An Ultra-Efficient Approach for High-Resolution MIMO Radar Imaging of Human Hand Poses

Johanna Bräunig, *Graduate Student Member, IEEE*, Vanessa Wirth,
Christoph Kammel, *Graduate Student Member, IEEE*,
Christian Schüßler, *Graduate Student Member, IEEE*,
Ingrid Ullmann, *Member, IEEE*, Marc Stamminger,
and Martin Vossiek, *Fellow, IEEE*

Abstract— The capturing of hands, including their poses, shapes and motions, has numerous potential applications, such as human-machine interfaces and medical use cases. However, in the radar context, most existing methods only allow for the recognition of dynamic hand gestures based on Doppler evaluations due to the respective systems' limited lateral resolution. Radar-based high-resolution three-dimensional (3D) imaging using multiple-input multiple-output (MIMO) radars is currently the state-of-the-art in personnel security scanning. However, the associated imaging techniques suffer from computationally burdensome reconstruction algorithms that sample the entire 3D space of interest, thereby making them less suitable for real-time applications. Moreover, their application in hand motion tracking scenarios is limited by low frame rates that result from a high number of transmit frequencies. Hence, we present an efficient and powerful approach for the radar-based 3D reconstruction of hand poses. The method extends the frequency shift keying continuous wave radar principle and reconstructs the hand surface using only two carrier frequencies. Instead of reconstructing an entire 3D volume, only two single-tone radar images are computed. Depth information is derived from phase differences between corresponding pixels in the images. The approach significantly reduces computational load by three orders of magnitude compared with the state-of-the-art and enables higher frame rates. Within this paper, this novel reconstruction principle is analyzed and compared to a state-of-the-art radar imaging approach using a MIMO radar system with 94 transmitting and 94 receiving antennas. Detailed simulations of point targets and comprehensive measurements demonstrate the excellent imaging performance of our approach.

Index Terms— Radar imaging, hand pose capture, 3D reconstruction, frequency shift keying.

I. Introduction

AUTOMATIC recognition of hand gestures, hand pose estimation, as well as hand tracking have recently become an attractive area of research, as contactless

Johanna Bräunig, Christoph Kammel, Christian Schüßler, Ingrid Ullmann, and Martin Vossiek are with the Institute of Microwaves and Photonics, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), 91058 Erlangen, Germany (e-mail: johanna.braeunig@fau.de).

Vanessa Wirth and Marc Stamminger are with the Chair of Visual Computing, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), 91058 Erlangen, Germany.

human–computer interaction becomes more important [1], [2]. In this regard, two popular research fields have emerged [3]. The first focuses on gesture classification based on the evaluation and extraction of high-level abstract motion or pose information. The second one builds on precise hand pose estimation, i.e., it aims at capturing the correct three-dimensional (3D) information of hand poses and motions. This is also of great interest in medical contexts, as the analysis of hand function can provide information regarding the current health of patients [4], [5], [6].

In the computer vision domain, markerless methods based on RGB cameras, depth cameras, or RGB-D cameras, i.e., a combination of the aforementioned are used for exact hand pose estimation, as well as static and dynamic automated gesture recognition [7], [8], [9]. Computer vision algorithms have a rich history and have been extensively researched. However, optical sensors have the disadvantage of their performance being highly dependent on the lighting conditions of the illuminated scene [10].

For this reason, radar-based methods for the capturing of hand poses or hand motion offer a promising alternative, as they do not suffer from this disadvantage on account of the wavelengths used; therefore, such methods enable a robust precise measurement of distances. Furthermore, unlike computer vision approaches, they are able to directly measure motion by analyzing the Doppler effect, which is beneficial when it comes to the evaluation of dynamic features. The majority of the work published in the radar context focuses on dynamic gesture classification. An overview of the state-of-the-art of hand gesture recognition approaches using radar sensors can be found in [11]. In this context, Google launched a project called Soli, that utilizes a frequency-modulated continuous wave (FMCW) radar in combination with machine learning algorithms for robust dynamic hand gesture recognition and gesture tracking with sub-millimeter accuracy [12]. Further research can be found in [13], [14], [15], [16], [17], [18], and [19], in which micro-Doppler signatures or range-Doppler maps are used to enable automatic hand gesture recognition in combination with machine learning approaches. In [20], multiple scattering points of the hand are extracted from the measurement data of an FMCW multiple-input-multiple-output (MIMO) radar, which are subsequently analyzed in the spectrum, spatial,

and time domains. In [21], an interferometric continuous wave (CW) radar is used to measure the angular velocity of dynamic hand gestures. The commonality among all these methods is that the radar utilized incorporates low lateral resolution, which is why they focus on dynamic rather than static features. In [22], a convolutional neural network is trained and applied to enable improved static hand gesture recognition based on radar data. The data acquisition is done using a mechanical MIMO FMCW radar scanner consisting of two transmitting (Tx) and receiving (Rx) antennas, which enables a multiperspective view and the generation of sufficient training samples. However, within this work as well as in the previously mentioned articles, the authors do not aim for a precise 3D imaging of different hand poses, as is the case in most computer vision approaches.

High-resolution 3D imaging of the human body has been the state-of-the-art for personnel security screening for many years, as radar waves in the millimeter-wave range (mmWave) can penetrate clothes and, therefore, enable the detection of forbidden items, such as weapons [23], [24], [25]. To ensure this, the used radar systems require a high 3D resolution, which can be achieved by large antenna apertures, comparably high carrier frequencies, and high signal bandwidths. As radar imaging requires measurements from many positions and viewing angles, MIMO radars are currently considered the state-of-the-art in building sufficient imaging systems. However, a high number of antennas and the use of high signal bandwidths lead to high requirements for the hardware [23]. Besides that, these radar systems commonly use a backprojection approach to precisely reconstruct a 3D object scene. This procedure is computationally expensive as it raises the need to, first, subdivide the 3D space into smaller sub-volumes (voxels) and, subsequently, reconstruct each of them individually (see Section III). Consequently, traditional imaging radar systems suffer from low measurement rates, computationally burdensome reconstruction algorithms, and high hardware requirements. All these limitations make them unsuitable for use in real-time capable and consumer-friendly gesture sensing systems, or hand-tracking applications.

A radar-based method that can be used in static and dynamic hand gesture classification tasks and is also suitable for precise hand pose tracking applications has not been researched yet. All previous studies either lack the ability to precisely reconstruct hand poses or their computational complexity makes them unattractive for real-time applications. Additionally, low frame rates associated with the latter present challenges for tasks involving direct motion evaluations. Therefore, we propose a new algorithm and measurement principle that is based on the frequency shift keying (FSK) continuous wave (CW) radar principle [26]. Our approach requires only two closely neighbored frequency steps (2FSK) and, thus, requires a small signal bandwidth, thereby reducing hardware requirements. Furthermore, we incorporate a novel 3D reconstruction approach that lowers computational complexity, as it is based on the computation of two single-tone images at an estimated target distance instead of sampling and reconstructing a full 3D space. In addition, measurement acquisition and data transfer times can be significantly decreased, as only two frequency
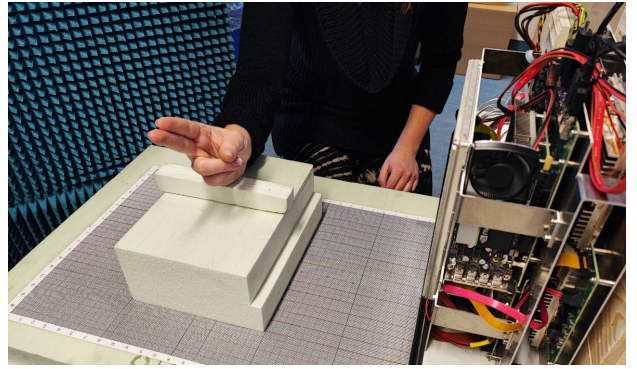

Fig. 1. Measurement setup to capture hand pose and surface. The hand is positioned in front of the MIMO radar using a styrodur structure placed on a reference coordinate system of the x-y plane.

steps are sent out by the antennas. By achieving a significant reduction in measurement time and computational complexity, this novel approach delivers high measurement rates and opens the way toward real-time hand-tracking applications. Furthermore, the precise 3D reconstruction of hands using radar technology allows for the application of thoroughly researched computer vision algorithms. This manuscript makes the following contribution to extant research. We present the theory underlying our fast, precise, and efficient 3D reconstruction technique of the human hand. To evaluate the performance of the novel 2FSK-based imaging principle, we compare its theoretical accuracy, 3D reconstruction results, and computational efficiency to a state-of-the-art radar imaging approach using a stepped frequency continuous wave (SFCW) signal form. Since automatic sign language recognition is an intensly researched application of static hand pose classification [27], [28], [29], we selected three exemplary hand poses of the American Sign Language (ASL) alphabet to compare the 3D imaging results of both approaches. To guarantee a static object scene for the comparison, all hand poses were 3D printed. In addition, we also included measurements that involved a real human hand to further prove the applicability of the novel approach. The manuscript is structured as follows: The measurement setup is described in Section II. In Section III and Section IV, we explain the theoretical background of the SFCW approach, as well as the theory behind our novel 2FSK based algorithm. In Section V, we define the signal and reconstruction parameters relevant for our setup. Thereafter, we analyze the performance of the proposed approach in Section VI. In Section VII, we discuss the results and in Section VIII, we present a summary.

