Design of a 3D Indoor Localization System Enabling Augmented Reality TV Applications

Francesco Sottile*, Shiva Ehsanibalajorshary†, Luigi Coriasco*, Claudio Pastrone*, Roberto Iacoviello‡, Davide Zappia‡
LINKS Foundation*, Politecnico di Torino†, Radiotelevisione Italiana‡
Turin, Italy
francesco.sottile@linksfoundation.com, shiva.ehsanibalajorshary@studenti.polito.it,
{luigi.coriasco, claudio.pastrone}@linksfoundation.com, {roberto.iacoviello, davide.zappia}@rai.it

***Abstract*—This paper focuses on the design of a robust Real Time Locating Systems (RTLS) based on the Ultra-Wide Band (UWB) technology for Augmented Reality (AR) applications in TV studios that require artists and/or a presenter to be accurately localized. According to a UWB-based measurement campaign, carried out in a TV studio environment, ranging measurements are heavily affected by the human body interference. Indeed, lots of outliers are present as the UWB receiver may synchronize to reflected paths, which result to be much stronger than the direct one. As a consequence, range errors are very large. In this context, to improve the localization performance, we increased the redundancy of the RTLS by employing more than one tag to localize the artists on the TV scene. In particular, we have applied the Extended Kalman Filter (EKF) algorithm to work with two and three tags. Moreover, an outlier detection and correction procedure have been defined and adopted for the ranging phase. The resulting localization performance, based on real range measurements, shows that the EKF with two tags outperforms by 83.5% the one with single tag.**

I. Introduction and Related Work

Real Time Locating Systems (RTLS), based on the Ultra-Wide Band (UWB) technology, have been extensively studied for many years and mainly proposed for indoor environments [1], where Global Navigation Satellite Systems (GNSS) do not work as the satellite signal is obstructed. Moreover, the development of RTLS has been proliferating thanks to commercially available off-the-shelf UWB kits, such as those provided by the Decawave company (now acquired by Qorvo) and recently by NXP, both compliant with the UWB standard IEEE 802.15.4-2015 [2]. As a result, UWB-based RTLS have been rapidly developed for various application domains, such as Industry 4.0 (*e.g.*, asset, forklift and employee tracking), Healthcare (*e.g.*, patient and asset tracking), Smart Home (*e.g.*, residential access control, object finding). Moreover, some applications are going to be developed for Augmented Reality (AR) gaming by the Spark Microsystem company also thanks to the recent integration of the UWB chip in some smartphones (*e.g.*, Samsung and Apple).

This paper focuses on the design of UWB-based RTLS for real-time AR television applications. In particular, some examples of applications envisaged by RAI (Radiotelevisione Italiana), the Italian public broadcasting company, for which the UWB localization acts as an enabling technology, are reported as follows. *Application A1: On TV screen display of real-time information associated to an artist in a television scene*. For example, the TV system could display a televoting code next to or above the tracked artist. *Application A2: Interaction of a presenter with some AR objects in the TV scene*. In this scenario, the system could display in real-time what the presenter draws with his index finger in the space in front of him. For example, the presenter of the *Giro d'Italia* TV programme could follow the profile of the mountains with his index finger while the system reports in real-time the corresponding trace on the TV screen.

Most of the time, AR effects are generated in post-production but this is in contrast to TV production where there is often a requirement for live use. As for live productions, AR in TV studio is made exploiting mechanical sensors with the drawback of a complex set up and a high budget. The proposed UWB solution overcomes these problems combining a reasonable budget with easy setup. Moreover, this technology allows interesting ways of interaction between the presenter and virtual elements in real-time during filming, enabling new TV formats. There is also the possibility to provide some form of visualization to the presenter by placing a TV monitor out-of shot but visible to the presenter, showing the composite image.

In the context of TV scenes, designing and developing an accurate UWB-based RTLS to track persons is not straightforward due to both the human body shadowing effect and reflections caused by the indoor environments. It has been observed that UWB tags mounted on the human body can raise Line-of-Sight (LoS), Quasi-Line-of-Sight (QLoS), and Non-Line-of-Sight (NLoS) scenarios leading to significant ranging errors [3]–[5]. In addition, it has been investigated that the ranging error strongly depends also on the relative heading angle between the human body, the tag and and anchors [6]. Also the presence of walls and other obstacles in indoor environments represents a significant challenge in terms of localization, as they can result in positively biased ranging estimates [7]. Some Machine Learning (ML)-based techniques able to perform NLoS identification and mitigation have been proposed [7], [8] that require a beforehand training process and thus a collection of a significant set of labelled data. Other ML techniques estimate ranging error directly from the received wave-form, without any a priori or a posteriori knowledge

ISSN 2305-7254

of the NLoS condition. However, not all UWB chips provide wave-form data.

This paper proposes a simple but efficient method to detect range outliers by employing two UWB tags to be placed on the human body. The tracking algorithm, which is based on Extended Kalman Filter (EKF) [9], has been properly designed to work with two and three tags.

The remainder of the paper is organized as follows. Section II introduces the EKF algorithm. Section III presents the application of the EKF for tracking 1-Tag. Sections IV and V present the extension of the EKF with 2-Tag and 3-Tag cases, respectively, for robust localization. Section VI presents the procedure for outliers detection and correction by using two tags while section VII shows the achieved localization performance in the TV studio. Finally, conclusions are given in section VIII.

II. Overview of the Extended Kalman Filter Algorithm

The Kalman Filter (KF) algorithm provides an efficient and practical solution for linear dynamic systems with Gaussian noisy measurements. Typically, localization problems are non linear, hence, some linearizations and approximations are needed. In this case, instead of KF, EKF can be applied [9].

The EKF algorithm, modelled by discrete-time state equations, can measure recursively the condition of a dynamic system:

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}) + \mathbf{w}_k, \tag{1}$$

where $\mathbf{x}_k$ is the state vector at time k, f is the state transition function that evolves the state in time given the previous state, and $\mathbf{w}_k = \mathcal{N}(0, \mathbf{Q}_k)$ is the random process noise vector that takes into account the non linearity and perturbations of the system. This vector is modelled with a normal distribution with zero mean and covariance matrix $\mathbf{Q}_k$.

