\,\mathrm{m}$.

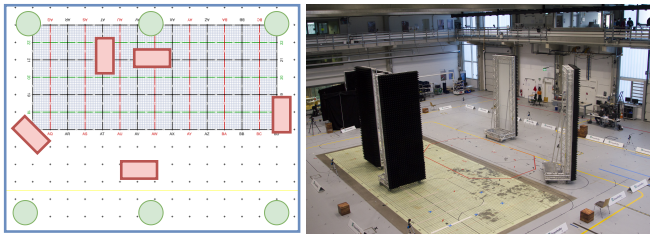


FIGURE 5. Schematic top view (left) of the environment (right). The red rectangles indicate reflective walls and the green dots are the stationary UWB transceiver modules. The recording area has a size of $18\,\mathrm{m} \times 11\,\mathrm{m}$.

499.2 MHz at a center frequency of 4 GHz with a recording frequency of 3 Hz. We designed a complex environment with walls, that reflect radio signals on the inner side (iron surface) and absorb them at the outside (black surface). Fig. 5 shows a schematic view and the real-world environment. The transceivers, indicated as green dots, are placed at the edges of the recording area, shown as blue rectangles. The reflective walls, indicated in red, are placed to block the LoS between the anchors and the robot platform, which causes ranging errors of the UWB radio system leading to high localization errors using classical positioning approaches. We recorded two different data sets: a training data set with 18,027 bursts, where one burst includes the 6 synchronized CIRs and ToF measurements, and a test data set with 3,382 bursts. Both data sets are recorded independently on different trajectories within the same area.

B. BASELINES

In the following, we describe all relevant state-of-the-art CC methods that we compare our solution with.

1) Non-parametric Approaches

We compare our approach with PCA, Sammon mapping [19], and Isomap [33]. For the pair-wise distances used in Sammon mapping, we applied the CIR distance defined in Section A. For Isomap, we used the geodesic distance defined in Section B.

2) Parametric Approaches

In general, we investigate three different parametric baseline methods, as follows.

First, we apply the two stage approach from Agostini et al. [44]. They used a convolutional autoencoder to compress the CSI to the quintessential information and used UMAP to create the channel chart. As autoencoder architecture, we use two convolutional layers with kernel size 3 for the first and 5 for the second layer. Both layers have 8 channels. As the decoder we use a fully connected layer to restore the input size with 2 consecutive transposed convolutional layers with kernel size 3 and 5 and also 8 channels. For all layers except the last layer, we used ReLU activation functions and for the convolutional layers batch normalization. The mean squared error (MSE) is used to evaluate the reconstruction error. We compressed the input data to 5 % of its size and applied UMAP on the latent variables with the physical distance as metric.

The second approach is the constrained autoencoder proposed by Huang et al. [39]. They used a fully connected autoencoder to compress the input to the size of two, which are the coordinates of the channel chart. Besides the reconstruction error (MSE), they added a constraint which minimizes the pair-wise distances between two compressed input instances. As pair-wise distances, we use the CIR distance defined in Section A between the raw input vectors. The architecture consists of 4 fully connected layers with [500, 100, 50, 20] neurons, while the encoder is the mirrored structure.

The third approach is proposed by Ferrand et al. [23] and uses triplets. The triplets are built using the recording time of the signals assuming that CSIs close in time are also close in space, due to the physical constraints of the movement, whereas samples that are farther in time are more distant in space. Hence, for the triplets, samples that are close in time (i.e., in our case within a window of $\pm 3\,\mathrm{s}$) are assigned as positive, and negative samples are assigned within a window of $\pm 2{,}400\,\mathrm{s}$ but not within the positive window. These parameters are estimated experimentally and depend on the movement pattern. As architecture, we use the same model as for our geodesic Siamese network defined in Section A.

C. PERFORMANCE METRICS

To measure the performance of a generated channel chart, we apply continuity (CT) and trustworthiness (TW) to measure its local similarity [19]. CT measures whether a chart introduces wrong nearest neighbors in the embedding and penalizes them proportional to the rank of the nearest neighbors

$$\mathrm{CT}(K) = 1 - \frac{2}{K(2N - 3K - 1)} \sum_{i=0}^{N-1} \sum_{j \in \mathcal{V}_K(v_i)} \left(\hat{r}(i,j) - K\right) . \tag{23}$$

$\mathcal{V}_K(v_i)$ are the K nearest neighbors in the original space, N is the total number of samples and $\hat{r}(i,j)$ is the rank among the pair-wise distances in the embedding. Conversely, TW measures whether the original space has different nearest

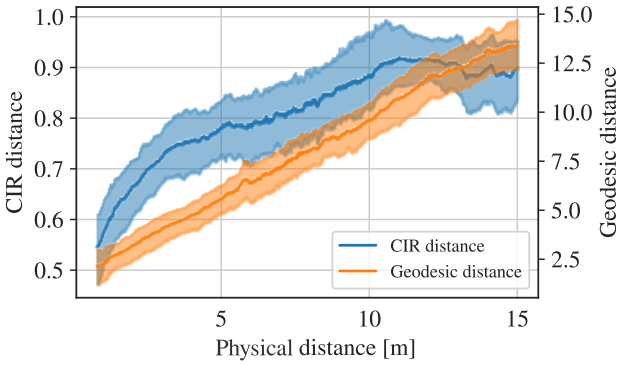


FIGURE 6. **The proposed local CIR distance (blue) and the global geodesic distance (orange) for random combinations of positions in the 5G environment.**

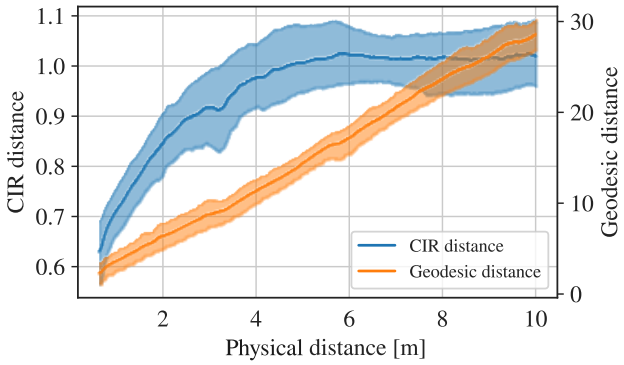


FIGURE 7. **The proposed local CIR distance (blue) and the global geodesic distance (orange) for random combinations of positions in the UWB environment.**

neighbors as modelled in the channel chart

$$\mathrm{TW}(K) = 1 - \frac{2}{K(2N-3K-1)} \sum_{i=0}^{N-1} \sum_{j \in \mathcal{U}_K(u_i)} \big(r(i,j) - K\big) \,. \tag{24}$$

Here, $\mathcal{U}_K(u_i)$ are the K nearest neighbors in the chart and $r(i,j)$ is the rank among the pair-wise distances in the original space. As proposed in [19], we use $K = 0.05N$ as the number of nearest neighbors to calculate TW and CT. Both metrics have a value range of $[0,1]$, where 1 means that the local geometry is perfectly preserved and 0 that the local geometry is not preserved.

As we mainly investigate CC for localization, we also estimate the positioning error after the affine transformation described in Section C. As the selection of points for the affine transformation affects the localization error, we use all reference positions to estimate the least squares solution. However, the effect of using different numbers of points for the affine transformation is investigated in Section D. We calculate the mean absolute error (MAE), with the error being the physical distance between the channel-chart points and ground truth and the 90$^{\text{th}}$ percentile of the cumulative distribution function (CDF) of the error (CE90) as performance metric.