II. Measurement Setup

Fig. 1 shows an overview of the experimental setup, which we used to capture different hand poses. To accurately position the hand in front of the radar, a styrodur structure, including a reference coordinate system of the x-y plane drawn on a sheet of paper, was placed on the table. Two absorber walls were placed behind the table. The hand was positioned roughly at a distance of 30 cm. In order to reliably resolve the fingers of the hand in lateral direction, an image resolution significantly less
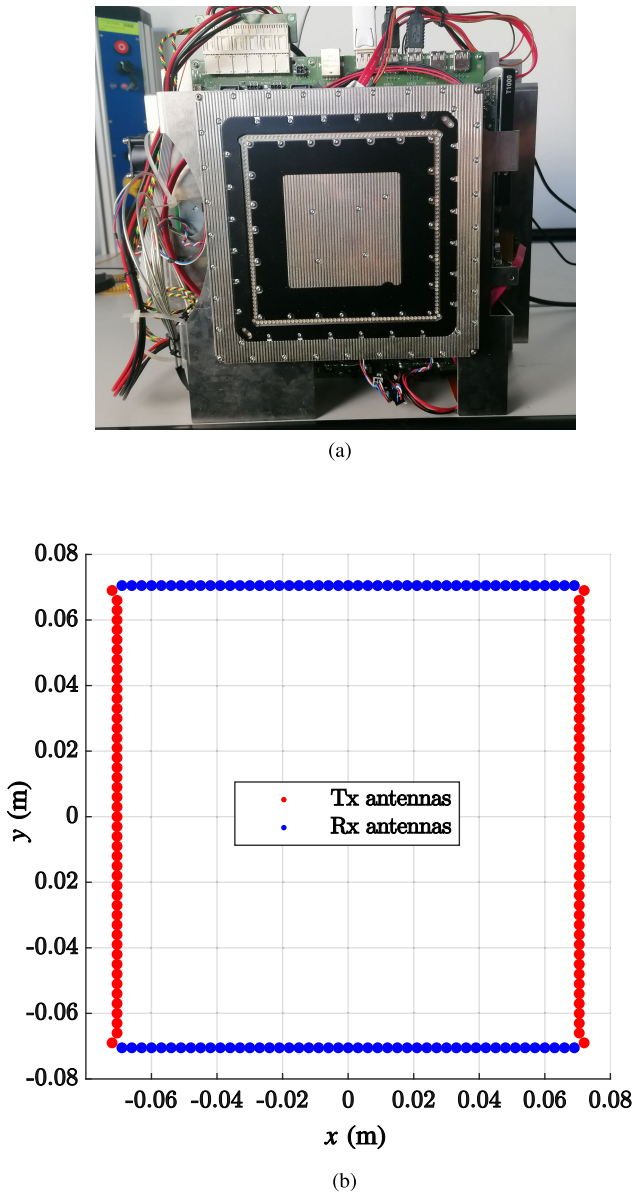
(a)

(b)

Fig. 2. MIMO radar used within the measurement setup. (a) Photo of hardware. (b) Antenna distribution of the respective MIMO radar.

Fig. 3. State-of-the-art MIMO millimeter-wave imaging. The 3D shape of an object within the volume of interest O, with $\vec{r}_{\tilde{v}}$ pointing to all relevant voxels, is obtained by reconstructing a 3D volume $\hat{O}$ and projecting the voxel containing the highest intensity along the z-axis onto the x-y plane while extracting its z-index (maximum projection).

than 1 cm is necessary. Current state-of-the-art 3D radar imaging systems of the human body are typically broadband MIMO systems equipped with a two-dimensional (2D) antenna array. To achieve the same resolutions in both lateral dimensions, a square aperture is a viable choice. The MIMO array used for our setup is a submodul of a commercially available automotive radome tester [30] and is depicted in Fig. 2(a). It has 94 Tx and 94 Rx antennas, which are arranged on a square frame, as illustrated in Fig. 2(b). As a result, we acquire a squared, uniform, and fully equipped virtual array of 8836 antenna pairs. The spacing between adjacent antenna elements is 3 mm, thereby resulting in an array with a physical size of aproximately 14 cm x 14 cm. The size of the resulting virtual MIMO aperture is 28 cm x 28 cm (assuming a far-field approximation). Hence, the lateral resolution in 30 cm distance is 4 mm, as revealed in the next section. The used radar signal modulation is SFCW. The bandwidth and number of
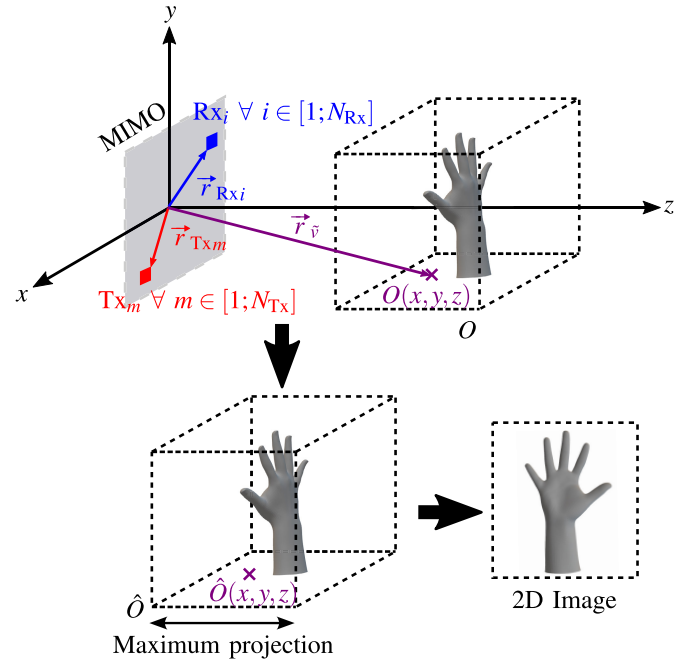
frequency steps are configurable within a range from 72 GHz to 82 GHz. Hence, it is well suited to compare the existing approaches and our novel 2FSK approach. The transmitting power per antenna and transmit frequency can be described by an effective isotropic radiated power of approximately 10 dBm.

III. MILLIMETER-WAVE IMAGING WITH MIMO RADARS

In the following, the state-of-the-art 3D mmWave imaging approach is discussed in more detail using the example of an SFCW signal modulation. The procedure is illustrated in Fig. 3 and explained in the following. Let us assume the volume of interest O containing an object whose 3D dimensions are to be detected. This volume is divided into smaller subvolumes, called voxels. Their size is determined by the sampling dimensions of the applied reconstruction. The vector $\vec{r}_{\tilde{v}}$ points to the different voxel positions $O(x, y, z)$. Within this volume there are N_v voxels located at $\vec{r}_v$ actually containing a scatterer. The scene is scanned with a MIMO radar consisting of N_{Tx} transmitting and N_{Rx} receiving antennas, sending out N_f frequency steps. To obtain a reconstructed voxel $\hat{O}(\vec{r}_{\tilde{v}})$ within the 3D object scene of interest, the baseband signals s_B need to be correlated with complex weights w to correct the respective signal delays before they are coherently summed up across all transmit-receive combinations [24]. This signal processing algorithm is also known as back propagation, backprojection, or digital beamforming [23]. Mathematically, this procedure can be described by [31]

$$\hat{O}(\vec{r}_{\tilde{v}}) = \sum_{k=1}^{N_f} \sum_{m=1}^{N_{Tx}} \sum_{i=1}^{N_{Rx}} s_B(f_k, \vec{r}_{Tx_m}, \vec{r}_{Rx_i})$$
$$\cdot w(f_k, \vec{r}_{Tx_m}, \vec{r}_{Rx_i}, \vec{r}_{\tilde{v}}), \quad (1)$$

where

$$s_B(f_k, \vec{r}_{Tx_m}, \vec{r}_{Rx_i}) \tag{2}$$

$$= \sum_{v=1}^{N_v} \exp(-j2\pi f_k(|\vec{r}_{Tx_m} - \vec{r}_v|$$

$$+ |\vec{r}_{Rx_i} - \vec{r}_v|)/c + \phi_{Rx}) \tag{3}$$

and

$$w(f_k, \vec{r}_{Tx_m}, \vec{r}_{Rx_i}, \vec{r}_{\tilde{v}}) = \exp(j2\pi f_k(|\vec{r}_{Tx_m} - \vec{r}_{\tilde{v}}|$$

$$+ |\vec{r}_{Rx_i} - \vec{r}_{\tilde{v}}|)/c). \tag{4}$$

All frequency steps that are transmitted by the Tx antennas are represented by f, and the speed of light is described by c. The vectors pointing from the origin, i.e., the center of the MIMO array, to any Tx or Rx antenna position are denoted as $\vec{r}_{Tx}$ and $\vec{r}_{Rx}$, respectively. A constant reflection phase shift is described by ϕ_{Rx}. By performing the backprojection algorithm for all relevant voxels $\vec{r}_{\tilde{v}}$, a 3D volume describing the object scene is obtained. The resolution of this volume depends on several parameters.

According to [24], the spatial resolution in front of the center of a square-shaped MIMO array with equal Tx and Rx apertures and side length D is given by

$$\delta_x = \delta_y = \frac{c}{4 f_{max}} \cdot \sqrt{4\left(\frac{L}{D}\right)^2 + 1} \tag{5}$$

and

$$\delta_z = \frac{c/2}{\Delta f + \left(1 - \frac{1}{\sqrt{1 + 0.5\left(\frac{D}{L}\right)^2}}\right) \cdot f_{min}}, \tag{6}$$

where δ_x and δ_y describe the cross-range (lateral) resolutions and the range resolution is depicted by δ_z. The distance between the focused spot and the antenna array is denoted by L. The minimum and maximum operating frequency as well as the signal bandwidth are indicated by f_{min}, f_{max}, and Δf. As high signal bandwidths lead to high range resolutions, one of the main applications of broadband mmWave imaging is personnel security screening, as this type of 3D imaging is able to resolve clothes, body surfaces, and possibly hidden items underneath the clothes. One disadvantage of this brute-force algorithm is the extremely high computational cost associated with the reconstruction of each voxel within the volume of interest.

If this approach is applied to the imaging of hands, it can be assumed that the strongest reflections are caused by the skin surface, as human skin mostly consists of water and, therefore, strongly reflects millimeter waves [24], [32]. To extract the physical dimensions of the hand from the reconstructed volume $\hat{O}$, a so-called maximum projection needs to be performed [33]. Hence, for each pixel position defined by x and y, the voxel $\hat{O}(x, y, z)$ with the strongest intensity along the z-axis within the reconstructed volume is extracted, which can then be used to locate the hand and its surface in 3D space and allows the projection of a 3D volume $\hat{O}(\vec{r}_v)$ onto a 2D image (see Fig. 3). This procedure has two main disadvantages. On the one hand, the frame rate depends on

the number of transmitted frequencies, as these influence the pure measurement duration and the time required for the data transfer. For the given hardware depicted in Fig. 2, the frame rate for $N_f = 128$ is approximately 70 Hz. To enable a precise radar-based tracking of hand motion, frame rates above 1 kHz are desirable; therefore, an increase in the frame rate of factor 14 is required to enable precise hand motion tracking. On the other hand, brute-force reconstruction of the complete volume is computationally expensive, as all iterations depicted in (1) have to be run for every voxel of interest. This is particularly a problem with regard to real-time applications. However, when it comes to imaging of the hand, reconstructing an entire volume is unnecessary, as only the visible parts of the hand surface create a radar response. Hence, there is no need to resolve multiple targets along the z-axis. Therefore, instead of reconstructing a volume and obtaining the contour of the hand by performing a maximum projection, new approaches should be identified that focus on efficient extraction of the hand surface from the radar data.