The system is observed through the following measurement equation:

$$\mathbf{z}_k = h(\mathbf{x}_k) + \mathbf{v}_k, \tag{2}$$

where $\mathbf{z}_k$ is the measurements vector at time k, h is the observation function that estimates the expected measurements of the state $\mathbf{x}_k$ and $\mathbf{v}_k = \mathcal{N}(0, \mathbf{R}_k)$ is the random observation noise vector assumed normally distributed with zero mean and covariance matrix $\mathbf{R}_k$.

The EKF is composed of two phases, the *predict* phase and the *update* phase. The *predict* phase takes the previous a posteriori state vector estimate, $\hat{\mathbf{x}}_{k-1|k-1}$, and computes the a priori state vector estimate:

$$\hat{\mathbf{x}}_{k|k-1} = f(\hat{\mathbf{x}}_{k-1|k-1}). \tag{3}$$

Also the covariance matrix associated to the predicted state vector is updated as follows:

$$\mathbf{P}_{k|k-1} = \mathbf{F}_k \mathbf{P}_{k-1|k-1} \mathbf{F}_k^T + \mathbf{Q}_k, \tag{4}$$

where $\mathbf{F}_k$ is the Jacobian matrix of the state transition function f calculated around the previous a posteriori state estimate $\hat{\mathbf{x}}_{k-1|k-1}$. After finding these a priori values, the algorithm proceeds with the *update* phase, by correcting the a priori estimates with a weighted factor of the measurements taking into account their covariance matrix and the a priori state estimate.

$$\mathbf{S}_k = \mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k, \tag{5}$$

$$\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}_k^T \mathbf{S}_k^{-1}, \tag{6}$$

$$\mathbf{s}_k = \mathbf{z}_k - h(\hat{\mathbf{x}}_{k|k-1}), \tag{7}$$

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k \mathbf{s}_k, \tag{8}$$

$$\mathbf{P}_{k|k} = (\mathbf{I}_n - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k|k-1}, \tag{9}$$

where $\mathbf{s}_k$ is the innovation vector and $\mathbf{S}_k$ its covariance matrix, $\mathbf{K}_k$ is the Kalman gain and $\mathbf{H}_k = \frac{\partial h}{\partial \mathbf{x}}\big|_{\hat{\mathbf{x}}_{k|k-1}}$ is the Jacobian matrix of the observation function h calculated around the a priori state estimate $\hat{\mathbf{x}}_{k|k-1}$. In general, the performance of the KF depends on how well the system is modeled and the initial parameters chosen, so a survey of some models inside the scope of this research is presented in section III.

III. EKF Tracking with One Tag

The first step in designing a good EKF is to formulate a state model that appropriately describes the system dynamics [10]. In particular, this research is focused on indoor localization in a 3D space, so two basic models are studied, P and PV models, which are presented in the following subsections.

A. P model

In this model the state vector includes position coordinates only:

$$\mathbf{x} = [\boldsymbol{x}_1] = [x_1, y_1, z_1]^T, \tag{10}$$

where $\boldsymbol{x}_1$ represents the 3D coordinates of a tag. This model assumes near constant position between two consecutive estimation. In fact, the state transition function f is an identity matrix and the process noise vector can be modelled with small white noise speeds:

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}) = \boldsymbol{F}_k \mathbf{x}_{k-1} = \boldsymbol{I}_{3,3} \mathbf{x}_{k-1}, \tag{11}$$

$$\mathbf{Q}_k = [\Delta_T \boldsymbol{I}_{3,3}][\Delta_T \boldsymbol{I}_{3,3}]^T \sigma_v^2, \tag{12}$$

where $\boldsymbol{F}_k = \boldsymbol{I}_{3,3}$ is a 3×3 identity matrix. $\mathbf{Q}_k$ is also a 3×3 matrix presented as function of Δ_T, which is the time elapsed between two consecutive measurements, and σ_v, which is the standard deviation of the random speed components of the process noise.

The $\mathbf{Q}$ matrix has a significant impact on the tracking performance. For instance, a lower σ_v value will smooth the output obtained at the cost of larger convergence time, while faster filters usually provide noisy outputs. Thus, this value is a key factor in the KF performance optimization.

B. PV model

It is a dynamic EKF and assumes near constant velocity between the estimation intervals Δ_T. Again the process noise matrix is a key factor for the performance. In particular, smooth tracking will be achieved if we consider small white accelerations. However, on non-linear trajectory, the velocity is no longer constant, thus, slow response or even divergence will be appreciated. The state vector is expressed as:

$$\mathbf{x} = [\boldsymbol{x}_1, \boldsymbol{v}_1] = [x_1, y_1, z_1, v_{x_1}, v_{y_1}, v_{z_1}]^T, \quad (13)$$

where $\boldsymbol{v}_1 = [v_{x_1}, v_{y_1}, v_{z_1}]$ represents the 3D velocity components of a tag.

For this model, the state transition function f and the $\mathbf{Q}$ matrix of the process noise vector can be written as:

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}) = \boldsymbol{F}_k \mathbf{x}_{k-1} = \begin{bmatrix} \boldsymbol{I}_{3,3} & \Delta_T \boldsymbol{I}_{3,3} \\ \boldsymbol{0}_{3,3} & \boldsymbol{I}_{3,3} \end{bmatrix} \mathbf{x}_{k-1}, \quad (14)$$

$$\mathbf{Q}_k = \begin{bmatrix} 1/2\Delta_T^2 \boldsymbol{I}_{3,3} \\ \Delta_T \boldsymbol{I}_{3,3} \end{bmatrix} \begin{bmatrix} 1/2\Delta_T^2 \boldsymbol{I}_{3,3} \\ \Delta_T \boldsymbol{I}_{3,3} \end{bmatrix}^T \sigma_a^2, \quad (15)$$

where $\boldsymbol{0}_{3,3}$ is a 3×3 null matrix.

The PV model is feasible in low acceleration movement scenarios, *i.e.*, in almost constant speed circumstances. Thus, the process noise is modeled as an independent random acceleration normally distributed with zero mean and covariance matrix $\mathbf{Q}_k$, which allows to track different forces that could temporally affect target's dynamics [10]. σ_a is the standard deviation of the random acceleration components of the process noise. This value is a key factor in the EKF performance optimization.