VI. EVALUATION

We first evaluate the proposed global geodesic distance metric. Next, we compare the geodesic Siamese network with the relevant state-of-the-art of CC for different radio setups w.r.t. positioning. Finally, we compare CC-assisted FP with supervised FP. The CC algorithms are trained only on the training data set and are tested on both the training and test data to show their generalization capabilities on unseen data.

To ensure the reproducibility of our results, we have made both the source code and the data sets used in our experiments publicly available.[3]

[3] https://github.com/mutschcr/cc_tmcn

A. DISTANCE METRIC EVALUATION

For the evaluation of the proposed geodesic distance metric described in Section B, we use the UWB and 5G data sets, where we randomly sample measurements and calculate the distance metrics and the corresponding physical distance in space. Fig. 7 shows the local CIR distance (blue) and the global geodesic distance (orange) for the UWB data set, where the physical distance between two positions is shown on the x-axis and the proposed distances are shown on the y-axis. As described in Section A, we can clearly see that the CIR distance is only (linearly) correlated to the physical distance for points that are close in space. From the CIR distance we can neither reliably (as the relation is constantly changing) nor unambiguously (as the CIR distance is not invertible) derive an physical distance for two points that are farther apart.

The mean and the standard deviation of the error increases monotonically with the physical distance until about 6 m. For physical distances larger then 6 m, the CIR distance has no more correlations to the physical distance, which restricts the measurements to their neighbors. We can see a similar behaviour for the 5G data set shown in Fig. 6. The CIR distance is correlated to the physical distance until about 10 m. Due to the lower bandwidth of the 5G signals, the lobe of the received signals are wider, which leads correlations of signals for higher physical distances compared to signals with lower bandwidths.

However, this also leads to a higher noise level due to power variations in the CIR because of fast fading effects, which is a violation of our distance metric model. To have a correlation of the radio signals also for larger distances, we propose to create geodesic distances from a linear combination of local CIR distances as described in Section B. For both radio systems, the geodesic distance shows a high linear correlation to the physical distance, shown in orange in Fig. 7 and Fig. 6. This means that the spatial density of CSI measurements is high enough to provide a linear correlation of the CIR distance to the physical distance in its neighborhood. This also means that the geodesic distance, i.e., a sum of locally linear CIR

TABLE 2. Results of CC methods on the 5G data set, tested on both the training and test data sets. The unit of CE90 and MAE is meter.

	Training				Test			
Method	CT	TW	CE90	MAE	CT	TW	CE90	MAE
Ours	0.986	0.986	2.35	1.40	0.983	0.982	2.48	1.46
Isomap [33]	0.985	0.986	2.46	1.42	0.984	0.984	2.58	1.47
C. UMAP [44]	0.966	0.961	3.87	2.12	0.958	0.947	4.55	2.37
Triplet [23]	0.920	0.824	6.90	4.17	0.919	0.842	6.69	3.89
Con. AE [39]	0.919	0.895	8.19	4.66	0.919	0.895	7.71	4.37
Sammon [19]	0.899	0.880	7.41	4.04	0.897	0.890	6.68	3.77
PCA [19]	0.918	0.860	7.59	4.16	0.921	0.869	7.27	4.02

TABLE 3. Results of CC methods on the UWB data set, tested on both the training and test data sets. The unit of CE90 and MAE is meter.

	Training				Test			
Method	CT	TW	CE90	MAE	CT	TW	CE90	MAE
Ours	0.997	0.997	1.30	0.69	0.997	0.996	1.28	0.72
Isomap [33]	0.997	0.997	1.33	0.72	0.996	0.996	1.34	0.80
C. UMAP [44]	0.996	0.996	1.37	0.72	0.995	0.994	1.24	0.68
Triplet [23]	0.975	0.951	3.98	2.29	0.973	0.948	3.94	2.25
Con. AE [39]	0.965	0.928	4.76	2.55	0.966	0.939	4.38	2.30
Sammon [19]	0.943	0.880	5.49	2.98	0.943	0.888	5.49	2.98
PCA [19]	0.934	0.825	7.41	3.73	0.941	0.874	6.18	3.46

distances, is linearly correlated to the physical distance. We therefore have a globally valid distance, which correlates to the physical distance for arbitrary large distances. This allows us to optimize channel charts with a global similarity to the real-world radio geometry, which is necessary for CC-assisted localization.

B. 5G SETUP

Table 2 lists the results of CC methods on the 5G setup, where *training* means that the models are evaluated on training data set and *test* that the models are evaluated on the test data set. Our method shows along with Isomap the highest local similarities of $\mathrm{TW} > 0.98$, $\mathrm{CT} > 0.98$ and a global similarity of $\mathrm{CE90} < 2.46\,\mathrm{m}$, $\mathrm{MAE} < 1.42\,\mathrm{m}$, while our approach achieves a slightly better global accuracy. The main difference of our approach to the Isomap is mainly the runtime improvement, as Isomap has to recalculate the geodesic distance matrix for every unseen data point, while our approach learned a transformation function enabling efficient prediction on unseen data. The combination of a convolutional autoencoder and UMAP (C. UMAP) still achieves reasonable results of TW and $\mathrm{CT} = 0.96$ for the local similarity, while the global similarity is low with a $\mathrm{CE90} = 3.87\,\mathrm{m}$ and $\mathrm{MAE} = 2.12\,\mathrm{m}$. We also tested UMAP on the raw input embeddings with the CIR distance as pair-wise distance and achieved $\mathrm{CT} = 0.947$ and $\mathrm{TW} = 0.972$ for the local similarity and $\mathrm{CE90} = 8.74\,\mathrm{m}$ and $\mathrm{MAE} = 5.26\,\mathrm{m}$ for the global similarity.