The aim of this paper is to present a radar measurement concept and image reconstruction approach that not only drastically reduces the measurement and computational effort but still provides pleasing, high-quality results of the reconstructed hand shell in the form of an image. The results of that study reveal that an efficient extraction of the body shell and a reduction of the computational complexity of three orders of magnitude compared to classical radar imaging methods is achieved with the proposed concept.

IV. 2FSK MIMO RADAR IMAGING PRINCIPLE

In this section, we describe our novel signal processing approach that enables efficient radar imaging of the hand. This approach combines the theory of the 2FSK radar concept [26] with the previously described principle of mmWave imaging. Compared to CW radar, 2FSK radar can significantly increase the unambiguous target range measured [34]. This can be achieved by subsequently sending out two neighbored CW frequencies f_1 and f_2 and evaluating the phase difference $\Delta\phi = \phi_2 - \phi_1$ between both baseband signals. As a result, the distance d to the target can be calculated by [26]

$$d = \frac{c}{4\pi(f_2 - f_1)} \Delta\phi. \tag{7}$$

The maximum unambiguous range is

$$d_{max} = \frac{c}{2(f_2 - f_1)}. \tag{8}$$

This formula can also be used to calculate the maximum unambiguous range achievable with the SFCW signal form. In this case, $f_2 - f_1 = \Delta f$ describes the step size between two frequencies.

When applied to radar imaging, the presented 2FSK approach reconstructs two single-tone images $(\hat{O}_{z_e}^{1/2})$ of one slice of the volume of interest O at an estimated depth z_e and performs the phase evaluation described in (7) for every pixel of that slice. This is an iterative process, as we first approximate the distance to our target and then adjust it based on an initial 2FSK evaluation. Thereafter, a final phase evaluation
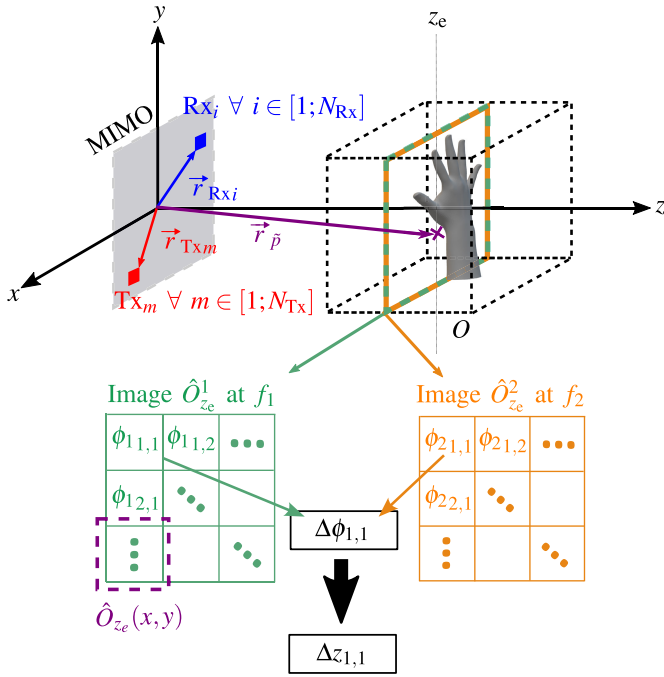
Fig. 4. The 2FSK MIMO radar imaging principle. The 3D shape of the object within the volume of interest O is obtained by reconstructing two 2D images $\hat{O}_{z_e}^1$ and $\hat{O}_{z_e}^2$ of one z-slice at an estimated depth z_e based on two neighbored frequencies. $\vec{r}_{\tilde{p}}$ indicates the different pixel positions of interest. The actual depth of the target is calculated by evaluating the phase difference $\Delta\phi$ between respective pixels of both images and correcting the initially estimated depth z_e.

is performed to estimate the contour of the respective object. Based on this approach, only two frequencies need to be transmitted by every Tx antenna of the MIMO array. The operating principle of this approach is depicted in Fig. 4 and is described in more detail in the following account.

If we assume that there is a target at position $\vec{r}_{\tilde{p}} = (x, y, z_e)$, the estimated distance $d_{e_{im}}$ from the mth Tx antenna ($m \in [1; N_{Tx}]$) to the respective target back to the ith Rx antenna ($i \in [1; N_{Rx}]$) can be calculated by

$$d_{e_{im}}(\vec{r}_{\tilde{p}}) = |\vec{r}_{Tx_m} - \vec{r}_{\tilde{p}}| + |\vec{r}_{Rx_i} - \vec{r}_{\tilde{p}}|, \tag{9}$$

whereas the scatterer is actually located at $\vec{r}_p = (x, y, z)$. Therefore, the correct distance to the point scatterer is

$$d_{im}(\vec{r}_p) = |\vec{r}_{Tx_m} - \vec{r}_p| + |\vec{r}_{Rx_i} - \vec{r}_p| \tag{10}$$

and

$$\Delta d_{im}(\vec{r}_p, \vec{r}_{\tilde{p}}) = d_{e_{im}}(\vec{r}_{\tilde{p}}) - d_{im}(\vec{r}_p), \tag{11}$$

where Δd_{im} describes the error of the distance estimate. For one Tx-Rx antenna pair, the two baseband signals

$$s_{B1_{im}}(\vec{r}_p) = A_1 \exp\left(-j2\pi f_1 \frac{d_{im}}{c} + \phi_{Rx}\right) \tag{12}$$

$$s_{B2_{im}}(\vec{r}_p) = A_2 \exp\left(-j2\pi f_2 \frac{d_{im}}{c} + \phi_{Rx}\right) \tag{13}$$

generated from this point target are obtained. For reasons of simplicity, the amplitudes of the baseband signals A_1 and A_2 are set to one. Furthermore, we note that the baseband signals are influenced by the reflections of multiple point scatterers. The impact of all scatterers in the volume of interest, located at x and y coordinates different from the pixel

being reconstructed, can be expressed as an extra unknown phase term in addition to ϕ_{Rx}. For simplicity, the additional phase term and the unknown reflection phase are summarized by ϕ_{Rx} below. The complex weights $w_{1_{im}}$ and $w_{2_{im}}$ corresponding to (12) and (13), which are needed to reconstruct a 2D image for both carrier frequencies, are based on the first estimate $d_{e_{im}}$ and described by

$$w_{1_{im}}(\vec{r}_{\tilde{p}}) = \exp\left(j2\pi f_1 \frac{d_{e_{im}}}{c}\right) \tag{14}$$

$$w_{2_{im}}(\vec{r}_{\tilde{p}}) = \exp\left(j2\pi f_2 \frac{d_{e_{im}}}{c}\right). \tag{15}$$

Correlating the baseband signals from (12) and (13) with their individual weights, with respect to (11), yields

$$s_{corr1_{im}}(\vec{r}_p, \vec{r}_{\tilde{p}}) = \exp\left(j2\pi f_1 \frac{\Delta d_{im}}{c} + \phi_{Rx}\right) \tag{16}$$

$$s_{corr2_{im}}(\vec{r}_p, \vec{r}_{\tilde{p}}) = \exp\left(j2\pi f_2 \frac{\Delta d_{im}}{c} + \phi_{Rx}\right). \tag{17}$$

To obtain pixel $\hat{O}_{z_e}^{1/2}(x, y) = \hat{O}^{1/2}(\vec{r}_{\tilde{p}})$ of the image of slice z_e, these results are added up over all possible Tx-Rx combinations:

$$\hat{O}_{z_e}^1(x, y) = \sum_{m=1}^{N_{Tx}} \sum_{i=1}^{N_{Rx}} s_{corr1_{im}}(\vec{r}_p, \vec{r}_{\tilde{p}}) \tag{18}$$

$$\hat{O}_{z_e}^2(x, y) = \sum_{m=1}^{N_{Tx}} \sum_{i=1}^{N_{Rx}} s_{corr2_{im}}(\vec{r}_p, \vec{r}_{\tilde{p}}). \tag{19}$$

To evaluate Δd, the 2FSK principle is followed. Hence, the phase difference between corresponding pixels in the two images is measured by muliplying $\hat{O}_{z_e}^2(x, y)$ with the complex conjugate of $\hat{O}_{z_e}^1(x, y)$. Therefore, the differential complex pixel information is obtained by

$$\begin{aligned}
\hat{O}_{z_e}(x, y) &= \hat{O}_{z_e}^2(x, y) \cdot \hat{O}_{z_e}^1(x, y)^* \\
&= \sum_{m=1}^{N_{Tx}} \sum_{i=1}^{N_{Rx}} \exp\left(j2\pi f_2 \frac{\Delta d_{im}}{c} + \phi_{Rx}\right) \cdot \\
&\quad \cdot \sum_{m=1}^{N_{Tx}} \sum_{i=1}^{N_{Rx}} \exp\left(-j2\pi f_1 \frac{\Delta d_{im}}{c} + \phi_{Rx}\right).
\end{aligned} \tag{20}$$

The phase $\Delta\phi$ of this differential complex pixel $\hat{O}_{z_e}(x, y)$ is then used to calculate Δd by applying (7), which results in

$$\Delta d = \frac{c}{2\pi \Delta f} \Delta\phi. \tag{21}$$

In (21), there is only a division by two, as Δd is defined as the complete distance error related to the path from Tx antenna to point scatterer and from point scatterer to Rx antenna, which is illustrated in Fig. 5, using the example of a monostatic antenna and a point target $P(x, y, z)$. In contrast, the distance d in (7) only covers the forward distance from a monostatic antenna to the target. To estimate Δz, we simplify

$$\Delta z = \frac{\Delta d}{2}, \tag{22}$$

which can then be used to correct the first estimated guess z_e. In Fig. 5, it also becomes evident that the difference between
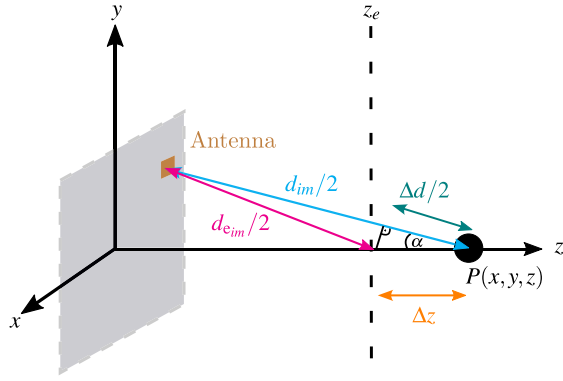
Fig. 5. Difference between Δd and Δz in case of a monostatic antenna.

