Spatial Calibration of Millimetre-Wave Radar for Close-Range Object Location

José A. Paredes, Miles Hansard, Khalid Z. Rajab and Fernando J. Álvarez

***Abstract*— Accurate object detection and location systems are essential for many robotic applications, including autonomous grasping and manipulation systems. In some cases, the target object may be obscured from view, in clutter, packaging, or debris. Millimetre wave radar is a potential alternative to visual sensing in such scenarios, owing to its ability to penetrate typical low-density non-metallic materials. However, this approach requires accurate spatial calibration of the radar signal, over the robot workspace. We propose to achieve this with reference to visual data, which provides ground truth locations for initial training of the system. Specifically, we describe a commodity millimetre wave radar system for detecting and localizing static metallic objects, over a 2D workspace. We compare similarity, affine, and thin-plate spline models of the spatial transformation from radar estimates to actual locations. Experiments were performed with a frequency modulated continuous wave (FMCW) multiple-input multiple-output (MIMO) device, using a starting frequency of 60 GHz and a bandwidth of 3.4 GHz. It is shown that the spline model performs best, achieving an average spatial error of 7 mm, which is an order of magnitude lower than that of the uncalibrated system.**

***Index Terms*— mmWave radar, RGB camera, mapping methods, spatial calibration**

I. Introduction

THE use of robot arms is well established for industrial tasks, such as assembly and welding, in controlled environments. More recently, it has become desirable to use robotic systems in less predictable environments, such as warehouses and farms. In all cases, a robust and accurate object detection and localization system is required. Visual sensors are often used in such contexts, as reviewed by Du et al. [1]. These systems may be based on fiducial markers [2], 3D scene analysis [3], or machine learning approaches [4]. For example, a calibrated RGB camera can be used to estimate the target pose [5], after an automatic object detection process [6]. Additional information can be obtained from RGBD cameras, for grasping and manipulation tasks [7].

Despite these advances in visual sensing, there remain many challenges in autonomous detection and manipulation, including specular or translucent object detection, tasks requiring high precision, or the improvement of grasping in clutter, as discussed in [8] and [9]. Furthermore, the detection and manipulation of concealed objects is essentially impossible using ordinary visual sensors. These difficulties motivate the use of radar sensors in robotic applications. In particular, millimetre wave (mmWave) radar, which operates in frequencies of several GHz, is able to penetrate a range of low-density materials, such as those used for packaging. Furthermore, typical mmWave radar devices are small, lightweight, and readily available.

Please, cite this as: *J. A. Paredes, M. Hansard, K. Z. Rajab and F. J. Álvarez, "Spatial Calibration of Millimetre-Wave Radar for Close-Range Object Location," in IEEE Sensors Journal (2024), 10.1109/JSEN.2024.3393030*

It is possible to calibrate a given mmWave radar, in a reference environment, so that the location of a target object can be estimated from the returned signal. In practice, however, this is not a complete solution. In particular, the reference calibration cannot account for the varying effects of unknown environmental clutter, device pose, and manufacturing tolerances. In this paper, rather than trying to analyze such effects, we simply model the overall spatial transformation between the radar signal maxima and the true physical locations. This is done using a rapid calibration procedure, which is applied to the target device, in the target environment. Our calibration method minimizes the differences between measurements obtained from the mmWave radar and a well-established camera-based method, in an offline training procedure.

A. Related Work

The use of optical systems in scene analysis for robotics is well established [1], [10], [11]. The corresponding vision systems may be based on fiducial markers [2], geometric principles [3], or machine learning approaches [4]. Radar-based systems have appeared more recently in robotics [12], and so a more detailed review will be given below, focusing on the mmWave case.

Wang et al. [13] performed a theoretical analysis of mmWave position estimation. They demonstrated that attaining millimetre-level accuracy is achievable in principle, given a suitable antenna array, substantial bandwidth, and high signal-to-noise ratio. The achievable precision has also been investigated experimentally, by Ahmad et al. [14]. They show that a 79 GHz device can localize a corner reflector with sub-millimetre precision, at ranges less than 5 m, in an anechoic chamber. In the context of autonomous driving [15], target identification (as well as localization) has been demonstrated

with mmWave radar [16], including the ability to distinguish between pedestrians, cyclists, and vehicles [17].

More generally, the *combination* of optical and radar data is an attractive approach in the automotive field [18], where sensor systems must be able to operate in adverse weather conditions [19], [20]. The mmWave subsystem can be used for object identification [21]–[23], as well as localization [24], [25]. For example, Peršić et al. present a method for spatial calibration of radar and LiDAR devices [26]. This work is based on specially designed targets, which can be detected and localized by all sensors. The calibration process involves two steps. Firstly, the reprojection error is minimized in azimuth and range, in the absence of radar elevation data. Secondly, the radar cross section is analyzed, and correlated with the LiDAR elevation estimates. This system was subsequently extended [27], to avoid the assumption of a specific static target, and to incorporate a Gaussian process model of the multi-sensor trajectories [28]. Other work has focused on spatial and temporal calibration between radar and LiDAR sensors [29]. Spatial alignment is achieved by compensating possible deviations in the elevation axis with 3D radar cross-section distribution measurements, whereas the temporal alignment is achieved by estimating the time-delay between radar and LiDAR measurements as the time difference obtained after aligning the target azimuth angle for both sensors.

In contrast, Oh et al. [30] address the more restricted problem of mapping estimated locations in the radar ground-plane into a camera image-plane. They show good performance, over scene depths of up to 50 m, using retro-reflector radar targets. While this approach enables fusion of the image and radar data, it does not address the issue of radar calibration, because there are no reference 3D estimates. Cheng et al. [31] demonstrate the detection of plastic water bottles, floating on water, using mmWave radar. Their fusion strategy addresses the limitations of each sensor, such as highly variable reflectivity for cameras and clutter for radar. The algorithm begins by transforming the radar 3D point cloud into 2D, which is then fused with the RGB camera data by projecting each radar point onto the image plane. Another solution to this problem can be found in [32], where the rotation between sensors is calibrated using a convolutional neural network (CNN). Wise et al. [33] also describe a continuous-time 3D radar-to-camera extrinsic calibration algorithm, for robotics, which does not require the use of retro-reflectors. This approach uses velocity (rather than position) estimates from the radar signal, in order to estimate the sensor pose, with respect to an attached camera system. This work has been extended, to avoid the dependence on special retroreflective radar targets [34], [35].

In recent years, there has been an increasing interest in the use of mmWave radar in the context of robotics applications [36]. For example, Stetco et al. [37] introduce a simulation approach for frequency modulated continuous wave (FMCW) radar sensors operating in cluttered environments. Although this simulator accounts for only a simplified approximation of reflections, the predictions are in good agreement with the experimental results. The authors continued their work in [38], where they present an advanced simulation environment that provides real-time raw data, incorporates multi-antenna configurations and a wave penetration model for non-conductive objects, accounts for various beam patterns, and incorporates realistic radar configurations. This proposed simulation framework was validated in real-world scenarios, including those with a single object in different static positions, radar occlusion caused by diverse materials, and dynamic human activity.