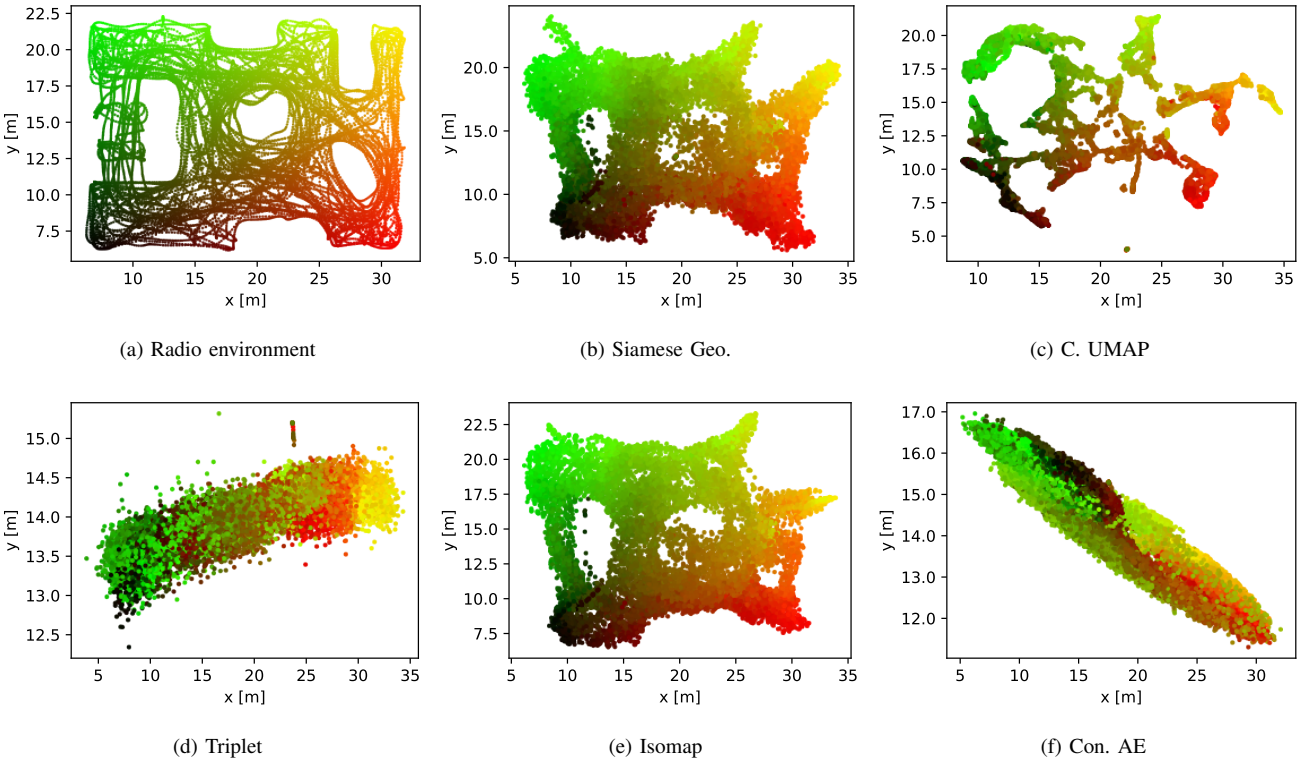
However, as already investigated in [44], we could not achieve higher results especially for the global similarity. The worst performance is achieved by PCA, Sammon mapping and the triplets approach with a local similarity $\mathrm{TW} < 0.89$ and $\mathrm{CT} < 0.93$ and a global similarity of $\mathrm{CE90} > 6.90\,\mathrm{m}$ and $\mathrm{MAE} > 4.00\,\mathrm{m}$. The PCA only learns a linear transformation and Sammon mapping considers the distances between all points, including the points that are far away. However, as distances between two CSI measurements are only valid in the proximity of two points, considering also far points introduces errors in the channel chart generation. The worse performance of the triplet approach can be explained by the assumption that far points in time are also far in space. This assumption depends highly on the movement pattern, and is therefore error prone, especially in the 5G data set as the receiver is moved by a person. The constrained autoencoder (Con. AE) faces the same problem as Sammon mapping, as also distances from far points are considered during optimization. However, we think due to the unique mapping from the input space into the embedding, the autoencoder identifies wrong distance labels implicitly and mitigates the effect of wrong distance approximations, which leads to slightly better results as in Sammon mapping.

Fig. 8 visualizes the channel charts for the 5G data set. Graph (a) shows the reference radio environment with a color gradient, while the Graphs (b)–(f) show the results of the channel charts for different methods. We can clearly see that our method, cf. 8b, and the Isomap, cf. 8e, preserve the global structure of the environment very well. However, it overlaps slightly on the left-hand side and drifts away on the right-hand side on the top. As both methods rely on the nearest neighbors to create the geodesic distances, only vertical neighbors are available on the tail on the upper right-hand side, which leads to a overestimation of the physical distance and therefore to a horizontal drift therein. The channel chart of the C. UMAP algorithm, cf. Fig. 8c, does not reflect the global similarity very well, as its shape is curved. However, the local similarity is still reasonable as the color gradient is valid in most areas. The other algorithms have a much lower quality of the channel charts: The triplet approach, cf. Fig. 8d, and the constrained autoencoder, cf. Fig. 8f, fail in recovering the geometry.

The generalization abilities are good for all the algorithms, showing no significant differences in performance using the test data set. This indicates that the algorithms are not overfitting to the training trajectories and can predict also well for unseen data in the same environment.

C. UWB SETUP

Table 3 lists the results of the CC methods on the UWB data set. The majority of the state-of-the-art CC methods show high accuracy for the local similarity with TW and $\mathrm{CT} > 0.95$ for our approach, Isomap, C. UMAP and the Triplet approach. Interestingly, C. UMAP achieves similarly high local similarity of TW and $\mathrm{CT} = 0.996$ and global



(a) Radio environment (b) Siamese Geo. (c) C. UMAP

(d) Triplet (e) Isomap (f) Con. AE

FIGURE 8. Results CC methods on the 5G data set. Graph (a) shows the recorded training data in the real-world environment. The color gradient is used to to show the relation of the positions of the channel charts to the real positions. Graphs (b)–(f) show the channel charts after the affine transformation to the global coordinate frame.

similarity of $\mathrm{CE90} = 1.37\,\mathrm{m}$ and $\mathrm{MAE} = 0.72\,\mathrm{m}$ to ours and the Isomap. Also the triplet based algorithm achieves higher local accuracies of $\mathrm{TW} = 0.951$ and $\mathrm{CT} = 0.975$. We think the improvement comes from the slow and consistent movement of the robot platform with less dynamics in the movement pattern, which leads to a more valid time based triplet selection. However, recovering the global structure of the environment is still not possible with a weak global accuracy of $\mathrm{CE90} = 3.98\,\mathrm{m}$ and $\mathrm{MAE} = 2.29\,\mathrm{m}$. Like in the 5G evaluation, Con. AE, Sammon mapping and PCA achieve the worst results in local and global similarity of $\mathrm{TW} < 0.93$, $\mathrm{CT} > 0.97$, $\mathrm{CE90} > 4.70\,\mathrm{m}$ and $\mathrm{MAE} > 2.50\,\mathrm{m}$.