$\Delta d/2$ and Δz is reduced for increasing values of z as the angle α becomes smaller. The theoretical maximum unambiguous value for $\Delta d/2$ is defined by (8), thereby resulting in $\pm c/(\Delta f \cdot 2)$.

To further illustrate the mechanism of this 2FSK imaging algorithm, (20) can be rewritten as

$$\hat{O}_{z_e}(x, y) = \left[\exp\left(j2\pi f_2 \frac{\Delta d_{11}}{c} + \phi_{Rx}\right) \right.$$
$$+ \exp\left(j2\pi f_2 \frac{\Delta d_{12}}{c} + \phi_{Rx}\right)$$
$$+ \ldots$$
$$\left. + \exp\left(j2\pi f_2 \frac{\Delta d_{N_{Rx}N_{Tx}}}{c} + \phi_{Rx}\right)\right]$$
$$\cdot \left[\exp\left(-\left(j2\pi f_1 \frac{\Delta d_{11}}{c} + \phi_{Rx}\right)\right)\right.$$
$$+ \exp\left(-\left(j2\pi f_1 \frac{\Delta d_{12}}{c} + \phi_{Rx}\right)\right)$$
$$+ \ldots$$
$$\left. + \exp\left(-\left(j2\pi f_1 \frac{\Delta d_{N_{Rx}N_{Tx}}}{c} + \phi_{Rx}\right)\right)\right]. \quad (23)$$

Multiplying out the parentheses results in

$$\hat{O}_{z_e}(x, y) = \exp\left(j2\pi f_2 \frac{\Delta d_{11}}{c} + \phi_{Rx}\right)$$
$$\cdot \exp\left(-\left(j2\pi f_1 \frac{\Delta d_{11}}{c} + \phi_{Rx}\right)\right)$$
$$+ \exp\left(j2\pi f_2 \frac{\Delta d_{11}}{c} + \phi_{Rx}\right)$$
$$\cdot \exp\left(-\left(j2\pi f_1 \frac{\Delta d_{12}}{c} + \phi_{Rx}\right)\right)$$
$$+ \exp\left(j2\pi f_2 \frac{\Delta d_{12}}{c} + \phi_{Rx}\right)$$
$$\cdot \exp\left(-\left(j2\pi f_1 \frac{\Delta d_{11}}{c} + \phi_{Rx}\right)\right)$$
$$+ \exp\left(j2\pi f_2 \frac{\Delta d_{12}}{c} + \phi_{Rx}\right)$$
$$\cdot \exp\left(-\left(j2\pi f_1 \frac{\Delta d_{12}}{c} + \phi_{Rx}\right)\right)$$
$$+ \ldots \quad (24)$$

which can be simplified to

$$\hat{O}_{z_e}(x, y) = \exp\left(j2\pi \Delta f \frac{\Delta d_{11}}{c}\right)$$
$$+ \exp\left(j2\pi \Delta f \frac{\Delta d_{12}}{c}\right)$$
$$+ \exp\left(j2\pi \frac{f_2 \Delta d_{11} - f_1 \Delta d_{12}}{c}\right)$$
$$+ \exp\left(j2\pi \frac{f_2 \Delta d_{12} - f_1 \Delta d_{11}}{c}\right)$$
$$+ \ldots \quad (25)$$

This evaluation leads to the cancelling out of the unknown ϕ_{Rx}. If all Δds were equal, this formula would result in an in-phase summation of the individual antenna contributions. For most imaging radars, Δd differs between Tx-Rx combinations. However, as Δf is set to comparably small values for 2FSK applications and, hence, small variations in Δd cause only small phase shifts, we can assume that lines 1 and 2 of (25) cause the same phase shift and lead to an in-phase summation of these summands. Moreover, the cross terms of (25) (e.g. in line 3 and 4) remain. To evaluate their influence, line 3 of (25) can be rewritten as

$$\exp\left(j2\pi \frac{f_2 \Delta d_{11} - f_1 \Delta d_{12}}{c}\right)$$
$$= \exp\left(j2\pi \frac{f_2 \Delta d_{11} - f_1(\Delta d_{11} + \delta d)}{c}\right)$$
$$= \exp\left(j2\pi \frac{\Delta f \Delta d_{11} - f_1 \delta d}{c}\right), \quad (26)$$

with $\Delta d_{12} = \Delta d_{11} + \delta d$ and δd being an offset between the two Δds. Therefore, line 4 of (25) can be rewritten as

$$\exp\left(j2\pi \frac{f_2 \Delta d_{12} - f_1 \Delta d_{11}}{c}\right)$$
$$= \exp\left(j2\pi \frac{\Delta f \Delta d_{11} + f_2 \delta d}{c}\right). \quad (27)$$

The term $\Delta f \Delta d_{11}/c$ of (26) and (27) causes an additional in-phase summand with respect to (25). In the following, we assume that

$$f_1 \delta d \approx f_2 \delta d, \quad (28)$$

which implies that the phase shifts caused by these terms in (26) and (27) cancel themselves out, so that only the in-phase summations of the complex pointers of (25) remain. In other words, the residual pixel phase of the multiplication described in (20) is influenced by the average Δd over all possible Tx-Rx combinations. The phase error that is made by this assumption is approximately $\exp\left(j2\pi \frac{\Delta f \delta d}{c}\right)$.

The overall workflow of the 2FSK MIMO imaging principle is depicted in Fig. 6. The phase difference between the corresponding pixels of the two complex images is evaluated for each pixel (x,y) (see Fig. 4), where a target can be assumed after applying a suitable threshold. To obtain the best possible focused image, the image reconstruction is divided into two steps. First, two images are reconstructed at z_e, and the pixel containing the strongest scatterer is evaluated.
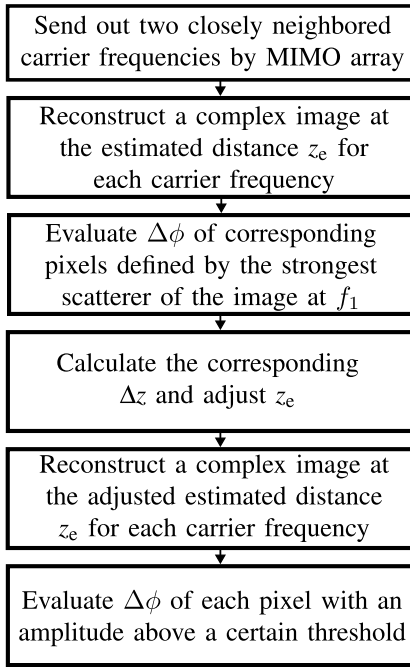
Fig. 6.   2FSK MIMO imaging workflow.

$P\,(x, y, z)$	$(|\delta x|, |\delta y|, |\delta z|)$ - 2FSK	$(|\delta x|, |\delta y|, |\delta z|)$ - 2FSK z_e adjusted
(0.00, 0.00, 0.30)	(0.000, 0.000, 0.001)	(0.000, 0.000, 0.000)
(0.00, 0.05, 0.30)	(0.000, 0.005, 0.001)	(0.000, 0.000, 0.000)
(0.00, 0.10, 0.30)	(0.000, 0.009, 0.000)	(0.000, 0.000, 0.000)
(0.10, 0.10, 0.30)	(0.009, 0.009, 0.002)	(0.001, 0.001, 0.000)

VI. PERFORMANCE OF THE 2FSK IMAGING APPROACH

Now that the theory behind the novel approach has been explained and the signal and reconstruction parameters have been established, the following section evaluates its reconstruction performance.

A. Simulation of Point Scatterers

Within this section, the theoretical accuracy of both techniques is evaluated by simulating the reflected signal of one point scatterer. First, the theoretical accuracy of the novel 2FSK MIMO imaging approach is analyzed for different point scatterer positions $P(x, y, z)$, as seen in Fig. 5 and for the MIMO geometry presented in Fig. 1. To estimate the point target position, two complex images are reconstructed at an estimated distance z_e using the two different carrier frequencies. At first, z_e is estimated at 30 cm. To adjust z_e, the phase difference of the pixel containing the strongest scatterer is evaluated (see Fig. 6). Afterward, again two complex images are reconstruced at the adjusted value for z_e, and the phase evaluation of the strongest scatterer is repeated.

Within the first simulation set, the x and y positions of the point target are varied, whereas the true z coordinate and z_e are kept constant, with $z = 0.3$ m and $z_e = 0.27$ m leading to $\Delta z = 0.03$ m. In Table I, the true position P of the point target and the absolute deviations per coordinate — δx, δy, δz — compared to the detected value using the 2FSK imaging principle before and after adjusting z_e can be seen. Furthermore, the 2FSK algorithm was evaluated for different z values, which can be seen in Table II. Here, Δz was kept constant at 0.03 m. In addition, various values for Δz were evaluated while x, y, and z were kept constant at (0,0,0.3). The results are presented in Table III. These results indicate that the accuracy of the 2FSK algorithm decreases the closer the target lies to the radar. This can be illustrated by Fig. 5. The closer the point target is located with respect to the radar, the greater the difference between $\Delta d/2$ and Δz becomes. It is also evident, that the readjustment of z_e significantly improves the overall accuracy of the detection of all coordinates.

Second, the theoretical accuracy of the SFCW approach is analyzed. For all point target positions listed in Table I and Table II the state-of-the-art approach yielded 100 % accuracy (results were rounded to the millimeter). The remarkable theoretical accuracy and the widespread use of this algorithm provide strong motivation to utilize its reconstruction results as a benchmark for evaluating the performance of the 2FSK approach for the upcoming measurement section.

After calculating the corresponding Δz, z_e is corrected by Δz, and two complex images are reconstructed at the adjusted z_e coordinate. Finally, the phase difference of all pixels with an amplitude above a certain threshold is evaluated by applying (21) and (22), and the 3D information is smoothed by an averaging filter.