B. Contributions and organization

Our principal contribution is a practical method for mapping radar-based object location estimates into the true physical workspace. As described in the introduction, this method accounts for the combined effects of unknown environmental clutter, device pose, and manufacturing tolerances. Note that this is *not* the same as simply mapping the radar estimates into the 2D image plane [30], because we use *both* sensors to estimate the scene structure (if required, we can easily map the radar estimates into the image, at the end our procedure).

We use a commodity mmWave radar system, which operates in the unlicensed 60 GHz band, over a limited range (3.4 GHz). Although compact and inexpensive, such devices have limited spatial resolution. In principle, this can be improved by standard methods, such as zero-padded fast Fourier transform (FFT) interpolation. In practice, however, the additional computational cost (in both CPU-cycles and memory) limits on-chip implementation. Our method implicitly interpolates the location estimates, in the spatial domain, thereby shifting the computational burden to the initial (off-chip) calibration procedure.

In order to estimate the spatial mapping, we use visual measurements as training data for the system. We evaluate three possible 2D geometric mappings, with increasing generality: similarity (rotation, scale, and translation), affine (linear transformation and translation), and thin-plate spline. Through a comparative analysis of these models, we determine that the thin-plate spline yields the best results without over-fitting the training data. Overall, this method not only achieves extrinsic calibration by aligning the radar measurements with the visual measurements, but also accounts for intrinsic calibration, effectively addressing any inherent biases introduced by the radar device itself. To do that, a single-target system has been developed, analogous to the setup of typical robotic grasping benchmarks [39].

The paper is organized as follows. Section II gives a self-contained summary of the necessary radar signal processing. Sections III and IV describe the reduction of the 3D object workspace to the 2D workbench surface, and the proposed spatial transformations, respectively. Section V describes our experimental evaluation, including the accuracy of the system and the determination of the parameters. Finally, our conclusions and suggestions for future research are stated in section VI.

II. Radar fundamentals

This section reviews the recovery of geometric information from the radar signal, based on reflections from the surrounding objects. Specifically, a FMCW-based multiple-input

multiple-output (MIMO) radar is used in this work, in which a continuous chirp-like mmWave signal is employed to achieve a wide bandwidth and thus calculate range and azimuth.

Briefly, two consecutive operations of spectral analysis are required to obtain spatial information from radar data, as described in [40]. If the received signals are laid out in an complex array $\mathbf{Q}$ according to the MIMO antenna layout [41], [42], then the range-azimuth intensity array $\mathbf{F}$ can be expressed as:

$$\mathbf{F}(r,\theta) = \Big|\mathrm{FFT}\big[\mathrm{FFT}\left(\mathbf{Q}\right)\big]\Big|. \tag{1}$$

To accomplish this, we conducted a two-dimensional fast Fourier transform (FFT) for the range (r) and azimuth (θ) dimensions, with zero-padding used to extend the azimuthal dimension from the 8 virtual antennas to 64 bins. From this construction, the achievable range resolution Δr in the far-field region is limited by:

$$\Delta r \geq \frac{c}{2B} \tag{2}$$

where c is the speed of light and B, the bandwidth. Numerically, for $B = 4$ GHz, the resolution is around 4 cm. And the the minimum angle separation $\Delta\theta$ for two objects to be detected in the angular FFT can be expressed as:

$$\Delta\theta \geq \frac{\lambda}{N_v\, L\, \cos\theta} \tag{3}$$

where N_v is the number of antennas and L the distance between them.

Consider now the range and azimuth profiles as those vectors passing through the intensity array peak, found in $[r_i, \theta_i]$ at each different measurement i:

$$\mathbf{f}_i(r) = \mathbf{F}(r_i, \theta) \tag{4}$$

$$\mathbf{f}_i(\theta) = \mathbf{F}(r, \theta_i). \tag{5}$$

The range and angular resolutions in these profiles may not be sufficient for fine positioning tasks, e.g. robotic grasping. We therefore apply bicubic spline interpolation to the $\mathbf{F}$ array, so that the maxima can be estimated more accurately from the interpolated version, $\widehat{\mathbf{F}}$. Fig. 1 shows an example of a range-azimuth intensity array acquired in the scene. The performance of the interpolation method can now be examined. Both peaks can be more accurately calculated by evaluating the curves extracted from the interpolated array $\widehat{\mathbf{F}}$, and its interpolated profiles $\widehat{\mathbf{f}}_i(r)$ and $\widehat{\mathbf{f}}_i(\theta)$, rather than from the raw array, as compared in Fig. 2.

After the interpolation procedure, the target coordinates in the frame i are defined as:

$$r_i = \arg\max_r \widehat{\mathbf{f}}_i(r) \tag{6}$$

$$\theta_i = \arg\max_\theta \widehat{\mathbf{f}}_i(\theta). \tag{7}$$

Having taken into account that the azimuth angle is referenced to the y-axis, the locations of the maxima can be transformed into Cartesian coordinates as

$$\mathbf{y}_i = r_i \begin{bmatrix} \sin\theta_i \\ \cos\theta_i \end{bmatrix} \tag{8}$$

in order to facilitate comparison with the visual estimates, in the next section.

III. Ground Plane Visual Coordinates

Ground-truth target locations are obtained from an ordinary camera, in conjunction with a set of *AprilTag* markers [2]. This system gives millimetre accuracy, subject to minimal uncertainty from the feature detection process, owing to the known structure of the marker patterns. The camera was fully calibrated, using standard methods [43]. This device is positioned above the workspace, both to avoid interference with the radar signals and to provide a clear view of the targets. In addition to the marker tops, five fixed markers were attached to the workspace surface, in order to transfer the 3D visual measurements to the 2D ground plane, as explained below.

The following procedure is used to project the 3D visual marker positions, which are on the tops of the objects, into the 2D workspace surface. Firstly, as in the radar case, it will be assumed that the workspace marker points $\mathbf{p}_i$ and target points $\mathbf{y}_i$ are appropriately *centred*, by subtraction of the mean marker location $\bar{\mathbf{p}}$:

$$\mathbf{p}_i \leftarrow \mathbf{p}_i - \bar{\mathbf{p}} \tag{9}$$

$$\mathbf{x}_i \leftarrow \mathbf{x}_i - \bar{\mathbf{p}}. \tag{10}$$

The objective now is to obtain an optimal estimate of the workspace plane, and then to project all 3D target points $\mathbf{x}_i$ into this plane. Let M be the number of points taken all over the surface. Then, the task can be approached by stacking the M workspace points $\mathbf{p}_i$ as the rows of an $M \times 3$ matrix $\mathbf{P}$, and then performing the singular value decomposition (SVD):

$$\mathbf{P}_{[M\times 3]} = \mathbf{U}_{[M\times M]}\, \mathbf{S}_{[M\times 3]}\, \mathbf{V}^\top_{[3\times 3]} \tag{11}$$

where $\mathbf{V}^\top$ denotes the transpose, and the subscripts indicate the dimensions of the corresponding matrices. The singular values σ_i are in the upper part of the block matrix, as follows

$$\mathbf{S} = \begin{bmatrix} \mathrm{diag}(\sigma_1, \sigma_2, \sigma_3) \\ \mathbf{0} \end{bmatrix}. \tag{12}$$

The five ground-plane points in $\mathbf{P}$ will not be exactly co-planar, in practice, owing to uncertainty in the visual estimates. This can be addressed by setting $\sigma_3 = 0$ in (12), thereby defining a projection matrix $\mathbf{S}_\Pi$. The orthogonal projection of the noisy ground points onto the best-fitting plane Π is then given by:

$$\mathbf{P}_\Pi = \mathbf{U}\, \mathbf{S}_\Pi \mathbf{V}^\top \tag{13}$$

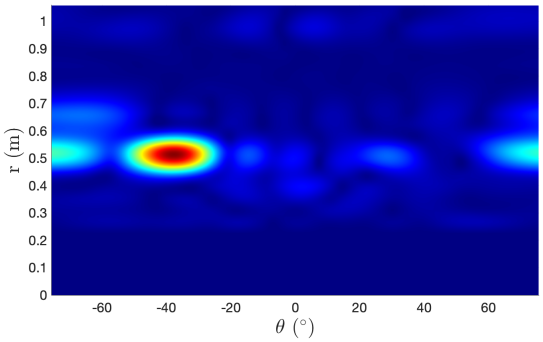
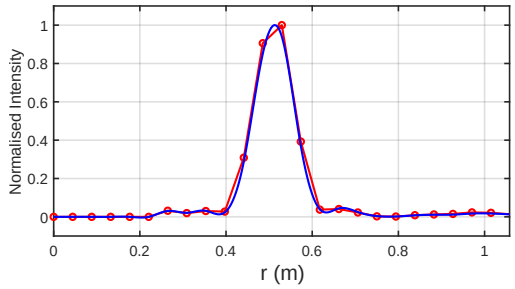
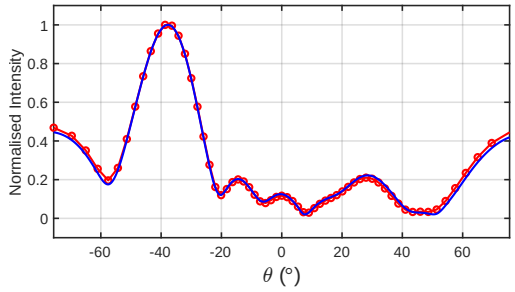


Fig. 1: Example of range-azimuth intensity array after spline interpolation. The peak position indicates the location of a metallic target.



(a) Range Profile



(b) Azimuth Profile

Fig. 2: Normalized (a) range r and (b) azimuth θ profiles, extracted from the intensity array in Fig. 1. Raw profiles are plotted in red, with circles representing the samples (which are nonuniform in θ). Note the improved definition of the maxima, after interpolation (blue).

where the exactly co-planar points are in the rows of the rank-two matrix $\mathbf{P}_\Pi$. If equation (13) is transposed, then $\mathbf{V}$ appears as a rotation matrix, acting on column vectors:

$$\mathbf{P}_\Pi^\top = \mathbf{V}\,\mathbf{S}_\Pi^\top\,\mathbf{U}^\top \tag{14}$$

where $\mathbf{S}_\Pi^\top\mathbf{U}^\top$ has dimensions $3 \times M$. Note that if $\det(\mathbf{V}) = -1$, then both $\mathbf{U}$ and $\mathbf{V}$ should be negated, to ensure that $\mathbf{V}$ does not involve a reflection. Then, the *standardized* coordinates are defined as

$$\widehat{\mathbf{P}}_\Pi^\top = \mathbf{S}_\Pi^\top\,\mathbf{U}^\top \tag{15}$$

where the third row (containing coordinates perpendicular to the plane) is zero. Conversely, in order to standardize the object points $\mathbf{x}_j$, in the rows of $N \times 3$ matrix $\mathbf{X}$, the inverse of transformation $\mathbf{V}$ should be applied to the column-vector points

$$\widehat{\mathbf{X}}^\top = \mathbf{V}^\top\mathbf{X}^\top \tag{16}$$

where $\mathbf{V}^\top = \mathbf{V}^{-1}$. Having performed these transformations, the 2D projections of the N standardized object points, in the columns of $\widehat{\mathbf{X}}^\top$, onto the standardized optimal plane spanned by $\widehat{\mathbf{P}}_\Pi^\top$ are simply

$$\left[\mathbf{x}_1,\ \mathbf{x}_2,\ \ldots,\ \mathbf{x}_N\right] \leftarrow \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\end{bmatrix}\,\widehat{\mathbf{X}}^\top. \tag{17}$$

The notation for these transformations is shown in Fig. 3, and the procedure for the real dataset is illustrated in Fig. 4. As can be observed, the final projections lie in the standardized workspace plane $z = 0$, in which the targets are placed.

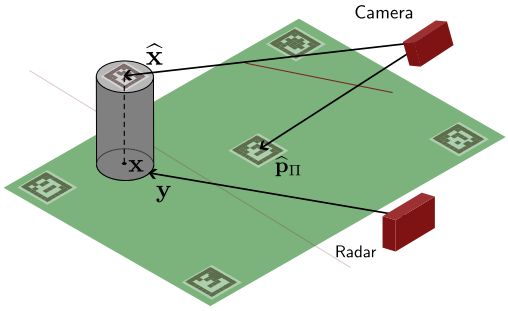


Fig. 3: Workspace and target configuration. The standardized visual coordinates are $[\hat{x}_1, \hat{x}_2, \hat{x}_3]^\top$ extracted from the camera, with direction $\hat{x}_3$ perpendicular to the estimated plane. The 2D radar coordinates $[y_1, y_2]^\top$ will be mapped to the visual coordinates $[x_1, x_2]^\top$, orthogonally projected into the plane determined by $\hat{\mathbf{p}}_\pi$.

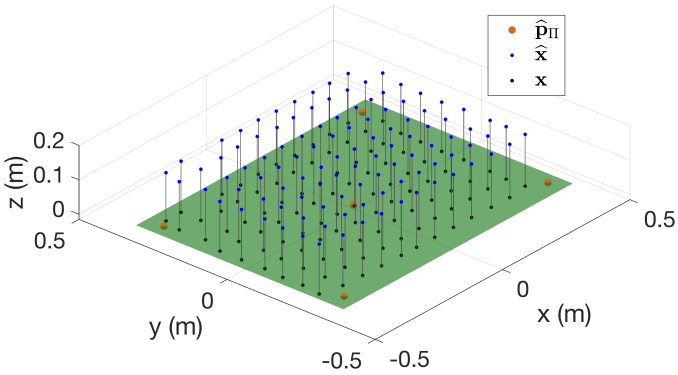


Fig. 4: Complete set of target locations, acquired sequentially from the visual marker system, as in Figure 3. The estimated positions $\widehat{\mathbf{x}}$ are shown, along with their orthogonal projections $\mathbf{x}$ onto the optimal estimate of the workspace plane.

IV. Mapping Methods

Once the radar and visual coordinates have been acquired and pre-processed, as described in sections II and III, we wish to perform a *calibration* of the radar device. This consists of a 2D spatial mapping from the visual estimates $\mathbf{x}_i$ to the corresponding radar coordinates $\mathbf{y}_i$. This mapping is to be estimated once, after which it can be used with or without visual information (e.g. when the target is obscured). The calibration task is treated as a fitting problem, in which a function $\mathbf{y} = \mathbf{g}(\mathbf{x})$ is to be estimated. In particular, note that the accuracy of the *AprilTag* estimates is on the order of millimetres [44], whereas the radar accuracy is on the order of centimetres, as noted in Table I. Hence it is natural to treat the visual data as the ground truth, which is used to predict the uncertain radar data, during the calibration process. We now develop three transformation models, and corresponding estimates, in increasing order of generality. All methods can be extended to 3D, although we are concerned with the 2D case.

A. Linear models

In this section we develop two linear models of the spatial mapping between visual and radar estimates. The simplest

model is a 2D **similarity** transformation, comprising a rigid motion consisting of a translation (t_1, t_2) and a rotation θ, subject to an overall scale factor s (which also absorbs any change of measurement units). The corresponding homogeneous matrix representation is

$$\begin{bmatrix} x'_1 \\ x'_2 \\ 1 \end{bmatrix} = \begin{bmatrix} s\,\cos\theta & -s\,\sin\theta & t_1 \\ s\,\sin\theta & s\,\cos\theta & t_2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix}. \tag{18}$$

More generally, we consider the 2D **affine** model, comprising a general linear transformation, and a translation. The corresponding homogeneous matrix representation is

$$\begin{bmatrix} x''_1 \\ x''_2 \\ 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & t_1 \\ a_{21} & a_{22} & t_2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix}. \tag{19}$$

Using matrix notation, the similarity (18) and affine (19) models can be expressed more concisely by assigning them $\mathbf{G}_S$ and $\mathbf{G}_A$ respectively as

$$\mathbf{x}' = \mathbf{G}_S\,\mathbf{x} \quad \text{and} \quad \mathbf{x}'' = \mathbf{G}_A\,\mathbf{x} \tag{20}$$

with four and six parameters respectively. If there is no shear or anisotropic scaling, then the similarity model may be preferred; this is an empirical question, to be addressed in section V.

The solutions for $\mathbf{G}_A$ or $\mathbf{G}_S$ will be over-determined as the dataset will be composed of around 70 point correspondences (see section V). An optimal estimate is obtained by minimizing the sum of squared point differences given by:

$$E = \sum_{i=1}^{n} \left\| \mathbf{y}_i - \mathbf{G}\mathbf{x}_i \right\|^2. \tag{21}$$

This corresponds to an isotropic Gaussian noise model for the marker locations, with $\mathbf{G}$ being taken according to the desired transformation model, e.g. $\mathbf{G}_S$ or $\mathbf{G}_A$.

The constrained solution for $\mathbf{G}_S$ is obtained by orthogonal Procrustes analysis with isotropic scaling. Here, we explicitly estimate the rotation matrix $\mathbf{R}$, the translation vector $\mathbf{t}$ and the scaling parameter s from (18). As stated in [45], the rotation can be determined by applying SVD to the outer product of the mean-centred points:

$$\mathbf{U}\mathbf{D}\mathbf{V}^\top = \sum_{i=1}^{n} \mathbf{x}_i\,\mathbf{y}_i^\top. \tag{22}$$

where $\mathbf{U}$ is orthogonal, $\mathbf{D}$ is diagonal and $\mathbf{V}$ is also orthogonal. The optimal rotation matrix is then constructed from the product

$$\mathbf{R}_\star = \mathbf{V} \begin{bmatrix} 1 & 0 \\ 0 & d \end{bmatrix} \mathbf{U}^\top \tag{23}$$

where $d = \pm 1$ is defined by $\det(\mathbf{V}\,\mathbf{U}^\top)$, in order to ensure that the solution is not reflected. The optimal scaling parameter is obtained from the scatter ratio of the mean-centred points

$$s_\star = \left(\frac{\sum_{i=1}^{n} \|\mathbf{y}_i\|^2}{\sum_{i=1}^{n} \|\mathbf{x}_i\|^2} \right)^{1/2} \tag{24}$$

as shown in [46], and where the $\|\cdot\|^2$ operator is the square of the length of the vector. Finally, the optimal translation is the mean offset, after rotation and scaling:

$$\mathbf{t}_\star = \frac{1}{n} \sum_{i=1}^{n} \big(\mathbf{y}_i - s_\star\,\mathbf{R}_\star\,\mathbf{x}_i \big). \tag{25}$$

The estimates $\mathbf{R}_\star$, $s_\star$ and $\mathbf{t}_\star$ are finally substituted into (18), to give the optimal similarity matrix $\mathbf{G}_S$.

The unconstrained solution for the affine model $\mathbf{G}_A$ in (19) is obtained by standard least squares methods, after vectorizing the equation $\mathbf{x}'' = \mathbf{G}_A\mathbf{x}$ as follows:

$$\begin{bmatrix} x''_{11} \\ x''_{12} \\ \vdots \\ x''_{n1} \\ x''_{n2} \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_{11} & x_{12} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_{n1} & x_{n2} & 1 \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{12} \\ t_1 \\ a_{21} \\ a_{22} \\ t_2 \end{bmatrix}. \tag{26}$$

This becomes an approximation to the correspondingly vectorized radar measurements $\mathbf{Y}$, in the presence of measurement and other errors, which can be expressed in terms of the $2n \times 6$ design matrix $\mathbf{B}$ and 6×1 parameter vector $\mathbf{a}$:

$$\mathbf{Y} \approx \mathbf{X}'' \quad \text{where} \quad \mathbf{X}'' = \mathbf{B}\,\mathbf{a} \tag{27}$$

is the vectorization of the transformed visual locations $\mathbf{x}''_i$, as above. The least squares solution $\mathbf{a}_\star$, which minimizes (21) is obtained directly from the matrix pseudoinverse

$$\mathbf{a}_\star = \mathbf{B}^+\,\mathbf{Y}. \tag{28}$$

This approach cannot, however, be used if there are additional geometric constraints on the transformation.

B. Nonlinear model

It is also desirable to consider nonlinear models of the spatial transformation, but these must be subject to practical constraints. The **thin-plate spline** (TPS) model is arguably the most natural nonlinear model for spatial data, without making any assumptions about the sampling pattern. This can be argued by physical analogy; if the basic affine mapping is interpreted as a plane, relating input to output coordinates, then the TPS allows a controlled deformation of the plane, defined by minimal bending energy [47]. Indeed, the affine mappings are a special case (zero deformation) of the TPS, as represented in Fig. 5. Finally, the optimization problem for the TPS has a closed-form solution, which avoids any ambiguities in the procedure.

If the third possible mapping function is $\mathbf{x}''' = \mathbf{g}(\mathbf{x})$, then the following combination of data fidelity and regularization energy is minimized, by standard methods [47]:

$$E_p = p \sum_{i=1}^{n} \left\| \mathbf{y}_i - \mathbf{g}(\mathbf{x}) \right\|^2 + (1-p) \iint \left\| \nabla^2 \mathbf{g}(\mathbf{x}) \right\|^2 \mathrm{d}\mathbf{x}. \tag{29}$$

Note the resemblance of the data term to (21), and also note that the Laplacian term is zero for any linear model. In fact, if $p \ll 1$, then the previous estimate (28) for the

Fig. 5: Spatial transformation models. The similarity transformation $\mathbf{G}_S$ can be considered a special case of the affine transformation $\mathbf{G}_A$. The latter is, in turn, a special case of the thin-plate spline $\mathbf{g}$ with regularization parameter $p = 0$ (maximal smoothness).

affine model $\mathbf{G}_A$ is recovered. This is because the effective regularization weight $(1-p)/p$ becomes dominant as $p \to 0$, thereby prohibiting any model for which the Laplacian term is nonzero.

We now have three properly nested classes of transformation: similarity, affinity, and thin-plate spline, as depicted in Fig. 5. This structure will simplify the experimental evaluation, as described below.

V. Evaluation

This section presents an evaluation of the proposed system, starting with a description of the hardware and experimental design, followed by a comparison of the proposed spatial transformation models.

A. Experimental setup

The experimental setup has been arranged to resemble the workspace of a typical robot arm. As explained in the introduction, this work pursues the development of a precise detection and positioning system for metallic targets, e.g. for grasping and manipulation.

The tests were carried out on a horizontal surface of dimensions 40 cm $\times$ 60 cm. The target is a hollow aluminium cylinder of 5 cm diameter and 10 cm height (material thickness 2 mm). This target object was chosen for the following four reasons. Firstly, it has a well-defined geometric centroid, which is constant in the vertical direction, making it suitable for our 2D localization experiments. Secondly, the object has a well-defined radar phase-centre which coincides with the geometric centroid. Thirdly, the axial symmetry of the object means that the location signal (which is of primary interest) is not affected by the axial orientation of the object. Finally, the metal cylinder resembles a typical component or container, to be found in an industrial setting.

An *AprilTag* marker on top is used to determine the visual position, as described in section III. Although visual marker systems are well established in the robotics literature (see e.g. [2]), we conducted an experimental assessment of the accuracy of the visual system for our particular setup, as the practical performance depends on the physical size of the markers, the camera resolution, and the overall viewing distance. Specifically, we prepared marker configurations with known pairwise separations, which we compared to the corresponding vision-based length estimates. The observed variance is consistent with a mean location error of approximately 1 mm, for each marker, according to standard uncertainty propagation methods [48].

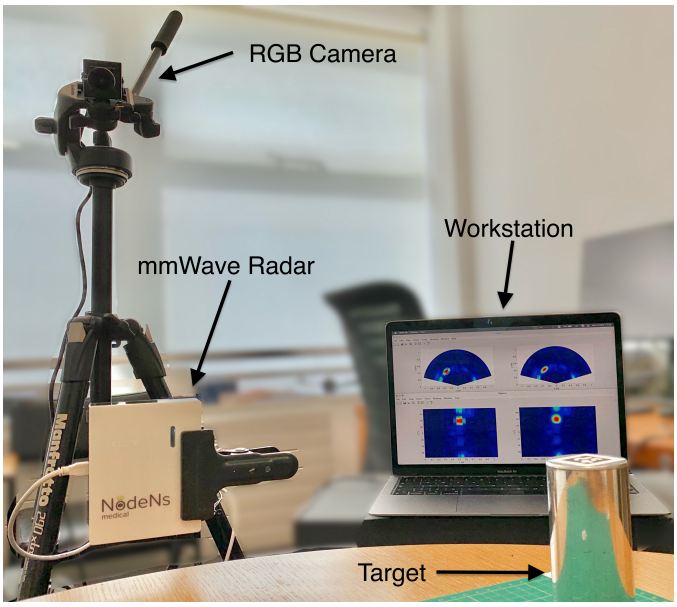


Fig. 6: Experimental setup. This image shows all the components of the proposed system, including the mmWave radar, the RGB camera, the metallic target, and the workstation.

The calibration dataset comprises 106 measurements, with the target placed in different positions across the entire workspace. A rectangular grid was used, in order to ensure an even distribution of data across the workspace (but note that any perturbation will affect both the visual and radar signals). In fact, the method can be applied for unevenly distributed markers, but we prefer to avoid any arbitrary asymmetries in our experimental setup. The surface and object marker positions are estimated from RGB images, taken by an overhead camera. Meanwhile, the mmWave radar is positioned at the workspace level, looking across the scene, parallel to the surface. This configuration tends to optimize the workspace visibility, with respect to the two devices. Fig. 6 shows the complete setup.

The radar was manufactured by NodeNs Medical Ltd [49] and is based on the Texas Instruments IWR6843 chipset. It is a versatile FMCW radar, as shown in other works like [50], which operates at the (unlicensed) 60 GHz band. Its configuration is shown in Table I. The selection and design of the 2×4 antenna configuration was based on the proposed system's requirement for in-plane detection, and because it makes a suitable tradeoff between cost/complexity and performance. The principles of MIMO radar and spatial beamforming are used for object localization (the current work involves a single target but this could be extended to multiple targets), while the technique presented in this study further enhances spatial resolution on the existing hardware [42].

It is important to note that the presence of low-density materials as potential blockages in our application could give rise to diffraction and delay effects. Nevertheless, these effects can be safely neglected in our application due to the inherent properties of the materials, which result in minimal shift of radar array peaks, typically no more than a few millimetres. However, if deemed necessary, a new calibration can be

TABLE I: Configuration of mmWave radar sensor.

