IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 19, 2019, accepted October 2, 2019, date of publication October 7, 2019, date of current version October 21, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2945968

Microwave Imaging of Non-Rigid Moving Target Using 2D Sparse MIMO Array

ZHANYU ZHU[1], LEI KUANG[2], (Member, IEEE), AND FENG XU[1], (Senior Member, IEEE)

[1]Key Laboratory for Information Science of Electromagnetic Waves (MoE), Fudan University, Shanghai 200433, China
[2]Shanghai Key Laboratory of Multidimensional Information Processing, Engineering Center of SHMEC for Space Information and GNSS, School of Information Science and Technology, East China Normal University, Shanghai 200241, China

Corresponding author: Lei Kuang (lkuang@ee.ecnu.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFB0502703, in part by the Natural Science Foundation of China under Grant 61822107, Grant 61571134, and Grant 61571190, and in part by the Research Fund of Shanghai Academy of Space Technology under Grant SAST2017-078.

ABSTRACT This paper develops a microwave/mmw (millimeter-wave) imaging method for moving objects of non-rigid body which can be used in the heavy traffic scenarios. This method is proposed with a 2D sparse MIMO (multiple-input multiple-output) array composed of two orthogonal linear arrays. In comparison to the RX antennas, a smaller number of TX antennas is used to decrease the data acquisition time of this system when using TDM (time division multiplexing) signal transmitting technique. The 2D sparse MIMO array is deployed with wide-band signals transmitted to reconstruct the backscattering coefficient in 3D space using synthetic aperture imaging technique. Then a series of synthesized 3D snapshot images are generated during the motion of the target. The non-rigid human body target is decomposed into 25 joints tracked by an optical device which is Kinect in this paper. After the decomposition, the movement of each joint can be described as a space-time trajectory tracked by Kinect, providing a preliminary indication of the 3D position of each joint. With the indication, we divide every microwave/mmw image into 25 sub-images corresponding to the joints' position to form an image sequence of each joint. Subsequently, a segmental joint estimation method is proposed to estimate the accurate motion parameters of each joint from the image sequence. The decorrelation of the images in large angle observation can be significantly reduced by the segmental estimation which leads global optimized motion parameters. Consequently, the estimated parameters are taken into the effective movement compensation of each component. Meanwhile, the image refocusing method is presented to generate the well-focused image of each component. Finally, all the images of components are fused into one image of the whole target to reconstruct the 3D high-resolution image. Imaging simulations and measuring data experiments show that this method is efficient.

INDEX TERMS Microwave imaging, non-rigid moving target, 2D sparse MIMO array.

I. INTRODUCTION

The increasing incidents of terrorist activity have driven development of advanced security screening technologies, where microwave/millimeter-wave (mmw) imaging is one of the most important techniques [1]. In comparison with X-ray scanner [2], microwave/mmw imaging is considered a safer method for the operators and the targets. Meanwhile, microwave/mmw radiation can penetrate through many commonly used nonpolar dielectric materials and ceramic with little attenuation, and obtain high-resolution imagery for the detection of concealed weapons, explosives and other threats. The range resolution of microwave/mmw image is proportional to the frequency bandwidth, while the cross-range resolution is proportional to the aperture size. In order to achieve high resolution in cross-range, the general method is to increase the aperture size, which can be either achieved by using a large size antenna array or taking advantage of relative motion between antenna and target to form large synthetic aperture.

The most successful microwave/mmw imaging systems are currently deployed in airport and other port security scenarios. One example of the commercial usages is the security system with two 1D electrical scanning arrays for

The associate editor coordinating the review of this manuscript and approving it for publication was Huawei Chen.

cylindrically scanning developed by Pacific Northwest National Laboratory [3]. Measurement time of this system is about several seconds and thus each person must stand in the system for several seconds. This leads a limited range of application because the system is not suitable for heavy traffic scenarios, such as public facilities and public gatherings.

However, the major threats of violent terrorist usually occur in the places with a large flow of people, such as subway stations, squares and open public spaces. In recent years, a number of major terrorist attacks happened around the world fall into this category. To address this problem, we first need to implement security screening of non-cooperative targets, i.e. an ideal screening system should not interfere the normal traffic. Specifically, the system should not apply any limitation to the behaviors of human body and conduct screen without awareness of the target. Secondly, the system should be compact or even portable so that the operator may be able to inspect at the vicinity of any suspicious target.

The conventional approach of non-cooperative target microwave imaging takes fast snapshot of moving targets, where the target is considered stationary in a short period of time. It requires rapid parallel data acquisition over a very large array to realize high resolution imaging. One such method is imaging with large planar antenna array and millimeter-wave focusing lens [4]. The system is bulky and expensive because a huge 2D array is required behind the lens. Therefore, it is not widely deployed in practical applications. Recently, Rohde & Schwarz company in German has developed a fast imaging system based on MIMO sparse array [5], [6]. The system has powerful digital signal processing (DSP) capability with a highly integrated active electronic scanning array. The system integrates 3072 TX channels and 3072 RX channels, in which the TX channels transmit signals via time division multiplexing (TDM) and it requires about 25ms to complete a whole frame data acquisition. The operating frequency of the system is between 70 and 80 GHz and the lateral resolution is about 2 mm in both azimuth and elevation. Although the number of array elements has been reduced greatly by using a sparse MIMO array, thousands of antennas and channels means extremely high system cost and complexity. In addition, TDM of thousands TXs still takes sub-second of time which means the quality of the image will degrade if the target moves too fast. Many other studies try to build security imaging systems through MIMO array [7]–[10]. These methods aim at cooperative target and their main purpose is to reduce the array size and thus reduce the system complexity and construction cost.

Suppose a MIMO system works in TDM mode has N TX antennas. The switching time between two TX channels is τ and the total measurement time T can be calculated by

$$T = N \cdot \tau \tag{1}$$

Large aperture size requires a large N which results in long measurement time. Thus, the requirement of high resolution is inherently contradictory with the requirement of fast imaging. To solve the fundamental contradiction between high resolution and fast imaging, MIMO systems transmitting orthogonal frequency division multiplexing (OFDM) signals has been proposed to replace TDM systems in order to reduce measurement time. Some methods for orthogonal signals design of MIMO radar have been proposed [11]–[13], and the purpose is to find optimized signals whose sidelobe of auto-correlation function and energy of cross-correlation function both are as low as possible. It is a NP (non-deterministic polynomial) problem to obtain OFDM signals through optimizing the objective function. For a MIMO system with a large number of channels, a set of sufficient orthogonal signals is usually not tractable. In practical applications, the suppression of cross-correlation function in OFDM is limited, which leads to a limited dynamic range of the system. Hence, new methods should be proposed for imaging of fast moving non-cooperative target in real time.

This paper addresses this issue with inverse synthetic aperture radar (ISAR) imaging taking advantage of target motion to form a synthetic aperture in azimuth. That means the target is moving and the system is stationary. Thus, the number of antennas can be reduced greatly and the measurement time of the system can be improved while still obtaining the high resolution images through ISAR imaging. Note that the actual array size along the moving direction is also reduced accordingly. Thus, we can purposely design a MIMO system in such a way that the number of TX elements is small along the moving direction. Such a system would have a slim physical aperture and thus can meet the requirement of compact system size.

The key of ISAR imaging is to estimate the target motion parameters. Once target motion parameters are known, the equivalent aperture can be synthesized to generate a well-focused image [14]–[16]. Jaime *et al.* [17] develop a 3D microwave imaging system equipped with an optical depth camera, synchronizing acquisitions from different views of the moving human body. They applied non-coherent integration method into the imaging procedure and the microwave image is blurry so that only the profile of the concealed weapon can be obtained.

In this paper, human body is modeled as a non-rigid moving target consisting of several rigid bodies. First, auxiliary optical sensor, Kinect, is used to track the moving human skeleton joints. With the tracking information, a coarse motion estimation of each rigid body component is obtained. Then a small real aperture sparse MIMO array is used to capture a local snapshot of each rigid body component. Subsequently, the accurate trajectory of each component is estimated by cross-correlation of consecutive snapshots, with which the final ISAR image of each component can be obtained by refocusing all snapshots data. At last, the human target image can be obtained by fusing the images of all components.

This paper is organized as follows. Section 2 reviews some related studies of detecting and imaging of moving targets. Section 3 describes the system design and analyzes the design principle and the imaging flow. Section 4 demonstrates the

segmental joint estimation method for the motion parameters of non-rigid body target and the refocus imaging method for the target. Imaging simulations and measuring data experiments are given in section 5. Finally, this paper is concluded in section 6.

II. RELEVANT STUDIES

For motion compensation of ISAR, targets are often treated as rigid body composed of point scatterers [18]. Chen et al. present a method of range alignment and motion compensation to image a moving aircraft from a stationary ground-based radar, using the correlation information of the adjacent echoes [19]. In the case that the rotate angle of the target is less than 0.01°, the correlation coefficient of adjacent echoes is generally over 0.95. The envelop shifts are obtained through the operation of adjacent echoes correlation and the motion of the target can be compensated by shifting the echoes such that the signals from the same scatterer are centered at the same range bin in different echoes. The principle of envelope correlation of two adjacent echoes is sensitive to noise and target scintillation. In real cases, the scintillation of the target causes a sudden change between the echo envelopes and leads to inaccurate estimation. Meanwhile, the existence of noise leads to the defect of error accumulation which will get worse during the operation of continuous adjacent echoes correlation. Delisle et al. present a range-bin alignment method for moving target imaging with a two-receiver ISAR radar in use of the envelope correlation feature of different echoes, while the reference envelope is obtained by exponentially averaging the previously received envelopes [20]. The reliability of the system is improved by the joint use of the correlation information of two antennas. However, this method does not effectively solve the defects of target scintillation and error accumulation. Wang et al. propose a polynomial model for the shifts made to the echoes, and the coefficients of this polynomial are chosen to optimize a global quality measurement of range alignment. The global method shows better performance against noise, target scintillation and error accumulation than the method of envelope correlation with adjacent echoes [21].

The envelope alignment method cannot be used to adjust the Doppler phase caused by target movement. The general way is to keep the signal structure unchanged during the envelope shifting and then remove the translational component by extracting the initial phase of the range bin. Considering the effect on the phase compensation, the envelope alignment must not exceed $1/2 \sim 1/4$ range bin. In practical applications, the signal is usually not an ideal rectangular wave signal and the waveform is distorted after bandpass filter. Therefore, the error of envelope alignment is required not to exceed $1/4 \sim 1/8$ range bin. The phase gradient autofocus (PGA) technique is the most classical method for phase error compensation [22], [23]. PGA provides a good performance over the entire scene content but has to rely on the existence of dominant scattering points to estimate the phase error of the image with the assumption that the phase error is not spatial-variant. For ISAR imaging of maneuvering targets, a spatial-variant autofocus algorithm is proposed in [24], based on optimizing a polynomial model to maximize the image contrast.

Conventional 2D ISAR systems has the inherent ambiguity of the image projection plane and image scale in azimuth direction. A system generating 3D ISAR images can eliminate the problem completely. 3D images also provide robust characterization and feature identification of the point scatterers on the target. In theory, the 2D images of different projection planes produced by the 3D motion of the target are equivalent to the 2D images observed from different angles of view. The positions of point scatterers in 3D space can be determined through connecting the same point scatterer in the 2D ISAR image sequence of the target. In literatures [25], [26], 3D ISAR images of moving target are formed by exploiting single antenna system. The system has the advantage of requiring a single antenna, but it relies on long observation time and require the target to have motion in all dimensions.

Interferometric inverse synthetic aperture radar (InISAR) imaging is a 3D imaging technique combining the interforemetric technique and ISAR. The 3D image of target scattering centers is reconstructed by interferometric phase processing of different 2D ISAR images from the various projection planes. The principle of InISAR is similar to InSAR (Interferometric synthetic aperture radar) and the third dimension information can be obtained through the interforemetric technique when the moving direction of projection planes is perpendicular to the direction of the baseline of the antennas. A minimally required number of antennas is two. However, the projection plane is not necessarily perpendicular to the baseline and the height information cannot always be obtained by just two antennas [27], [28]. In [29], a 3D ISAR imaging system is proposed and a series of images with different view angles are obtained by using the same received data. Then, the 3D information is obtained from the series of images. A multichannel InISAR system composed of three TX/RX antennas lying on two orthogonal baselines with an L-shape is proposed in [30], [31]. Ma et al. analyze the 3D imaging method using the cross MIMO array, the square MIMO array and the interforemetric MIMO array in [32] and generate the ISAR image of the moving target using the square MIMO array in [33]. The simulation shows the 3D imaging of moving target can be effectively carried out using MIMO+ISAR method. More studies [34]–[38] have shown that MIMO+ISAR method can effectively perform 3D imaging of the target.