C. Measurement Model

Most of indoor positioning systems are based on the UWB technology that provides distance measurements between tags and anchors.

Let $\mathcal{M}_{1,k}$ be the set of anchors visible by Tag 1 at time k, and $|\mathcal{M}_{1,k}| = M_1$ be the corresponding number of visible anchors. The observation vector can be written as:

$$\mathbf{z}_k = \mathbf{z}_{1,k} \triangleq \{z_{T_1,A_j,k} | A_j \in \mathcal{M}_{1,k}\}^T, \quad (16)$$

where $z_{T_1,A_j,k}$ is the distance measurement between Tag 1 and Anchor j at time k. Typically, UWB distance measurements are modelled with an additive Gaussian noise:

$$z_{T_i,A_j,k} = d_{T_i,A_j} + n_d, \quad (17)$$

where d_{T_i,A_j} is the exact distance between tag i and anchor j, $n_d \sim \mathcal{N}(0, \sigma_d^2)$ and σ_d is the standard deviation of the ranging noise. Depending on the UWB platform, typically, in Line of Sight (LoS) condition, σ_d has a value of about 5-10 cm.

Given the observation vector defined in equation (16), the corresponding observation function h can be written as:

$$h(\mathbf{x}_k) = h_1(\boldsymbol{x}_{1,k}) = \begin{bmatrix} \text{dist}(\boldsymbol{x}_{1,k}, \boldsymbol{x}_{A_1}) \\ \text{dist}(\boldsymbol{x}_{1,k}, \boldsymbol{x}_{A_2}) \\ \vdots \\ \text{dist}(\boldsymbol{x}_{1,k}, \boldsymbol{x}_{A_{M_1}}) \end{bmatrix}, \quad (18)$$

where dist$(\cdot)$ is the Euclidean distance, which is defined as follows:

$$\text{dist}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}. \quad (19)$$

As it can be observed, the h function is non-linear with respect to the state vector. Thus, some linearizations have to be performed to compute the $\mathbf{H}_k$ matrix as the Jacobian matrix defined in section II.

For the P model, $\mathbf{H}_k = \mathbf{H}_{1,k}$ where:

$$\mathbf{H}_{1,k} = \begin{bmatrix} \frac{\hat{x}_{1,k|k-1} - x_{A_1}}{\hat{d}_{T_1,A_1,k}} & \frac{\hat{y}_{1,k|k-1} - y_{A_1}}{\hat{d}_{T_1,A_1,k}} & \frac{\hat{z}_{1,k|k-1} - z_{A_1}}{\hat{d}_{T_1,A_1,k}} \\ \frac{\hat{x}_{1,k|k-1} - x_{A_2}}{\hat{d}_{T_1,A_2,k}} & \frac{\hat{y}_{1,k|k-1} - y_{A_2}}{\hat{d}_{T_1,A_2,k}} & \frac{\hat{z}_{1,k|k-1} - z_{A_2}}{\hat{d}_{T_1,A_2,k}} \\ \vdots & \vdots & \vdots \\ \frac{\hat{x}_{1,k|k-1} - x_{A_{M_1}}}{\hat{d}_{T_1,A_{M_1},k}} & \frac{\hat{y}_{1,k|k-1} - y_{A_{M_1}}}{\hat{d}_{T_1,A_{M_1},k}} & \frac{\hat{z}_{1,k|k-1} - z_{A_{M_1}}}{\hat{d}_{T_1,A_{M_1},k}} \end{bmatrix}, \quad (20)$$

which has size $M_1 \times 3$, and $\hat{d}_{T_i,A_j,k} = \text{dist}(\hat{\boldsymbol{x}}_{i,k|k-1}, \boldsymbol{x}_{A_j})$.

For the PV model, the $\mathbf{H}_k$ matrix is given by:

$$\mathbf{H}_k = \begin{bmatrix} \mathbf{H}_{1,k} & \boldsymbol{0}_{M_1,3} \end{bmatrix}, \quad (21)$$

where $\boldsymbol{0}_{M_1,3}$ is a $M_1 \times 3$ null vector. Thus, the size of $\mathbf{H}_k$ for the PV model is $M_1 \times 6$.

Finally, for both P and PV models, the $\mathbf{R}_k$ matrix is a $M_1 \times M_1$ diagonal matrix with all the diagonal elements equal to σ_d^2.

IV. EKF Tracking with Two Tags

In this section, two independent tags worn by the presenter are considered for the tracking problem.

A. P model

For the P model, the state vector for 2 tags would be:

$$\mathbf{x} = [\boldsymbol{x}_1, \boldsymbol{x}_2] = [x_1, y_1, z_1, x_2, y_2, z_2]^T, \quad (22)$$

where sub-indexes 1 and 2 refer to tags 1 and 2, respectively. Similar to the P model for one tag, reported in section III-A, the state equation would be:

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}) = \boldsymbol{F}_k \mathbf{x}_{k-1} = \boldsymbol{I}_{6,6} \mathbf{x}_{k-1}, \quad (23)$$

$$\mathbf{Q}_k = [\Delta_T \boldsymbol{I}_{6,6}][\Delta_T \boldsymbol{I}_{6,6}]^T \sigma_v^2, \quad (24)$$

where $\boldsymbol{F}_k = \boldsymbol{I}_{6,6}$ is a 6×6 identity matrix.