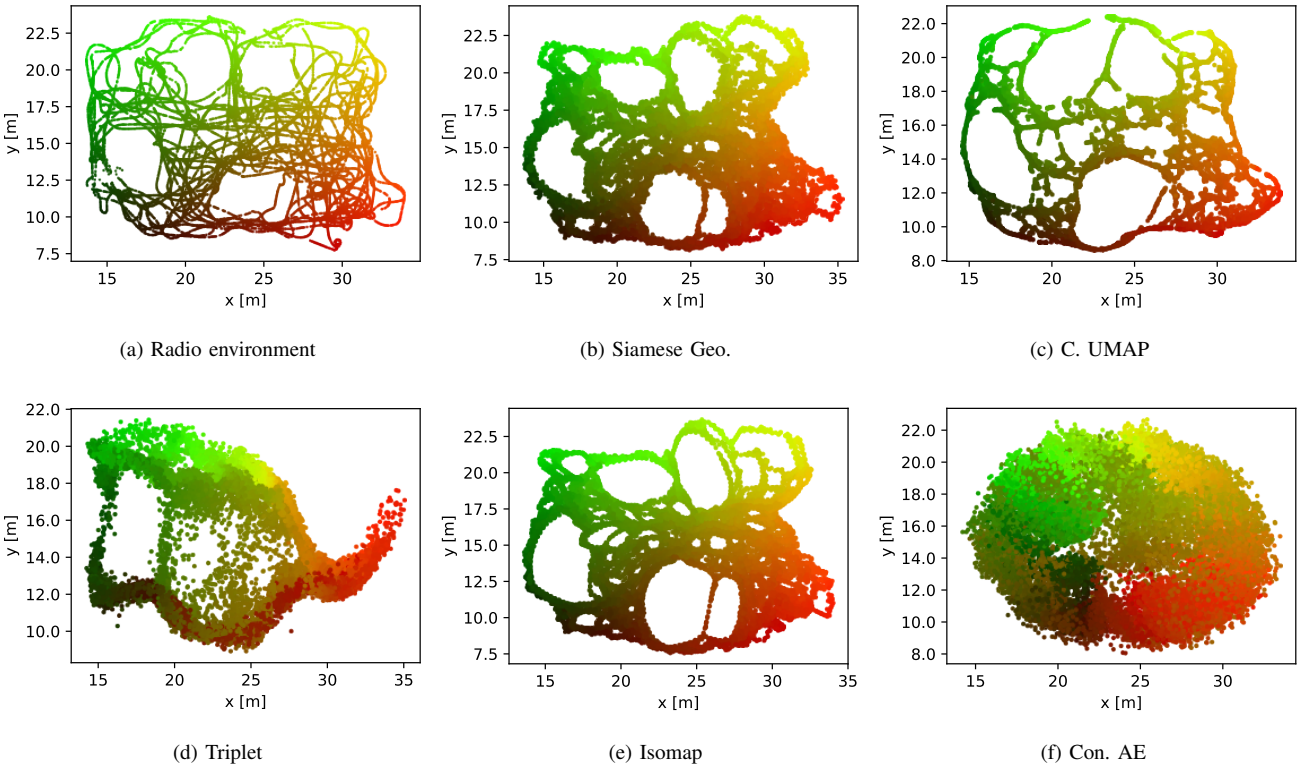
Fig. 9 visualizes the channel charts for the UWB data set. Graph (a) shows the reference radio environment with a color gradient, while the Graphs (b)–(f) show the results of the channel charts for different methods. We can clearly see that our method and Isomap model the global structure of the environment well. There is no overlap of the gradients and the majority of the empty areas in the original environment are also present. However, due to the empty areas in the environment, the geodesic distances are overestimated as no direct physical connection via the neighborhood is possible for all data points, which leads to several over-estimations of the geodesic distance to the physical distances. The effect can be seen in the area at $x = 15\,\mathrm{m}$ and $y = 15\,\mathrm{m}$, where the channel-chart coordinates are pushed towards the left-hand side.

The channel chart generated by C. UMAP is similarly accurate as the one generated by our approach and Isomap. The chart is smoother but maps near datapoints on similar trajectories, with some discontinuities. In contrast, our geodesic Siamese network shows no discontinuities of the trajectories. While C. UMAP can achieve a globally consistent map in the UWB scenario UMAP does not ensure this behaviour as shown in the 5G evaluation, cf. 8c. We therefore think that C. UMAP is not ideal for CC-assisted localization, which needs a continuous globally consistent chart. The other channel charts, i.e., triplet and Con. AE, cannot recover the global geometry well, while the triplet algorithm provides more structure of the environment. This may be caused by the restrictive assumptions of the triplet generation and the limitations of the distance metric for far points for Con. AE.

Also here, the generalization abilities are good for all of the algorithms, showing no significant differences of the performance using the test data set. This indicates that the algorithms are not overfitting to the training trajectories and can predict also well for unseen data in the same environment.

D. SUPERVISED FP vs. CC-ASSISTED FP

In this section, we compare CC-assisted FP with supervised FP. We employed the supervised FP approach proposed in



(a) Radio environment (b) Siamese Geo. (c) C. UMAP

(d) Triplet (e) Isomap (f) Con. AE

FIGURE 9. Results of CC for the UWB data set. Graph (a) shows the recorded training data in the real-world environment, whereas the color gradient indicates the relation of the positions of the channel charts to the real positions. Graphs (b)–(f) show the channel charts after the affine transformation to the global coordinate frame.

[13], which uses the same data and preprocessing as described in Section B. As a neural-network architecture, we exploit the same as for the Siamese network described in Section A. In this evaluation, all models are trained on the training data set and tested on the test data set. For our CC-assisted FP, we used labeled data samples to estimate the linear transformation of the channel chart to the real-world coordinate frame.

The results for the 5G radio system are shown in Fig. 10 (left), where the supervised FP is shown in blue and the CC-assisted FP is depicted in orange. The x-axis shows the employed number of labeled training samples, whereas the y-axis shows the CE90. The training was repeated for 5 times to avoid any bias from the sample selection. The mean results are shown as a solid line, and the upper and lower limit of the estimations are also given. It can clearly be seen that CC-assisted FP outperforms supervised FP for a low number of labeled samples. Only 20 labeled samples are needed to achieve the maximum accuracy, while also noteworthy, 10 samples are sufficient such that the accuracy is very close to the optimal performance. With only 3 and 5 samples, the estimation of the affine transformation is very sensitive to errors in the selected points in the channel chart. In contrast to CC-assisted FP, supervised FP can achieve higher accuracies of a $\mathrm{CE90} = 1.03\,\mathrm{m}$ with a high amount of samples (10,240). However, with only a few samples, the supervised FP accuracy is very low with, e.g., an $\mathrm{CE90} = 11.0\,\mathrm{m}$ for 20 samples. About 1,500 samples are needed at least to achieve similar results as CC-assisted FP.

We can see similar results for the UWB radio setup, see Fig. 10 (right). CC-assisted FP is more stable with 3 and 5 labeled data samples, as the channel chart is less noisy compared to the 5G experiment. This leads to a lower likelihood for the selection of erroneous channel-chart points. However, similarly, 10 samples are sufficient to achieve a stable linear transformation to the real-world coordinate frame. The supervised FP shows the same behaviour as in the 5G experiment. Higher accuracies of a $\mathrm{CE90} = 0.42\,\mathrm{m}$ can be achieved with a high number of labeled training samples (10,240), while the localization fails with only very few annotated training data. To achieve similar results with supervised FP about 1,000 labeled samples are needed. In contrast to the 5G radio system, fewer samples are needed in this case as the environment is smaller.

VII. DISCUSSION & FUTURE WORK

The experiments have shown that neighborhood-graph-based algorithms, e.g., C. UMAP, Isomap, and the proposed geodesic Siamese network, are superior in the generation of channel charts compared to the other algorithms. We think that this is due to the fact that they only consider local distances, which is advantageous as the CSI distance is proportional to the physical distance in the proximity of

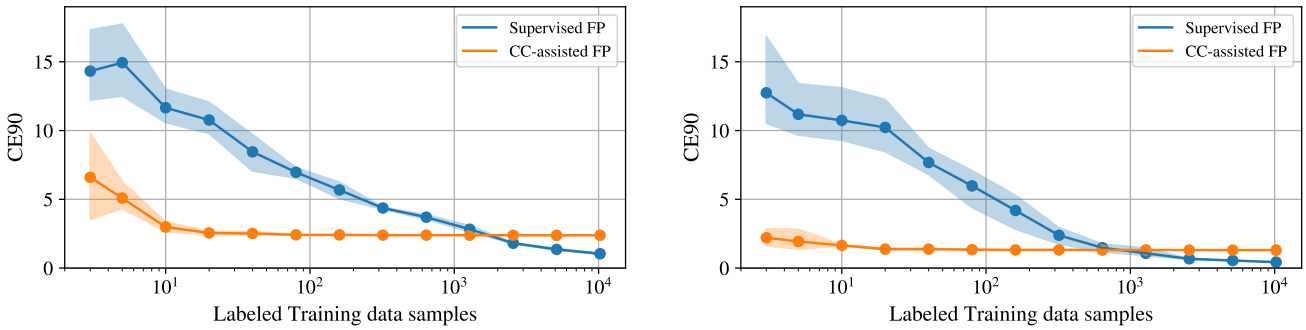


FIGURE 10. CE90 error of supervised FP and CC-assisted for the 5G (left) and the UWB (right) experiment with different amount of labeled training samples.