The aforementioned methods are based on the principle that the target is rigid body. That means the target is treated as several permanent scattering centers between which the relative positions are fixed. However, this condition cannot be satisfied in the case of non-rigid body. Especially the permanent scattering centers are hardly to be extracted without obvious geometric structure in human body imaging. Then these methods are not suitable for imaging moving

target of non-rigid body (such as the moving human body). In this paper, a microwave/mmw imaging method for moving objects of non-rigid body using 2D sparse MIMO array is proposed. A space-time trajectory model is introduced to describe the motion of the non-rigid target which is decomposed into several joints. Auxiliary optical sensor, Kinect, is used to obtain preliminary estimation of the 3D positions of human skeleton. With the sparse MIMO array, the 3D snapshot of each component dominated by a certain joint is synthesized and the image sequence is generated during the motion of the component. The accurate motion parameters of each component is estimated from the image sequence by the segmental joint estimation method and the refocused image of each component is reconstructed with movement compensation. Finally, all the images of components are fused into a 3D focused image of the whole target. The major innovations and contributions of this paper are summarized as follows.

(1) This system combines 2D sparse MIMO array and ISAR technique to perform 3D microwave/mmw imaging of the moving targets of non-rigid body.

(2) The 3D image of the non-rigid target is reconstructed by fusing together all the decomposed components which are considered as rigid bodies and imaged with movement compensation and refocusing.

(3) The imaging process uses Kinect as auxiliary optical sensor to obtain a preliminary coarse estimation of the positions of scatterers. Such coarse estimation is used to determine the imaging regions of each scatterer, which can reduce the load and improve the efficiency of data processing.

(4) The joint estimation method is proposed to estimate global optimal motion parameters of each scatterer from the image sequence.

III. SYSTEM DESIGN

A. OVERALL ARCHITECTURE

We design the system with a 2D planar array to realize the 3D imaging for moving targets of non-rigid body. The system design is an improvement on our related study [39]. This paper takes moving human body as an example. Suppose that the human target moves in horizontal direction. Respectively, x, y and z denotes the human walking direction, the vertical direction and the radar looking direction, as depicted in Fig. 1. We adopt a 2D sparse MIMO array with 'cross' shape arrangement in the system whose geometry is shown in Fig. 1. This MIMO array is consisted of two perpendicular linear arrays and the TX antennas are located in the x direction while the RX antennas are located in the y deriction. TDM is employed to switch among the TX antennas radiating electromagnetic signals sequentially while the RX antennas receive signals simultaneously. Noted that the array size in the x direction (the direction of human motion) is much shorter. Thus, the number of TX is greatly reduced resulting in short observation time and high frame rate. Note that the final imaging resolution along x direction would not be reduced as the final synthesized aperture will be increased due to human movement in the x direction, thereby balancing the resolution of the refocused image in the x and z directions.

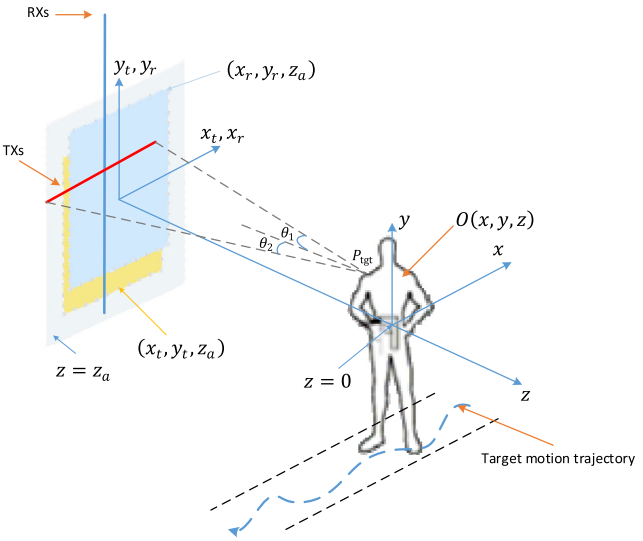

FIGURE 1. 3D geometry of MIMO array imaging system.

In terms of system size, the structure of such antenna array can be easily made up of two linear arrays which are detachable and reassemble in the field. Therefore, the system can be used in a temporary check point if needed. It is applicable to detect concealed threats carried by pedestrian. The ISAR image of the human body is generated through the motion parameter estimation, movement compensation and refocusing. The advantages of such system includes transportable, deployable in temporary place, without interfering human traffic, etc.

Mathematically, the backscattering coefficient of the target can be derived with the time domain algorithm [39], described as

$$\hat{O}(x, y, z) = \sum_{l=1}^{L} \sum_{n=1}^{N} \sum_{m=1}^{M} s\left(x_{t_m}, y_{t_m}, x_{r_n}, y_{r_n}, k_l\right) \cdot \exp\left(jk_l R_{t_m}\right) \cdot \exp\left(jk_l R_{r_n}\right) \quad (2)$$

where M, N and L represent the number of TX antennas, Rx antennas and frequency sample points respectively. z_a is the planar array position in z axis. The TX and Rx antenna positions are represented as (x_t, y_t, z_a) and (x_r, y_r, z_a) respectively. $O(x, y, z)$ is the backscattering coefficient of the target. k is the wavenumber of the transmitted signal. $\left(x_{t_m}, y_{t_m}, z_a\right)$ and $\left(x_{r_n}, y_{r_n}, z_a\right)$ are the positions of the mth TX antenna and the nth RX antenna respectively. k_l is the lth wavenumber of the transmitted signal. $\hat{O}(x, y, z)$ represents the processed backscattering coefficient of the target. $R_{t_m} = \sqrt{\left(x_{t_m} - x\right)^2 + \left(y_{t_m} - y\right)^2 + (z_a - z)^2}$ represents the range from the mth TX antenna to the target. $R_{r_n} = \sqrt{\left(x_{r_n} - x\right)^2 + \left(y_{r_n} - y\right)^2 + (z_a - z)^2}$ represents the range from the nth RX antenna to the target.

The backscattering coefficient of the target also can be solved with the frequency domain algorithm, described as

$$\hat{O}(x, y, z) = \mathrm{IFT}_{3\mathrm{D}}\left\{S\left(k_x, k_y, k_z\right) \cdot \exp\left(-jk_z z_a\right)\right\} \quad (3)$$

where $S(\cdot)$ is the wavenumber domain expression of $s(\cdot)$ and can be written as

$$S\left(k_x, k_y, k_z\right) = \exp\left(jk_z z_a\right) \cdot \int_x \int_y \int_z O(x, y, z) \cdot \exp\left(-jk_x x\right) \cdot \exp\left(-jk_y y\right) \cdot \exp\left(-jk_z z\right) \cdot dx\, \mathrm{d}y\, \mathrm{d}z \quad (4)$$

with

$$\begin{aligned} k_x &= k_{x_t} + k_{x_r} \\ k_y &= k_{y_t} + k_{y_r} \\ k_z &= \sqrt{k^2 - k_{x_t}^2 - k_{y_t}^2} + \sqrt{k^2 - k_{x_r}^2 - k_{y_r}^2} \end{aligned} \quad (5)$$

where k_{x_t}, k_{y_t}, k_{x_r} and k_{y_r} represent the wavenumber of TX antennas and RX antennas in x direction and y direction respectively.

The frequency domain algorithm utilizes a dimension-reducing rearrangement operation to transfer the 5D wavenumber domain $\left(k_{x_t}, k_{y_t}, k_{x_r}, k_{y_r}, k\right)$ data into 3D wavenumber domain $\left(k_x, k_y, k_z\right)$. To facilitate the rearrangement, it is required that the sampling steps should be the same in k_{x_t} and k_{x_r} and also the same in k_{y_t} and k_{y_r}. Moreover, k_x and k_y are required uniform along the x and y direction respectively.

The planar cross MIMO array is adopted in 3D microwave imaging since it has spatial degrees of freedom in three dimensions with wideband illumination [40], [41]. According to the frequency domain algorithm, the image resolution of the planar cross MIMO array system can be written as

$$\begin{aligned} \delta_x &= \frac{2\pi}{\Delta k_x} \\ \delta_y &= \frac{2\pi}{\Delta k_y} \\ \delta_z &= \frac{2\pi}{\Delta k_z} \end{aligned} \quad (6)$$

Especially, suppose a point target P_{tgt} in front of the antenna array and its resolution in vertical direction can be further calculated as

$$\delta_x \approx \frac{2\pi}{k_{\mathrm{c}}\left(\sin\theta_1 + \sin\theta_2\right)} \quad (7)$$

where k_{c} represents the wavenumber of the carrier frequency. θ_1 and θ_2 are the angles between the z axis and the line from P_{tgt} to each edge of the TX antennas array along the x axis, as described in Fig. 1.

According to the derivation above, the target resolution in xy plane can be simplified as

$$\begin{aligned} \delta_x &\approx \frac{\pi}{k_{\mathrm{c}}\sin\left(\theta/2\right)} \\ \delta_y &\approx \frac{\pi}{k_{\mathrm{c}}\sin\left(\varphi/2\right)} \end{aligned} \quad (8)$$

where θ and φ are the accumulated angles of the point target along vertical and horizontal direction respectively, with the assumption that the target is located in the center of the scene.

In the far field condition, the target resolution in xy plane can be derived as

$$\begin{aligned} \delta_x &\approx \frac{\lambda}{\theta} \\ \delta_y &\approx \frac{\lambda}{\varphi} \end{aligned} \quad (9)$$

Thus it is can be seen that the relationship between the resolution of the planar cross MIMO array and the monostatic radar in far filed condition is

$$\begin{aligned} \delta_{x\_\mathrm{mimo}} &= 2\delta_{x\_\mathrm{mono}} \\ \delta_{y\_\mathrm{mimo}} &= 2\delta_{y\_\mathrm{mono}} \end{aligned} \quad (10)$$

In literature [42], the resolution $\delta_{x,y}$ in xy plane is given with aperture size as

$$\delta_{x,y} \approx \frac{\lambda}{L_{x,y}} R \quad (11)$$

where $L_{x,y}$ represents the length of the aperture in the corresponding direction. λ is the wavelength of the transmitted signal. R denotes the distance between the target and the antenna. The range resolution δ_z is determined by the frequency bandwidth B, and can be written as

$$\delta_z \approx \frac{c_0}{2B} \quad (12)$$

where c_0 is the speed of light.

Obviously, (10) is adapted in the condition $R \gg L_{x,z}$. According to the principle of phase center approximation (PCA) [53], the phase center of a TX and RX antenna pair in the MIMO system can be approximated as the phase of a virtual antenna at the middle position of the TX antenna and the RX antenna. With this approximation, the planar cross MIMO array is equivalent to a rectangular monostatic array in half size. Thus, (10) and (11) can also be derived from PCA principle and can be adopted in the condition $R \gg L_{x,y}$.

In order to obtain high resolution image of target, the resolution in range direction can be improved by increasing bandwidth B while the resolution in azimuth can be improved by increasing the electrical size of the antenna apertures which are L_x/λ and L_y/λ. The increase of the electrical size of the apertures means that more antenna units are required to satisfy the half-wavelength space sampling principle. In this paper, we obtain the resolution in x and y directions using 2D sparse MIMO array. Sparse MIMO array can reduce the cost and weight of the system and break the half-wavelength space sampling condition but without degrading the image quality.

When imaging the moving human target, it is assumed that the target moves along the x axis. Therefore, the resolution δ_x in x direction is

$$\delta_x \approx \frac{\pi}{k_{\mathrm{c}}\sin\left(\theta_{\mathrm{acc}}/2\right)} \quad (13)$$

where θ_{acc} is the accumulated angle caused by the human movement during the MIMO array observation time. Using the inverse synthetic aperture technique, a larger virtual antenna array can be generated from the human movement to obtain a larger accumulated angle θ_{acc}. Note that (13) is taken from (8) with the assumption that any TX and RX antenna pair forms a channel, which means the RX antenna can receive the signal radiated from the TX antenna.

Simultaneously, according (11) and (12) we can see that the resolutions in y direction and z direction are λ/φ and $c_0/2B$, respectively. Using sparse MIMO array, 3D real-aperture snapshot image of the target can be obtained. Equation (13) shows the resolution in target moving direction can be improved by synthesizing larger aperture through target motion.