B. PV model

For the PV model, the state vector for 2 tags would be:

$$\begin{aligned} \mathbf{x} &= [\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{v}_1, \boldsymbol{v}_2] \\ &= [x_1, y_1, z_1, x_2, y_2, z_2, v_{x_1}, v_{y_1}, v_{z_1}, v_{x_2}, v_{y_2}, v_{z_2}]^T. \end{aligned} \quad (25)$$

Similar to the PV model for one tag, reported in section III-B, the state equation would be:

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}) = \boldsymbol{F}_k \mathbf{x}_{k-1} = \begin{bmatrix} \boldsymbol{I}_{6,6} & \Delta_T \boldsymbol{I}_{6,6} \\ \boldsymbol{0}_{6,6} & \boldsymbol{I}_{6,6} \end{bmatrix} \mathbf{x}_{k-1}, \quad (26)$$

$$\mathbf{Q}_k = \begin{bmatrix} 1/2\Delta_T^2 \boldsymbol{I}_{6,6} \\ \Delta_T \boldsymbol{I}_{6,6} \end{bmatrix} \begin{bmatrix} 1/2\Delta_T^2 \boldsymbol{I}_{6,6} \\ \Delta_T \boldsymbol{I}_{6,6} \end{bmatrix}^T \sigma_a^2. \quad (27)$$

C. Measurement Model

Similar to the case with one tag, let $\mathcal{M}_{2,k}$ be set of anchors visible by Tag 2 at time k and $|\mathcal{M}_{2,k}| = M_2$ be the corresponding number of visible anchors. Thus, the observation vector for Tag 2 can be defined as:

$$\mathbf{z}_{k,2} \triangleq \{z_{T_2,A_j,k} | A_j \in \mathcal{M}_{2,k}\}^T. \tag{28}$$

Thus, the global observation vector can be written as:

$$\mathbf{z}_k = [\mathbf{z}_{1,k}, \mathbf{z}_{2,k}]. \tag{29}$$

Similar to tag 1, we can define the observation function for tag 2 as:

$$h_2(\boldsymbol{x}_{2,k}) = \begin{bmatrix} \text{dist}(\boldsymbol{x}_{2,k}, \boldsymbol{x}_{A_1}) \\ \text{dist}(\boldsymbol{x}_{2,k}, \boldsymbol{x}_{A_2}) \\ \vdots \\ \text{dist}(\boldsymbol{x}_{2,k}, \boldsymbol{x}_{A_{M_2}}) \end{bmatrix}. \tag{30}$$

Thus, the global observation function h can be written as:

$$h(\mathbf{x}_k) = \begin{bmatrix} h_1(\boldsymbol{x}_{1,k}) \\ h_2(\boldsymbol{x}_{2,k}) \end{bmatrix}. \tag{31}$$

Note that, given the set of visible anchors for Tag 1, denoted as $\mathcal{M}_{1,k} = \{A_1, A_2, \ldots, A_{M_1}\}$, and the set of visible anchors for Tag 2, denoted as $\mathcal{M}_{2,k} = \{A_1, A_2, \ldots, A_{M_2}\}$, there is not necessary a one to one correspondence between the two sets. For the sake of clarity, let's suppose to have a network of three anchors, labeled as B_{12}, B_{15} and B_{20}. At time k, Tag 1 could have anchors connectivity $\mathcal{M}_{1,k} = \{A_1, A_2\} = \{B_{15}, B_{20}\}$ while Tag 2 could have connectivity $\mathcal{M}_{2,k} = \{A_1, A_2, A_3\} = \{B_{12}, B_{15}, B_{20}\}$. As it can be observed, in this particular case, $A_1 \in \mathcal{M}_{1,k}$ does not coincide with $A_1 \in \mathcal{M}_{2,k}$. The same applies for A_2.

Similar to the definition of $\mathbf{H}_{1,k}$ in equation (20), it is possible to define $\mathbf{H}_{2,k}$ for Tag 2 as:

$$\mathbf{H}_{2,k} = \begin{bmatrix} \frac{\hat{x}_{2,k|k-1} - x_{A_1}}{\hat{d}_{T_2,A_1,k}} & \frac{\hat{y}_{2,k|k-1} - y_{A_1}}{\hat{d}_{T_2,A_1,k}} & \frac{\hat{z}_{2,k|k-1} - z_{A_1}}{\hat{d}_{T_2,A_1,k}} \\ \frac{\hat{x}_{2,k|k-1} - x_{A_2}}{\hat{d}_{T_2,A_2,k}} & \frac{\hat{y}_{2,k|k-1} - y_{A_2}}{\hat{d}_{T_2,A_2,k}} & \frac{\hat{z}_{2,k|k-1} - z_{A_2}}{\hat{d}_{T_2,A_2,k}} \\ \vdots & \vdots & \vdots \\ \frac{\hat{x}_{2,k|k-1} - x_{A_{M_2}}}{\hat{d}_{T_2,A_{M_2},k}} & \frac{\hat{y}_{2,k|k-1} - y_{A_{M_2}}}{\hat{d}_{T_2,A_{M_2},k}} & \frac{\hat{z}_{2,k|k-1} - z_{A_{M_2}}}{\hat{d}_{T_2,A_{M_2},k}} \end{bmatrix}, \tag{32}$$

which has size $M_2 \times 3$.

For the P model with two tags, the global $\mathbf{H}_k$ matrix is given by:

$$\mathbf{H}_k = \begin{bmatrix} \mathbf{H}_{1,k} & \boldsymbol{0}_{M_1,3} \\ \boldsymbol{0}_{M_2,3} & \mathbf{H}_{2,k} \end{bmatrix}, \tag{33}$$

which has size $(M_1 + M_2) \times 6$.

For the PV model with two tags, the global $\mathbf{H}_k$ matrix is given by:

$$\mathbf{H}_k = \begin{bmatrix} \mathbf{H}_{1,k} & \boldsymbol{0}_{M_1,3} & \boldsymbol{0}_{M_1,6} \\ \boldsymbol{0}_{M_2,3} & \mathbf{H}_{2,k} & \boldsymbol{0}_{M_2,6} \end{bmatrix}, \tag{34}$$

which has size $(M_1 + M_2) \times 12$.

For both P and PV models, the $\mathbf{R}_k$ matrix is a $(M_1+M_2) \times (M_1 + M_2)$ diagonal matrix with all diagonal elements equal to σ_d^2.

In this case, at each time step k, the presenter position can be estimated as the average value of the two tags' a posteriori state vector positions $\hat{\boldsymbol{x}}_{1,k|k}$ and $\hat{\boldsymbol{x}}_{2,k|k}$.

V. EKF Tracking with Three Tags

In this section, three independent tags worn by the presenter are considered for the tracking problem.

