Evaluation Metrics for xDR Challenge 2018

0. List of Evaluation Indices

Metrics related to accuracy (modified in this year)

- E_{median_error}: Evaluating absolute error of the integrated localization system
- E_{accum_error} :Evaluating PDR/VDR accumulating error

Metrics related to the trajectory naturalness (same as last year)

- $E_{velocity}$: Metric related to the naturalness of travel speed
- $E_{frequency}$: Metric related to position measurement output frequency

Specific metrics for warehouse scenario (same as last year)

- $E_{obstacle}$: Metric related to collision with obstacles
- $E_{picking}$: Metric related to motions during picking work

Note:

The dataset provided for the competitors can be separated into two type. First one is the sample data for allowing them to test their algorithm with the sample data in same format with the real competition data. Second one is the real test dataset which is actually used for calculating the score.

Most of all evaluations are based on the comparisons between ground truth points provided by WMS and points at the nearest neighbor time in estimated (submitted) trajectories.

1. Overview of metrics related to accuracy

Basically, the concept of evaluation metrics for PDR and VDR will be adhered the PDR Challenge 2017. We plan to adopt new method for error evaluation related to E_{median_error} and E_{accum_error}. In the PDR Challenge 2017, the BLE signals from the tags are always provided. The methods evaluated in Es in the PDR Challenge2017 not pure PDR, but integrated localization. In order to evaluating “pure PDR/VDR” errors, in the xDR Challenge 2018, the BLE signals are partially hided as the BLE unreachable period (BUP in short). The error evaluations are going to be combination of accumulated error evaluation in BUP by E_{accum_error} and evaluation of magnitude of absolute error of integrated localization system in other periods by E_{median_error}.

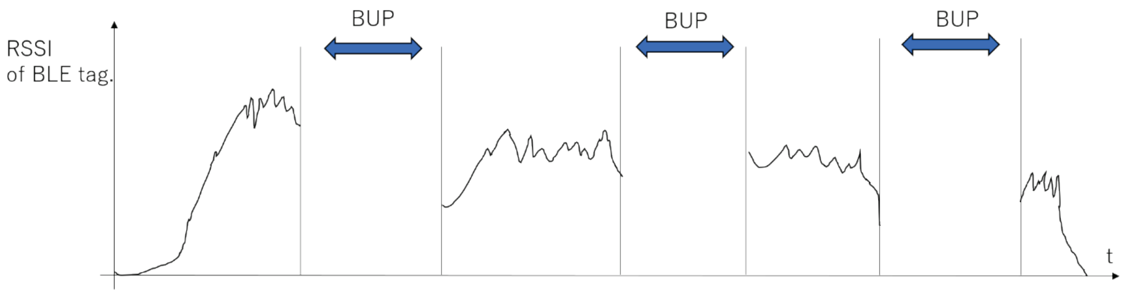

Figure 1. Conceptual illustration of BLE un-reachable period (BUP)