In the real aperture radar system, the imaging resolution in x direction can be derived from (13). When k_{c} is established in system design, we can obtain that $\delta_x \propto 1/\theta_{\text{acc}} \approx \text{R}/L_x \propto 1/L_x$. That is to say, the minimum of δ_x is limited by L_x. If the target is moving in x direction, the relative motion between the target and the system is equivalent to the case where system moving in x direction and the target is stationary. Using ISAR processing, we can synthesize virtual array positions over time and form a larger antenna array aperture of length $L_x + L_{\text{move}}$, where L_{move} is the movement length in the x direction. The improved resolution can be written as

$$\delta'_x \approx \frac{\pi}{k_{\text{c}} \sin\left[\left(L_x + L_{\text{move}}\right)/2R\right]} \tag{14}$$

Compared with the real aperture radar system, the same performance can be achieved with the combination of MIMO array and synthetic aperture method which requires less number of antenna elements. In this system, the synthetic aperture is formed by the motion of the human body. Meanwhile, a parameter model based on space-time trajectory [47] is introduced to describe the motion of human body during the observation time. Simultaneously, the human target is decomposed into a series of 3D points to describe the positions of corresponding scatterers. The target motion is extracted from the image sequence and recorded as the joint positions with space-time trajectory.

Based on the kinematic analysis, a human body can be modeled by a multicomponent system and each interconnected component segment is considered as a rigid body [48]. These component segments represent different range-Doppler information in return signals of radar system when the human body is moving. For a rigid target, the range and velocity information can be precisely measured from the radar signals with phase derivation [49]. For more complex cases, some researches have been studied about extracting parameters of multicomponent target from radar signals in [50], [51].

In this paper, the human target is decomposed into 25 components such as the head, neck, shoulder and elbow in skeleton tracking to reconstruct the 3D motion positions, as illustrated in Fig. 2. With the introduction of some

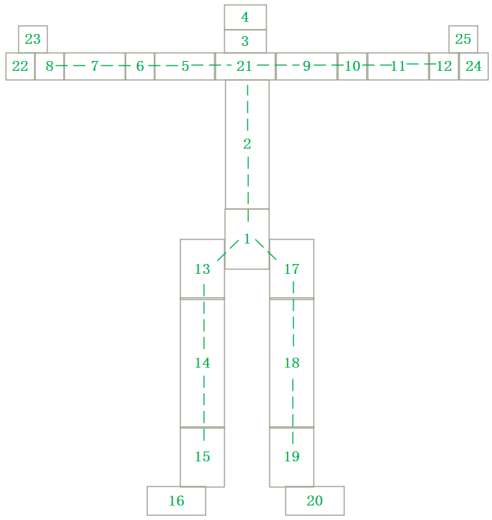

FIGURE 2. The model of human skeleton joints.

TABLE 1. The parameters of Kinect v2.

	Parameter	Value
Depth camera	Resolution	512 × 424 px
	Field of view (h × v)	70.6° × 60.0°
	Angular resolution	0.14° / px
	Operating range	0.5 – 4.5 m
Frame rate		30 Hz
Minimum latency		20 ms
Connection type		USB 3.0
Voltage		12 V DC
Power usage		~15 W
Price		200 USD

inexpensive 3D information acquisition system, such as Kinect [52], real-time 3D information extraction can be obtained. Kinect v2 reconstructs a 3D depth image using the depth camera [43] with parameters [44] listed in Table 1. Kinect v2 can detect 25 skeleton joints which satisfies the initial imaging requirement in our in-lab developed prototype system, while Kinect v1 can only detect 20. We consider each component dominated by the skeleton joints as a rigid body without deformation to simplify the motion model of the human body. In practice, the movement of human body is more complicated with the contraction and rotation of muscles. The accurate parameters of more joints can be acquired, the better imaging quality can be achieved. Since the good performance of Kinect v2 in the human motion tracking and its low price, it is easy to get and determined to be used as the auxiliary optical device.

Considering the field of view and system error of Kinect, the range from the target to Kinect should be ∼2 m. Then, the field of view achieves 3.6m × 2.0m(h × v) and the resolution of depth camera is ∼5mm. The accuracy of depth value

is effected by system error and environmental noise such as light condition. Therefore, the error of the skeleton information achieves centimeter-level. When deployed in large areas, an effective approach is to increase the number of the developed system to cover these areas and different sizes of the human body. Another choice is to introduce optical device with motion tracking function in longer distance as auxiliary.

Consistent with Kinect and as a preliminary validation, we use the method described in Fig. 2 directly to decompose the non-rigid target into 25 joints which can be seen as point satterers, with indices ranging from 1 to 25. Note that the point scatterers and the joints do not necessarily coincide with each other in practical applications. We can find that there are almost no obvious point scatterers at the positions of the joints from the 3D mm-wave images of human body in literature [6]. We assume the existence of these scattering points in order to interpolate them to obtain the accurate component positions of the human body. It is similar to the optical processing in literature [45], we consider the human body consists of several rigid body segments and dominated by the joints with degrees of freedom (DOFs). Using Kinect v2 as an auxiliary, 25 joints of the human body can be easily obtained. However, the relationship between the point scatterers and the joint points should be established in practice to effectively indicate the position of each component. It is not discussed in this paper which is rather a proof-of-concept study. Moreover, we just propose a preliminary discussion about microwave imaging of non-rigid body and the shade effect of the scatterers is not considered here.

Considering the normal speed of human walk is 1∼2 m/s, the frame rate of MIMO array snapshot is designed about 10Hz in our work, which makes a compact coverage in spatial sampling along x axis. The resolution δ'_x depends on the movement length L_{move} in x direction which can be derived from (14). That means the best resolution can be achieved when the movement spanned the entire observation area of Kinect in x direction, even if the human body does not walk very fast. Since the frame rate of Kinect is 30Hz which is faster than the MIMO array snapshot, we adapt linear interpolation method in Kinect skeleton data to obtain the skeleton position corresponding to the MIMO array snapshot.

The components represented by the scatterers are approximately regarded as rigid bodies. When the human body passes through the surveillance area, it is considered that the synthesized apertures are formed separately for each component, as they may move differently relative to radar and thus their motion parameters should be estimated independently. The motion parameter of each scatterer is estimated preliminarily using the correlation operation among the real-aperture snapshots at different times.

Consequently, the accurate parameters are estimated using the joint estimation method to suppress the accumulation error and target scintillation. Meanwhile, the segmental motion parameter estimation is used to reduce the decorrelation of the images in large angle observation. In practical applications, motion parameters of component connecting two scatterers, e.g. arm between two joints, can be obtained by interpolation. Every component is then imaged by refocusing all observation data. Finally, images of all components are fused to form an image of the whole target. Fig. 3 summarizes the flowchart of the 3D imaging method of moving non-rigid target. One case of the moving object imaging is depicted in the experiments part of this paper. The results of the single snapshot, the segmental-estimation, the joint-estimation and the final imaging are shown in figures.

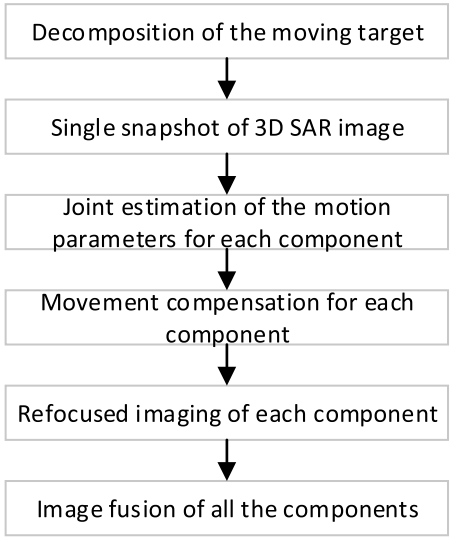

FIGURE 3. Flowchart of 3D imaging method of non-rigid moving target.

B. 2D SPARSE MIMO ARRAY

In MIMO system with M transmitters and N receivers, the number of virtual antennas is MN according to the principle PCA. As is demonstrated in [43], the equivalent array factor $a_E(x, y)$ can be used to evaluate the performance of the MIMO system, which is

$$a_E(x, y) = a_T(x, y) \circledast a_R(x, y) \tag{15}$$

where $a_T(x, y)$ and $a_R(x, y)$ denote the distribution of the transmitters and the receivers, respectively and denotes the convolution operation.

The cross MIMO array adopted in this paper is shown in Fig. 1. M transmitters are located along the x direction to form a linear periodic array with the interval of 0.75λ, denoted as $T_1, T_2, \ldots, T_M$. In the same way, N receivers are located along the y direction with the interval of 0.75λ, denoted as $R_1, R_2, \ldots, R_N$. The corresponding equivalent array is a uniform 2D square array according to the PCA principle. The equivalent aperture intervals in x and y directions both are 0.375λ. Therefore, the space sampling interval of the equivalent array satisfies the constraint of $< \lambda/2$. Since (15) is derived under the far-field condition, the equivalent array interval is designed as 0.375λ to avoid the grating lobe caused by under-sampling in the near-field region.

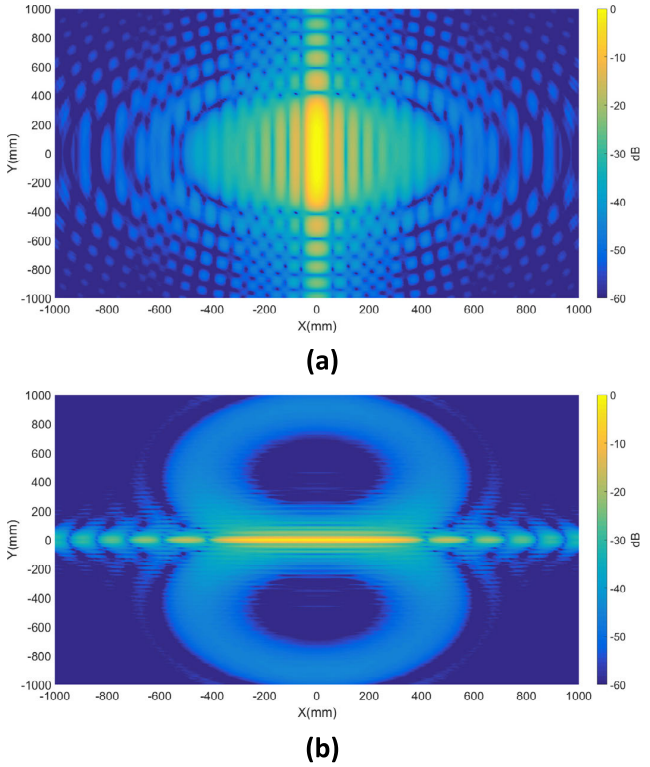

FIGURE 4. In the xz plane, (a) PSF of the transmitter arrays, (b) PSF of the receiver arrays.

The examples given in this paper are designed for the application of human security screening. The carrier frequency is 24GHz. The wavelength is 12.5mm. The bandwidth is 4GHz. The real aperture sizes in x and y directions are 0.2155m and 0.89m respectively. The conventional imaging distance is 2 m. The resolutions estimated according to (8) and (12) in x, y and z directions are 11.62cm, 2.88cm and 3.75cm respectively. The measurement time for one real-aperture snapshot is about 0.24ms where a pulse width 10us per channel. The human body target is seen as stationary when its movement is negligible during 0.24ms. The snapshot is taken every 96ms, which means the frame rate of MIMO array snapshot is about 10Hz. Note that all these parameters are chosen according to specifications of our in-lab developed prototype system.

Under the near-field condition, the received echoes are spatially non-uniform in the MIMO system and the frequency domain imaging methods based on Fourier transform are not directly applicable. In this paper, the back projection (BP) algorithm based on the time domain imaging method is used to image the target. 3D BP algorithm is implemented based on the graphic processing unit (GPU) which can be found in our related work [39]. The GPU implementation of 3D BP can be accelerated by over 400 times. With a powerful GPU station, it takes only $\sim$1s to finish the imaging of one person. The response function of the ideal point target in the imaging system is called point spread function (PSF), which can be used to measure the performance of the imaging system [45]. PSF of the antenna array in xy plane is analyzed using BP algorithm in Fig. 4, where PSF of the transmitter array is

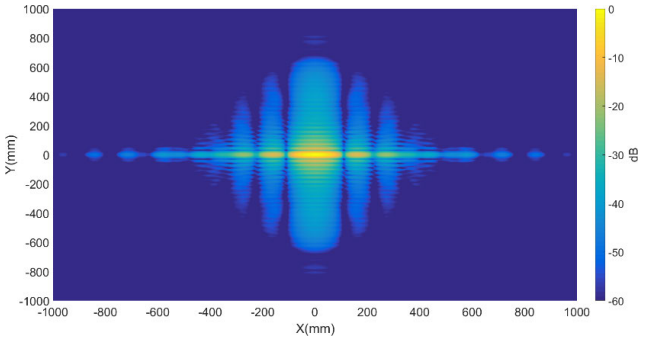

FIGURE 5. PSF of the 2D sparse MIMO array.

shown in Fig. 4(a) and PSF of the receiver array is shown in Fig. 4(b). In Fig. 4(a), the PSF has a narrow main lobe in the x direction along which the transmitter antennas are located to form a linear periodic array. Meanwhile, in Fig. 4(b), the PSF has a narrow main lobe in the y direction along which the transmitter antennas are located to form a linear periodic array. The resolution of the z direction is determined by the bandwidth of the transmit signals.