A. P model

For the P model, the state vector for 3 tags would be:

$$\begin{aligned} \mathbf{x} &= [\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3] \\ &= [x_1, y_1, z_1, x_2, y_2, z_2, x_3, y_3, z_3]^T, \end{aligned} \tag{35}$$

where sub-indexes 1, 2 and 3 refer to Tags 1, 2 and 3, respectively. Similarly to the P model for one tag, reported in section III-A, the state equation would be:

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}) = \boldsymbol{F}_k \mathbf{x}_{k-1} = \boldsymbol{I}_{9,9} \mathbf{x}_{k-1}, \tag{36}$$

$$\mathbf{Q}_k = [\Delta_T \boldsymbol{I}_{9,9}][\Delta_T \boldsymbol{I}_{9,9}]^T \sigma_v^2, \tag{37}$$

where $\boldsymbol{F}_k = \boldsymbol{I}_{9,9}$ is a 9×9 identity matrix.

B. PV model

For the PV model, the state vector for 3 tags would be:

$$\begin{aligned} \mathbf{x} =& [\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3] \\ =& [x_1, y_1, z_1, x_2, y_2, z_2, z_1, x_3, y_3, z_3, \\ & v_{x_1}, v_{y_1}, v_{z_1}, v_{x_2}, v_{y_2}, v_{z_2}, v_{x_3}, v_{y_3}, v_{z_3}]^T. \end{aligned} \tag{38}$$

Similar to the PV model for one tag, reported in section III-B, the state equation would be:

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}) = \boldsymbol{F}_k \mathbf{x}_{k-1} = \begin{bmatrix} \boldsymbol{I}_{9,9} & \Delta_T \boldsymbol{I}_{9,9} \\ \boldsymbol{0}_{9,9} & \boldsymbol{I}_{9,9} \end{bmatrix} \mathbf{x}_{k-1}, \tag{39}$$

$$\mathbf{Q}_k = \begin{bmatrix} 1/2\Delta_T^2 \boldsymbol{I}_{9,9} \\ \Delta_T \boldsymbol{I}_{9,9} \end{bmatrix} \begin{bmatrix} 1/2\Delta_T^2 \boldsymbol{I}_{9,9} \\ \Delta_T \boldsymbol{I}_{9,9} \end{bmatrix}^T \sigma_a^2. \tag{40}$$

C. Measurement Model

Similar to the case with two tags, let $\mathcal{M}_{3,k}$ be set of anchors visible by Tag 3 at time k and $|\mathcal{M}_{3,k}| = M_3$ be number of visible anchors. Thus, similarly, the observation vector for Tag 3 can be defined as:

$$\mathbf{z}_{k,3} \triangleq \{z_{T_3,A_j,k} | A_j \in \mathcal{M}_{3,k}\}^T. \tag{41}$$

The global observation vector can be written as:

$$\mathbf{z}_k = [\mathbf{z}_{1,k}, \mathbf{z}_{2,k}, \mathbf{z}_{3,k}]. \tag{42}$$

The global observation function can be written as:

$$h(\mathbf{x}_k) = \begin{bmatrix} h_1(\boldsymbol{x}_{1,k}) \\ h_2(\boldsymbol{x}_{2,k}) \\ h_3(\boldsymbol{x}_{3,k}) \end{bmatrix}, \tag{43}$$

where

$$h_3(\boldsymbol{x}_{3,k}) = \begin{bmatrix} \text{dist}(\boldsymbol{x}_{3,k}, \boldsymbol{x}_{A_1}) \\ \text{dist}(\boldsymbol{x}_{3,k}, \boldsymbol{x}_{A_2}) \\ \vdots \\ \text{dist}(\boldsymbol{x}_{3,k}, \boldsymbol{x}_{A_{M_3}}) \end{bmatrix} \tag{44}$$

Similar to the definition of $\mathbf{H}_{1,k}$ in equation (20), it is possible to define $\mathbf{H}_{3,k}$ for Tag 3 as:

$$\mathbf{H}_{3,k} = \begin{bmatrix} \frac{\hat{x}_{3,k|k-1} - x_{A_1}}{\hat{d}_{T_3,A_1,k}} & \frac{\hat{y}_{3,k|k-1} - y_{A_1}}{\hat{d}_{T_3,A_1,k}} & \frac{\hat{z}_{3,k|k-1} - z_{A_1}}{\hat{d}_{T_3,A_1,k}} \\ \frac{\hat{x}_{3,k|k-1} - x_{A_2}}{\hat{d}_{T_3,A_2,k}} & \frac{\hat{y}_{3,k|k-1} - y_{A_2}}{\hat{d}_{T_3,A_2,k}} & \frac{\hat{z}_{3,k|k-1} - z_{A_2}}{\hat{d}_{T_3,A_2,k}} \\ \vdots & \vdots & \vdots \\ \frac{\hat{x}_{3,k|k-1} - x_{A_{M_3}}}{\hat{d}_{T_3,A_{M_3},k}} & \frac{\hat{y}_{3,k|k-1} - y_{A_{M_3}}}{\hat{d}_{T_3,A_{M_3},k}} & \frac{\hat{z}_{3,k|k-1} - z_{A_{M_3}}}{\hat{d}_{T_3,A_{M_3},k}} \end{bmatrix}, \tag{45}$$

which has size $M_3 \times 3$.

For the P model with three tags, the global $\mathbf{H}_k$ matrix is given by:

$$\mathbf{H}_k = \begin{bmatrix} \mathbf{H}_{1,k} & \boldsymbol{0}_{M_1,3} & \boldsymbol{0}_{M_1,3} \\ \boldsymbol{0}_{M_2,3} & \mathbf{H}_{2,k} & \boldsymbol{0}_{M_2,3} \\ \boldsymbol{0}_{M_3,3} & \boldsymbol{0}_{M_3,3} & \mathbf{H}_{3,k} \end{bmatrix}, \tag{46}$$

which has size $(M_1 + M_2 + M_3) \times 6$.

For the PV model with three tags, the global $\mathbf{H}_k$ matrix is given by:

$$\mathbf{H}_k = \begin{bmatrix} \mathbf{H}_{1,k} & \boldsymbol{0}_{M_1,3} & \boldsymbol{0}_{M_1,3} & \boldsymbol{0}_{M_1,9} \\ \boldsymbol{0}_{M_2,3} & \mathbf{H}_{2,k} & \boldsymbol{0}_{M_2,3} & \boldsymbol{0}_{M_2,9} \\ \boldsymbol{0}_{M_3,3} & \boldsymbol{0}_{M_3,3} & \mathbf{H}_{3,k} & \boldsymbol{0}_{M_3,9} \end{bmatrix}, \tag{47}$$

which has size $(M_1 + M_2 + M_3) \times 18$.

