# A JOINT TARGET LOCALIZATION AND CLASSIFICATION FRAMEWORK FOR SENSOR NETWORKS

*Kyunghun Lee*[⋆†] *Benjamin S. Riggan*[†] *Shuvra S. Bhattacharyya*[⋆‡]
[⋆]Department of Electrical and Computer Engineering, University of Maryland, College Park, MD, USA
[†]U.S. Army Research Laboratory, Adelphi, MD, USA
[‡]Department of Pervasive Computing, Tampere University of Technology, Tampere, Finland
{leekh3, ssb}@umd.edu {benjamin.s.riggan.civ}@mail.mil

## ABSTRACT

In this paper, we propose a joint framework for target localization and classification using a single generalized model for non-imaging based multi-modal sensor data. For target localization, we exploit both sensor data and estimated dynamics within a local neighborhood. We validate the capabilities of our framework by using a multi-modal dataset, which includes ground truth GPS information (e.g., time and position) and data from co-located seismic and acoustic sensors. Experimental results show that our framework achieves better classification accuracy compared to recent fusion algorithms using temporal accumulation and achieves more accurate target localizations than multilateration.

***Index Terms***— localization, classification, tracking, sensor fusion, sensor networks.

## 1. INTRODUCTION

Automatic target localization and discrimination is critical for border protection and surveillance settings, especially in remote locations where it can be costly or logistically difficult to employ human enforced security. A number of robust target classification and localization algorithms using cameras have been suggested (e.g., see [1]). However, computing both location and class information from these types of devices can be challenging. For example, image-based localization and tracking solutions have several challenges to consider, such as occlusions, fog, lighting variations, limited field of view, and processing/power requirements.

Alternatively, we can consider using non-imaging sensors, such as seismic and acoustic sensors, to perform both target localization and classification. The Doppler effect causes faster objects to generate signals with different signatures compared to those of slower or stationary objects. The Doppler effect in acoustic and seismic signals is significant because acoustic and seismic signals have a slower wave propagation speed compared to electromagnetic signals [2]. Dragoset showed that seismic signals can have phase dispersion caused by the Doppler effect [3]. Target velocities can be estimated based on the Doppler shift and then used to discriminate between people and vehicles and potentially provide localization.

In this paper, we introduce a joint framework for tracking and classifying targets using acoustic and seismic signals from multiple, locally distributed sensor nodes. This framework is based on probabilistic confidence maps based on a spatial accumulative framework from acoustic and seismic signals to locate and classify target. Through extensive experiments, we demonstrate that our proposed framework provides better localization than that of baseline multilateration-based location estimation (e.g., beamforming). At the same time, we show that our framework achieves better classification performance than recent seismic and acoustic fusion approaches in [4].

A distinguishing aspect of our work is that we provide a framework that can be used for a classification and localization of targets simultaneously using multiple sensor nodes with a singular generalized model, which can be applied to every node in a sensor network. By employing probabilistic maps from acoustic, seismic, and estimated velocities, we improve classification performance compared to our recent work [4], and we provide a better localization (tracking) capability compared to multilateration-based methods.

Another distinguishing aspect of our work is that we validate it by using actual data collected in an outdoor setting, mimicking common operating environments and location information from GPS. For experiments, we employ acoustic and seismic data that are synchronized with GPS signals collected from the field. In contrast, many previous studies for localization of acoustic and seismic signals employ simulated signals to validate the work (e.g., see [2]), which can have different characteristics compared to real-time datasets.

## 2. RELATED WORKS

Various algorithms have been proposed for target localization using image-based or other modalities. These include vehicle detection and tracking using acoustic and video sensors [5]; location estimation using video, image, and audio signals [6]; location estimation based on Received Signal

Strength (RSS) [7–10]; confidence-based iterative localization [11]; and target location estimation from detection of dense sensor networks [12]. Our work differs by jointly performing both classification and localization, and validating our approach under more challenging conditions using sparsely distributed sensor nodes (e.g., 0.0025 sensors per square meter).