In this system, the antenna arrays of transmitters and receivers are sparsely arranged to form a series of transmitter/receiver pairs. The interaction of the transmitter/receiver pairs makes PSF have narrow main lobes both in two dimensions of the $x-y$ plane. It can be seen that the PSF of the equivalent array can be roughly seen as the product of the transmitter array PSF and the receiver array PSF by transforming the equivalent array factor in (15) to the angular domain through Fourier transform. This is basically consistent with the results shown in Fig. 5.

In this system, the MIMO array can perform snapshot imaging of the target, but the resolution along the x direction is low due to the small array size of the transmitter array. However, the response capability of the system is improved with the reduction of the number of transmitters. As a complement to the performance, the system uses the target motion along the x direction to synthesize a larger aperture, thereby increasing the resolution in x direction.

IV. JOINT ESTIMATION OF MOTION PARAMETERS

A. CORRELATION OF 3D REAL-APERTURE SNAPSHOTS

The design of the system proposed in this paper has high frame rate in terms of real-aperture snapshot imaging. Thus, the target movement between two successive snapshot frames is small relative to the imaging scale. High correlation between frames can be guaranteed. The extraction and tracking of the target point scatterers are realized using the human skeleton extraction and tracking technique assisted with Kinect v2. Then each scatterer is imaged separately using the method of 3D real aperture imaging. The positions estimated by Kinect provide coarse position information which can be used to determine the imaging area of each scatterer, which can reduce the data amount of processing and improve the efficiency. The estimated positions of Kinect usually have an error of centimeter level. In order to obtain accurate motion

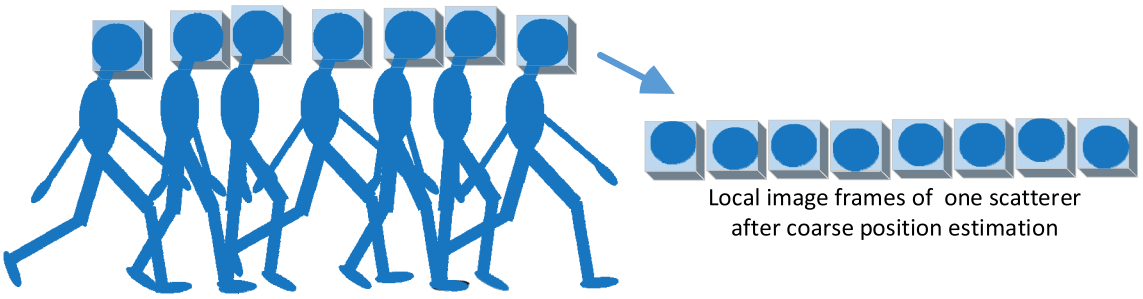

FIGURE 6. Local snapshot images of a single scatterer.

parameters, the 3D correlation of successive image frames is carried out. The image frames used for correlation are the localized image patches of the selected scatterer, as shown in Fig. 6.

The correlation function $R_{ij}\left(x',y',z'\right)$ between two 3D images $g_i(x, y, z)$ and $g_j(x, y, z)$ is defined as

$$R_{ij}\left(x', y', z'\right) = \int\int\int g_i(x, y, z)g_j(x - x', y - y', z - z')\mathrm{d}x\mathrm{d}y\mathrm{d}z \quad (16)$$

It is assumed that there only exists linear translation (m, n, l) between two frames, i.e.

$$g_j(x, y, z) = g_i(x - m, y - n, z - l) \quad (17)$$

which can be expressed via Fourier transform as

$$G_j(f_x, f_y, f_z) = G_i(f_x, f_y, f_z)e^{-2\pi j(mf_x+nf_y+lf_z)} \quad (18)$$

where G_j and G_i are the Fourier transforms of g_j and g_i, respectively. Then, (18) can be written as

$$\begin{aligned} R_{ij}(x, y, z) &= \mathrm{IFT}_{3\mathrm{D}}\left(G_i\left(f_x, f_y, f_z\right)G_j^*\left(f_x, f_y, f_z\right)\right) \\ &= \mathrm{IFT}_{3\mathrm{D}}\left(F\left(f_x, f_y, f_z\right)e^{2\pi j(mf_x+nf_y+lf_z)}\right) \\ &= f(x - m, y - n, z - l) \end{aligned} \quad (19)$$

where, $F\left(f_x, f_y, f_z\right) = G_i\left(f_x, f_y, f_z\right)G_i^*(f_x, f_y, f_z)$ is satisfied and f is the inverse Fourier transform of F. According to the characteristic of the autocorrelation function, the maximum value of $f(x, y, z)$ is located at $(0, 0, 0)$ and $R_{ij}(x, y, z)$ reaches the maximum value at (m, n, l). Hence, the peak of the cross-correlation function represents the offset between two frames.

The images operated in practical applications for cross-correlation consist of discrete values. Thus, interpolation is needed to improve the accuracy of m, n and l. The operation of FFT, zero-padding and IFFT is an efficient method to interpolate digital signals.

When K images are processed, using cross-correlation function of any two successive images, $K-1$ peak offsets are extracted. The offsets represent the movements of the target in the 3D space. In fact, due to the signal noise, the signal distortion caused by the system and the target scattering scintillation, the estimation of motion parameters using successive image correlation method leads to cumulative error and abrupt change of the motion parameters. This paper proposes a joint estimation method for the motion parameters. The least squares estimation of the motion parameters using any two images are taken to estimate the global optimal solution. Since this method degrades performance as the image correlation decreases, we propose to conduct joint estimation method in a segment-by-segment fashion.

B. SEGMENTAL JOINT ESTIMATION

Assuming the target is moving relative to the radar system and K images are obtained from the system, the cross-correlation coefficients R_{ij} and R_{ji} of 3D images $g_i(x, y, z)$ and $g_j(x, y, z)$ can be calculated through (16). According to the peak coordinates of the cross-correlation coefficients, the relative offsets of the images are $P_{ij} = \left[\Delta x_{ij}, \Delta y_{ij}, \Delta z_{ij}\right]^T$ and $P_{ji} = \left[\Delta x_{ji}, \Delta y_{ji}, \Delta z_{ji}\right]^T$. Thus, the cross-correlation matrix $\mathbf{R}$ of the any pair of images can be written as

$$\mathbf{R} = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1K} \\ P_{21} & P_{22} & \cdots & P_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ P_{K1} & P_{K2} & \cdots & P_{KK} \end{bmatrix} \quad (20)$$

Assume that $P_i = \left[\Delta x_i, \Delta y_i, \Delta z_i\right]^T$ is the offset of $g_i(x, y, z)$ relative to the reference image and $P_j = \left[\Delta x_j, \Delta y_j, \Delta z_j\right]^T$ is the offset of $g_j(x, y, z)$ relative to the reference image. We have $P_{ij} = P_i - P_j$ and $P_{ji} = P_j - P_i$. Thus, (20) can be written as

$$\begin{aligned} \mathbf{R} &= \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1K} \\ P_{21} & P_{22} & \cdots & P_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ P_{K1} & P_{K2} & \cdots & P_{KK} \end{bmatrix} \\ &= \begin{bmatrix} P_1 & P_1 & \cdots & P_1 \\ P_2 & P_2 & \cdots & P_2 \\ \vdots & \vdots & \ddots & \vdots \\ P_K & P_K & \cdots & P_K \end{bmatrix} \\ &\quad - \begin{bmatrix} P_1 & P_2 & \cdots & P_K \\ P_1 & P_2 & \cdots & P_K \\ \vdots & \vdots & \ddots & \vdots \\ P_1 & P_2 & \cdots & P_K \end{bmatrix} \end{aligned} \quad (21)$$

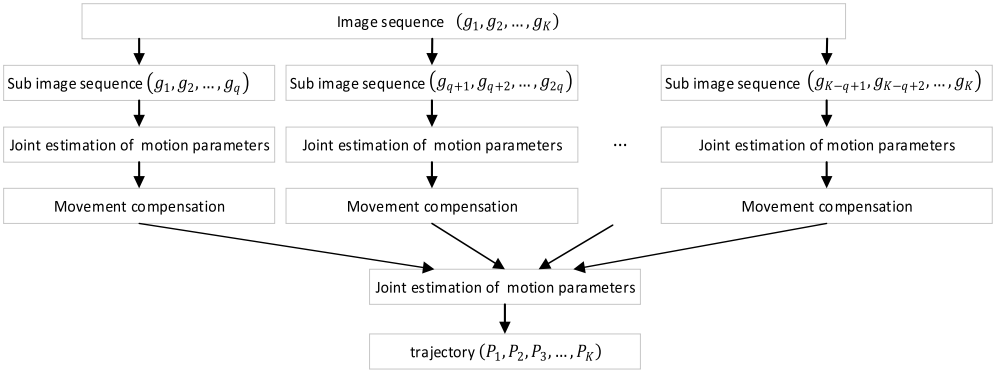

FIGURE 7. The segmental joint estimation method of target motion parameters.

The elements in $\mathbf{R}$ can be rearranged into a vector y as

$$y = [P_{11}, P_{12}, \ldots, P_{1K}, P_{21}, P_{22}, \ldots, P_{2K}, \ldots, P_{K1}, P_{K2}, \ldots, P_{KK}]^T \quad (22)$$

and we introduce x as

$$x = [P_1, P_2, \ldots, P_K]^T \quad (23)$$

Therefore, y can be written as

$$y = \mathbf{A}x \quad (24)$$

where $\mathbf{A}$ is a $K^2 \times K$ matrix defined as

$$\boldsymbol{A} = \left\{a_{ij}\right\}^{K^2 \times K},$$
$$a_{ij} = \begin{cases} 1 & mod\ (i, K) \neq 1 \& j = \mathrm{mod}(i, K) \\ -1 & mod\ (i, K) \neq 1 \& j = \mathrm{ceil}(i/K) \\ 0 & mod\ (i, K) = 1 \end{cases} \quad (25)$$

where mod$(\cdot)$ denotes modular operation and ceil$(\cdot)$ denotes round-up operation. Equation (24) is an underdetermined problem. In (24), x has infinite solutions because rank $(A) = K - 1$. Therefore, an additional constraint is introduced as

$$y' = \begin{bmatrix} 0 \\ y \end{bmatrix} \quad (26)$$

$$\mathbf{A}' = \begin{bmatrix} \mathbf{A}'' \\ \mathbf{A} \end{bmatrix}, \quad \mathbf{A}'' = [1, 0, \ldots 0] \quad (27)$$

Therefore, (24) can be written as

$$y' = \mathbf{A}'x \quad (28)$$

The solution of (27) can be obtained by least square method as

$$\hat{x} = \left(\mathbf{A}'^{\mathrm{T}}\mathbf{A}'\right)^{-1}\mathbf{A}'^{\mathrm{T}}y' \quad (29)$$

where $\hat{x}$ is the estimated value of x.

Due to target angular decorrelation effect, the correlation coefficient between two images decreases with increased separation of the two images in time and/or observation angle. However, in the near-field condition, the imaging system is usually operated with a wide-angle observation to obtain a high resolution image of the target. Thus, a segmental joint estimation method for target motion is proposed in this paper to solve the problem of angular decorrelation.

Suppose that K images are generated by snapshot imaging of the moving target, denoted as $(g_1, g_2, \ldots, g_K)$. The distance which the target moves respect to the radar system is L_{obj}. The distance between the imaging area center and the antenna array center is R_{obj}. The rotate angle of the target relative to the antenna array is θ_m and $\theta_m \approx \arcsin\left(L_{\mathrm{obj}}/R_{\mathrm{obj}}\right)$. The snapshot images are divided into k groups. In each group, there are $q = K/k$ images and the rotate angle of the target is $\theta_k \approx \theta_m/k$. The images in the ith group are denoted as $\left(g_{i1}, g_{i2}, \ldots, g_{iq}\right)$. The image acquisition interval is denoted as (t_{i-1}, t_i). After the joint estimation of the target motion parameters, the offsets of the images are $Q_i = (Q_{i1}, Q_{i2}, \ldots, Q_{iq})$,which can describe the trajectory of the target. Using the target motion parameters, the signals acquired in (t_{i-1}, t_i) can be compensated to generate the image g_i. Consequently, k images are generated, previously defined as $(g_1, g_2, \ldots, g_{\mathrm{k}})$. Then the joint-estimation method is used again in the image sequence $(g_1, g_2, \ldots, g_{\mathrm{k}})$ to obtain a new trajectory $Q' = (Q'_1, Q'_2, \ldots, Q'_k)$. Q' is used to correct the value of Q_i, and the corrected value P_i can be written as

$$P_i = \left(Q'_i + Q_{i1}, Q'_i + Q_{i2}, \ldots, Q'_i + Q_{iq}\right) \quad (30)$$

Therefore, the trajectory of the target during the observation is $(P_1, P_2, P_3, \ldots, P_K)$. The flow chart of the segmental joint-estimation method of target motion parameters is shown in Fig. 7.