For both P and PV models, the $\mathbf{R}_k$ matrix is a $(M_1 + M_2 + M_3) \times (M_1 + M_2 + M_3)$ diagonal matrix with all diagonal elements equal to σ_d^2.

In this case, at each time step k, the presenter position can be estimated as the average value of the three tags' a posteriori state vector positions $\hat{\boldsymbol{x}}_{1,k|k}$, $\hat{\boldsymbol{x}}_{2,k|k}$ and $\hat{\boldsymbol{x}}_{3,k|k}$.

VI. Detection and Correction of Distance Outliers

Using two or more tags in a tracking system allows to implement a simple rule for the detection of distance measurement outliers. Let's suppose to have two tags, T_1 and T_2, placed at a fixed distance $d_{T_1T_2}$, for instance, on the shoulders of a presenter to be localized, as depicted in the Fig. 1. The figure shows graphically the two noisy distance measurements with respect to anchor A_5. In particular, the $\tilde{d}_{T_1A_5}$, represented by the red line, is affected by a very large noise, for instance, due to the synchronization of the receiver to a reflected path. First of all, the outlier detection and correction procedure, selects the minimum distance measured by the two tags with respect to a specific anchor. In the example of Fig. 1, the minimum measured distance corresponds to $\tilde{d}_{T_2A_5}$, represented by the green line. With a good likelihood, we can assume that the minimum distance is not an outlier, *i.e.*, not affected by large noise. After that, a threshold denoted as d_{thr}, will be defined as the sum of the minimum distance and the fixed distance

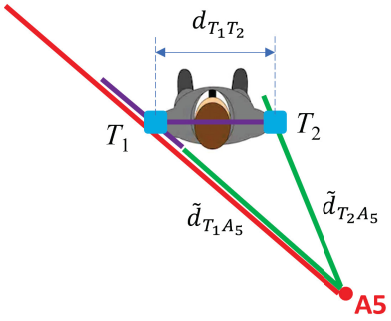

Fig. 1. Outlier detection with two tags

between the two tags, that is, $d_{\text{thr}} = \tilde{d}_{T_2A_5} + d_{T_1T_2}$. Finally, for the outlier detection, it will applied the triangles' rule stating that one side of a triangle cannot be larger than the sum of the other two. Thus, if the largest distance measurement, in the example $\tilde{d}_{T_1A_5}$, is larger than the threshold d_{thr}, as depicted in the Fig. 1, it will be considered as an outlier and corrected to d_{thr} plus some little noise, which depends on the UWB technology. Alternatively, if the maximum distance is less or equal than the threshold, nothing happens and the measurement will not be modified.

The above defined approach could be applied also in a scenario with three tags. Considering that, only pair of distances are processed at a time, the procedure for detection and correction will be applied three times.

VII. Evaluation of Localization Performance and Optimization

The designed EKF localization algorithms have been tested off-line via MATLAB but using real range measurements collected, using a UWB kit, in a RAI TV studio. The TV studio, as shown in Fig. 2, has a circular shape with a diameter of about 17 meters, and structured in three levels of bleachers.

Fig. 2. A picture of the RAI TV studio

For the measurement campaign, a UWB kit based on the DW1000 ScenSor integrated circuit [11] from Decawave have been used. In particular, as shown in Fig. 3, a total of eight anchors have been deployed (red dots), of which four anchors at a height of about 4 m, close to the external circle of the studio, while the other four anchors at a height of about 6.5 m, in correspondence of the four lower anchors. In addition,

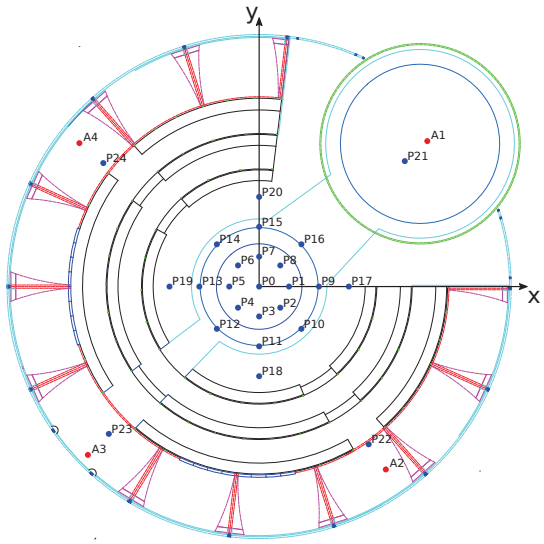

Fig. 3. Anchors and Test Points (TPs) positions

24 TPs (blue dots) have been selected for the tag's positions, where the localization performance has been evaluated.

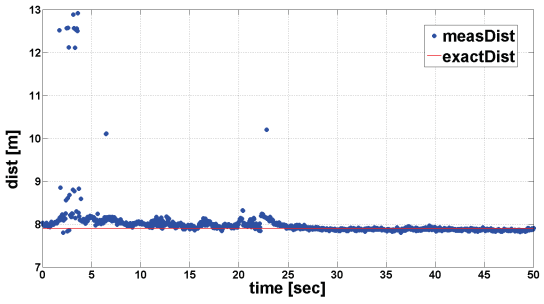

Fig. 4. Range measurements between anchor A6 (above A2) and the tag placed on the right shoulder positioned at TP P0

First of all, UWB-based range measurements have been carried out, stored and then analyzed. Firstly, we have performed measurements without the human body, for example, deploying the tag on a tripode. Considering range measurements on TPs P0, P12, P16, P21 and P23, globally, in this scenario, a ranging error with average zero and standard deviation equal to 14 cm have been achieved. After that, we have performed measurements with the human body. As an example, Fig. 4 shows graphically some range measurements between anchor A6 (deployed above A2) and a tag, placed on a shoulder, positioned at TP P0. As it can be observed, some range outliers occurred. Deploying the tag on the same five TPs (reported above), globally, an average range error equal to 12.2 cm and a standard deviation equal to 63.4 cm have been registered. This larger standard deviation value, 63.4cm, with respect to the one Without the human body, 14 cm, proofs the huge presence of range outliers in the measurements.