V. SIGNAL AND RECONSTRUCTION DESIGN

When choosing the optimum signal parameters for the 2FSK approach, the value for $\Delta d_{max}/2$ (see (8)) is one design criterion. We set Δf in the following to 200 MHz; in this manner, we acquire an unambiguous range for $\Delta d/2$ of approximately 0.75 m and ensure that Δf is still sufficiently small to enable the use of all simplifications described in Section IV. Further, the CW frequencies are set to $f_1 = 79.8$ GHz and $f_2 = 80$ GHz, thereby enabling high angular resolution. With regard to (5) and (6), the cross-range and range resolution for a CW signal at $f_1 = 79.8$ GHz at a distance $L = 0.3$ m are approximately $\delta_x = \delta_y \approx 4$ mm and $\delta_z \approx 4$ cm, respectively. With regard to the backprojection, two complex images for $x/y \in [-0.1\,\text{m}, 0.1\,\text{m}]$ are reconstructed. $N_x = N_y$ are set to 101 sampling points, thereby leading to a pixel dimension of 1 mm × 1 mm.

When applying the SFCW approach, in our case, a signal form that covers a bandwidth of 10 GHz from 72 GHz to 82 GHz with $N_f = 128$ is used. This enables a cross-range resolution at distance $L = 0.3$ m of approximately $\delta_x \approx \delta_y \approx 4$ mm as well as a range resolution $\delta_z \approx 1$ cm. With respect to the backprojection of the volume, x and y also lie within $[-0.1\,\text{m}, 0.1\,\text{m}]$ and $N_x = N_y$ are also set to 101 sampling points. The depth z is sampled from $[0.26\,\text{m}, 0.34\,\text{m}]$ in $N_z = 81$ steps, thereby yielding a voxel dimension of 1 mm × 1 mm × 1 mm.

TABLE II

DEVIATIONS PER COORDINATE — δx, δy, δz — COMPARED TO TRUE POINT TARGET POSITION P FOR DIFFERING VALUES OF z IN METERS. Δz IS KEPT CONSTANT AT 0.03 m

P (x, y, z)	$(\|\delta x\|, \|\delta y\|, \|\delta z\|)$ - 2FSK	$(\|\delta x\|, \|\delta y\|, \|\delta z\|)$ - 2FSK z_{e} adjusted
(0.00, 0.00, 0.05)	(0.000, 0.000, 0.009)	(0.000, 0.000, 0.004)
(0.00, 0.00, 0.15)	(0.000, 0.000, 0.003)	(0.000, 0.000, 0.000)
(0.00, 0.00, 0.50)	(0.000, 0.000, 0.000)	(0.000, 0.000, 0.000)
(0.00, 0.00, 1.00)	(0.000, 0.000, 0.000)	(0.000, 0.000, 0.000)

TABLE III

DEVIATIONS PER COORDINATE — δx, δy, δz — COMPARED TO TRUE POINT TARGET POSITION P FOR DIFFERING VALUES OF Δz IN METERS. P IS KEPT CONSTANT AT (0.00, 0.00, 0.30) METER

| Δz	$(\|\delta x\|, \|\delta y\|, \|\delta z\|)$ - 2FSK	$(\|\delta x\|, \|\delta y\|, \|\delta z\|)$ - 2FSK z_{e} adjusted
0.04	(0.000, 0.000, 0.002)	(0.000, 0.000, 0.000)
0.10	(0.006, 0.006, 0.004)	(0.000, 0.000, 0.000)
0.15	(0.017, 0.017, 0.009)	(0.000, 0.000, 0.000)
0.19	(0.025, 0.025, 0.014)	(0.000, 0.000, 0.000)

B. Measurements Using 3D Printed Hand Poses

This section presents measurement results regarding the 3D reconstruction of different hand poses of the ASL alphabet to further evaluate the performance of the novel 2FSK 3D imaging approach. In this regard, the results are compared to the state-of-the-art SFCW approach. To allow this comparison, three hand poses were 3D printed and coated with an electromagnetic interference shielding lacquer. In this manner, a static object scene can be ensured. The hand poses describe the ASL letters B, F, and U and are depicted in Fig. 7(a).

The MIMO radar array used for the measurements corresponds to the geometry described in Section II. The measurement setup can be seen in Fig. 7(b). The 3D printed hands were positioned roughly at a z coordinate distance of 30 cm to the radar. We placed two absorber walls behind the hands. To position the hands, we used a table with a styrodur plate and one absorber mat on top. For both the 2FSK and SFCW measurements, a measurement averaging across three successive measurements was applied to the measurement data.

For the 3D reconstruction of the hand poses following the novel 2FSK approach, we followed the steps depicted in Fig. 6. A threshold of -13 dB below the maximum amplitude to suppress sidelobes is applied; hence, every pixel within the image reconstructed at $f = 79.8$ GHz with a magnitude below this threshold is removed from both reconstructed images. The remaining pixels then define the region of interest (ROI) in x and y dimension.

To perform a 3D reconstruction of the hand poses using the SFCW algorithm, first, the volume of interest is reconstructed. Then, a 2D image is generated via the maximum projection and the corresponding z coordinates of the skin surface are extracted as described in Section III. Thereafter, a threshold of -15 dB below the maximum amplitude is applied to suppress side lobes.

The result of the image reconstruction normalized to the corresponding maximum magnitude with applied thresholding
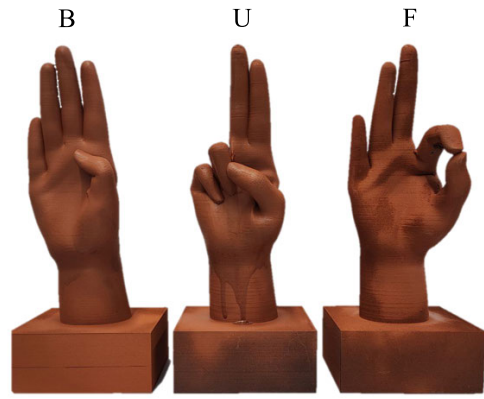
(a)
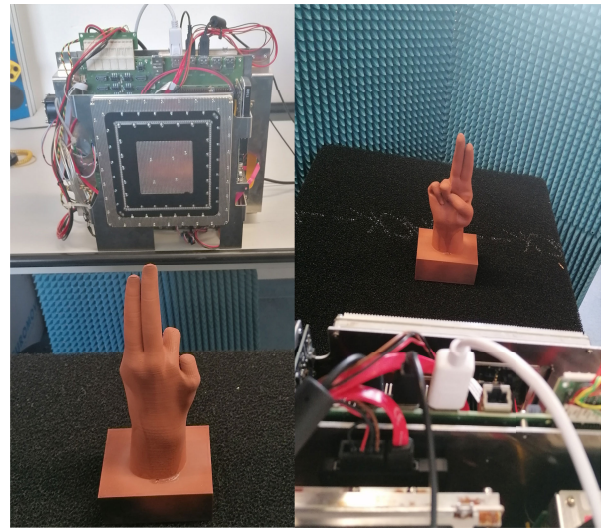
(b)

Fig. 7. Measurements using 3D printed hand poses. (a) 3D printed hand poses (left: B; middle: U; right: F). (b) Multiperspective view of measurement setup (left: front, right: back).

for both algorithms is presented in Fig. 8(a). At this point it is noted that the absolute maximum magnitude for the SFCW maximum projection image is higher than that of the reconstructed 2FSK images, as the SFCW algorithm runs a backprojection for multiple frequencies.

To determine the z coordinate of the skin surface for each pixel using the novel 2FSK approach, the phase of both complex images is compared for each pixel within the extracted ROI. With respect to (21) and (22), Δz as well as $z = z_{\mathrm{e}} + \Delta z$ are calculated. Afterwards, the values for z are smoothed with a 2D running mean filter with a kernel size of 15. The z coordinates extracted during the maximum projection in case of the SFCW evaluation, are also smoothed with a 2D running mean filter with a kernel size of 15.

The results for the detection of the skin surface of the hand for both algorithms can be seen in Fig. 8(b)–(h). The 3D reconstruction of the hand surface yields rather similar results when both approaches are compared. The estimated z coordinate fluctuates slightly more in the 2FSK case, whereas the surface estimate using the SFCW approach appears smoother. The small fluctuations of the depth estimate in the 2FSK algorithm
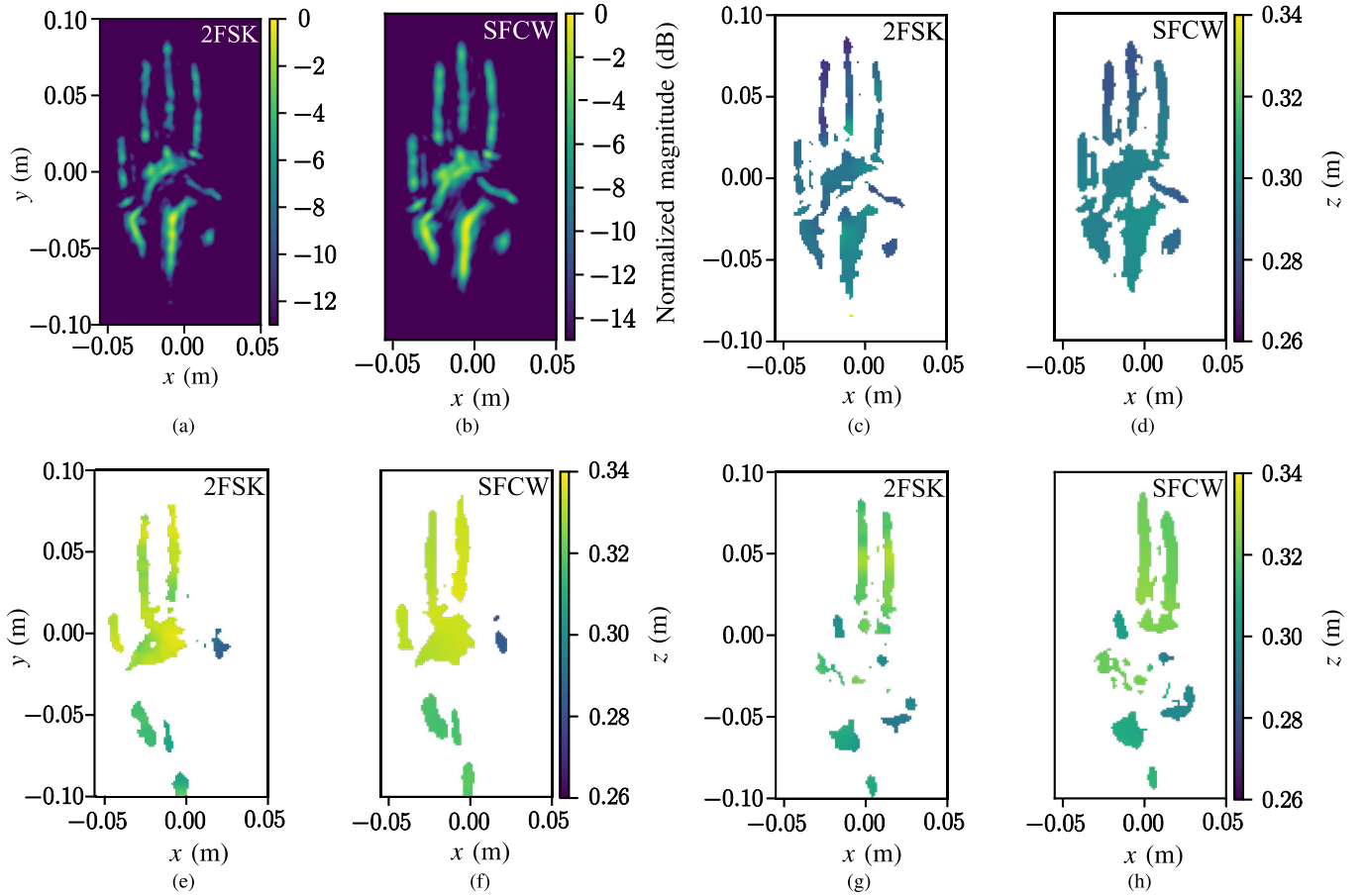
Fig. 8.   3D imaging results of three printed hand poses from the ASL alphabet. Radar image of 3D printed hand pose B (a) reconstructed at f_1 based on 2FSK approach and (b) based on SFCW approach after maximum projection. Estimated hand surface coordinates based on the novel 2FSK approach for hand poses (c) B, (e) F, and (g) U. Estimated hand surface coordinates based on the state-of-the-art SFCW approach for hand poses (d) B, (f) F, and (h) U.