C. REFOCUSING

The system performs snapshot imaging during the target motion, and the K images obtained have relatively low resolution in the x direction. After the joint estimation of the motion parameters, the offsets of these images are $(P_1, P_2, \ldots, P_K)$. Assuming that the measurement time of one frame is short enough, the target can be considered stationary in each frame time. The frame offsets $(P_1, P_2, \ldots, P_K)$ are relative to the initial position $\bar{P}_1$.

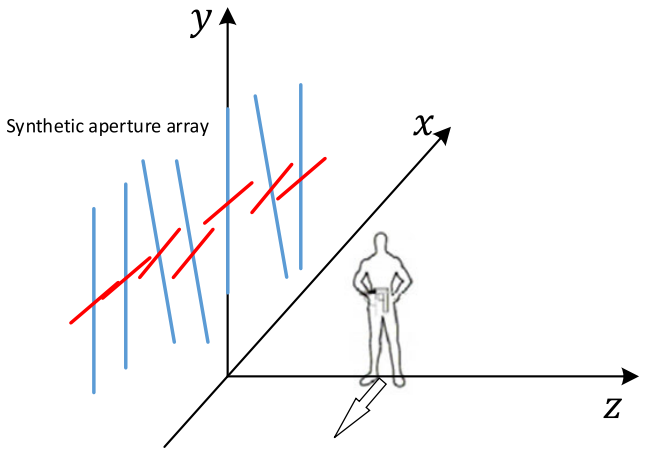

FIGURE 8. Synthetic aperture array formed by compensating target's relative motion.

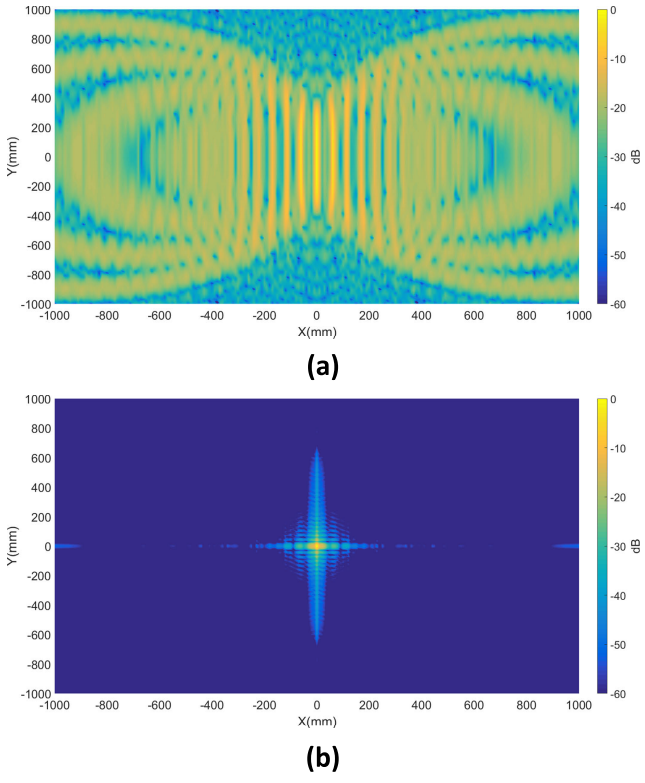

FIGURE 9. The PSF of, (a) ISAR, (b) MIMO+ISAR.

The system uses M transmitters and N receivers to form a planar MIMO array. The positions of the transmitter array and the receiver array are known as $\left(P_{T_1}, P_{T_2}, \ldots, P_{T_M}\right)$ and $\left(P_{R_1}, P_{R_2}, \ldots, P_{R_N}\right)$, respectively. Where, $P_{T_i} = \left[x_{T_i}, 0, 0\right]^T, i = 1, 2, \ldots, M$ and $P_{R_j} = \left[0, y_{R_j}, 0\right]^T, j = 1, 2, \ldots, N$. The target position is P_{t_i} relative to the initial position $\bar{P}_1$ at time t_i. Assuming that the target position is stationary and the radar is moving, the antenna positions of the system are compensated. The compensated transmitter antenna positions are $\left(P_{T_1} - P_{t_i}, P_{T_2} - P_{t_i}, \ldots, P_{T_M} - P_{t_i}\right)$ and the compensated receiver antenna positions are $\left(P_{R_1} - P_{t_i}, P_{R_2} - P_{t_i}, \ldots, P_{R_N} - P_{t_i}\right)$. As a result, a complex

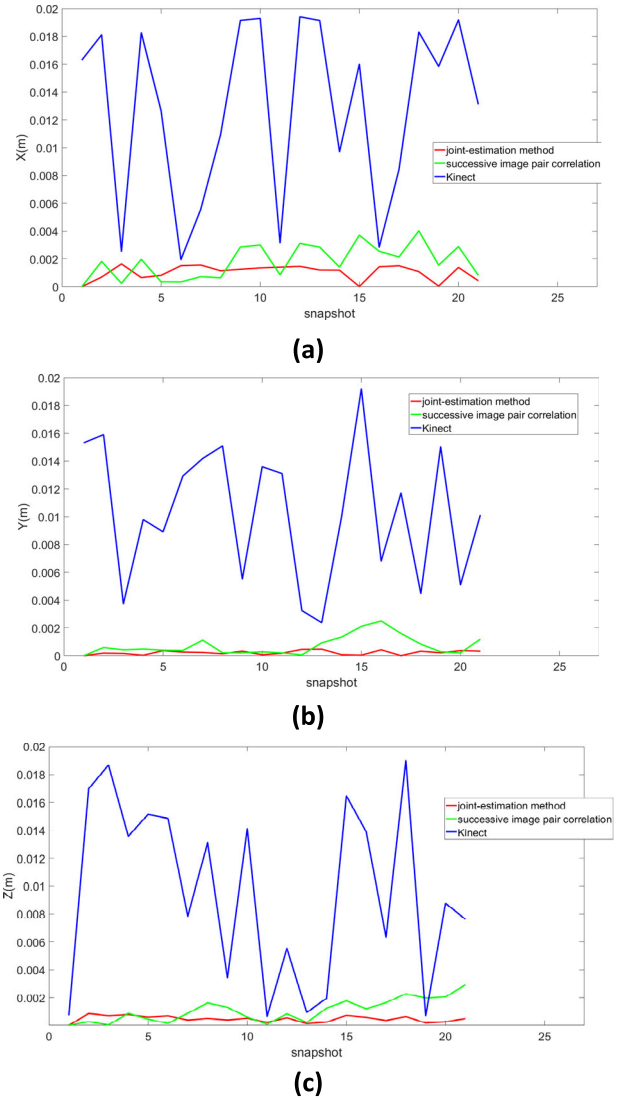

FIGURE 10. With 2cm perturbation error, the absolute estimation error comparison in (a) x direction, (b) y direction and (c) z direction.

virtual aperture array is synthesized during the target motion, where the positions of the virtual apertures may be irregularly located, as shown in Fig. 8.

The quality of the final image depends on the accuracy of the target motion parameter estimation. However, the phase error caused by the target motion is not fully compensated after movement compensation, because its accuracy at the sub resolution level might not satisfactory to compensate phase errors. Thus, the target image still has the residual phase error after imaging procedure.

In this paper, the motion parameters obtained by joint estimation are used as preliminary compensation for the target motion, and K 3D images are obtained, where the envelope of the point scatterer is aligned in three dimensions. Based on the BP algorithm, the image can be coherently synthesized. Especially, when the peak point of the scatterer is added in phase, the best performance can be achieved. To implement the synthesis, K images are denoted as $g_k\ (x, y, z)\ , k = 1, 2, \ldots, K$

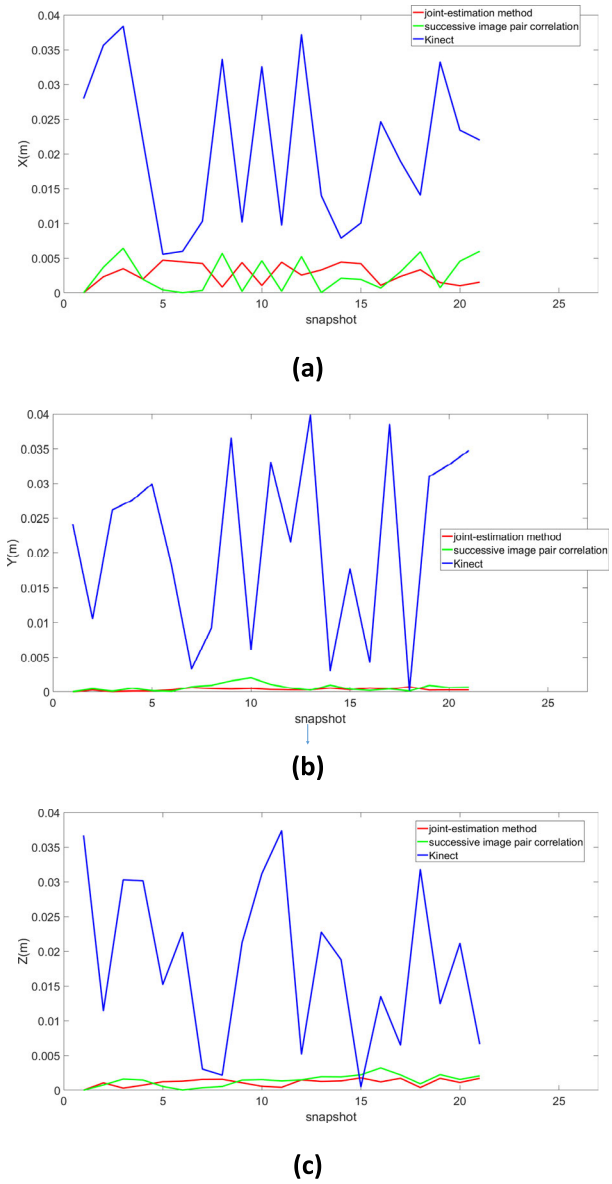

FIGURE 11. With 4cm perturbation error, the absolute estimation error comparison in (a) x direction, (b) y direction and (c) z direction.

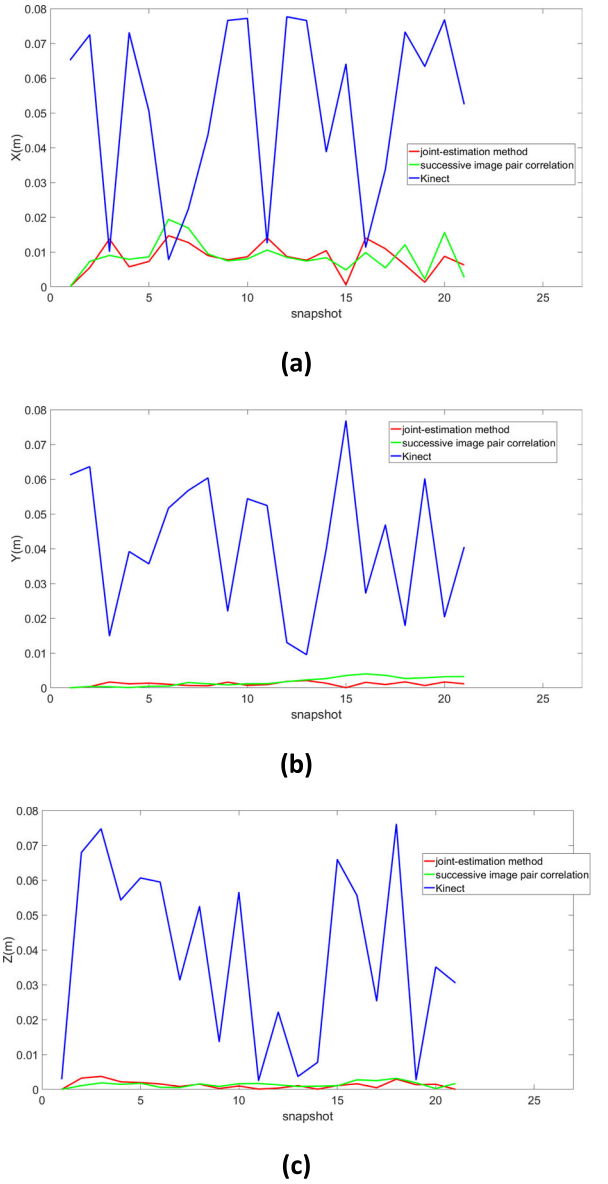

FIGURE 12. With 8cm perturbation error, the absolute estimation error comparison in (a) x direction, (b) y direction and (c) z direction.

and $g_k\,(x, y, z)$ can be written as

$$g_k\,(x, y, z) = A_k\,(x, y, z)\, e^{j\xi_k(x,y,z)} \tag{31}$$

where, $A_k\,(x, y, z)$ is the amplitude of the kth image. $\xi_k(x, y, z)$ is the residual phase error caused by the target motion. In the near-field imaging MIMO system, $\xi_k(x, y, z)$ has the spatial variability. x, y and z are assigned from a small range near the peak point of the scatterer. At this point, $A_k\,(x, y, z)$ is considered not vary with k and $\xi_k(x, y, z)$ is independent of the values x, y and z. $\gamma_k\,(x, y, z)$ is introduced and written as

$$\gamma_k\,(x, y, z) = g_k\,(x, y, z)\, g^*_{k-1}\,(x, y, z) \tag{32}$$

Assuming that the image noise obeys the Gaussian distribution, the phase error is estimated based on the maximum likelihood estimation. Then

$$\Delta\hat{\xi}_k = \arg\left[\sum_x \sum_y \sum_z \gamma_k\,(x, y, z)\right] \tag{33}$$

where $\Delta\hat{\xi}_k$ is the estimated value of $\Delta\xi_k$. ξ_k can be obtained by accumulating the value of $\Delta\xi_k$. The estimated phase error is compensated to the 3D image and the refocused image is obtained through the coherent synthesis.