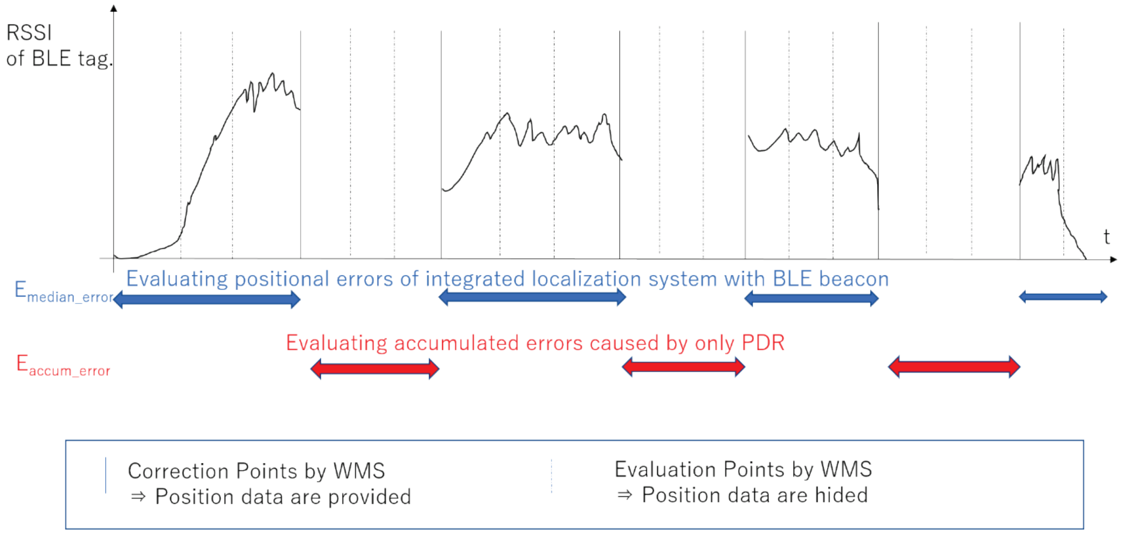

Figure 2. Terms for evaluating each metric for error evaluations.

1.1 Definition of E_{median_error}

This metric is an error metric of absolute positional error of integrated positioning with PDR/VDR, BLE beacon, WMS picking positions, and MAP constraint. These all cues for localizations are available in out-side of BUPs, which are shown as blue arrows in Figure2. In these terms, distance between the estimated and true position is measured as PosErr at the checking points when the ground truth positions are available from WMS. We calculate median of PosErr in the checking points.

【Detailed metric for integrated positioning error: E_{median_error}】

$E_{ave\_error} = 100$ (if median(PosErr) < 1)

$E_{ave\_error} = 100 - \frac{100}{29}(median(PosErr) - 1)$ (if 30>median(PosErr) > 1)

$E_{ave\_error} = 0$ (if median (PosErr) > 30)

1.2 Definition of E_{accum_error}

This metric indicates speed of error accumulation caused by PDR/VDR. We partially provided the ground truth of the target's positions for error correction. The BUPs start at the timing of the error correction, because we can assume that errors are 0 at the correction points. We also assume that the error increases as time proceeds from the correction points and quantify the speed of error accumulation from the correction points. In order to evaluate the pure result by PDR/VDR, target term for this metric are BUPs when the BLE signals are not available.

We term the slope of the error accumulation as error accumulation gradient (EAG). EAG can be calculated for every checking point by error value divided by the time from the closest correction points. We use median of EAG for calculating E_{accum_error}..

【Detailed metric for integrated positioning error: $E_{accum\_error.}$】

$E_{accum\_error} = 100$ (if median(EAGs)< 0.05)

$E_{accum_{error}} = 100 - \frac{100}{1.95}(median(EAGs) - 0.05)$ (if 2.0> median(EAGs)> 0.05)

$E_{accum\_error} = 0$ (if median(EAGs) > 2.0)

2. Overview of the metrics for trajectory naturalness

We believe that ideal localization method outputs natural trajectories for the targets. We deal with naturalness of the trajectory for two aspects; naturalness of the speed, and naturalness of the output frequency.

2.1 Definition of $E_{velocity}$

This metric checks whether the velocity in the trajectory is within range of natural human movements. Some algorithm may be able to accurately estimate the position with a noticeable frequent jitter. We prefer natural and smooth trajectories. In order to deduct such artifact by a jitter, we check the velocity of all submitted points. We define 1.5 m/sec as the threshold of natural human speed and 4.5m/sec as the threshold of forklift's max. speed in the warehouse. The $E_{velocity}$ is calculated as the percentage of the checking points within the speed limit.

【Detailed metric for velocity naturalness: $E_{velocity.}$】

Given, $Flag_i^v$ is assigned 0 when target moves faster than the threshold (1.5 m/sec for PDR, 4.5m/sec for VDR) and is assigned 1 when it is below the threshold at the i^{th} frame in the trajectory's frames.

$$E_{velocity} = \frac{\sum_{i=1}^{N_{traj\_frame}} Flag_i^v}{N_{traj\_frame}} \times 100$$

where N_{traj_frame} is the number of total frames of the trajectory.

2.2 Definition of $E_{frequency}$

This metric checks the update frequencies of the trajectories submitted by the competitors. If there are two trajectories with same accuracy of the positional estimation and different update frequencies, the trajectory with the higher frequency would have a greater value than the one with a lower frequency. Given the application of localization for the employees during the work, we define the minimum update frequency as 1 Hz. To calculate the update frequency for every frame, the elapsed time from the previous submitted frame is measured for all submitted frames.