may be caused by clutter that can have a stronger effect compared to the SFCW case, as only one z-slice is reconstructed. However, the depth profiles of the respective hand poses are clearly recognizable in the 2FSK case. By means of these figures based on both approaches, the different hand poses can clearly be distinguished from each other and assigned to the corresponding letter. This remains the case if some parts of the hand do not return a signal due to their orientation to the radar — for example, the thumb saddle joint in the case of the letter F.

Table IV presents the absolute maximum ($|\delta z_{\text{SFCW}-\text{FSK}_{\max}}|$) and mean value ($\overline{|\delta z_{\text{SFCW}-\text{FSK}}|}$) for $\delta z_{\text{SFCW}-\text{FSK}}$, describing the deviation between SFCW- and 2FSK-based z coordinate estimation. The maximum deviations $|\delta z_{\text{SFCW}-\text{FSK}_{\max}}|$ lie in the range of 1 cm to 2 cm, whereas the mean value of the deviation lies approximately around 3 mm to 5 mm. It becomes evident that the algorithms generate similar results for all analyzed sign language letters.

To further evaluate the performance of the 2FSK approach, the reconstruction results for letter F are compared to the state-of-the-art for increased distances to the radar. The comparison of the 3D reconstructed hand surface are depicted in Fig. 9. The average and maximum deviation in case of letter F for increasing distances to the radar are compared within Table V. At 40 cm and 50 cm, the 2FSK algorithm

TABLE IV
DEVIATION $\delta z_{\text{SFCW}-\text{FSK}}$ OF ESTIMATED z COORDINATE BASED ON SFCW COMPARED TO 2FSK ALGORITHM IN MILLIMETERS

Letter	$|\delta z_{\text{SFCW}-\text{FSK}_{\max}}|$	$\overline{|\delta z_{\text{SFCW}-\text{FSK}}|}$
B	20	5
F	16	3
U	14	4

still yields pleasing results that are comparable to the SFCW outcome. However, with increased distances to the radar, the lateral resolution reduces for both approaches. With regard to the 2FSK algorithm, this implies that the depth information within one lateral resolution cell is overlayed. In case of the SFCW algorithm, two targets at differing depths that lie in the same lateral resolution cell might still be distinguishable due to the high radial resolution. Nevertheless, when applying the maximum projection, only the depth information with the higher amplitude is retained. Hence, as for the respective application individual fingers need to be seperated, a lateral resolution below 1 cm is required. At a distance of 50 cm, the reduction in lateral resolution is clearly recognizable, which is why this distance should not be exceeded. In Table V, an increase of the deviation between SFCW- and 2FSK-based results with the increase in distance to the radar is evident. There are two main reasons for this behavior.
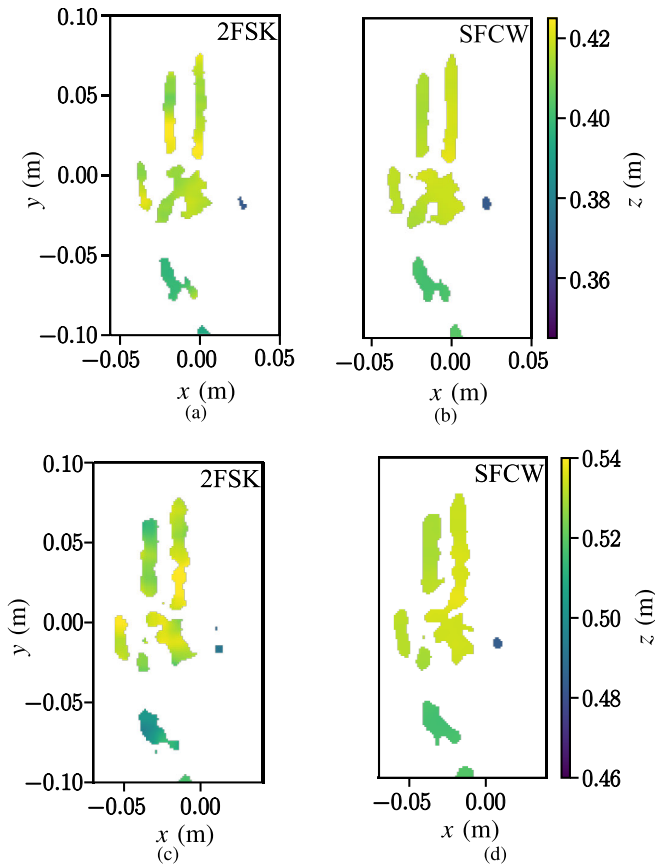
Fig. 9. Comparison of 3D imaging results of printed hand pose F for increased distances. Estimated hand surface coordinates at a distance of 40 cm to the radar based on the (a) 2FSK and (b) SFCW approach. (c) 2FSK and (d) SFCW based 3D reconstruction results at a further increased distance of 50 cm.

TABLE V

DEVIATION $\delta z_{\text{SFCW-FSK}}$ OF ESTIMATED z COORDINATE IN CASE OF LETTER F AND INCREASED DISTANCES TO THE RADAR

Distance	$|\delta z_{\text{SFCW-FSK}_{\max}}|$	$\overline{|\delta z_{\text{SFCW-FSK}}|}$
30 cm	16	3
40 cm	15	4
50 cm	21	5

First, the received signal amplitude decreases, which affects the 2FSK earlier compared to an SFCW signal generation, as only two transmit frequencies are sent, thereby making it more difficult to seperate the main lobe from sidelobes. Second, a decrease in lateral resolution affects both algorithms in a different manner, which can cause an increase in the deviation between the two.

In summary, based on the prior point target simulations and the comparison of both approaches using real measurements for distances up to 50 cm, it can be stated that the novel 2FSK approach has a performance comparable to the state-of-the-art radar imaging algorithm with a theoretical accuracy in the sub-millimeter to millimeter range.

C. Efficiency of the 2FSK Imaging Approach

Now that we have shown in the previous section that the results of the novel 2FSK 3D reconstruction are comparable to the results of the state-of-the-art approach, the aspect of

TABLE VI

COMPUTATIONAL COMPLEXITY OF PROPOSED 2FSK ALGORITHM

Step	Complexity
Backprojection of image at z_{e}	$2 \cdot N_{\text{Tx}} \cdot N_{\text{Rx}} \cdot N_{\text{x}} \cdot N_{\text{y}}$
Strongest scatterer	$N_{\text{x}} \cdot N_{\text{y}}$
Backprojection after adjusting z_{e}	$2 \cdot N_{\text{Tx}} \cdot N_{\text{Rx}} \cdot N_{\text{x}} \cdot N_{\text{y}}$
Phase difference per pixel	$N_{\text{x}} \cdot N_{\text{y}}$
Total number of calculation steps (for $N_{\text{Tx}} = N_{\text{Rx}} = 94$; $N_{\text{x}} = N_{\text{y}} = 201$)	approx. $1.43 \cdot 10^9$

TABLE VII

COMPUTATIONAL COMPLEXITY OF SFCW (STATE-OF-THE-ART) ALGORITHM

Step	Complexity
Backprojection for volume	$N_{\text{Tx}} \cdot N_{\text{Rx}} \cdot N_{\text{x}} \cdot N_{\text{y}} \cdot N_{\text{z}} \cdot N_{\text{f}}$
Maximum projection	$N_{\text{x}} \cdot N_{\text{y}} \cdot N_{\text{z}}$
Total number of calculation steps (for $N_{\text{Tx}} = N_{\text{Rx}} = 94$; $N_{\text{x}} = N_{\text{y}} = 201, N_{\text{z}} = 81$; $N_{\text{f}} = 50$)	approx. $1.45 \cdot 10^{12}$

efficiency is now addressed. First, the new algorithm presented here, requires much less bandwidth than the broadband state-of-the-art reconstruction approaches. This reduces the requirements for the necessary hardware and increases bandwidth efficiency. Further, the presented reconstruction algorithm is computationally more efficient. In order to quantify this, the computational complexity — represented by calculation steps — of the novel approach is compared to the exemplary SFCW state-of-the-art algorithm in Table VI and Table VII. We assume the scenario as depicted in Section VI apart from setting $N_{\text{f}} = 10\,\text{GHz}/200\,\text{MHz} = 50$ to ensure the same unambiguous range for both approaches. The advantage of the novel 2FSK algorithm lies in the reconstruction of two single-tone images at one z coordinate, whereas the brute-force state-of-the-art approach reconstructs a full volume by evaluating a high number of frequency steps paired with a computationally expensive maximum projection. The results reveal the immense reduction in computational effort by three orders of magnitude, thereby opening the way toward real-time reconstruction of hand poses. It should be noted that the SFCW-based algorithm requires that one is aware of the approximate position of the hand, otherwise huge 3D volumes have to be reconstructed, thereby resulting in a strong increase in computational burden. Using the iterative localization of the hand as proposed by the novel algorithm, this problem does not occur. In addition, this novel approach holds the possibility of considerably increasing the frame rate, as only two frequency steps are required. With regard to the hardware used here, the frame rate can be theoretically increased to $70 \cdot \frac{128}{2} \approx 4.5\,\text{kHz}$, allowing radar-based hand motion tracking even for fast motions.