Since the major direction of the target motion is along x axis, the final image gets an improved resolution in the x direction. Fig. 9(a) shows the PSF of the virtual array synthesized by the target motion using a single aperture. Fig. 9 (b) shows the PSF of the virtual array synthesized by the target motion using MIMO aperture array. Compared

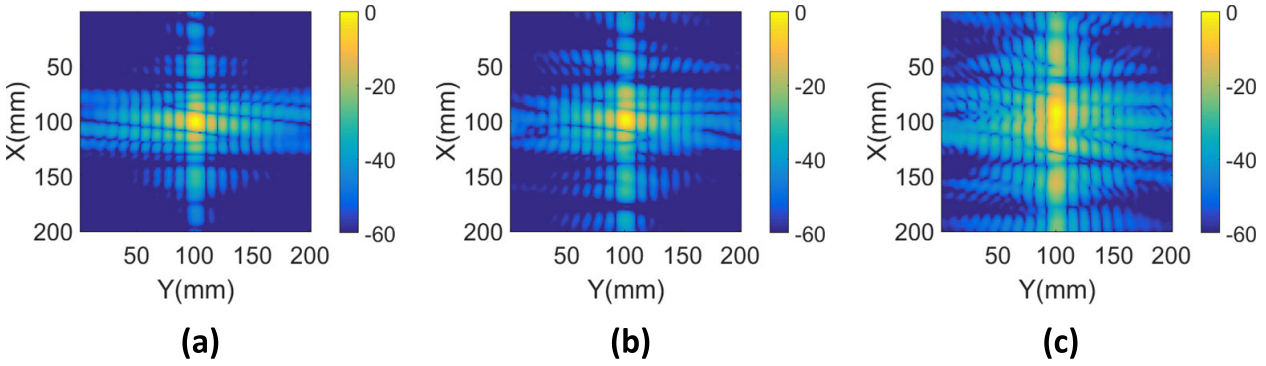

FIGURE 13. The comparison of the point target image in xy plane and without phase adjustment with perturbation error of (a) 2cm, (b) 4cm and (c) 8cm.

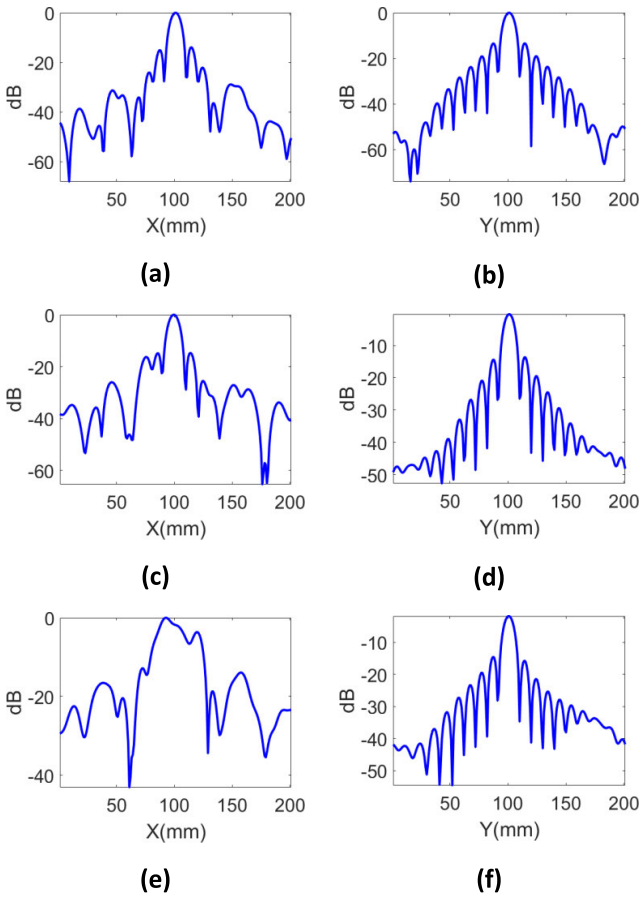

FIGURE 14. 1D profile of the point target in (a) x direction with 2cm perturbation error, (b) y direction with 2cm perturbation error, (c) x direction with 4cm perturbation error, (d) y direction with 4cm perturbation error, (e) x direction with 8cm perturbation error and (f) y direction with 8cm perturbation error.

with the PSF of the real aperture MIMO system in Fig. 5, the synthetic array using MIMO and target motion has higher resolution in the target motion direction.

D. COMPUTATION COMPLEXITY

In this security imaging system, the moving target is decomposed into 25 components and the components are imaged separately with guiding of Kinect. The computational of BP algorithm is [54]

$$C_{\mathrm{BP}} \approx 8N_T N_R N_x N_y N_z \mathrm{FLOP} \tag{34}$$

where N_T and N_R are the numbers of TX antennas and RX antennas. N_x, N_y and N_z are the pixel numbers of final image in x, y and z directions respectively.

Then the computation complexity of 25 components is

$$C_1 \approx 8N_T N_R \sum_{i=1}^{25} N_{x_i} N_{y_i} N_{z_i} \mathrm{FLOP} \tag{35}$$

where N_{x_i}, N_{y_i} and N_{z_i} are the pixel numbers of ith component image in x, y and z directions respectively. Note that $\sum_{i=1}^{25} N_{x_i} N_{y_i} N_{z_i} < N_x N_y N_z$ and the values of N_{x_i}, N_{y_i} and N_{z_i} depend on the sizes of these components.

When the motion parameters are estimated using joint-estimation method, there are 3D FFTs, IFFTs, matrix inversions and matrix multiplication which are included in (19) and (29). Therefore, the computational load for this step is

$$C_2 \approx 25 \times \left[2K \log_2 \left(N_x N_y N_z\right) + 9K^6\right] \mathrm{FLOP} \tag{36}$$

The computational load of refocusing process depends on a small range near the peak point and can be considered as a constant to be neglected. Then the image fusion can be seen as 25 imaging processes and the computational complexity is

$$C_3 \approx 8KN_T N_R \sum_{i=1}^{25} N_{x_i} N_{y_i} N_{z_i} \mathrm{FLOP} \tag{37}$$

In summary, the total computational complexity of the proposed method is

$$\begin{aligned} C_{\mathrm{total}} &\approx KC_1 + C_2 + C_3 \\ &= 16KN_T N_R \sum_{i=1}^{25} N_{x_i} N_{y_i} N_{z_i} + 50K \log_2 \left(N_x N_y N_z\right) \\ &\quad + 225K^6 \mathrm{FLOP} \end{aligned} \tag{38}$$

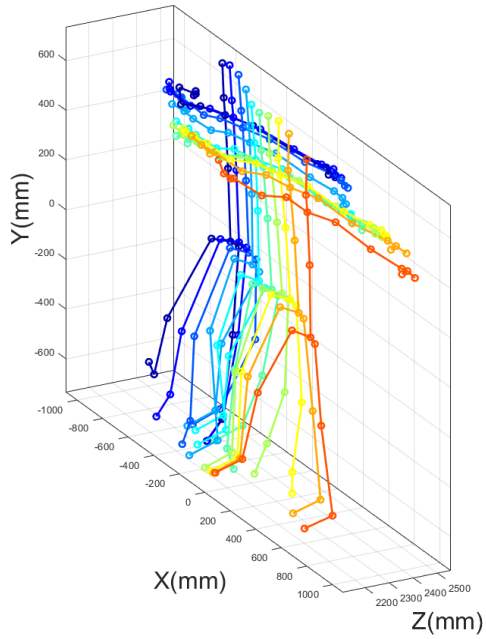

FIGURE 15. Human skeleton in motion.

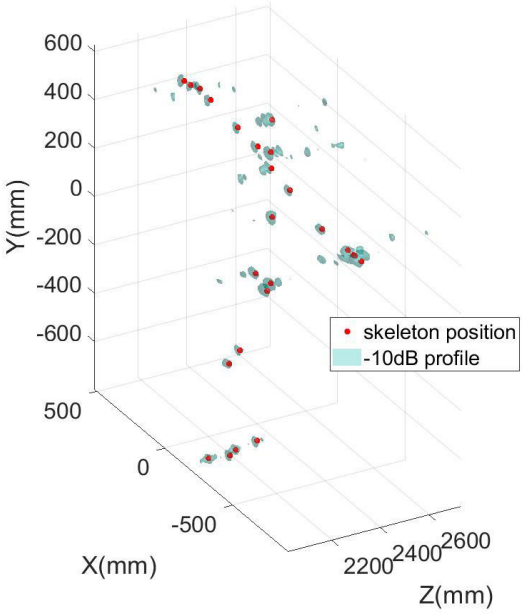

FIGURE 16. The simulation result of the human skeleton joints.

V. EXPERIMENTS

A. POINT TARGET

In this section, the performance of target imaging using sparse MIMO array is verified by point target simulation. System configuration is the same as mentioned in Section 3. The point target moves in sinusoidal trajectory in the $x-y$ plane and $z=2$m. Stochastic perturbations up to 2cm, 4cm and 8cm are added to the trajectory respectively to make it realistic.

The baseline sinusoidal trajectory with perturbation in centimeter-level is assumed to be obtained via Kinect. Motion estimation is only tasked to estimate the perturbations. The joint estimation method proposed in this paper is used to estimate the trajectory of the target which requires a preliminary motion information to image the local component.

In order to compare the performance of successive image pair correlation and joint estimation method, the absolute value of motion errors is analyzed in Fig. 10, Fig. 11 and Fig. 12 with perturbation error of 2cm, 4cm and 8cm respectively. The conventional successive image pair correlation method can obtain the target trajectory but the cumulative error is increasing with the movement of the target. The proposed joint-estimation method effectively reduce the accumulated error and shows a better performance. The major motion of the target is in the x direction where the cumulative error increases along x direction. The sinusoidal motion of the target in the z direction leads sinusoidal oscillation error which becomes larger with the target motion. The target motion in the y direction causes the error to vary randomly. Both of the conventional successive image pair correlation method and the joint-estimation method can obtain more accurate motion parameters than the simulated Kinect error. The performance of joint estimation method is the best in all three dimensions when the stochastic perturbation error is 2cm, 4cm and 8cm respectively. From Fig. 10, Fig. 11 and Fig. 12, we can see the estimation error of joint estimation method is increasing when the perturbation error added is growing, and achieves 2cm in x direction when the perturbation error is 8cm as shown in Fig. 12(a).

All signals acquired during the target motion are compensated simultaneously. Finally, the refocused images of the point target in $x-y$ plane at $z=2$m with different perturbation errors are shown in Fig. 13. According to the assumption that the phase error is spatially invariant in small area, the phase adjustment is taken. The 1D image profiles in x and y dimensions are shown in Fig. 14 with different perturbation errors. The profiles in z direction depend on the signal bandwidth and are not analyzed here. Since the joint-estimation error achieves 2cm in x direction when the perturbation error added is 8cm, the 1D profile of the point target is spread. Therefore, we can conclude that the joint-estimation method is valid when the perturbation error is less than 4cm and becomes not valid when the perturbation error is larger than 8cm.