Multilateration (see Figure 1) is a common approach used for localization using wireless sensor networks (e.g., see [13, 14]). By using measured (or estimated) distances from multiple sensors, the target position can be estimated. The location of an unknown node (or target) can be estimated based on the intersection of the distance from multiple nodes. For example, Hefeeda et al. [15] provide an approach for early detection of forest fires using multilateration. Damarla et al. [16] provide a sniper localization method using the time-difference-of-arrival (TDOA) between the muzzle blast and the shock wave using multiple single-acoustic-sensor nodes. Our work differs from these works in that we employ a probabilistic score model instead of using estimated distances from nodes, and as emphasized previously, we provide a joint framework for both localization and classification.

Several works estimate the motion of a target (mainly for vehicles) from acoustic signals (e.g., see [17–19]) based on the Doppler effect. While we employ estimated dynamics from the Doppler Effect, our work differs from this earlier work in that we use both acoustic and seismic signals, and we provide not only localization but also classification. Moreover, we consider both people and vehicles. Incorporating detection of peoples makes the problem significantly more challenging. This is because humans generate very small acoustic and seismic signatures compared to vehicles.

In [4], an accumulative model is proposed to combine multiple evidences over time to improve discriminability. However, this introduces an unnecessary latency. Instead, we propose an accumulative method using spatially distributed sensors for improving discriminability.

## 3. JOINT LOCALIZATION AND CLASSIFICATION FRAMEWORK

In this section, we describe a joint target localization and classification framework using estimated dynamics (velocities) and multimodal data. We refer to this approach as Classification and Localization using Estimated Dynamics and Multimodal data (CLEDM). In this paper, we use acoustic and seismic sensing modalities. However, the CLEDM framework is not dependent on these modalities and we envision that it can readily be adapted to other ones. Investigating such adaptations is a useful direction for future work.

CLEDM is motivated by the Doppler effect, which enables us to effectively estimate the velocity of the target that is being tracked. Using a probabilistic confidence map to assess the movement of the target using estimated velocities, we estimate the next target location. We also apply the estimated velocities to improve the performance of classification.

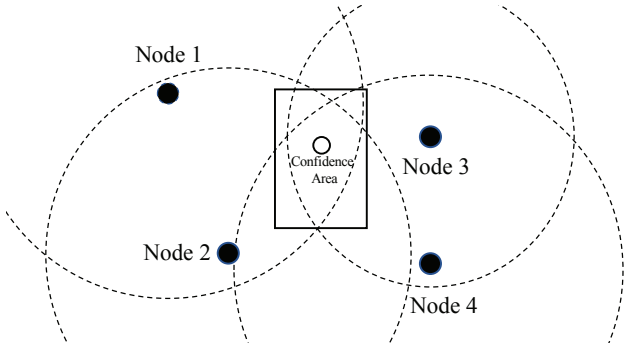


**Fig. 1**: Localization using multilateration.

When training, modalities that exploit the Doppler effect are required. We need the ground truth class and location of the target corresponding to each signal segment in the training dataset.

CLEDM decomposes time into windows (*segments*) of some fixed duration $t_s$. We use $t_s = 1\text{sec}$ in our experiments. For $i = 1, 2, \ldots$, we denote the starting time $(t_s \times (i-1))$ of the $i$th segment by $t_i$. Let $D_{\alpha,i}(\tau)$ and $D_{\sigma,i}(\tau)$ denote the acoustic and seismic signal, respectively, for the $i$th time segment $(0 \leq \tau < t_s)$.