D. Measurements Including a Real Human Hand

To further support our results, we also include an exemplary evaluation of a measurement setup in which a real
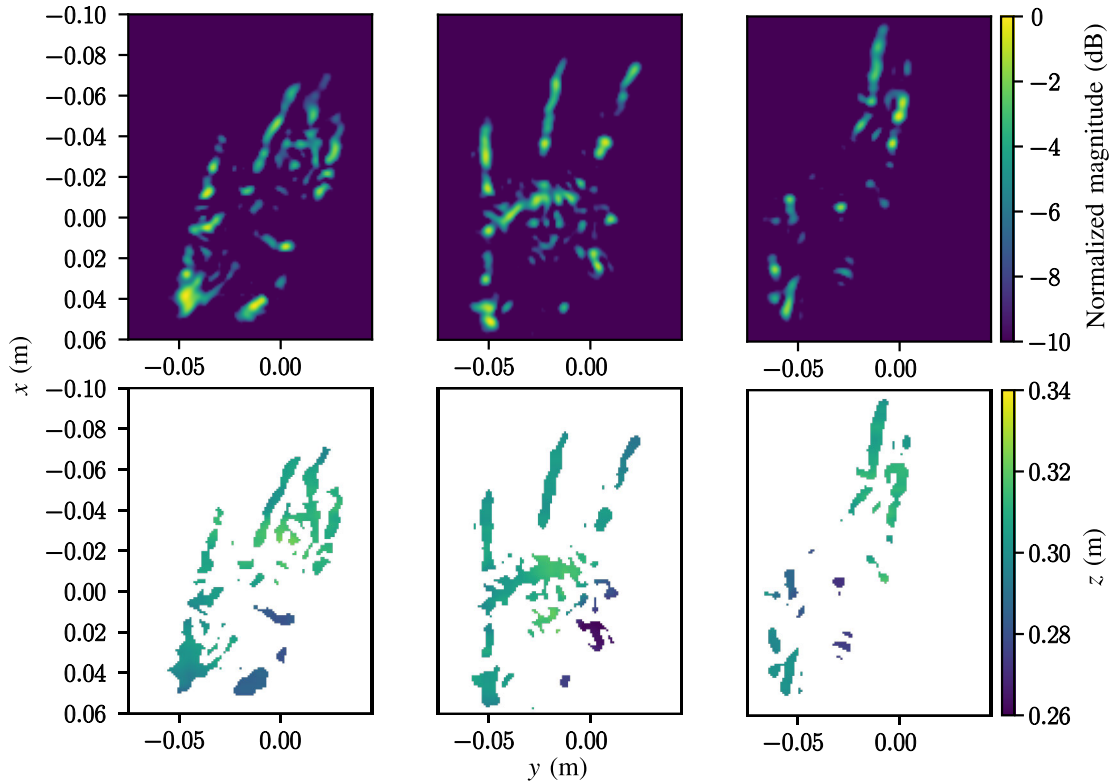
Fig. 10. 3D imaging results of a real human hand based on the 2FSK approach for three exemplary hand poses from the ASL alphabet (left: B; middle: F; right: U). Images reoconstructed at f_1 (top) and the respective estimated hand surface coordinates of the hand poses based on the 2FSK algorithm (bottom). The extracted 3D information accurately represents the corresponding hand pose and allows a clear assignment.

hand was positioned in front of the sensor (as depicted in Fig. 1 and described in Section II), and the 2FSK signal and reconstruction approach was applied. Again, we roughly positioned the hand at a distance of 30 cm and imitated the three letters from the ASL alphabet. For these measurements, we gained optimum results setting the measurement averaging to a factor of 1, as a real human hand cannot be static like the 3D printed versions. The reconstructed images as well as the results for the 3D reconstruction are depicted in Fig. 10. The threshold was further adjusted to −10 dB. The results of the 3D reconstruction of a real human hand provide further proof that the application of the 2FSK-based approach for 3D imaging of the human hand is valid and yields promising results. Their specific depth profiles can be clearly assigned to the respective hand poses. In addition, the radar image as well as the depth profile strongly resemble the measurement results when using the 3D printed hands.

VII. Discussion

Our findings have clearly revealed that the proposed fast and efficient 2FSK-based imaging approach yields promising results for the 3D hand pose reconstruction that are comparable to the outcome of the state-of-the-art SFCW-based radar imaging approach. Simultaneously, our novel approach reduces computational complexity by three orders of magnitude and significantly increases possible frame rates. However, there are also a few drawbacks of the proposed method that appear to cause the small differences when comparing the 3D reconstruction outcome of both methods. One limitation

exists due to the fact that in our proposed method, we only reconstruct one z-slice of our object scene. This leads to clutter having a stronger impact compared to the backprojection of an entire volume, followed by a maximum projection. In the future, it would be interesting to research how the robustness against clutter could be improved. Empty space measurements could be a viable approach in this context. Furthermore, the reconstruction of only one z-slice also makes it more difficult to seperate the side lobes from the main lobe. This is particularly important for hand areas that reflect only a small signal amplitude. An interesting approach to reduce this effect would be a stepwise subtraction of the point spread function. Moreover, the approach would benefit from one additional 2FSK evaluation round. This implies that the 3D coordinates are estimated by our proposed approach; thereafter, we reconstruct the depth profile obtained by our prior evaluation, instead of one z-slice, to increase reconstruction accuracy. In addition, given that only two transmit frequencies are evaluated, the impact of phase errors also must be considered. In general, radar-based approaches, unlike optical methods, suffer from numerous specular reflections, thereby limiting the hand areas, where a 3D reconstruction is possible. In the future, this could be addressed by multi-perspective imaging, which would also reduce the effect of occlusion.

VIII. Conclusion

In this manuscript, we present a novel algorithm for the 3D reconstruction of hand poses that incorporates the theory of

frequency shift keying as part of the concept of millimeter-wave imaging. For this approach, only little bandwidth is required, as only two closely spaced transmit frequencies are needed. The concept of frequency shift keying is first used to iteratively locate the hand within a 3D space. Thereafter, the phase difference of corresponding pixels between two single-tone images is evaluated according to the 2FSK principle to precisely estimate the contour of the hand. The big advantage of this approach is that it enables a direct reconstruction of the body shell, i.e., a 2D depth profile, and avoids the backprojection of a 3D volume, as is currently the case with millimeter-wave imaging. This fact brings great advantages. The presented approach is characterized by high efficiency in terms of computational effort and hardware requirements. Furthermore, image reconstruction complexity can be reduced by a factor of 1000, which brings the potential to reduce reconstruction time, while frame rates can be significantly increased. In this manner, we were able to not only raise the efficiency of radar-based static hand pose estimation but also open the way to precise real-time radar-based hand motion tracking. In this study, suitable simulations revealed that the algorithm yields a theoretical accuracy in the sub-millimeter to millimeter range. In real measurement setups, the novel 2FSK approach reveals comparable results to state-of-the-art 3D radar imaging principle, with deviations lying in the millimeter range. Finally, it can be concluded that the novel 2FSK imaging approach has great potential to improve radar-based methods for capturing hand poses, including static and dynamic gesture recognition and hand-tracking applications. In the future, this methodology should be implemented by leveraging the advantages of radar-based methods and the advancements made in computer vision algorithms to enhance their performance.

ACKNOWLEDGMENT

REFERENCES

[1] S. S. Rautaray and A. Agrawal, "Vision based hand gesture recognition for human computer interaction: A survey," *Artif. Intell. Rev.*, vol. 43, no. 1, pp. 1–54, Jan. 2015.

[2] H.-S. Yeo, B.-G. Lee, and H. Lim, "Hand tracking and gesture recognition system for human-computer interaction using low-cost hardware," *Multimedia Tools Appl.*, vol. 74, no. 8, pp. 2687–2715, Apr. 2015.

[3] A. Erol, G. Bebis, M. Nicolescu, R. D. Boyle, and X. Twombly, "Vision-based hand pose estimation: A review," *Comput. Vis. Image Understand.*, vol. 108, nos. 1–2, pp. 52–73, Oct. 2007.

[4] A. Anaz, M. Skubic, J. Bridgeman, and D. M. Brogan, "Classification of therapeutic hand poses using convolutional neural networks," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 3874–3877.

[5] N. Yang et al., "Automatic detection pipeline for accessing the motor severity of Parkinson's disease in finger tapping and postural stability," *IEEE Access*, vol. 10, pp. 66961–66973, 2022.

[6] U. Phutane et al., "Evaluation of optical and radar based motion capturing technologies for characterizing hand movement in rheumatoid arthritis—A pilot study," *Sensors*, vol. 21, no. 4, pp. 1208–1226, Feb. 2021.

[7] W. Chen et al., "A survey on hand pose estimation with wearable sensors and computer-vision-based methods," *Sensors*, vol. 20, no. 4, pp. 1074–1098, Feb. 2020.

[8] T. Chatzis, A. Stergioulas, D. Konstantinidis, K. Dimitropoulos, and P. Daras, "A comprehensive study on deep learning-based 3D hand pose estimation methods," *Appl. Sci.*, vol. 10, no. 19, pp. 6850–6876, Sep. 2020.

[9] S. Yuan et al., "Depth-based 3D hand pose estimation: From current achievements to future goals," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2636–2645.