two CSI measurements. This leads to the optimization of the channel chart considering only the valid local connections between CSIs. However, this also means that, ideally, a uniform distribution of high data density should be available to create a valid global neighborhood graph. As the idea of CC is that every user in the environment can contribute to the underlying data set, the amount of collected data will grow rapidly, which leads to a high data density. However, this does not necessarily imply that the density of data has a uniform distribution in the radio environment. While this is not a problem, as long as the data density is high, methods that rely on proximity distances, such as nearest-neighbor approaches, have to calculate a distance matrix, which has a complexity of $\mathcal{O}(n^2)$.

While there are good approximations for neighborhood graphs [54] the complexity is still non-linear, which restricts the most promising algorithms to small data sets. The selection of the most necessary information of the recorded data is therefore important, but a non-trivial problem, as a random selection of samples might lead to redundant CSIs at high density regions of the environment and therefore also to a under representation of low-density regions. As the collection of data is unsupervised, a spatial selection of data is impossible, which makes the sample selection challenging.

Another challenge is to exploit CC for positioning. As the channel chart only reflects the local geometry, a transformation into the global real-world coordinate frame is necessary. In Section D, we have shown that a simple affine transformation, estimated using only 10 to 20 reference-coordinates samples, is already enough to provide high localization accuracies, while supervised FP needs 50 times more labeled data samples to achieve similar results. However, supervised FP can achieve an even higher localization performance if a high number of labeled data samples is available. As shown in [13], supervised FP needs a good coverage of labeled data samples in the radio environment due to the restricted interpolation abilities. This means that for larger areas the numbers of needed labeled training samples also increases, which is not the case for CC-assisted FP as only an alignment of coordinates frames is needed. We therefore think that CC-assisted FP is especially useful when only a low amount of labeled data samples is available and in large areas, as the linear transformation only requires a small amount of annotated data, which does not scale with the size of the environment.

A further problem is that the consistency of the channel chart mainly depends on the data collected in a certain scenario. Environmental changes may alter the environment-specific CSI, which leads to errors in the prediction. This is a common problem in FP models [13] and makes an identification of environmental changes and an update necessary over time. This might also be needed for channel charts rendering a CC life-cycle management essential for a robust sage for downstream tasks.

VIII. CONCLUSION

In this work, we propose a CC-assisted localization method for synchronized multi-base-station SISO radio setups with high bandwidth, a 5G and a UWB setup. We derived a novel CSI distance metric, which allows the modeling of globally consistent channel charts, enabling high precision CC-assisted FP. Our proposed method, based on a Siamese network can achieve localization accuracies of $0.69\,\mathrm{m}$ for the UWB and $1.4\,\mathrm{m}$ for the 5G setup in challenging NLoS-dominated indoor environments by only applying a linear transformation after the generation of the channel chart.

An extensive study has shown that our proposed method outperforms the state-of-the-art algorithms, since neighborhood-graph-based algorithms like our method can achieve the best results. In contrast to CC-assisted FP, supervised FP needs 50 times more labeled data samples to achieve similar results. This indicates that CC has a very high potential for unsupervised FP-based localization methods to lower the effort for data recording, as only very few ground-truth data are needed for the coordinate-system transformation of the chart from the local to the real-world frame.

APPENDIX A
MODEL FOR TDOA-BASED MEASUREMENTS

The CIR distance for TDoA-based localization system is derived in the following. Without loss of generality, we assume B_{ref} as the time-reference base station for TDoA-based measurements. Moreover, B_{ref} is always chosen such that it provides the shortest ToA. The difference between two TDoA measurements ($T_{\mathsf{bias},\ell}$ and $T_{\mathsf{bias},\ell'}$ are some measurement bias) is given by

$$\begin{aligned}\Delta T_{n,ij}^{(k)} &\stackrel{\text{def}}{=} \big|T_{0,i}^{(k)} + T_{\mathsf{bias},\ell} - (T_{0,i}^{\mathsf{ref}} + T_{\mathsf{bias},\ell}) \\ &\qquad - (T_{0,j}^{(k)} + T_{\mathsf{bias},\ell'} - (T_{0,j}^{\mathsf{ref}} + T_{\mathsf{bias},\ell'}))\big| \\ &= \big|T_{0,i}^{(k)} - T_{0,i}^{\mathsf{ref}} - (T_{0,j}^{(k)} - T_{0,j}^{\mathsf{ref}})\big| \\ &= \big|T_{0,i}^{(k)} - T_{0,j}^{(k)} + T_{0,j}^{\mathsf{ref}} - T_{0,i}^{\mathsf{ref}}\big| \\ &= \big|\Delta t_{n,ij}^{(k)} + \Delta t_{n,ji}^{\mathsf{ref}}\big| \,,\end{aligned}$$

which can be expressed as

$$\begin{aligned}&= \Bigg|\pm\sqrt{\left(d_{\mathsf{phy}}^{(k)}\right)^2/\mathrm{c}^2 - \left(\epsilon_{n,ij}^{(k)}\right)^2} \\ &\qquad \pm\sqrt{\left(d_{\mathsf{phy}}^{\mathsf{ref}}\right)^2/\mathrm{c}^2 - \left(\epsilon_{n,ij}^{\mathsf{ref}}\right)^2}\Bigg| \,. \end{aligned} \tag{25}$$

When $\epsilon_{n,ij}^{(k)}$ is maximum, then $\Delta T \in [0, d_{\mathsf{phy}}/\mathrm{c}]$ which implies $\epsilon_{n,ij}^{\mathsf{ref}} \in [0, d_{\mathsf{phy}}/\mathrm{c}]$. Similarly, when $\epsilon_{n,ij}^{\mathsf{ref}}$ is maximum, $\epsilon_{n,ij}^{(k)} \in [0, d_{\mathsf{phy}}/\mathrm{c}]$. This means that

$$2\, d_{\mathsf{phy}} > \mathrm{c}\left(\epsilon_{n,ij}^{(k)} + \epsilon_{n,ij}^{\mathsf{ref}}\right) \,, \tag{26}$$

since equality in (26) can only hold when both base stations are at the exact same coordinates.

Following the same derivation as in Section A, we now re-define the distance metric (13) using TDoAs, i.e., $\Delta T_{n,ij}^{(k)}$:

$$d_{\mathsf{s}}(\tilde{\boldsymbol{H}}_i, \tilde{\boldsymbol{H}}_j) \stackrel{\text{def}}{=} \sum_{k=0}^{N_{\mathsf{b}}-1} \sum_{n=0}^{N_{\mathsf{p}}-1} \Delta T_{n,ij}^{(k)} \,. \tag{27}$$

Again, we have a monotonically increasing mean and standard deviation, and a lower bound of the error.