For comparison, we have collected range measurements also with the tag placed on the belt and inside a pocket as shown in Fig. 5.

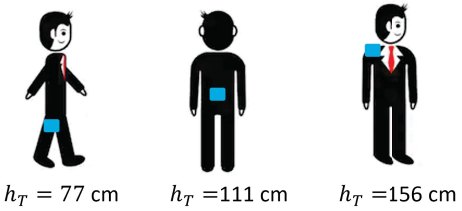

Fig. 5. Three different positions of the tag on the presenter

The resulting range performance have been listed in Table I. As it can be observed, although the tag placed on the shoulder and on the belt show similar range performance, the position on the shoulder is preferable as the range availability is the largest. In contrast, the tag in the pocket shows the worst range performance and the lowest range availability.

TABLE I. RANGING PERFORMANCE

	Ranging Error			
Tag Posit.	**avg [m]**	**std [m]**	**RMS [m]**	**Range Avail. [%.]**
Pocket	0.45	0.66	0.80	5.0
Belt	0.28	0.50	0.57	88.6
Shoulder	0.12	0.55	0.56	95.7

After that, we have evaluated localization performance with the tag placed on the shoulder using the three designed EKF algorithms, with 1 Tag, 2 Tags and 3 Tags. In the ranging phase of each algorithm, we have included the procedure for outlier detection and correction presented in section VI. Moreover, each EKF has been initialized by a simple Linear Least Squared (LLS) algorithm [12].

For the evaluation of the localization performance, we have calculated the following errors:

- **2DLocErr**: defined as the Euclidean distance between the exact Tag's position and the estimated one in the x, y plane.
- **3DLocErr**: defined as the Euclidean distance between the exact Tag's position and the estimated one in the x, y, z plane.
- **XErr**: defined as the difference between the exact Tag's x-coordinate and the estimated one.
- **YErr**: defined as the difference between the exact Tag's y-coordinate and the estimated one.
- **ZErr**: defined as the difference between the exact Tag's z-coordinate and the estimated one.

After that, for each set of the defined errors, the following statistics have been calculated: **avg**, defined as the average value, **std**, defined as the standard deviation value, and finally, the **RMS**, which is the root mean squared value.

For the evaluation of the localization performance, we have considered only the static scenarios, as the exact tags' positions were known. However, we have executed the algorithms also in mobility scenarios but we could not evaluate the performance as we had not the means to record the ground truth along the time.

As far as the static scenarios are concerned, the performance of the EKF algorithms have been optimized by tuning the σ_v and σ_a parameters of the $\mathbf{Q}$ matrix for P and PV models, respectively.

Tables II and III list the localization performance with one tag on the shoulder for P and PV models, respectively. Although, both P and PV models show similar results, the 3DLocErr is quite large, in the order of 67 cm, which is not suitable for AR applications.

TABLE II. 1-TAG EKF ALGORITHM LOCALIZATION PERFORMANCE - P MODEL ($\sigma_v = 0.01$)

Localization Performance
2DLocErr [avg=0.272, std=0.403, RMS=0.486] m
3DLocErr [avg=0.380, std=0.548, RMS=0.667] m
XErr [avg=-0.169, std=0.353, RMS=0.392] m
YErr [avg=-0.095, std=0.272, RMS=0.288] m
ZErr [avg=-0.184, std=0.417, RMS=0.456] m

TABLE III. 1-TAG EKF ALGORITHM LOCALIZATION PERFORMANCE - PV MODEL ($\sigma_a = 1.5$)

Localization Performance
2DLocErr [avg=,0.282 std=0.440, RMS=0.523] m
3DLocErr [avg=0.386, std=0.599, RMS=0.712] m
XErr [avg=-0.174, std=0.397, RMS=0.434] m
YErr [avg=-0.097, std=0.276, RMS=0.293] m
ZErr [avg=-0.214, std=0.434, RMS=0.484] m

Tables IV and V list the localization performance with two tags on the shoulder for P and PV models, respectively. Comparing these results with the EKF 1-Tag ones, we can observe a considerable improvement of the performance (83.6%). Indeed, the 3DLocErr is in the order of 11 cm, which is suitable to enable AR applications.

TABLE IV. 2-TAG EKF ALGORITHM LOCALIZATION PERFORMANCE - P MODEL ($\sigma_v = 0.5$)

Localization Performance
2DLocErr [avg=0.085, std=0.032, RMS=0.091] m
3DLocErr [avg=0.107, std=0.038, RMS=0.107] m
XErr [avg=-0.063, std=0.043, RMS=0.076] m
YErr [avg=-0.033, std=0.037, RMS=0.049] m
ZErr [avg=-0.019, std=0.065, RMS=0.072] m

TABLE V. 2-TAG EKF ALGORITHM LOCALIZATION PERFORMANCE - PV MODEL ($\sigma_a = 0.01$).

Localization Performance
2DLocErr [avg=0.088, std=0.031, RMS=0.093] m
3DLocErr [avg=0.103, std=0.034, RMS=0.109] m
XErr [avg=-0.059, std=0.044, RMS=0.288] m
YErr [avg=-0.038, std=0.042, RMS=0.057] m
ZErr [avg=-0.010, std=0.056, RMS=0.063] m

Finally, Tables VI and VII list the localization performance with three tags on the shoulder for P and PV models, respectively. As it can be observed, the performance are similar to the EKF 2-Tag ones, sometimes a little bit worse. Therefore, it seems that a kind of saturation effect on the performance is observed using more than two tags.

TABLE VI. 3-TAG EKF ALGORITHM LOCALIZATION PERFORMANCE - P MODEL ($\sigma_v = 3.5$)

Localization Performance
2DLocErr [avg=0.112, std=0.040, RMS=0.119] m
3DLocErr [avg=0.158, std=0.082, RMS=0.134 m
XErr [avg=-0.078, std=0.065, RMS=0.101] m
YErr [avg=0.009, std=0.062, RMS=0.062] m
ZErr [avg=-0.094, std=0.094, RMS=0.094] m

TABLE VII. 3-TAG EKF ALGORITHM LOCALIZATION PERFORMANCE - PV MODEL ($\sigma_a = 3$)

Localization Performance
2DLocErr [avg=0.112, std=0.039, RMS=0.118] m
3DLocErr [avg=0.158, std=0.082, RMS=0.137] m
XErr [-avg=0.077, std=0.064, RMS=0.101] m
YErr [avg=0.009, std=0.062, RMS=0.063] m
ZErr [avg=-0.093, std=0.094, RMS=0.094] m

For a better comparison, the 3D localization performance have been listed in Table VIII.