B. HUMAN SKELETON

In this subsection, the motion of non-rigid body target is simulated using multiple points. Real data of human skeleton joints measured by Kinect is used to simulate the human motion more realistically. Kinect decomposes the human target into 25 joints, as shown in Fig. 2. In this experiment, 10 frames are interpolated from the original 30 frames, synchronized with the microwave snapshotting time, to extract the skeleton joints of the human body. Meanwhile, 2cm stochastic error is added to the joint positions which is taken as an input parameter for the imaging of non-rigid body target. The human skeleton motion parameters are shown

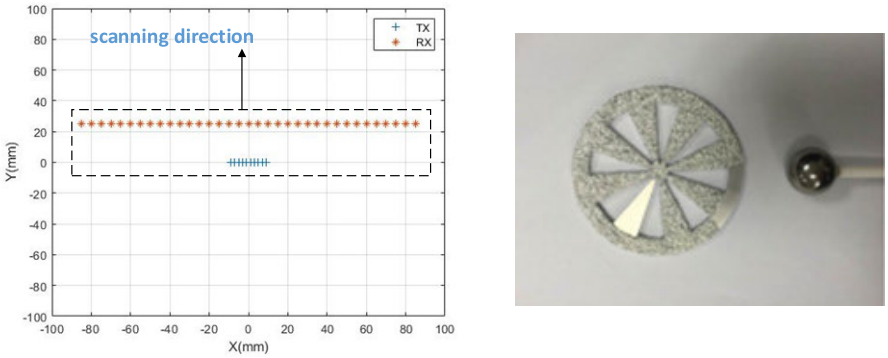

FIGURE 17. (a) The linear TX and RX arrays of measuring system and (b) the measured objects.

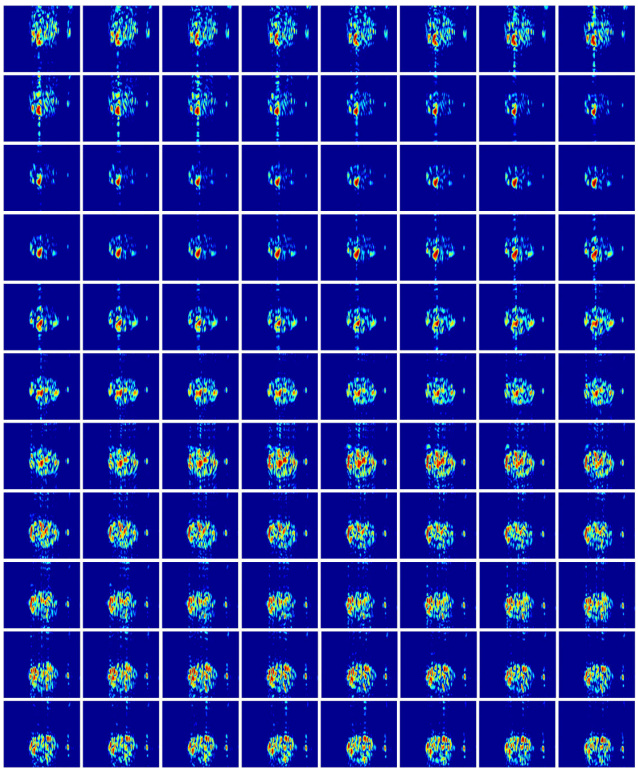

FIGURE 18. The multiple snapshots of the objects imaged with multiple scans.

in Fig. 15. The major movement of the human target is about 1m in the x direction.

The motion parameters of each skeleton scatterer are estimated using the joint estimation method and used to compensate the motion of each scatterer. Then, the images obtained are refocused separately and merged into one image to reconstruct the whole image of the human skeleton which is shown in Fig. 16. The red points present the real position of the human skeleton and the blue color presents the -10dB profile of the reconstructed 3D image. As can be seen in Fig. 16, 25 scatterers are imaged and a complete view of human skeleton are achieved.

C. MOVING OBJECTS

In this subsection, we use real data as measured by an equivalent MIMO scanning system as presented in [55] to further verify the proposed method in this paper. The system adopts stepped-frequency (SF) mode and the working frequency is 130GHz $\sim$ 150GHz. The number of sampling points is 201. This is an equivalent system without real 2D MIMO antenna array. As shown in Fig. 17(a), 10 virtual TX antennas and 35 virtual RX antennas are formed by changing the positions of one real TX antenna and one real RX antenna in the system. Then, the 2D virtual MIMO array is synthesized by repeating this formation along the scanning direction in Fig. 17(a). With the virtual MIMO array, the TX antennas and RX antennas are paralleled linear arrays with 2.5cm distance. The spaces of TX antennas and RX antennas are 2mm and 5mm respectively. The imaging area center is at 30cm from the system with two objects as shown in Fig. 17(b). The linear TX and RX arrays are controlled by a stepping motor to realize the scanning in y direction with accuracy of 1mm.

The scanning data of the objects are obtained and can be divided into multiple observations. In detail, we use 20 scans to image the objects as one snapshot to form multiple snapshots with large number of scans. As shown in Fig. 18, 107 scans are used to form 88 snapshots. These scans are performed sequentially with space of 1mm. For example, the first snapshot is formed by scans 1 to 20 and the next one is formed by scans 2 to 21. The snapshots are imaged using BP algorithm.

In Fig. 18, we can see that the snapshots present different features of the objects. These snapshots are randomly ordered to form a sequence which can be seen the moving sequence of objects in y direction. Then the successive image pair correlation and the joint-estimation method are used to derive the motion parameters of the objects. In order to avoid the decorrelation caused by the different features of the snapshots, the joint-estimation method divides the image sequence into three segments to estimation the motion parameters. The estimation error of three segments

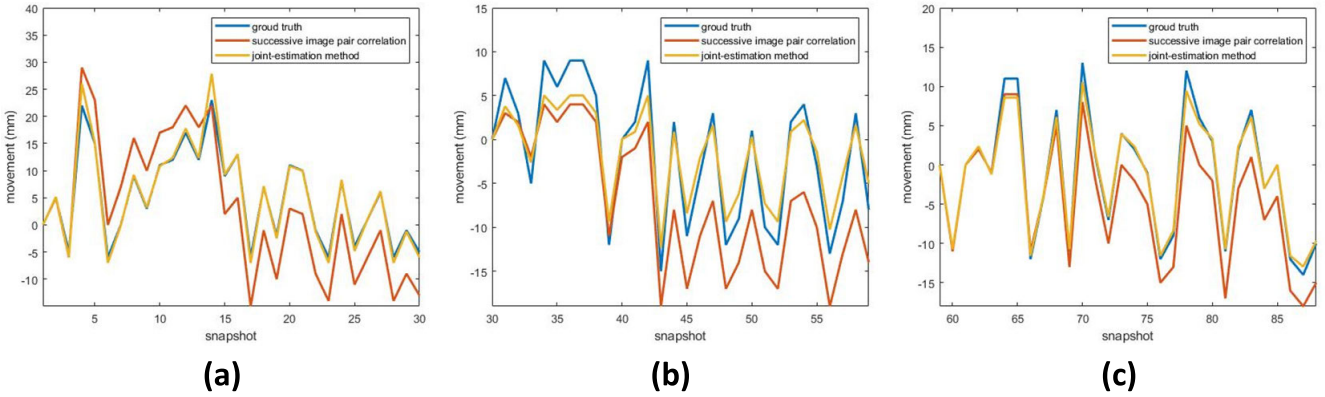

FIGURE 19. The estimation error of (a) the first, (b) the second and (c) the third segment in object moving direction.

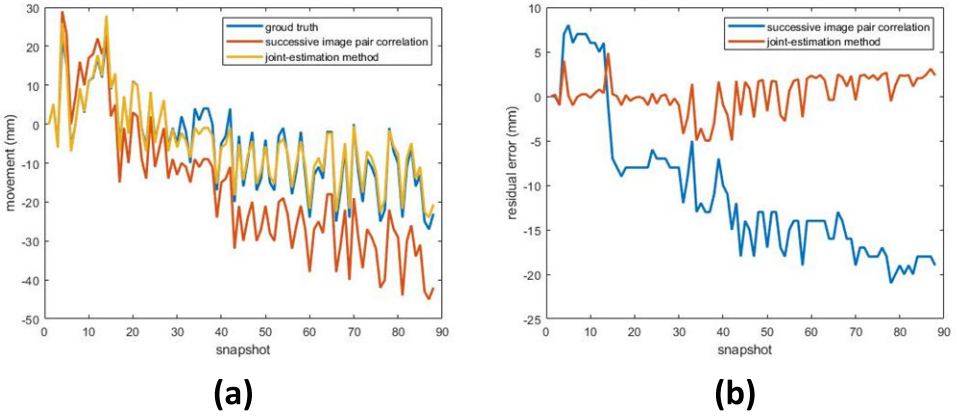

FIGURE 20. (a) The estimation error and (b) the absolute estimation error comparison in object moving direction.

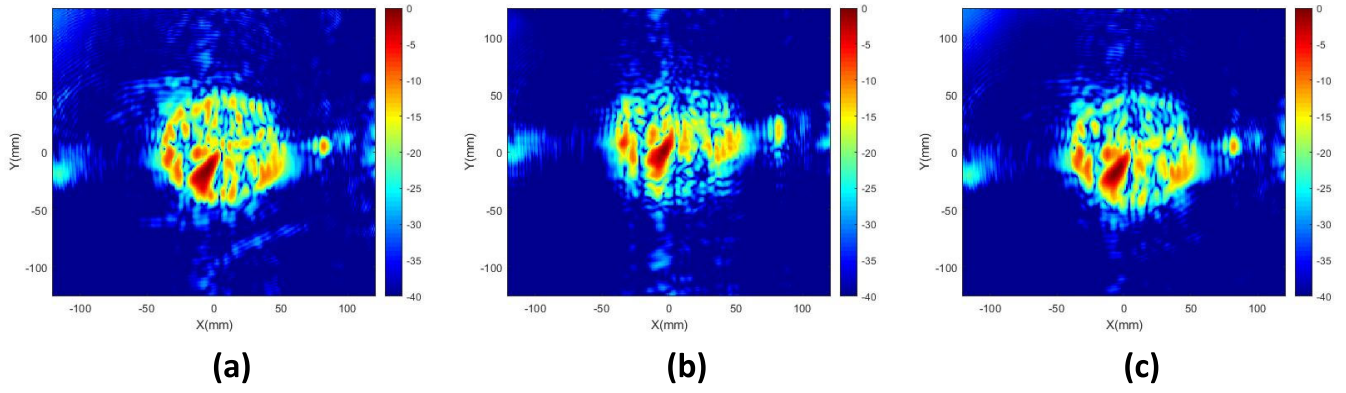

FIGURE 21. Imaging results of objects (a) without motion error and with motion parameters estimated by (b) successive image pair correlation and (c) joint-estimated method.

are shown in Fig. 19(a), Fig. 19(b) and Fig. 19(c) respectively. Then the joint-estimation is carried out and the estimation results are shown in Fig. 20.

Using the estimated motion parameters, the final images of the moving objects can be obtained using BP algorithm, which are shown in Fig. 21. Fig. 21(a) shows the image without motion error. Fig. 21(b) and Fig. 21(c) show the images after movement compensation with the estimated motion parameters of the successive image pair correlation and the joint-estimation method respectively.

From Fig. 21, we can see that the proposed method can effectively using the moving of the target to form larger equivalent array and archive high resolution imaging of the moving objects.

VI. CONCLUSION

This paper proposes a microwave/mmw imaging method for the motion target of non-rigid body based on 2D sparse MIMO array. In the 2D sparse MIMO array, a small number of transmitter antennas is used to improve the measurement

time of the system. The space-time trajectory model is introduced to describe the motion target which decompose the non-rigid body target as a series of point joints. The component of each joint is imaged separately and merged into one final image to reconstruct the 3D image of the whole target. The optical device Kinect is used in this method to measure the positions of human skeleton joints and the measurement are taken as a preliminary estimation of the target motion parameters, which is used to indicate the local image positions of the joint components. With the local image series of each joint component, an accurate motion parameters estimation is taken using the segmental joint estimation method proposed in this paper. Through movement compensation and refocus imaging, the high resolution 3D image of the moving target of non-rigid body is obtained. The system design, the imaging flow and the design principle are illustrated in this paper. The validity of this method is proved through the simulations of human skeleton and experiment on measuring data of real targets.

ACKNOWLEDGMENT

The authors would like to thank B. Cheng, L. Zhao, and Y. Zhao, of Microsystem and Terahertz Center for providing the measurement data used in this study.

REFERENCES

[1] M. C. Kemp, "Millimetre wave and terahertz technology for detection of concealed threats—A review," in *Proc. 32nd Int. Conf. Infr. Millim. Waves 15th Int. Conf. THz Electron.*, Sep. 2007, pp. 647–648.

[2] G. Donges, T. Herwig, C. Koch, and G. Geus, "X-Ray scanner," U.S. Patent 4 736 401, Apr. 5, 1988.

[3] D. M. Sheen, D. L. Mcmakin, and T. E. Hall, "Near field imaging at microwave and millimeter wave frequencies," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Jun. 2007, pp. 1693–1696.