APPENDIX B
LINEARITY OF THE CIR DISTANCE

If we could extract all multipath delay information from the received CSIs we could apply (13), which would lead to a linear correlation of the CIR distance to the physical distance. In this work, we use the approximation defined in (15), which is non-linear if we use a typical pulse shape like a sinc function.

However, we can show that the approximated CIR distance is linear for small time-differences $\Delta t_{n,ij}^{(k)}$. Without loss of generality, we assume to have only one base station, i.e., $N_{\mathsf{b}} = 1$ and a single path arriving at the receiver, i.e., $N_{\mathsf{p}} = 1$.[4] By using (15) and the inverse triangle inequality we have

$$d_{\mathsf{s}}(\tilde{\boldsymbol{h}}_i, \tilde{\boldsymbol{h}}_j) \le \sum_{t=0}^{T-1} \left|\tilde{h}_i^{(1)}(t) - \tilde{h}_j^{(1)}(t)\right| \,. \tag{28}$$

If we assume to have only a single path arriving, e.g., the LoS component, we can substitute $\tilde{h}$ by the sinc function

$$d_{\mathsf{s}}(\tilde{\boldsymbol{h}}_i, \tilde{\boldsymbol{h}}_j) \le \sum_{t=0}^{T-1} \left|\mathrm{sinc}(t) - \mathrm{sinc}(t - \Delta t_{0,ij}^{(1)})\right| \,. \tag{29}$$

Here, we omit the bandwidth of the signal and assume to have a equal amplitude, which is valid for small $\Delta t_{0,ij}^{(1)}$. By using the Taylor series expansion we get

$$\begin{aligned}d_{\mathsf{s}}(\tilde{\boldsymbol{h}}_i, \tilde{\boldsymbol{h}}_j) \le \sum_{t=0}^{T-1} \Bigg|&\sum_{n=0}^{\infty} \frac{(-1)^n t^{2n}}{(2n+1)!} \\ &- \sum_{n=0}^{\infty} \frac{(-1)^n (t - \Delta t_{0,ij}^{(1)})^{2n}}{(2n+1)!}\Bigg| \,. \end{aligned} \tag{30}$$

If we have $|\Delta t_{0,ij}^{(1)}| \ll |t|$ we can use the binomial approximation defined in Lemma B.1 to get

$$\begin{aligned}d_{\mathsf{s}}(\tilde{\boldsymbol{h}}_i, \tilde{\boldsymbol{h}}_j) &\le \sum_{t=0}^{T-1} \Bigg|\sum_{n=0}^{\infty} \frac{(-1)^n t^{2n}}{(2n+1)!} \\ &\qquad - \sum_{n=0}^{\infty} \frac{(-1)^n (t^{2n} - 2nt^{2n-1}\Delta t_{0,ij}^{(1)})}{(2n+1)!}\Bigg| \qquad (31) \\ &= \Big|\Delta t_{0,ij}^{(1)}\Big| \underbrace{\sum_{t=0}^{T-1} \left|\sum_{n=0}^{\infty} \frac{(-1)^n 2nt^{2n-1}}{(2n+1)!}\right|}_{=\text{constant}} \,, \qquad (32)\end{aligned}$$

which is linear for small $\Delta t_{0,ij}^{(1)}$ with a constant slope. This means that also the sum of $\Delta t_{n,ij}^{(k)}$ for the base stations k and the received paths n is linear, as long as $\Delta t_{n,ij}^{(k)}$ is small. As $\Delta t_{n,ij}^{(k)}$ is bounded by the displacements of the CSI measurements $\tilde{\boldsymbol{h}}_i$ and $\tilde{\boldsymbol{h}}_j$, we have a linear CIR distance for small displacements.

Lemma B.1:
For $|\Delta t| \ll |t|$ we can use the following approximation of the binomial function

$$\begin{aligned}f(\Delta t) &= (t - \Delta t)^{2n} && (33)\\ f'(\Delta t) &= -2n(t - \Delta t)^{2n-1} && (34)\\ f'(0) &= -2nt^{2n-1} && (35)\\ f(\Delta t) &\approx f(0) + f'(0)(\Delta t - 0) && (36)\\ &= t^{2n} - 2nt^{2n-1}\Delta t \,. && (37)\end{aligned}$$

The error of the approximation is defined as

$$e(\Delta t) = (t - \Delta t)^{2n} - (t^{2n} - 2nt^{2n-1}\Delta t) \,, \tag{38}$$

[4]The base station index $k = 1$ is used for the derivation and the same applies for all indices k.

where for small values of Δt, the error approaches 0, i.e.,

$$\lim_{\Delta t \to 0} e(\Delta t) = 0 \, . \tag{39}$$

REFERENCES

[1] C. Laoudias, A. Moreira, S. Kim, S. Lee, L. Wirola, and C. Fischione, "A Survey of Enabling Technologies for Network Localization, Tracking, and Navigation," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 3607–3644, 2018.

[2] Y. Wu, F. Tang, and H. Li, "Image-based camera localization: an overview," *Visual Computing for Industry, Biomedicine, and Art*, vol. 1, no. 8, 2018.

[3] M. Elhousni and X. Huang, "A Survey on 3D LiDAR Localization for Autonomous Vehicles," in *IEEE Intelligent Vehicles Symposium (IV)*. Las Vegas, NV, USA: IEEE, 2020, pp. 1879–1884.

[4] A. M. Rahman, T. Li, and Y. Wang, "Recent Advances in Indoor Localization via Visible Lights: A Survey," *Sensors*, vol. 20, no. 5, p. 1382, 2020.

[5] N. Saeed, H. Nam, T. Y. Al-Naffouri, and M.-S. Alouini, "A State-of-the-Art Survey on Multidimensional Scaling-Based Localization Techniques," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3565–3583, 2019.

[6] F. Pang, K. Doğançay, N. H. Nguyen, and Q. Zhang, "AOA Pseudolinear Target Motion Analysis in the Presence of Sensor Location Errors," *IEEE Transactions on Signal Processing*, vol. 68, pp. 3385–3399, 2020.

[7] H.-C. Yen, L.-Y. O. Yang, and Z.-M. Tsai, "3-d indoor localization and identification through rssi-based angle of arrival estimation with real wi-fi signals," *IEEE Transactions on Microwave Theory and Techniques*, vol. 70, no. 10, pp. 4511–4527, 2022.

[8] W. Gifford, D. Dardari, and M. Win, "The Impact of Multipath Information on Time-of-Arrival Estimation," *IEEE Transactions on Signal Processing*, vol. 70, pp. 31–46, 2020.

[9] C. E. O'Lone, H. S. Dhillon, and R. M. Buehrer, "Characterizing the First-Arriving Multipath Component in 5G Millimeter Wave Networks: TOA, AOA, and Non-Line-of-Sight Bias," *IEEE Transactions on Wireless Communications*, vol. 21, no. 3, pp. 1602–1620, 2021.

[10] M. Stahlke, S. Kram, C. Mutschler, and T. Mahr, "NLOS Detection using UWB Channel Impulse Responses and Convolutional Neural Networks," in *International Conference on Localization and GNSS (ICL-GNSS)*. Tampere, Finland: IEEE, 2020.

[11] M. Stahlke, S. Kram, F. Ott, T. Feigl, and C. Mutschler, "Estimating TOA Reliability With Variational Autoencoders," *IEEE Sensors Journal*, vol. 22, no. 6, pp. 5133–5140, 2021.