[10] S. Ahmed, F. Khan, A. Ghaffar, F. Hussain, and S. H. Cho, "Finger-counting-based gesture recognition within cars using impulse radar with convolutional neural network," *Sensors*, vol. 19, no. 6, pp. 1429–1442, Mar. 2019.

[11] S. Ahmed, K. D. Kallu, S. Ahmed, and S. H. Cho, "Hand gestures recognition using radar sensors for human-computer-interaction: A review," *Remote Sens.*, vol. 13, no. 3, pp. 527–550, Feb. 2021.

[12] J. Lien et al., "Soli," *ACM Trans. Graph.*, vol. 35, no. 4, pp. 1–19, Jul. 2016.

[13] Y. Kim and B. Toomajian, "Application of Doppler radar for the recognition of hand gestures using optimized deep convolutional neural networks," in *Proc. 11th Eur. Conf. Antennas Propag. (EUCAP)*, Mar. 2017, pp. 1258–1260.

[14] S. Skaria, A. Al-Hourani, M. Lech, and R. J. Evans, "Hand-gesture recognition using two-antenna Doppler radar with deep convolutional neural networks," *IEEE Sensors J.*, vol. 19, no. 8, pp. 3041–3048, Apr. 2019.

[15] S. Zhang, G. Li, M. Ritchie, F. Fioranelli, and H. Griffiths, "Dynamic hand gesture classification based on radar micro-Doppler signatures," in *Proc. CIE Int. Conf. Radar (RADAR)*, Oct. 2016, pp. 1–4.

[16] G. Zhang, S. Lan, K. Zhang, and L. Ye, "Temporal-range-Doppler features interpretation and recognition of hand gestures using mmW FMCW radar sensors," in *Proc. 14th Eur. Conf. Antennas Propag. (EuCAP)*, Mar. 2020, pp. 1–4.

[17] H. Liang, X. Wang, M. S. Greco, and F. Gini, "Enhanced hand gesture recognition using continuous wave interferometric radar," in *Proc. IEEE Int. Radar Conf. (RADAR)*, Apr. 2020, pp. 226–231.

[18] J.-T. Yu, L. Yen, and P.-H. Tseng, "MmWave radar-based hand gesture recognition using range-angle image," in *Proc. IEEE 91st Veh. Technol. Conf. (VTC-Spring)*, May 2020, pp. 1–5.

[19] L. Zheng et al., "Dynamic hand gesture recognition in in-vehicle environment based on FMCW radar and transformer," *Sensors*, vol. 21, no. 19, pp. 6368–6387, Sep. 2021.

[20] C. Liu, Y. Li, D. Ao, and H. Tian, "Spectrum-based hand gesture recognition using millimeter-wave radar parameter measurements," *IEEE Access*, vol. 7, pp. 79147–79158, 2019.

[21] E. Klinefelter and J. A. Nanzer, "Interferometric radar for spatially-persistent gesture recognition in human-computer interaction," in *Proc. IEEE Radar Conf. (RadarConf)*, Apr. 2019, pp. 1–5.

[22] J. W. Smith, S. Thiagarajan, R. Willis, Y. Makris, and M. Torlak, "Improved static hand gesture classification on deep convolutional neural networks using novel sterile training technique," *IEEE Access*, vol. 9, pp. 10893–10902, 2021.

[23] S. S. Ahmed, A. Schiessl, F. Gumbmann, M. Tiebout, S. Methfessel, and L.-P. Schmidt, "Advanced microwave imaging," *IEEE Microw. Mag.*, vol. 13, no. 6, pp. 26–43, Sep. 2012.

[24] S. S. Ahmed, "Microwave imaging in security—Two decades of innovation," *IEEE J. Microw.*, vol. 1, no. 1, pp. 191–201, Jan. 2021.

[25] L. Qiao, L. Wang, Z. Zhao, and Z. Chen, "Exact reconstruction for near-field three-dimensional planar millimeter-wave holographic imaging," *J. Infr., Millim., Terahertz Waves*, vol. 36, no. 12, pp. 1221–1236, Oct. 2015.

[26] E. J. Barlow, "Doppler radar," *Proc. IRE*, vol. 37, no. 4, pp. 340–355, Apr. 1949.

[27] C. Keskin, F. Kiraç, Y. E. Kara, and L. Akarun, "Real time hand pose estimation using depth sensors," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCV Workshops)*, Nov. 2011, pp. 1228–1234.

[28] W. Tao, M. C. Leu, and Z. Yin, "American sign language alphabet recognition using convolutional neural networks with multiview augmentation and inference fusion," *Eng. Appl. Artif. Intell.*, vol. 76, pp. 202–213, Nov. 2018.

[29] C. Dong, M. C. Leu, and Z. Yin, "American sign language alphabet recognition using Microsoft Kinect," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2015, pp. 44–52.

[30] "Rohde & Schwarz QAR50 radome tester," Mar. 2022. [Online]. Available: https://www.rohde-schwarz.com/qar50/

[31] S. S. A. Ahmed, *Electronic Microwave Imaging with Planar Multistatic Arrays*. Berlin, Germany: Logos Verlag, 2014.

[32] M. Zhadobov, N. Chahat, R. Sauleau, C. Le Quement, and Y. Le Drean, "Millimeter-wave interactions with the human body: State of knowledge and recent advances," *Int. J. Microw. Wireless Technol.*, vol. 3, no. 2, pp. 237–247, Apr. 2011.

[33] I. Ullmann, J. Adametz, D. Oppelt, A. Benedikter, and M. Vossiek, "Non-destructive testing of arbitrarily shaped refractive objects with millimetre-wave synthetic aperture radar imaging," *J. Sensors Sensor Syst.*, vol. 7, no. 1, pp. 309–317, Apr. 2018.

[34] H. Rohling and C. Moller, "Radar waveform for automotive radar systems and applications," in *Proc. IEEE Radar Conf.*, May 2008, pp. 1–4.

Johanna Bräunig (Graduate Student Member, IEEE) was born in Bamberg, Germany, in 1993. She received the M.Sc. degree in electrical engineering from Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany, in 2020. She is currently pursuing the Ph.D. degree with the Institute of Microwaves and Photonics. She is particularly interested in radar-based motion tracking as well as the use of radar technology in medical engineering and machine learning applications. Her research interests include radar imaging and signal processing.

Vanessa Wirth was born in Schwabach, Germany, in 1996. She received the M.Sc. degree in computer science from Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany, in collaboration with the Technical University of Munich, Munich, Germany. She is currently pursuing the Ph.D. degree with the Chair of Visual Computing, FAU. Her research interests include computer graphics, computer vision, and machine learning. She is particularly interested in surface reconstruction and motion tracking of non-rigid objects with radar-based and optical sensor systems.

Christoph Kammel (Graduate Student Member, IEEE) was born in Nuremberg, Germany, in 1996. He received the M.Sc. degree in electrical engineering from Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany, in 2020, and joined the Institute of Microwaves and Photonics, FAU, where he is currently pursuing the Ph.D. degree. His current research interests include radar imaging and signal processing in the space sector.

Christian Schüßler (Graduate Student Member, IEEE) was born in Fürth, Germany, in 1991. He received the B.Eng. and M.Eng. degrees in electrical engineering and information technology from Technische Hochschule Nürnberg Georg Simon Ohm, Nuremberg, Germany, in 2014 and 2015, respectively. He is currently pursuing the Ph.D. degree with the Institute of Microwaves and Photonics (LHFT). From 2015 to 2019, he was a Software Engineer in the field of computed tomography with the Algorithmic Group, Fraunhofer Center for X-Ray Development (EZRT). His research interests include simulation, signal processing, and data fusion, especially for large MIMO arrays in automotive applications.

Ingrid Ullmann (Member, IEEE) received the M.Sc. degree in electrical engineering and the Ph.D. degree from Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany, in 2016 and 2021, respectively. She is currently a Post-Doctoral Researcher and the Head of the Research Group "Wave-Based Imaging Systems," Institute of Microwaves and Photonics, FAU. In 2022, she spent one month as a Visiting Researcher with the Microwave Sensing, Signals and Systems Group, Delft University of Technology, Delft, The Netherlands. Her research interests include radar imaging and radar signal processing for nondestructive testing, security screening, medical radar, and automotive applications. She is a member of the IEEE CRFID Technical Committee on Motion Capture and Localization. She received the Argus Science Award (sponsored by Airbus Defense and Space, now Hensoldt) in 2016 and the EuRAD Conference Prize in 2019. Since 2022, she has been an Associate Editor of the IEEE TRANSACTIONS ON RADAR SYSTEMS. She serves as a reviewer for the European Radar Conference (EuRAD) and various journals in the field of microwaves.

Marc Stamminger is currently the Chair of Visual Computing, Friedrich-Alexander-Universität Erlangen-Nürnberg. He has published about 200 papers in leading journals and conference proceedings. His research interests include visual computing, such as real-time rendering, visualization, 3D reconstruction, object tracking, video manipulation, and virtual and augmented reality. He has been elected to the Eurographics Executive Committee and serves currently as the Eurographics Professional Board Chair.

Martin Vossiek (Fellow, IEEE) received the Ph.D. degree from Ruhr-Universität Bochum, Bochum, Germany, in 1996. In 1996, he joined Siemens Corporate Technology, Munich, Germany, where he was the Head of the Microwave Systems Group, from 2000 to 2003. Since 2003, he has been a Full Professor with Clausthal University, Clausthal-Zellerfeld, Germany. Since 2011, he has been the Chair of the Institute of Microwaves and Photonics (LHFT), Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany. He has authored or coauthored more than 250 articles. His research has led to more than 90 granted patents. His current research interests include radar, transponder, RF identification, communication, and locating systems. He has been a member of organizing committees and technical program committees of many international conferences. He is a member of the German IEEE Microwave Theory and Techniques (MTT)/Antennas and Propagation (AP) Chapter Executive Board and the IEEE MTT-S Technical Coordinating Committees MTT-24, MTT-27, and MTT 29. He was the Founding Chair of the MTT IEEE Technical Coordinating Subcommittee MTT-27 Wireless-Enabled Automotive and Vehicular Application. He was a recipient of several international awards. Recently, he was awarded the 2019 Microwave Application Award from the IEEE MTT Society (MTT-S). He has served on the review boards for numerous technical journals. From 2013 to 2019, he was an Associate Editor of IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES.