A Deep Learning Approach for Robust Target Tracking in a Cluttered Environment

Jaeuk Baek†, SungHo Lee*†, Changeun Lee† and Sangjoon Park†
†Intelligent Convergence Research Laboratory, ETRI, Daejeon 34129, Korea
Email: {jubaek, lsh31973, celee, sangjoon}@etri.re.kr
*Information Technology Management, Seoul National University of Science&Technology, Gongneung-ro 232, Seoul, Korea
Email: sean03101@seoultech.ac.kr

2021 International Conference on Information and Communication Technology Convergence (ICTC) | 978-1-6654-2383-0/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICTC52510.2021.9620770

***Abstract*—In a challenging environment with severe measurement errors, one of the most important factor to improve target tracking accuracy is to understand the channel characteristics of the environment and to estimate the target position in data-oriented manner. In this paper, we propose a deep learning model to estimate target position by processing time difference of arrival (TDOA) measurements in an end-to-end manner, especially for cluttered environments. The proposed TDOA image based target tracking (TITT) model converts each TDOA measurement into TDOA image, then a mask is created to handle TDOA errors. Then, by accumulating and treating TDOA images with masks as input data, we train a CNN model with fully connected layer to estimate target position through back propagation. Simulation results demonstrate that our proposed TITT model outperforms the simple deep learning model in challenging environment with many clutters.**

***Index Terms*—TDOA; Deep Learning; Real time localization; Target tracking; Radar system**

I. Introduction

In recent years, many studies have demonstrated that deep learning outperforms traditional algorithms in terms of classification, prediction, and inference. For example, [1] proposes a deep learning model to estimate the movement of human in real time, which has been treated as very difficult problem to solve with traditional algorithms. In the field of computer vision, [2] and [3] utilize generative adversarial network (GAN) to generate large scale data sets and classify images with high accuracy. Based on these recent advance in deep learning architecture, radar systems try to replace the traditional target tracking process (e.g., (i) preprocess of radar measurements, (ii) estimation of target position for each timestamp, (iii) correction of time-series estimation results by kalman filter [4]) with end-to-end deep learning models because the former is subject to performance degradation in challenging environment with a lot of clutters (obstacles).

In radar systems, time Difference of Arrival (TDOA) is one of the most actively used measurements, which can be obtained every few milliseconds [5]–[8]. TDOA is defined as the difference of time of arrival (TOA) at multiple sensors, and can eliminate the need for measurement of global time. Also, it can reduce the effect of clock bias of each sensor on estimation. However, in order to obtain the estimation results from TDOA measurements, complex non-linear equations need to be solved, but this is challenging because the closed-form solution does not exist in general. In addition, due to a time-subtraction operation to make TDOAs, high uncertainty is observed in the value and such TDOA errors severely degrade the estimation accuracy.

In order to relieve inherent shortcomings of TDOA measurements, deep learning architecture has been proposed to estimate target position in data-oriented aspects. [9] corrects abnormalities of TDOA measurement by learning channel impulse response (CIR) and [10] proposes a DeepTAL to eliminate TDOA error through asynchronous localization algorithm. These prior studies try to increase the reliability of TDOA measurements, but do not apply deep learning architecture to the entire target tracking process. Also, it is still difficult to estimate target position in challenging environments with severe TDOA errors where none line of sight (NLOS) path by clutters dominates ling of sight (LOS) path between sensors.

In this paper, we propose a deep learning model to estimate target position in an end-to-end manner. Specifically, the proposed TDOA image based target tracking (TITT) model converts each TDOA measurement into TDOA image, then mask is created to reduce the effect of TDOA errors on estimation results. Then, we train a CNN model with fully connected layer to estimate target position through back propagation. Simulation results demonstrate that our proposed TITT model outperforms the simple deep learning model in challenging scenarios. The main contributions of TITT model are as follows.

- We construct TDOA dataset composed of pure data and data with errors, where TDOA errors are determined by the performance of Decawave DW1000 chips in real environment [11].
- The proposed TITT model is composed of three steps: (i) measurement to image conversion (ii) mask generation (ii) CNN model for estimation. Note that the image size of TDOA measurement is hyperparameter and can be determined by the desired estimation accuracy. Also, the mask is generated to each image to have information on TDOA errors, so it can deal with TDOA errors and is critical to achieve high estimation accuracy in challenging environment.

The paper is organized as follows. Sec. II presents our proposed deep learning model to estimate target position in an end-to-end manner. Sec. III provides simulation results to

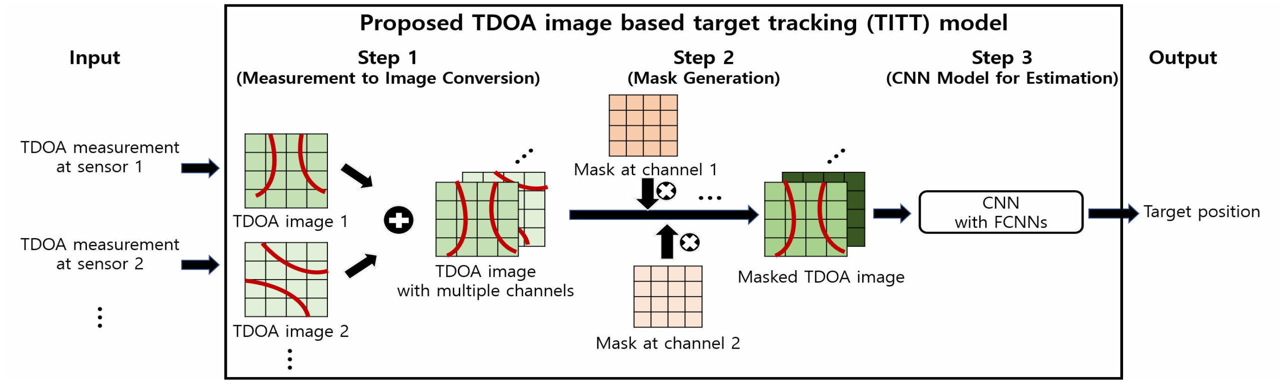


Fig. 1. System architecture for our propose TDOA image based target tracking (TITT) model.

validate the performance of the TITT model. Finally, Sec. IV concludes this paper.

II. The Proposed TITT Model

A. System Models

We consider radar systems with multiple sensors to estimate a single target position. We assume that sensors are deployed at the boundary of surveillance area, and one of them is selected as the reference sensor. Based on TOA at each sensor, TDOA is computed in central computing unit (CCU) as the difference of TOAs between the reference and remaining sensors, and the measurement error is modelled as white Gaussian noise with variance σ. Note that TDOA measurements are obtained for each several millisecond and the delay of communication between each sensor and CCU is neglected.

In conventional estimation methods (e.g., least square (LS) method [12]), more than three TDOAs need to be combined and processed to estimate two dimensional (2D) position of targets. However, we consider a challenging scenario with cluttered environments, where some of TDOA measurements are severely noisy or not obtained, resulting that less than three TDOAs are obtained at some timestamp.

B. Overall Procedures of TITT model

We propose a TITT model to provide robust target tracking results in cluttered environment. Fig. 1 represents the overall procedure of TITT model, which is composed of three steps; (i) measurement to image conversion (ii) mask generation (ii) CNN model for estimation. When TDOA measurements are obtained at CCU, each TDOA measurement is converted into image. Then, a mask is created to deal with measurement errors, where the value of mask can be determined by human (based on observation of environments) or deep learning architecture. By combining and treating TDOA images with masks as input data, we train a CNN model with fully connected layer to estimate target position. Details of each step are illustrated in the following subsections.

C. Step1: Measurement to Image Conversion

In the field of computer vision, convolution filed is utilized to extract correlation information about pixels of multiple

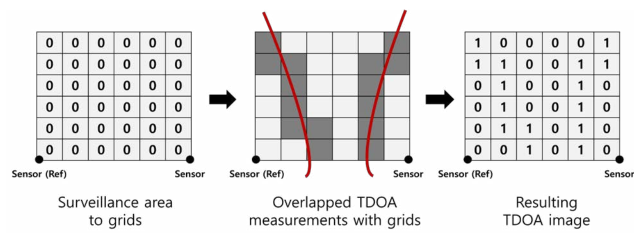


Fig. 2. Example of TDOA measurement to image conversion.

channels, which is the key of image classification. Motivated by this approach, we convert each TDOA measurement into image of single channel. Then, by stack multiple images of single channel into *TDOA image* with multiple channels, we can analyze correlation of TDOA measurements through convolution filter, which is critical in target tracking.

Fig. 2 illustrates how to create TDOA image for given TDOA measurement. We divide the surveillance area into grids [13], then TDOA measurement is overlapped with each grid. Since the overlapped grid can be the possible position of target, we assign a value one to the overlapped grid, and otherwise zero. Note that the value assigned to the grid can be designed more specifically depending on the measurement errors, which is our future work.

D. Step2: Mask Generation

Based on TDOA images obtained in previous subsection, we can train a CNN model with fully connected layer to estimate target position. However, in cluttered environment with severe measurement errors, each TDOA measurement cannot be reliable and the corresponding TDOA image can be erroneous. To deal with this problem, we create and apply a mask to each channel of TDOA images.

By adopting zero or one as a value of mask, we can apply them to each pixel of TDOA images, where mask with value one can be used for TDOA measurement with less error. Although deep learning model can learn the value of mask by itself through analyzing correlation of TDOA measurements, we simply adopt the fixed value of mask, and more studies will be conducted in our future work.

TABLE I

DISTANCE DIFFERENCE (IN m) OF PREDICTED AND TRUE TARGET POSITIONS IS ADOPTED AS PERFORMANCE METRIC.

	Mean	Std	Min	25%	50%	75%	Max
FCNN	0.6266	0.7843	0.0073	0.2481	0.428	0.613	7.966
TITT	0.5821	0.3867	0.0021	0.3004	0.4943	0.7671	2.6014

E. Step3: CNN Model for Estimation

We design a deep learning model to estimate target position, where three CNN layers are constructed to analyze the input data (i.e., TDOA images), followed by two fully connected layers for estimation. The kernel size of all CNN layers is the same as two with stride one. After each CNN layer, we apply max pooling with kernel size two, and Relu function [14] is used as activation function.

III. SIMULATION RESULTS

A. Dataset

In order to generate TDOA dataset, we conduct 40,000 simulations with four sensors, where a single target is randomly deployed for each iteration over surveillance area $[0, 25] \times [0, 25]$ (in m). TDOA error is modelled based on the performance of Decawave DW1000 chips in real environment [11]. Also, in order to model a cluttered environment, we assume that only 70 $\%$ of sensors can obtain ToA without any distortions. As for setups of our proposed TITT model, grid size is 0.5 (m) in steps 1 and 2. For training a deep learning model in step 3, we divided TDOA dataset into 80% training set and 20% test set.

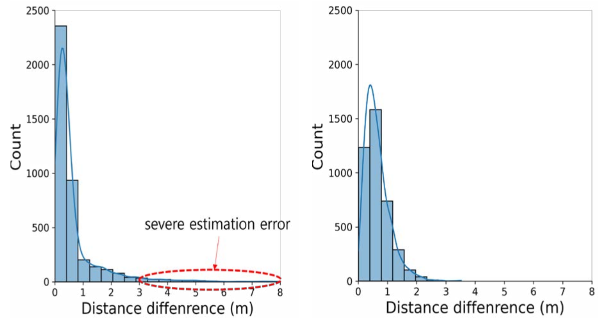
B. Performance Evaluation

To evaluate performance of our proposed TITT model, we adopt a fully connected neural network (FCNN) as a baseline for comparison. Table I illustrates performance comparison of algorithms, where distance difference of predicted and true target positions is adopted as performance metric. It is observed from Table I that our proposed TITT model has a lower mean distance difference compared to the FCNN model. Also, by observing max distance difference, it can be concluded that the FCNN model cannot always provide accurate estimates, but not in our proposed TITT model.

Fig. 3 represents distribution of distance difference of predicted and true target positions. It is observed from Fig. 3 that our TITT model can provide robust target tracking results compared to the FCNN model in cluttered environment.

IV. CONCLUSION

In this paper, we have proposed a deep learning model to estimate target position by processing TDOA measurements in an end-to-end manner. To tackle challenging scenarios with a lot of clutters, we have proposed a TDOA image based target tracking (TITT) model composed of three important steps. Simulation results demonstrate that our proposed TITT model outperforms the simple deep learning model. In our future work, we will extend our proposed TITT model to consider multiple types of measurements for hybrid target tracking process.



(a) FCNN model (b) Our TITT model

Fig. 3. Distribution of distance difference of predicted and true target positions.

ACKNOWLEDGMENT

This work was supported by the ICT R&D program of MSIP/IITP. [2020-0-00114, Problem Solving in Military Camp Life Safety and Reliability of Human Resources Management]

REFERENCES

[1] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7291–7299.

[2] D. Castelvecchi, "Astronomers explore uses for AI-generated images," *Nature News*, vol. 542, no. 7639, p. 16, 2017.

[3] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[4] J. Baek, D. Kim, H. Shim, and Y. Han, "Improved target-tracking process in PCL," in *MILCOM 2017 - 2017 IEEE Military Communications Conference (MILCOM)*, 2017, pp. 513–518.

[5] J. F. Valente and J. C. Alves, "Real-time TDOA measurements of an underwater acoustic source," in *OCEANS 2016 MTS/IEEE Monterey*. IEEE, 2016, pp. 1–7.

[6] D. Šoštarić and G. Mester, "Drone localization using ultrasonic TDOA and RSS signal: Integration of the inverse method of a particle filter," *FME Transactions*, vol. 48, no. 1, pp. 21–30, 2020.

[7] L. Huang, J. Li, H. Hao, and X. Li, "Micro-seismic event detection and location in underground mines by using convolutional neural networks (CNN) and deep learning," *Tunnelling and Underground Space Technology*, vol. 81, pp. 265–276, 2018.

[8] J. Baek, J. Lee, H. Shim, S. Im, and Y. Han, "Target tracking initiation for multi-static multi-frequency PCL system," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 10, pp. 10 558–10 568, 2020.

[9] A. Niitsoo, T. Edelhäuβer, and C. Mutschler, "Convolutional neural networks for position estimation in TDOA-based locating systems," in *2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, 2018, pp. 1–8.

[10] Y. Xue, W. Su, H. Wang, D. Yang, and Y. Jiang, "Deeptal: Deep learning for TDOA-based asynchronous localization security with measurement error and missing data," *IEEE Access*, vol. 7, pp. 122 492–122 502, 2019.

[11] A. R. J. Ruiz and F. S. Granja, "Comparing ubisense, bespoon, and decawave uwb location systems: Indoor performance analysis," *IEEE Transactions on instrumentation and Measurement*, vol. 66, no. 8, pp. 2106–2117, 2017.

[12] J. Smith and J. Abel, "Closed-form least-squares source location estimation from range-difference measurements," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 12, pp. 1661–1669, 1987.

[13] J. Baek, J. Lee, Y. Han, H. Shim, and S. Im, "Grid-based target estimation scheme for passive coherent location system," in *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*, 2019, pp. 1–5.

[14] A. F. Agarap, "Deep learning using rectified linear units (relu)," *arXiv preprint arXiv:1803.08375*, 2018.