[4] K. J. Roe and C. W. Gregory, "Wave-based sensing and imaging for security applications," in *Proc. Eur. Conf. Antennas Propag.*, Apr. 2015, pp. 1–5.

[5] A. Schiessl, A. Genghammer, S. S. Ahmed, and L. P. Schmidt, "Hardware realization of a 2 m × 1 m fully electronic real-time mm-wave imaging system," in *Proc. Eur. Conf. Synth. Aperture Radar*, 2012, pp. 40–43.

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

[7] E. L. Jacobs, "Reflect-array based mm-wave people screening system," *Proc. SPIE*, vol. 8900, Oct. 2013, Art. no. 890002.

[8] T. Derham, H. Kamoda, T. Iwasaki, and T. Kuki, "Active MMW imaging system using the frequency-encoding technique," in *Proc. Microw. Conf.*, Nov. 2007, pp. 181–184.

[9] S. Bertl and J. Detlefsen, "Effects of a reflecting background on the results of active MMW SAR imaging of concealed objects," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3745–3752, Oct. 2011.

[10] S. A. Lang, M. Hägelen, J. Ender, S. Hantscher, and H. Essen, "A new approach for fast security scanning with millimetre-waves: SARGATE," *Proc. SPIE*, vol. 8022, May 2011, Art. no. 802208.

[11] H. Deng, "Polyphase code design for Orthogonal Netted Radar systems," *IEEE Trans. Signal Process.*, vol. 52, no. 11, pp. 3126–3135, Nov. 2004.

[12] J. Li, P. Stoica, and X. Zheng, "Signal synthesis and receiver design for MIMO radar imaging," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3959–3968, Aug. 2008.

[13] H. He, P. Stoica, and J. Li, "Designing unimodular sequence sets with good correlations—Including an application to MIMO radar," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4391–4405, Nov. 2009.

[14] J. Fortuny, "An efficient 3-D near-field ISAR algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 34, no. 4, pp. 1261–1270, Oct. 1998.

[15] J. Fortuny and A. J. Sieber, "Three-dimensional synthetic aperture radar imaging of a fir tree: First results," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 2, pp. 1006–1014, Mar. 1999.

[16] K. Suwa, K. Yamamoto, M. Iwamoto, and T. Kirimoto, "Reconstruction of 3-D target geometry using radar movie," in *Proc. Eur. Conf. Synth. Aperture Radar*, Jun. 2008, pp. 1–4.

[17] L. Jaime, M. T. Ghasr, M. López-Portugués, F. Las-Heras, and R. Zoughi, "Real-time multiview SAR imaging using a portable microwave camera with arbitrary movement," *IEEE Trans. Antennas Propag.*, vol. 66, no. 12, pp. 7305–7314, Dec. 2018.

[18] V. C. Chen and H. Ling, *Time-Frequency Transforms for Radar Imaging and Signal Analysis*. Boston, MA, USA: Artech House, 2001.

[19] C.-C. Chen and H. C. Andrews, "Target-motion-induced radar imaging," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-16, no. 1, pp. 2–14, Jan. 1980.

[20] G. Y. Delisle and H. Wu, "Moving target imaging and trajectory computation using ISAR," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 30, no. 3, pp. 887–899, Jul. 1994.

[21] J. Wang and D. Kasilingam, "Global range alignment for ISAR," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 1, pp. 351–357, Jan. 2003.

[22] P. H. Eichel, D. C. Ghiglia, and C. V. Jakowatz, Jr., "Speckle processing method for synthetic-aperture-radar phase correction," *Opt. Lett.*, vol. 14, no. 1, pp. 1–3, 1989.

[23] D. E. Wahl, F. H. Eichel, D. C. Ghiglia, and C. V. Jakowatz, "Phase gradient autofocus-a robust tool for high resolution SAR phase correction," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 30, no. 3, pp. 827–835, Jul. 1994.

[24] S. Shao, L. Zhang, H. Liu, and Y. Zhou, "Spatial-variant contrast maximization autofocus algorithm for ISAR imaging of maneuvering targets," *Sci. China Technol. Sci.* vol. 53, Apr. 2010, Art. no. 40303.

[25] T. Cooke, "Ship 3D model estimation from an ISAR image sequence," in *Proc. IEEE Radar Conf.*, Sep. 2003, pp. 36–41.

[26] M. Stuff, M. Biancalana, G. Arnold, and J. Garbarino, "Imaging moving objects in 3D from single aperture synthetic aperture radar," in *Proc. IEEE Radar Conf.*, Apr. 2004, pp. 94–98.

[27] C. Ma, T. S. Yeo, C. S. Tan, and H. S. Tan, "Sparse array 3-D ISAR imaging based on maximum likelihood estimation and clean technique," *IEEE Trans. Image Process.*, vol. 19, no. 8, pp. 2127–2142, Aug. 2010.

[28] C. Ma, T. S. Yeo, H. S. Tan, J. Wang, and B. Chen, "Three-dimensional ISAR imaging using a two-dimensional sparse antenna array," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 378–382, Jul. 2008.

[29] G. Wang, X.-G. Xia, and V. C. Chen, "Three-dimensional ISAR imaging of maneuvering targets using three receivers," *IEEE Trans. Image Process.*, vol. 10, no. 3, pp. 436–447, Mar. 2001.

[30] M. Martorella, D. Stagliano, F. Salvetti, and N. Battisti, "3D interferometric ISAR imaging of noncooperative targets," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 4, pp. 3102–3114, Oct. 2014.

[31] N. Battisti and M. Martorella, "Intereferometric phase and target motion estimation for accurate 3D reflectivity reconstruction in ISAR systems," in *Proc. IEEE Radar Conf.*, May 2010, pp. 108–112.

[32] C. Ma, T. Yeo, and C. Tan, "Three-dimensional imaging of targets using colocated MIMO radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 8, pp. 3009–3021, Aug. 2011.

[33] C. Ma, T. S. Yeo, C. S. Tan, J.-Y. Li, and Y. Shang, "Three-dimensional imaging using colocated MIMO radar and ISAR technique," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 8, pp. 3189–3201, Aug. 2012.

[34] Y. Zhu, Y. Su, and W. Yu, "An ISAR imaging method based on MIMO technique," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 8, pp. 3290–3299, Aug. 2010.

[35] D. Pastina, M. Bucciarelli, and P. Lombardo, "Multistatic and MIMO distributed ISAR for enhanced cross-range resolution of rotating targets," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 8, pp. 3300–3317, Aug. 2010.

[36] D.-W. Wang, X.-Y. Ma, A.-L. Chen, and Y. Su, "High-resolution imaging using a wideband MIMO radar system with two distributed arrays," *IEEE Trans. Image Process.*, vol. 19, no. 5, pp. 1280–1289, May 2010.

[37] C. Ma, T. S. Yeo, Y. Zhao, and J. Feng, "MIMO radar 3D imaging based on combined amplitude and total variation cost function with sequential order one negative exponential form," *IEEE Trans. Image Process.*, vol. 23, no. 5, pp. 2168–2183, May 2014.

[38] X. Hu, N. Tong, Y. Zhang, and D. Huang, "MIMO radar imaging with nonorthogonal waveforms based on joint-block sparse recovery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 5985–5996, Oct. 2018.

[39] Z. Zhu and F. Xu, "Demonstration of 3D ISAR security imaging at 24GHz with a sparse MIMO array," Aug. 2019, *arXiv:1908.06619*. [Online]. Available: https://arxiv.org/abs/1908.06619
[40] Z. Rongqiang, J. Zhou, B. Cheng, Q. Fu, and G. Jiang, "Interpolation-free method for near-field cross MIMO array imaging," *Electron. Lett.* vol. 54, no. 14, pp. 890–892, Jul. 2018.
[41] Z. Jianxiong, R. Zhu, G. Jiang, L. Zhao, and B. Cheng, "A precise wavenumber domain algorithm for near range microwave imaging by cross MIMO array," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 4, pp. 1316–1326, Apr. 2019.
[42] D. R. Wehner, *High Resolution Radar*. Norwood MA, USA: Artech House, 1987.
[43] J. Shotton, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake, M. Cook, and R. Moore, "Real-time human pose recognition in parts from single depth images," *Commun. ACM*, vol. 56, no. 1, pp. 116–124, Jan. 2013.
[44] P. Fankhauser, M. Bloesch, D. Rodriguez, R. Kaestner, M. Hutter, and R. Siegwart, "Kinect v2 for mobile robot navigation: Evaluation and modeling," in *Proc. Int. Conf. Adv. Robot. (ICAR)*, Jul. 388–394.
[45] R. T. Hoctor and S. A. Kassam, "The unifying role of the coarray in aperture synthesis for coherent and incoherent imaging," *Proc. IEEE*, vol. 78, no. 4, pp. 735–752, Apr. 1990.
[46] M. Vondrak, L. Sigal, and O. C. Jenkins, "Dynamical simulation priors for human motion tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 52–65, Jan. 2013.
[47] J. K. Aggarwal and M. S. Ryoo, "Human activity analysis: A review," *ACM Comput. Surv.*, vol. 43, no. 3, Apr. 2011, Art. no. 16.
[48] V. C. Chen, *The Micro-Doppler Effect in Radar*. Norwood, MA, USA: Artech House, 2011.
[49] H. Fan, L. Ren, T. Long, and E. Mao, "A high-precision phase-derived range and velocity measurement method based on synthetic wideband pulse Doppler radar," *Sci. China Inf. Sci.*, vol. 60, no. 8, Aug. 2017, Art. no. 082301.
[50] Z. Sun, J. Wang, Y. Zhang, J. Sun, C. Yuan, and Y. Bi "A multicomponent micro-Doppler signal decomposition and parameter estimation method for target recognition," *Sci. China Inf. Sci.*, vol. 58, no. 12, pp. 1–13, Dec. 2015.
[51] Z. Sun, J. Wang, Y. Zhang, J. Sun, C. Yuan, and Y. Bi, "Multiple walking human recognition based on radar micro-Doppler signatures," *Sci. China Inf. Sci.* vol. 58, no. 12, pp. 1–13, Dec. 2015.
[52] Z. Zhang, "Microsoft Kinect sensor and its effect," *IEEE MultiMedia*, vol. 19, no. 2, pp. 4–10, Feb. 2012.
[53] A. Bellettini and M. A. Pinto, "Theoretical accuracy of synthetic aperture sonar micronavigation using a displaced phase-center antenna," *IEEE J. Ocean. Eng.*, vol. 27, no. 4, pp. 780–789, Oct. 2002.
[54] R. Zhu, J. Zhou, G. Jiang, and Q. Fu, "Range migration algorithm for near-field MIMO-SAR imaging," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2280–2284, Dec. 2017.
[55] B. Cheng, G. Jiang, C. Wang, C. Yang, Y. Cai, Q. Chen, X. Huang, G. Zeng, J Jiang, X. Deng, and J. Zhang, "Real-time imaging with a 140 GHz inverse synthetic aperture radar," *IEEE Trans. THz Sci. Technol.*, vol. 3, no. 5, pp. 594–605, Sep. 2013.

ZHANYU ZHU received the B.E. degree in information and computing science and the M.E. degree in computer science and technology from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2008 and 2011, respectively. He is currently pursuing the Ph.D. degree with the School of Information Science and Technology, Fudan University, Shanghai, China. From 2011 to 2016, he was an Engineer with the Aviation Industry Corporation of China, for radar signal processing. His current research interests include SAR simulation and radar signal processing.

LEI KUANG received the B.S. and M.S. degrees from Anhui University, Hefei, China, in 2001 and 2004, respectively, and the Ph.D. degree from Fudan University, Shanghai, China, in 2008. He was a Visiting Scholar with Department of Electrical and Computer Engineering, Duke University, NC, USA, from 2016 to 2017, supported by China Scholarship Council. He is currently an Associate Professor with the School of Information Science and Technology, East China Normal University, Shanghai. His current research interests include computational electromagnetics, electromagnetic scattering, and microwave imaging. Dr. Kuang received the "Excellent Instructor Award for Practical Innovation from East China Normal University," in 2013.

FENG XU (S'06–M'08–SM'14) received the B.E. degree (Hons.) in information engineering from Southeast University, Nanjing, China, in 2003, and the Ph.D. degree (Hons.) in electronic engineering from Fudan University, Shanghai, China, in 2008. From 2008 to 2010, he was a Postdoctoral Fellow with the NOAA Center for Satellite Application and Research (STAR), Camp Springs, MD, USA. From 2010 to 2013, he was with Intelligent Automation Inc., Rockville, MD, while partly working for NASA Goddard Space Flight Center, Greenbelt, MD as a Research Scientist. Since 2013, he has been with Fudan University as a Professor with the School of Information Science and Technology and the Vice Director of the MoE Key Laboratory for Information Science of Electromagnetic Waves.