Number of TX antennas	2
Number of RX antennas	4
Initial Frequency	60 GHz
Bandwidth	3.4 GHz
Range Resolution	4.4 cm
Angular Field of View	120°

performed for convenience, which would account for any new adverse effects that may arise. The RF circuitry which is used for chirp synthesis may exhibit non-linearities which could affect the performance of the radar. Hence, actions have been taking to mitigate against these, including a closed-loop phase-locked loop (PLL) implemented on-chip, and appropriate design of the chirp profiles. For the latter, an idle time period is incorporated between successive chirps, to allow time for circuit oscillations to dampen, and an ADC ramp-up time is allowed, so that the chirp begins in the linear region of circuit operation. Further details are provided in the TI IWR6843 datasheet [51].

B. Regularization parameter analysis

As explained in Section IV, the thin-plate spline (TPS) optimization includes a regularization weight $1-p$, which controls the smoothness of the estimate (29). An appropriate value for this parameter can be obtained by cross-validation, based on the median squared error, which is chosen for robustness. We evaluate 30 random training/test splits, for $p \in [0,1]$, with resolution $\Delta p = 0.001$. A plot of the median squared error suggests a quite abrupt transition into overfitting, as the regularization weight $1-p$ approaches zero. We propose the following piecewise rational/linear model for the residuals, with break-point at $p=q$, and slope δ for the linear component:

$$F(p) = \begin{cases} \alpha + \beta/(\gamma + p) & \text{if } p < q \\ F(q) + \delta(p-q) & \text{otherwise.} \end{cases} \tag{30}$$

A good fit to the observations is obtained by nonlinear minimization over q and $[\alpha, \beta, \gamma, \delta]$, as shown in Fig. 7 (red curve). The optimal p value corresponds to the estimated break-point $q = 0.934$, as indicated. Note that setting $p=0$ corresponds to the affine model (19), while setting $p=1$ corresponds to an interpolating spline, which overfits the data.

C. Results

We now compare the proposed mapping methods, based on the Euclidean distance between estimates and reference values. In order to avoid overfitting, we divide the whole dataset into training and test groups with a $70\%/30\%$ split, employing a random sampling process (without replacement).

The general pattern of results is indicated in the histograms of Fig. 8, which show the residual errors for one random training/test division, for each mapping type. As can be seen, the affinity outperforms the similarity, which suggests that shear and anisotropic scaling is required. In addition, the TPS model outperforms both of the other methods (given an appropriate estimate of the regularization parameter), meaning that the spline has modelled systematic nonlinear errors, without overfitting.

TABLE II: Root mean square spatial error (cm), computed over 1000 train/test splits.

	Similarity		Affinity		TPS	
	Median	Mean	Median	Mean	Median	Mean
Training	2.75	2.70	1.01	1.06	0.45	0.50
Test	2.83	2.78	1.06	1.11	0.62	0.70

Table II shows the estimation errors, having performed 1000 random training/test splits. The average errors are approximately 1 cm for the affine transformation and around 7 mm for the TPS process. Finally, the performance on a typical train/test trial is visualized in Fig. 9, which shows the TPS performance. Here, the connecting lines correspond to the estimation errors, for comparison with Table II.

Finally, the efficiency of the proposed algorithms will be considered. In the offline stage, the TPS model (discussed in Section IV-B) has the highest computational complexity, being $\mathcal{O}(n^3)$, where n denotes the total number of points in the training stage. However, with $n = 70$ in the reported experiments, the average computation time in this process is less than 2 ms. If very large numbers of calibration points are required, then fast TPS approximations are available [52]. In all cases, the mapping need only be estimated once, and does not contribute to the run-time cost. The most demanding run-time task is the spline interpolation, which is applied to the radar intensity array. This can be implemented by separable convolution with a fixed kernel, resulting in complexity $\mathcal{O}(K_r K_\theta)$ for a radar array of size $K_r \times K_\theta$. The convolution could be implemented on-chip, although we used a software implementation in the

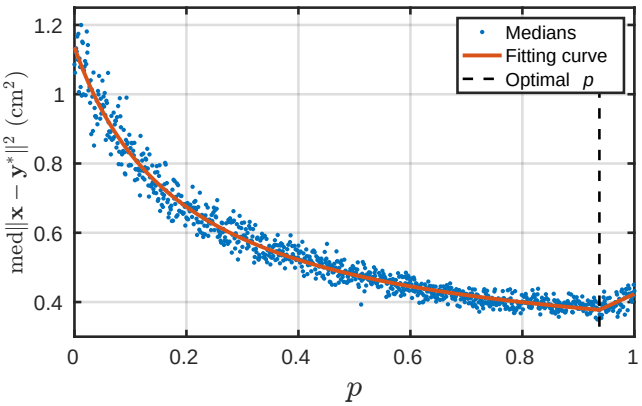


Fig. 7: Cross-validation of the thin-plate spline regularization weight $1-p$. The blue dots represent the medians of squared spatial errors, over 30 random training/test data divisions, for each trial p-value. The optimal p-value (dashed line) is estimated as the break-point of the fitted piecewise model (red line) given by (30).

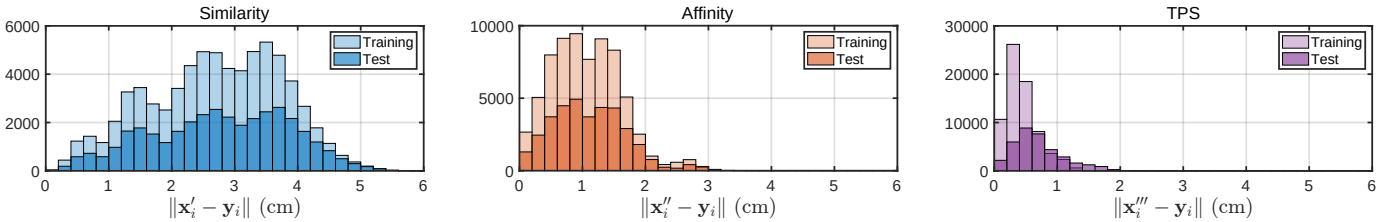


Fig. 8: Histograms depicting the distances between reference values and estimates in all 1000 random training/test splits, for all proposed transformations: (a) Similarity transformation, (b) Affine Transformation, and (c) TPS.

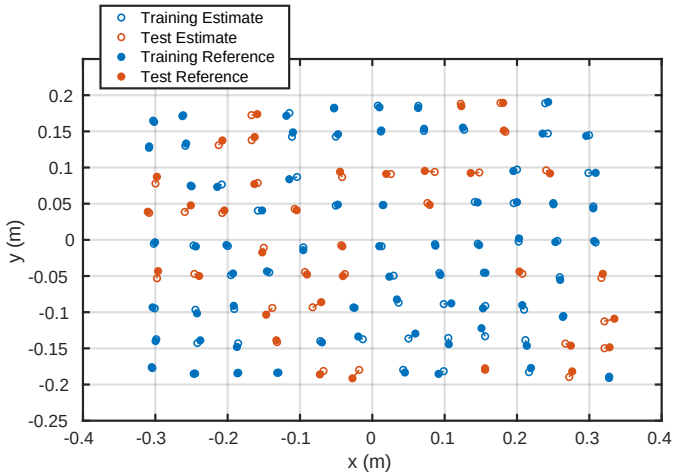


Fig. 9: Example system performance on one training/test data split. The filled circles represent the radar data, while the empty circles correspond to the estimates obtained from the TPS model. Colours indicate the training (blue circles) and test (orange circles) data, on this trial. The connecting lines represent the estimation errors for each measurement.

present experiments.