[12] A. Niitsoo, T. Edelhäußer, E. Eberlein, N. Hadaschik, and C. Mutschler, "A Deep Learning Approach to Position Estimation from Channel Impulse Responses," *Sensors*, vol. 19, no. 5, p. 1064, 2019.

[13] M. Stahlke, T. Feigl, M. H. C. García, R. A. Stirling-Gallacher, J. Seitz, and C. Mutschler, "Transfer Learning to adapt 5G AI-based Fingerprint Localization across Environments," in *IEEE 95th Vehicular Technology Conference (VTC2022-Spring)*. Helsinki, Finland: IEEE, 2022.

[14] K. Liu, H. Zhang, J. K.-Y. Ng, Y. Xia, L. Feng, V. C. Lee, and S. H. Son, "Toward Low-Overhead Fingerprint-Based Indoor Localization via Transfer Learning: Design, Implementation, and Evaluation," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 3, pp. 898–908, 2017.

[15] S. De Bast, A. P. Guevara, and S. Pollin, "CSI-based Positioning in Massive MIMO systems using Convolutional Neural Networks," in *IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*. Antwerp, Belgium: IEEE, 2020.

[16] M. Widmaier, M. Arnold, S. Dorner, S. Cammerer, and S. ten Brink, "Towards Practical Indoor Positioning Based on Massive MIMO Systems," in *IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*. Honolulu, HI, USA: IEEE, 2019.

[17] N. Singh, S. Choe, and R. Punmiya, "Machine learning based indoor localization using wi-fi rssi fingerprints: an overview," *IEEE Access*, 2021.

[18] B. Jia, J. Liu, T. Feng, B. Huang, T. Baker, and H. Tawfik, "Ttsl: An indoor localization method based on temporal convolutional network using time-series rssi," *Computer Communications*, vol. 193, pp. 293–301, 2022.

[19] C. Studer, S. Medjkouh, E. Gonultaş, T. Goldstein, and O. Tirkkonen, "Channel Charting: Locating Users Within the Radio Environment Using Channel State Information," *IEEE Access*, vol. 6, pp. 47 682–47 698, 2018.

[20] A. Schwartz and R. Talmon, "Intrinsic isometric manifold learning with application to localization," *SIAM Journal on Imaging Sciences*, vol. 12, no. 3, pp. 1347–1391, 2019.

[21] F. Ghazvinian Zanjani, I. Karmanov, H. Ackermann, D. Dijkman, S. Merlin, M. Welling, and F. Porikli, "Modality-agnostic topology aware localization," *Advances in Neural Information Processing Systems*, vol. 34, pp. 10 457–10 468, 2021.

[22] E. Lei, O. Castañeda, O. Tirkkonen, T. Goldstein, and C. Studer, "Siamese Neural Networks for Wireless Positioning and Channel Charting," in *57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. Monticello, IL, USA: IEEE, 2019, pp. 200–207.

[23] P. Ferrand, A. Decurninge, L. G. Ordoñez, and M. Guillaud, "Triplet-Based Wireless Channel Charting: Architecture and Experiments," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 8, pp. 2361–2373, 2021.

[24] J. Deng, O. Tirkkonen, J. Zhang, X. Jiao, and C. Studer, "Network-side Localization via Semi-Supervised Multi-point Channel Charting," in *International Wireless Communications and Mobile Computing (IWCMC)*. Harbin City, China: IEEE, 2021, pp. 1654–1660.

[25] Q. Zhang and W. Saad, "Semi-Supervised Learning for Channel Charting-Aided IoT Localization in Millimeter Wave Networks," in *IEEE Global Communications Conference (GLOBECOM)*. Madrid, Spain: IEEE, 2021.

[26] L. Ribeiro, M. Leinonen, H. Djelouat, and M. Juntti, "Channel Charting for Pilot Reuse in mMTC with Spatially Correlated MIMO Channels," in *IEEE Globecom Workshops*. Taipei, Taiwan: IEEE, 2020.

[27] H. Al-Tous, O. Tirkkonen, and J. Liang, "Adaptive Sector Splitting based on Channel Charting in Massive MIMO Cellular Systems," in *IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)*. Helsinki, Finland: IEEE, 2021.

[28] H. Al-Tous, T. Ponnada, C. Studer, and O. Tirkkonen, "Multipoint channel charting-based radio resource management for v2v communications," *EURASIP Journal on Wireless Communications and Networking*, no. 132, 2020.

[29] T. Ponnada, H. Al-Tous, and O. Tirkkonen, "Location-Free Beam Prediction in mmWave Systems," in *IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)*. Helsinki, Finland: IEEE, 2021, pp. 1–6.

[30] T. Ponnada, P. Kazemi, H. Al-Tous, Y.-C. Liang, and O. Tirkkonen, "Best Beam Prediction in Non-Standalone mm Wave Systems," in *Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*. Porto, Portugal: IEEE, 2021, pp. 532–537.

[31] P. Kazemi, T. Ponnada, H. Al-Tous, Y.-C. Liang, and O. Tirkkonen, "Channel Charting Based Beam SNR Prediction," in *Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*. Porto, Portugal: IEEE, 2021, pp. 72–77.

[32] J. Deng, S. Medjkouh, N. Malm, O. Tirkkonen, and C. Studer, "Multipoint Channel Charting for Wireless Networks," in *52nd Asilomar Conference on Signals, Systems, and Computers*. Pacific Grove, CA, USA: IEEE, 2018, pp. 286–290.

[33] L. Le Magoarou, "Efficient Channel Charting via Phase-Insensitive Distance Computation," *IEEE Wireless Communications Letters*, vol. 10, no. 12, pp. 2634–2638, 2021.

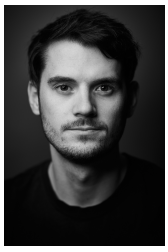
[34] P. Agostini, Z. Utkovski, and S. Stańczak, "Channel Charting: an Euclidean Distance Matrix Completion Perspective," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Barcelona, Spain: IEEE, 2020, pp. 5010–5014.

[35] R. Schmidt, "Multiple Emitter Location and Signal Parameter Estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[36] D. Spielman, A. Paulraj, and T. Kailath, "Performance Analysis of the MUSIC Algorithm," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 11. Tokyo, Japan: IEEE, 1986, pp. 1909–1912.

[37] T. Ponnada, H. Al-Tous, O. Tirkkonen, and C. Studer, "An Out-of-Sample Extension for Wireless Multipoint Channel Charting," in *International Conference on Cognitive Radio Oriented Wireless Networks*. Poznan, Poland: Springer, 2019, pp. 208–217.

[38] C. Geng, H. Huang, and J. Langerman, "Multipoint Channel Charting With Multiple-Input Multiple-Output Convolutional Autoencoder," in *IEEE/ION Position, Location and Navigation Symposium (PLANS)*. Portland, OR, USA: IEEE, 2020, pp. 1022–1028.
[39] P. Huang, O. Castañeda, E. Gönültaş, S. Medjkouh, O. Tirkkonen, T. Goldstein, and C. Studer, "Improving Channel Charting with Representation-Constrained Autoencoders," in *IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. Cannes, France: IEEE, 2019.
[40] P. Ferrand, A. Decurninge, L. G. Ordoñez, and M. Guillaud, "Triplet-Based Wireless Channel Charting," in *IEEE Global Communications Conference (GLOBECOM)*. Taipei, Taiwan: IEEE, 2020.
[41] B. Rappaport, E. Gönültaş, J. Hoydis, M. Arnold, P. K. Srinath, and C. Studer, "Improving Channel Charting using a Split Triplet Loss and an Inertial Regularizer," in *17th International Symposium on Wireless Communication Systems (ISWCS)*. Berlin, Germany: IEEE, 2021.
[42] F. Euchner, P. Stephan, M. Gauger, S. Dörner, and S. Ten Brink, "Improving Triplet-Based Channel Charting on Distributed Massive MIMO Measurements," in *IEEE 23rd International Workshop on Signal Processing Advances in Wireless Communication (SPAWC)*. Oulu, Finland: IEEE, 2022.
[43] T. Yassine, L. Le Magoarou, S. Paquelet, and M. Crussière, "Leveraging triplet loss and nonlinear dimensionality reduction for on-the-fly channel charting," in *2022 IEEE 23rd International Workshop on Signal Processing Advances in Wireless Communication (SPAWC)*. IEEE, 2022, pp. 1–5.
[44] P. Agostini, Z. Utkovski, S. Stańczak, A. A. Memon, B. Zafar, and M. Haardt, "Not-Too-Deep Channel Charting (N2D-CC)," in *IEEE Wireless Communications and Networking Conference (WCNC)*. Austin, TX, USA: IEEE, 2022, pp. 2160–2165.
[45] J. Pihlajasalo, M. Koivisto, J. Talvitie, S. Ali-Löytty, and M. Valkama, "Absolute Positioning with Unsupervised Multipoint Channel Charting for 5G Networks," in *IEEE 92nd Vehicular Technology Conference (VTC2020-Fall)*. Victoria, BC, Canada: IEEE, 2020.
[46] S. Kram, C. Kraus, M. Stahlke, T. Feigl, J. Thielecke, and C. Mutschler, "Delay Estimation in Dense Multipath Environments using Time Series Segmentation," in *IEEE Wireless Communications and Networking Conference (WCNC)*. Austin, TX, USA: IEEE, 2022, pp. 1671–1676.
[47] J. B. Tenenbaum, V. d. Silva, and J. C. Langford, "A Global Geometric Framework for Nonlinear Dimensionality Reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.
[48] E. W. Dijkstra, "A Note on Two Problems in Connexion with Graphs," *Numerische mathematik*, vol. 1, no. 1, pp. 269–271, 1959.
[49] D. E. Kirk, *Optimal Control: An Introduction*. Prentice-Hall, 1970.
[50] V. De Silva and J. B. Tenenbaum, "Sparse multidimensional scaling using landmark points," Technical report, Stanford University, Tech. Rep., 2004.
[51] G. Pai, R. Talmon, A. Bronstein, and R. Kimmel, "DIMAL: Deep Isometric Manifold Learning Using Sparse Geodesic Sampling," in *IEEE Winter Conference on Applications of Computer Vision (WACV)*. Waikoloa, HI, USA: IEEE, 2019, pp. 819–828.
[52] D. Chicco, "Siamese Neural Networks: An Overview," *Artificial Neural Networks*, pp. 73–94, 2021.
[53] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Deep learning for time series classification: a review," *Data mining and knowledge discovery*, vol. 33, no. 4, pp. 917–963, 2019.
[54] M. Wang, X. Xu, Q. Yue, and Y. Wang, "A Comprehensive Survey and Experimental Comparison of Graph-Based Approximate Nearest Neighbor Search," *Proc. VLDB Endow.*, vol. 14, no. 11, p. 1964–1978, jul 2021.



Maximilian Stahlke received his Master's degree in Electronic and Mechatronic Systems at the Institute of Technology Georg Simon Ohm, Germany, in 2020. Since 2020 he works at the precise positioning and analytics department at Fraunhofer IIS in the Hybrid Positioning & Information Fusion group. His research interests are hybrid positioning for radio-based localization systems with the focus on model- and data-driven information fusion.

George Yammine received the B.E. degree in 2010 in computer and communications engineering from Notre Dame University, Louaize, Lebanon, the M.Sc. in communications technology and the Dr.-Ing. degrees from Ulm University, Ulm, Germany, in 2014 and 2020, respectively. His dissertation was on noncoherent detection in massive MIMO systems. From 2005 until 2020, he was a Research Assistant at the Institute of Communications Engineering, Ulm University.

In 2021, he joined the Positioning in Wireless Networks group at Fraunhofer Institute for Integrated Circuits IIS, Nürnberg, Germany. Currently, he is a Senior Scientist working on different aspects of 6G mobile communications and positioning. His research interests include noncoherent detection techniques, signal processing, advanced positioning techniques, multi-user detection, machine learning applied in communications and signal processing, reconfigurable intelligent surfaces and massive MIMO systems.

Dr. Yammine was the recipient of the ARGUS Science Award 2014 from Airbus Defence and Space.

Tobias Feigl received his Ph.D. degree in Computer Science from the Friedrich-Alexander-University Erlangen-Nuremberg (FAU) in 2021 and his Masters degree from the University of Applied Sciences Erlangen-Nuremberg, Germany, in 2017. In 2009 he joined the Fraunhofer Institute for Integrated Circuits (IIS) Nuremberg. In parallel, since 2017 he is a lecturer at the Computer Science department (Programming Systems lab) at FAU, where he gives courses on machine and deep learning and supervises related qualification work. His research combines human-computer interaction and machine learning to improve localization.

Bjoern M. Eskofier heads the Machine Learning and Data Analytics (MaD) Lab at the Friedrich-Alexander-University Erlangen-Nürnberg (FAU). He is also the current speaker of FAU's Department Artificial Intelligence in Biomedical Engineering (AIBE) and of the German Ministry of Economic Affairs and Climate Action GAIA-X usecase project "TEAM-X" and the co-speaker of the German Research Foundation collaborative research center "EmpkinS" (SFB 1483).

Dr. Eskofier studied Electrical Engineering at the FAU and graduated in 2006. He then studied under the supervision of Prof. Dr. Benno Nigg at the University of Calgary (Canada). He authored more than 300 peer reviewed articles, submitted 7 patent applications, started three spinoff startup companies, and is in a supporting role for further startups. He won several medical-technical research awards, including the "Curious Minds" award in "Life Sciences" by Manager Magazin and Merck. In 2016, he was a visiting professor in Dr. Paolo Bonato's Motion Analysis Lab at Harvard Medical School (February-March), and in 2018, he was a visiting professor in Dr. Alex "Sandy" Pentland's Human Dynamics group at MIT Media Lab (March-August).

Bjoern Eskofier has defined his research and entrepreneurial agenda to revolve around contributions to a "Digital Health Ecosystem", where patients are connected to other stakeholders within the Healthcare system using digital support tools. His digital health research philosophy is that only multidisciplinary teams of engineers, medical experts, industry representatives and entrepreneurs will have the tools to actually implement changes in Healthcare.

Christopher Mutschler leads the precise positioning and analytics department at Fraunhofer IIS. Prior to that, Christopher headed the Machine Learning & Information Fusion group. He gives lectures on machine learning at the FAU Erlangen-Nürnberg (FAU), from which he also received both his Diploma and Ph.D. in 2010 and 2014, respectively. Christopher's research combines machine learning with radio-based localization.