TABLE VIII. PERFORMANCE COMPARISON WITH DIFFERENT NUMBER OF TAGS

# Tags	Model	$\sigma_{v,a}$ (param. of $\mathbf{Q}$)	3DLoc RMSE [m]
1	P	0.01	0.667
1	PV	1.5	0.731
2	P	0.5	0.107
2	PV	0.01	0.109
3	P	3.5	0.134
3	PV	3	0.137

Finally, Fig. 6 shows an example of localization in a mobility scenario, where only one tag has been used, carried on the shoulder. The performer has moved, at a walking pace, along the three levels of bleachers. In this case, to optimize both the estimated trajectory smoothness and the convergence of the estimation (mainly in the areas with high change of direction), we have used a PV model with $\sigma_a = 3$ that, as expected, resulted to be larger than to the one used for the static case (PV model, $\sigma_a = 1.5$).

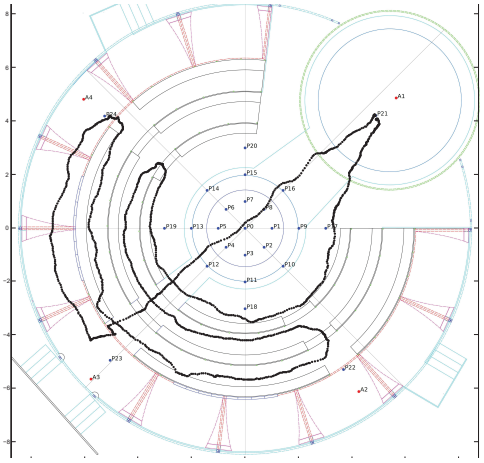

Fig. 6. Example of tracking with real UWB measurements from 1 tag (PV model, $\sigma_a = 3$)

VIII. Conclusion

This paper presented a robust RTLS based on the UWB technology suitable for real-time AR applications in TV studios. To support the design process, first of all, a UWB-based measurement campaign was performed in a TV studio, where rangings resulted to be heavily affected by the human body interference, thus, causing outliers. During the measurements, the tag was placed in three different potions of the human body, on a shoulder, on the belt, and inside a pocket. According to the range analysis, the shoulder was the best position as the ranging error was the lowest (similar to the belt position) and the range availability was the largest.

To improve the localization performance and meet the AR requirements, we adopted more than one tag to be worn by the presenter/performer on the shoulders for accurate tracking. In particular, we applied the EKF algorithm to work with two and three tags. P and PV models were considered for the state vector. In addition, exploiting the system redundancy, an outlier detection and correction procedure was defined and then applied in the ranging phase of the RTLS.

The designed EKF localization algorithms have been tested off-line via MATLAB but using the real range measurements collected in the RAI TV studio, where 8 anchors have been deployed for 3D localization. The resulting localization performance showed that the EKF with two tags outperformed the one with single tag by 83.5%. In addition, the achieved level of accuracy is suitable for applications in a TV studio. We think that this technology could foster the developments of real-time AR for TV programmes with an easy setup and a reduction in production costs.

Acknowledgment

The authors wish to thank Professor Roberto Garello for his revision of the paper.

References

[1] S. Gezici, Z. Tian, G. Giannakis, H. Kobayashi, A. Molisch, H. Poor, and Z. Sahinoglu, "Localization via ultra-wideband radios: a look at positioning aspects for future sensor networks," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 70–84, 2005.

[2] "IEEE standard for low-rate wireless networks," *IEEE Std 802.15.4-2015 (Revision of IEEE Std 802.15.4-2011)*, pp. 1–709, 2016.

[3] Q. Tian, K. I.-K. Wang, and Z. Salcic, "Human body shadowing effect on uwb-based ranging system for pedestrian tracking," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 10, pp. 4028–4037, 2019.

[4] Y. Kiliç, A. J. Ali, A. Meijerink, M. J. Bentum, and W. G. Scanlon, "The effect of human-body shadowing on indoor uwb toa-based ranging systems," in *2012 9th Workshop on Positioning, Navigation and Communication*, 2012, pp. 126–130.

[5] T. Otim, A. Bahillo, L. E. Díez, P. Lopez-Iturri, and F. Falcone, "Fdtd and empirical exploration of human body and uwb radiation interaction on tof ranging," *IEEE Antennas and Wireless Propagation Letters*, vol. 18, no. 6, pp. 1119–1123, 2019.

[6] ——, "Impact of body wearable sensor positions on uwb ranging," *IEEE Sensors Journal*, vol. 19, no. 23, pp. 11 449–11 457, 2019.

[7] S. Maranò, W. M. Gifford, H. Wymeersch, and M. Z. Win, "Nlos identification and mitigation for localization based on uwb experimental data," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 7, pp. 1026–1035, 2010.

[8] X. Yang, "Nlos mitigation for uwb localization based on sparse pseudo-input gaussian process," *IEEE Sensors Journal*, vol. 18, no. 10, pp. 4311–4316, 2018.

[9] G. F. Welch and G. Bishop, "An introduction to the kalman filter," University of North Carolina, Chapel Hill, NC, USA, Tech. Rep., 1995.

[10] M. A. Caceres, F. Sottile, and M. A. Spirito, "Adaptive location tracking by kalman filter in wireless sensor networks," in *2009 IEEE International Conference on Wireless and Mobile Computing, Networking and Communications*, 2009, pp. 123–128.

[11] Decawave ScenSor, "DW1000, 2021. [online] available: www.decawave.com."

[12] I. Guvenc, S. Gezici, and Z. Sahinoglu, "Fundamental limits and improved algorithms for linear least-squares wireless position estimation," *Wirel. Commun. Mob. Comput.*, vol. 12, no. 12, p. 1037–1052, Aug. 2012.