VI. CONCLUSION

This work has presented an accurate detection and localization system for metallic objects, based on a combination of mmWave radar and visual sensors. First, the radar signal was interpolated, to an appropriate resolution for typical robotic manipulation tasks. Next, a system of visual markers was used, in order to obtain ground truth position data, across the entire workspace. Finally, the relationship between the radar and ground truth visual estimates was modelled, using three types of spatial transformation. It has been shown that the TPS model is substantially more accurate than the default linear models. The complete system achieves sub-centimetre accuracy in the radar estimates. This is sufficient for basic robotic grasping tasks [39], which can now be performed in the absence of visual information, such as when the target is obscured by packaging.

The introduction of a robot arm and gripper would raise new questions, in a practical system. Firstly, it may be necessary to account for radar interference, if object localization is to be performed while (as opposed to before) the arm moves in the scene. Indeed, this analysis could be combined with the extension of the present methods to a 3D workspace, in which the robot is used to position the calibration targets. Furthermore, it may be useful to estimate the shape and pose of the target objects, as well as their locations, from the radar data. Finally, the proposed system could potentially be used to detect multiple and/or moving objects. These extensions will be explored in future work.

REFERENCES

[1] G. Du, K. Wang, S. Lian, and K. Zhao, "Vision-based Robotic Grasping From Object Localization, Object Pose Estimation to Grasp Estimation for Parallel Grippers: A Review," *Artificial Intelligence Review*, vol. 54, no. 3, pp. 1677–1734, 2021.

[2] E. Olson, "AprilTag: A Robust and Flexible Visual Fiducial System," in *2011 IEEE International Conference on Robotics and Automation*. IEEE, 2011, pp. 3400–3407.

[3] D. Kragic and M. Vincze, "Vision for robotics," *Foundations and Trends® in Robotics*, vol. 1, no. 1, pp. 1–78, 2009.

[4] Z. Xu, Z. He, J. Wu, and S. Song, "Learning 3D Dynamic Scene Representations for Robot Manipulation," in *Proceedings of the 2020 Conference on Robot Learning*. PMLR, 2021, pp. 126–142.

[5] H. Cheng, Y. Wang, and M. Q.-H. Meng, "A Vision-Based Robot Grasping System," *IEEE Sensors Journal*, vol. 22, no. 10, pp. 9610–9620, 2022.

[6] H. Karaoguz and P. Jensfelt, "Object Detection Approach for Robot Grasp Detection," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 4953–4959.

[7] S. Kumra and C. Kanan, "Robotic Grasp Detection using Deep Convolutional Neural Networks," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 769–776.

[8] K. Kleeberger, R. Bormann, W. Kraus, and M. F. Huber, "A Survey on Learning-Based Robotic Grasping," *Current Robotics Reports*, vol. 1, no. 4, pp. 239–249, 2020.

[9] Y. Sun, J. Falco, M. A. Roa, and B. Calli, "Research Challenges and Progress in Robotic Grasping and Manipulation Competitions," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 874–881, 2022.

[10] R. Tsai and R. Lenz, "A new technique for fully autonomous and efficient 3D robotics hand/eye calibration," *IEEE Transactions on Robotics and Automation*, vol. 5, no. 3, pp. 345–358, 1989.

[11] R. B. Rusu, N. Blodow, Z. C. Marton, and M. Beetz, "Close-range scene segmentation and reconstruction of 3D point cloud maps for mobile manipulation in domestic environments," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2009, pp. 1–6.

[12] B. van Berlo, A. Elkelany, T. Ozcelebi, and N. Meratnia, "Millimeter Wave Sensing: A Review of Application Pipelines and Building Blocks," *IEEE Sensors Journal*, vol. 21, no. 9, pp. 10 332–10 368, 2021.

[13] D. Wang, M. Fattouche, and X. Zhan, "Pursuance of mm-Level Accuracy: Ranging and Positioning in mmWave Systems," *IEEE Systems Journal*, vol. 13, no. 2, pp. 1169–1180, 2019.

[14] W. A. Ahmad, A. Ergintav, J. Wessel, D. Kissinger, and H. J. Ng, "Experimental Evaluation of Millimeter-Wave FMCW Radar Ranging Precision," in *2021 IEEE Radio and Wireless Symposium (RWS)*, 2021, pp. 70–72.

[15] A. Pandharipande, C.-H. Cheng, J. Dauwels, S. Z. Gurbuz, J. Ibanez-Guzman, G. Li, A. Piazzoni, P. Wang, and A. Santra, "Sensing and Machine Learning for Automotive Perception: A Review," *IEEE Sensors Journal*, vol. 23, no. 11, pp. 11 097–11 115, 2023.

[16] S. Gupta, P. K. Rai, A. Kumar, P. K. Yalavarthy, and L. R. Cenkeramaddi, "Target Classification by mmWave FMCW Radars Using Machine Learning on Range-Angle Images," *IEEE Sensors Journal*, vol. 21, no. 18, pp. 19 993–20 001, 2021.

[17] B. Tan, Z. Ma, X. Zhu, S. Li, L. Zheng, S. Chen, L. Huang, and J. Bai, "3-D Object Detection for Multiframe 4-D Automotive Millimeter-Wave Radar Point Cloud," *IEEE Sensors Journal*, vol. 23, no. 11, pp. 11 125–11 138, 2023.

[18] Z. Wei, F. Zhang, S. Chang, Y. Liu, H. Wu, and Z. Feng, "MmWave Radar and Vision Fusion for Object Detection in Autonomous Driving: A Review," *Sensors*, vol. 22, no. 7, 2022.

[19] J. Guan, S. Madani, S. Jog, S. Gupta, and H. Hassanieh, "Through Fog High-Resolution Imaging Using Millimeter Wave Radar," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 11 461–11 470.

[20] M. Sheeny, E. De Pellegrin, S. Mukherjee, A. Ahrabian, S. Wang, and A. Wallace, "RADIATE: A Radar Dataset for Automotive Perception in Bad Weather," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. Xi'an, China: IEEE, 2021, pp. 1–7.

[21] P. Liu, G. Yu, Z. Wang, B. Zhou, and P. Chen, "Object Classification Based on Enhanced Evidence Theory: Radar–Vision Fusion Approach for Roadside Application," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–12, 2022.