During the training process, since ground truth target location information is given, we are able to calculate two types of dynamics: a relative speed $v_r$ and an absolute speed $v_a$. We use an $x-y$ coordinate system to model the spatial layout of the region of interest that is monitored by the given sensor network. We assume that the origin in this coordinate system is the location of an active sensor node $\nu$ that acquires the signals $D_\alpha$ and $D_\sigma$. If we denote the target location relative to $\nu$ at $t_k$ by $\vec{r}_k = (x_k, y_k)$, then the following expressions can be used to determine $v_r$ and $v_a$:

$$v_r = \frac{|\vec{r}_{i+1}| - |\vec{r}_i|}{t_s}, \text{ and} \tag{1}$$

$$v_a = \frac{|\vec{r}_{i+1} - \vec{r}_i|}{t_s}. \tag{2}$$

Intuitively, $v_r$ is the estimated rate of change of the distance between the target and the sensor node $\nu$. This rate of change can be positive or negative. Similarly, $v_a$ represents the absolute speed of the target, which is independent of individual sensor nodes.

Now suppose that $F_{\alpha,i}$ and $F_{\sigma,i}$ represent extracted features, such as cepstral features, from $D_{\alpha,i}$ and $D_{\sigma,i}$, respectively, and let $F_{fs,i}$ represent the concatenation $[F_{\alpha,i}, F_{\sigma,i}]$ of these feature vectors. In this paper, we used 50 cepstral features extracted from acoustic and seismic signals for $F_{\alpha,i}$ and $F_{\sigma,i}$.

Then we formulate the following composite feature vector $X_i$ for time $t_i$:

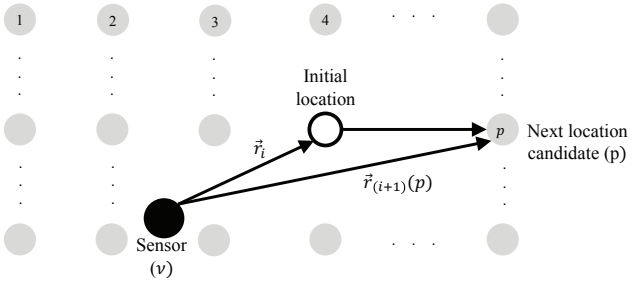


**Fig. 2**: Grid map for estimating the next target location.

$$X_i = [F_{fs,i}, v_a, v_r]. \tag{3}$$

During the training process, we compute $X_i$ for each time segment $i$ of available training data. We assume that a ground truth class label $Y_i$ is available as part of the training data for each time segment $i$. $Y_i$ indicates whether the target is of class person ($A$) or vehicle ($B$). After computing the $X_i$ values, we train a model $H$ to classify between classes $A$ and $B$. In our experiments, we use a support vector machine (SVM) as the model $H$, but the CLEDM methodology is not restricted to SVMs and can readily be adapted to use other types of models.

Based on the trained model $H$, a real-valued classification score $\Gamma(X_i)$ can be calculated if $F_{fs}$, $v_a$, and $v_r$ are given. The score is formulated such that $sgn(\Gamma)$ represents the classification decision between classes $A$ and $B$, and $abs(\Gamma)$ represents the classification "confidence" of the associated prediction. Here, $sgn$ and $abs$ represent the sign and absolute value functions, respectively.

After training $H$, we assume that the initial target location and the neighborhood $\mathcal{N}$ of sensors are known to initialize the tracking component of our framework. From the current target position at time $t_i$, we can extract $F_{\alpha,i}$, $F_{\sigma,i}$ from $D_{\alpha,i}$, $D_{\sigma,i}$.

Around the current target position, which is located at a distance of $\vec{r}_i$ from a particular sensor, we define a grid $G$ discrete locations that represent potential target locations at time $t_{i+1}$. This grid map is illustrated in Figure 2. For a given candidate target location $p = (p_x, p_y)$ at time $t_{i+1}$, we set $\vec{r}_{i+1} = (p_x, p_y)$, and then estimate $v_a$ and $v_b$ from Eq. 2 and 1, respectively.

Next, using Eq. 3 and our trained model, we calculate the feature vector $X_i(p;\nu)$ for the candidate next point $p$ and sensor $\nu$.

Then, we calculate the classification score:

$$\gamma(p;\nu) = \Gamma(X_i(p;\nu)). \tag{4}$$

We repeat this process to determine $\gamma$ for all points $p \in G$. For a sensor $\nu$, we define the class-specific score functions $\delta_A$ and $\delta_B$ over the domain $G$ as follows:

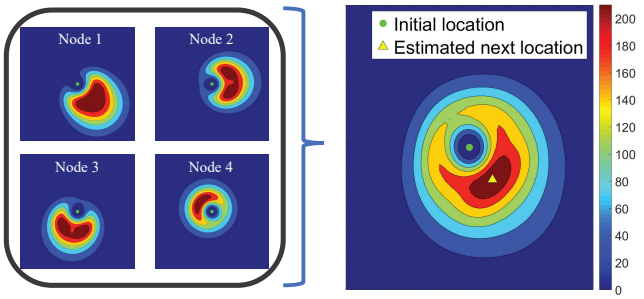


**Fig. 3**: Confidence maps for each node are shown on the left. The spatial accumulation of these maps are shown on the right.

$$\delta_A(p;\nu) = \begin{cases} |\gamma(p;\nu)| & \text{if } \gamma(p;\nu) \geq 0 \\ 0 & \text{if } \gamma(p;\nu) < 0 \end{cases} \tag{5}$$

$$\delta_B(p;\nu) = \begin{cases} 0 & \text{if } \gamma(p;\nu) \geq 0 \\ |\gamma(p;\nu)| & \text{if } \gamma(p;\nu) < 0 \end{cases} \tag{6}$$

Eq. 5 and 6 are derived based on a specific sensor node, whose position is taken to be the origin of the coordinate system in the associated derivations. When multiple sensor nodes are present, these class- and node-specific score functions can be summed across all of the nodes to yield probabilistic confidence maps $M_A = \sum_{\nu \in \mathcal{N}} \delta_A(p;\nu)$ and $M_B = \sum_{\nu \in \mathcal{N}} \delta_B(p;\nu)$ for the two classes $A$ and $B$, respectively.

This is a form of spatial accumulation (across the available sensor nodes), which is different from the temporal accumulation applied in [4]. Figure 3 illustrates an example of a probabilistic confidence map that is derived using this form of spatial accumulation.

To predict the class of the target, we first determine for each $c \in \{A, B\}$ the maximum absolute value $Z_c$ within $M_c$ over all points in the grid $G$. Then if $Z_A \geq Z_B$, the predicted class is $A$; otherwise, it is $B$. Here, we arbitrarily select $A$ as the predicted class in case of a tie ($Z_A = Z_B$).

After classification, we calculate the centroid of the score map $M_\kappa$ that is associated with the predicted class $\kappa$. This centroid is the next estimated location.

The whole process of joint classification and localization described above is repeated iteratively to provide continuous tracking.

## 4. EXPERIMENTS AND RESULTS

In this section, we present an experimental evaluation of the proposed CLEDM framework. In our evaluation, we employ 16 multimodal sensor nodes that each collect acoustic and seismic data. The nodes are grouped into 4 sets of 4 nodes each and placed in a $20\text{m} \times 20\text{m}$ square. The sets are denoted as Set 1 through Set 4. The placement of the nodes is illustrated in Figure 4.

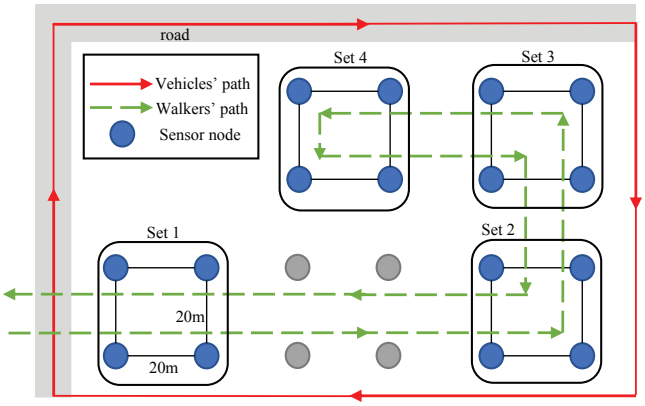


**Fig. 4**: Layout of sensor nodes.

Set 1 is used for training, and contains 1523 frames of data. Sets 2–4 are used for testing, and contain a total of 1620 frames. Each frame contains 1 second of acoustic and seismic data corresponding to a single target (person or vehicle). The data within each frame is sampled at 4096Hz. Each frame also contains GPS location data of the associated target. More details about this dataset can be found in [20].

To evaluate classification performance between people and vehicles, we compared classification accuracy among single-modality, SVM-based classification; a state-of-the-art multi-modal classification architecture called Accumulation of Local Feature-level Fusion Scores (ALFFS) for acoustic and seismic signals [4]; and our proposed CLEDM framework. Table 1 shows the results of this comparison. The columns labeled Acoustic and Seismic correspond to the single-modality results. The results in Table 1 show that the CLEDM framework provides superior accuracy.

**Table 1**: Accuracy comparison (%).

| | Acoustic | Seismic | ALFFS [4] | CLEDM |
|---|---|---|---|---|
| Accuracy (%) | 77.654 | 77.222 | 79.753 | **81.852** |

To evaluate the tracking performance for people and vehicles, we compared tracking using multilateration to our proposed CLEDM approach. For this comparison, we used 44 tracks composed of 1620 data frames in total. For both algorithms, only the first point is synchronized with ground truth, so error increases as time goes on. Table 2 and 3 summarize the results from our experiments on tracking performance.

For the multilateration-based baseline that we used in these experiments, we employed a convolutional neural network for the regression model. We used this model to estimate the distances required for localization. We employed a 2-D convolutional layer with 20 filters of size 25 each followed by a ReLU layer. We also employed a fully-connected output layer of size 1, and a regression layer. In this network model, the input data are formed by the concatenation of acoustic and seismic data segments (1 seconds each in duration), and the output is the estimated distance.

We compared the average error between ground truth and estimated location, and the average maximum errors from all 44 tracks. We also compared the percentage of data segments that have less than a certain amount of error: in Table 2 and 3, $\mathrm{err} < X\mathrm{m}$ gives the percentage of tracks for which the error was less than $X$ meters.

The results in the two tables show that CLEDM has significantly better tracking performance compared to the baseline localization approach overall, and especially favorable performance for tracking people.

**Table 2**: Tracking performance comparison (people)

| | avg. error (m) | avg. maximum error (m) | err <3m (%) | err <5m (%) | err <10m (%) |
|---|---|---|---|---|---|
| multilateration | 6.979 | 9.385 | 15.552 | 28.852 | 76.599 |
| CLEDM | **2.897** | **4.787** | **58.552** | **86.483** | **99.927** |

**Table 3**: Tracking performance comparison (vehicle)

| | avg. error (m) | avg. maximum error (m) | err <5m (%) | err <10m (%) | err <20m (%) |
|---|---|---|---|---|---|
| multilateration | 7.064 | 12.877 | 41.177 | **72.549** | 97.712 |
| CLEDM | **6.244** | **10.309** | **43.791** | 72.026 | **100** |

## 5. CONCLUSION

In this paper, we have introduced a spatial accumulative framework for both target classification and localization. We leveraged signal phenomenology to estimate dynamics and extract discriminative information, which is accumulated across multiple nodes within a local neighborhood. Experimental results have shown that our algorithm provides better localization performance compared to a baseline localization algorithm based on multilateration, while our algorithm also achieves better classification performance compared to relevant prior work. Specifically, CLEDM achieved an absolute improvement of 2.099% in accuracy compared to the baseline ALFFS approach on average. Also, CLEDM achieved 2.897 and 6.244 average error (meter) for people and vehicles. Whereas, the baseline approach achieved 6.679 and 7.064 average error (meter) for people and vehicles. Therefore, accumulating multiple evidences (e.g., dynamics, latent information) across multiple sensors enhances target discrimination and tracking capabilities.

## 6. REFERENCES


[1] L. Meng and J. P. Kerekes, "Object tracking using high resolution satellite imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 1, pp. 146–152, February 2012.

[2] D. Li, K. D. Wong, Y. H. Hu, and A. M. Sayeed, "Detection, classification, and tracking of targets," *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 17–29, March 2002.

[3] W. H. Dragoset, "Marine vibrators and the Doppler effect," *GEOPHYSICS*, vol. 53, no. 11, pp. 1388–1398, 1988.

[4] K. Lee, B. S. Riggan, and S. S. Bhattacharyya, "An accumulative fusion architecture for discriminating people and vehicles using acoustic and seismic signals," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, March 2017, pp. 2976–2980.

[5] R. Chellappa, G. Qian, and Q. Zheng, "Vehicle detection and tracking using acoustic and video sensors," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, May 2004, vol. 3, pp. iii–793–6 vol.3.

[6] G. Friedland, O. Vinyals, and T. Darrell, "Multimodal location estimation," in *Proceedings of the ACM International Conference on Multimedia*, 2010, pp. 1245–1252.

[7] B. Ferris, D. Hähnel, and D. Fox, "Gaussian processes for signal strength-based location estimation," in *Proceeding of Robotics: Science and Systems*, 2007, vol. 2, pp. 303–310.

[8] Y. Y. Cheng and Y. Y. Lin, "A new received signal strength based location estimation scheme for wireless sensor network," *IEEE Transactions on Consumer Electronics*, vol. 55, no. 3, pp. 1295–1299, August 2009.

[9] R. Niu and P. K. Varshney, "Target location estimation in sensor networks with quantized data," *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4519–4528, December 2006.

[10] C. Feng, W. S. A. Au, S. Valaee, and Z. Tan, "Received-signal-strength-based indoor positioning using compressive sensing," *IEEE Transactions on Mobile Computing*, vol. 11, no. 12, pp. 1983–1993, December 2012.

[11] Z. Yang and Y. Liu, "Quality of trilateration: Confidence-based iterative localization," *IEEE Transactions on Parallel and Distributed Systems*, vol. 21, no. 5, pp. 631–640, May 2010.

[12] A. Artes-Rodriguez, M. Lazaro, and L. Tong, "Target location estimation in sensor networks using range information," in *Processing of the IEEE Workshop on Sensor Array and Multichannel Signal Processing*, July 2004, pp. 608–612.

[13] K. Langendoen and N. Reijers, "Distributed localization in wireless sensor networks: a quantitative comparison," *Computer Networks*, vol. 43, no. 4, pp. 499 – 518, 2003, Wireless Sensor Networks.

[14] A. Awad, T. Frunzke, and F. Dressler, "Adaptive distance estimation and localization in wsn using rssi measures," in *10th Euromicro Conference on Digital System Design Architectures, Methods and Tools*, August 2007, pp. 471–478.

[15] M. Hefeeda and M. Bagheri, "Wireless sensor networks for early detection of forest fires," in *2007 IEEE International Conference on Mobile Adhoc and Sensor Systems*, October 2007, pp. 1–6.

[16] T. Damarla, L. M. Kaplan, and G. T. Whipps, "Sniper localization using acoustic asynchronous sensors," *IEEE Sensors Journal*, vol. 10, no. 9, pp. 1469–1478, September 2010.

[17] B. G. Quinn, "Doppler speed and range estimation using frequency and amplitude estimates," *The Journal of the Acoustical Society of America*, vol. 98, no. 5, pp. 2560–2566, 1995.

[18] V. Cevher, R. Chellappa, and J. H. McClellan, "Vehicle speed estimation using acoustic wave patterns," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 30–47, January 2009.

[19] C. Couvreur and Y. Bresler, "Doppler-based motion estimation for wide-band sources from single passive sensor measurements," in *1997 IEEE International Conference on Acoustics, Speech, and Signal Processing*, April 1997, vol. 5, pp. 3537–3540 vol.5.

[20] S. M. Nabritt, T. Damarla, and G. Chatters, "Personnel and vehicle data collection at Aberdeen proving ground (APG) and its distribution for research," Tech. Rep. ARL-MR-0909, US Army Research Laboratory, October 2015.