[22] A. Kosuge, S. Suehiro, M. Hamada, and T. Kuroda, "mmWave-YOLO: A mmWave Imaging Radar-Based Real-Time Multiclass Object Recognition System for ADAS Applications," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–10, 2022.

[23] Y. Song, Z. Xie, X. Wang, and Y. Zou, "MS-YOLO: Object Detection Based on YOLOv5 Optimized Fusion Millimeter-Wave Radar and Machine Vision," *IEEE Sensors Journal*, vol. 22, no. 15, pp. 15 435–15 447, 2022.

[24] Y. Li, J. Deng, Y. Zhang, J. Ji, H. Li, and Y. Zhang, "EZFusion: A Close Look at the Integration of LiDAR, Millimeter-Wave Radar, and Camera for Accurate 3D Object Detection and Tracking," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 11 182–11 189, 2022.

[25] X. Tang, Z. Zhang, and Y. Qin, "On-Road Object Detection and Tracking Based on Radar and Vision Fusion: A Review," *IEEE Intelligent Transportation Systems Magazine*, vol. 14, no. 5, pp. 103–128, 2022.

[26] J. Peršić, I. Marković, and I. Petrović, "Extrinsic 6DoF calibration of a radar–LiDAR–camera system enhanced by radar cross section estimates evaluation," *Robotics and Autonomous Systems*, vol. 114, pp. 217–230, 2019.

[27] J. Peršić, L. Petrović, I. Marković, and I. Petrović, "Online multi-sensor calibration based on moving object tracking," *Advanced Robotics*, vol. 35, no. 3-4, pp. 130–140, 2021.

[28] ——, "Spatiotemporal Multisensor Calibration via Gaussian Processes Moving Target Tracking," *IEEE Transactions on Robotics*, vol. 37, no. 5, pp. 1401–1415, 2021.

[29] C.-L. Lee, Y.-H. Hsueh, C.-C. Wang, and W.-C. Lin, "Extrinsic and Temporal Calibration of Automotive Radar and 3D LiDAR," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Las Vegas, NV, USA: IEEE, 2020, pp. 9976–9983.

[30] J. Oh, K.-S. Kim, M. Park, and S. Kim, "A Comparative Study on Camera-Radar Calibration Methods," in *15th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, 2018.

[31] Y. Cheng, H. Xu, and Y. Liu, "Robust Small Object Detection on the Water Surface through Fusion of Camera and Millimeter Wave Radar," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 15 243–15 252.

[32] C. Scholler, M. Schnettler, A. Krammer, G. Hinz, M. Bakovic, M. Guzet, and A. Knoll, "Targetless Rotational Auto-Calibration of Radar and Camera for Intelligent Transportation Systems," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. Auckland, New Zealand: IEEE, 2019, pp. 3934–3941.

[33] E. Wise, J. Peršić, C. Grebe, I. Petrović, and J. Kelly, "A Continuous-Time Approach for 3D Radar-to-Camera Extrinsic Calibration," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 13 164–13 170.

[34] E. Wise, Q. Cheng, and J. Kelly, "Spatiotemporal Calibration of 3D mm-Wavelength Radar-Camera Pairs," *IEEE Transactions on Robotics*, vol. 39, no. 6, pp. 4552–4566, 2023.

[35] Q. Cheng, E. Wise, and J. Kelly, "Extrinsic Calibration of 2D Millimetre-Wavelength Radar Pairs Using Ego-Velocity Estimates," in *2023 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, 2023, pp. 559–565.

[36] K. Harlow, H. Jang, T. D. Barfoot, A. Kim, and C. Heckman, "A New Wave in Robotics: Survey on Recent mmWave Radar Applications in Robotics," 2023, arXiv:2305.01135.

[37] C. Stetco, B. Ubezio, S. Mühlbacher-Karrer, and H. Zangl, "Radar Sensors in Collaborative Robotics: Fast Simulation and Experimental Validation," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 2020, pp. 10 452–10 458.

[38] C. Schoffmann, B. Ubezio, C. Bohm, S. Muhlbacher-Karrer, and H. Zangl, "Virtual Radar: Real-Time Millimeter-Wave Radar Sensor Simulation for Perception-Driven Robotics," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 4704–4711, 2021.

[39] Y. Bekiroglu, N. Marturi, M. A. Roa, K. J. M. Adjigble, T. Pardi, C. Grimm, R. Balasubramanian, K. Hang, and R. Stolkin, "Benchmarking Protocol for Grasp Planning Algorithms," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 315–322, 2020.

[40] M. A. Richards, *Fundamentals of Radar Signal Processing*, 2nd ed. New York: McGraw-Hill Education, 2014.

[41] E. Fishler, A. Haimovich, R. Blum, D. Chizhik, L. Cimini, and R. Valenzuela, "MIMO Radar: An Idea Whose Time Has Come," in *2004 IEEE Radar Conference*, 2004, pp. 71–78.

[42] J. Li and P. Stoica, "MIMO Radar with Colocated Antennas," *IEEE Signal Processing Magazine*, vol. 24, no. 5, pp. 106–114, 2007.

[43] J.-Y. Bouguet, "Camera Calibration Toolbox for Matlab," CaltechDATA, 2022.

[44] S. M. Abbas, S. Aslam, K. Berns, and A. Muhammad, "Analysis and Improvements in AprilTag Based State Estimation," *Sensors*, vol. 19, no. 24, p. 5480, 2019.

[45] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-Squares Fitting of Two 3-D Point Sets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-9, no. 5, pp. 698–700, 1987.

[46] B. K. P. Horn, "Closed-Form Solution of Absolute Orientation Using Unit Quaternions," *Journal of the Optical Society of America A*, vol. 4, no. 4, pp. 629–642, 1987.

[47] F. L. Bookstein, *Morphometric Tools for Landmark Data*. United Kingdom: Cambridge University Press, 1991.

[48] S. Yi, R. M. Haralick, and L. G. Shapiro, "Error Propagation in Machine Vision," *Machine Vision and Applications*, vol. 7, no. 2, pp. 93–114, 1994.

[49] NodeNs Medical Ltd., "Millimetre Wave Radar," https://nodens.eu, 2021.

[50] Z. Yu, A. Taha, W. Taylor, A. Zahid, K. Rajab, H. Heidari, M. A. Imran, and Q. H. Abbasi, "A Radar-Based Human Activity Recognition Using a Novel 3-D Point Cloud Classifier," *IEEE Sensors Journal*, vol. 22, no. 19, pp. 18 218–18 227, 2022.

[51] Texas Instruments, "IWR6843, IWR6443 Single-Chip 60- to 64-GHz mmWave Sensor," https://www.ti.com/lit/ds/symlink/iwr6843.pdf, 2021.

[52] G. Donato and S. Belongie, "Approximate Thin Plate Spline Mappings," in *7th European Conference on Computer Vision (ECCV)*, 2002, pp. 21–31.