【Detailed metric for frequency naturalness: $E_{frequency}$】

Given, $Flag_i^f$ is assigned 1 when the temporal frequency is more than 1 Hz and assigns 0 when the frequency is less than 1 Hz at the i^{th} frames in the submitted trajectory.

$$E_{frequency} = \frac{\sum_{i=1}^{N_{traj\_frame}} Flag_i^f}{N_{traj\_frame}} \times 100$$

3. Overview of specific metrics for warehouse scenario

In this xDR Challenge, we adopt many possible evaluation metrics required for the warehouse scenario. We deal with obstacle interference, and naturalness of behavior during the picking operations.

3.1 Definition of $E_{obstacle}$

This metric checks whether the trajectories enter forbidden areas. We assume that ideal trajectories do not enter the area where employees cannot walk inside or pass over because of the existence of obstacles such as shelves, poles, and walls. Inconsistency between the trajectory and environments also affects the analysis of the trajectories such as traffic line analysis in warehouses. We introduce a metric $E_{obstacle}$ for quantifying the degree of incursion of the trajectories into the forbidden area as shown in Figure 3. We shared the location and size of shelves, poles, walls as reference information about the warehouse. Therefore, the competitors were informed about the forbidden areas. $E_{obstracle}$ is the ratio of the length of trajectories which enters the forbidden area.

【Detailed metric for obstacle interference: $E_{obctacle}$】

Given, $Flag_i^o$ is assigned 1 when the target pixel on the trajectory is not in obstacle areas, and is assigned 0 when the target pixel is in the obstacle areas at the i^th^ pixels in the trajectory.

$$E_{obstacle} = \frac{\sum_{i=1}^{N_{traj\_frame}} Flag_i^o}{N_{traj\_frame}} \times 100$$

Note that we define a tolerance area with 0.17 m width around borders of the forbidden area for ignoring small amount of the incursion.

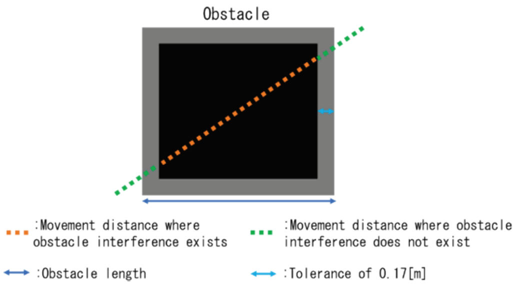

Figure 3. Conceptual illustration of obstacle interference metric

3.2 Definition of $E_{picking}$

This metric evaluates naturalness of the trajectories during picking. As observed in the warehouse, employees who move by foot stop in front of the shelves when they pick items from the shelves. Order picking fork which is a type of the forklift can directly lift the driving base for allowing the drivers to pick the items from high position shelfs by hands. We pick up the data measured with the order picking fork for the competition. Therefore, we can assume that forklifts also stop in front of the shelves when they the drivers pick items from the shelves. This characteristic should be reflected in the estimated trajectory at the time of picking. We check whether the target of the trajectories stop or not with this metric. More specifically, we check the total length of the movement measured from 1.5 sec before the picking to 1.5 sec after the picking. If the movement is more than 1 meter, index of this metric is deducted. The final $E_{picking}$ is calculated as the percentage of the checking points without the movement out of all the checking points.

【Detailed metric for picking naturalness : $E_{picking}$】

Given, $Flag_i^p$ is assigned 1 when target stops and is assigned 0 when target does not stop at the i[th] checking points.

$$E_{picking} = \frac{\sum_{i=1}^{N_{traj\_frame}} Flag_i^p}{N_{picking}} \times 100$$

, where $N_{picking}$ is the number of checking points used for the evaluation.

4. Comprehensive Evaluation as the Integrated Indicator for Determining the Winner

In order to rank the submitted trajectories by quantifying the performance as an integrated indicator, the elemental evaluation metrics are integrated with the following weights. We term the integrated index as “comprehensive evaluation” (C.E.), calculated by equation.

$$C.E. = 0.3 \times E_{ave\_error} + 0.3 \times E_{accum\_error} + 0.05 \times E_{picking} + 0.1 \times E_{velocity} + 0.15 \times E_{obstacle} + 0.1 \times E_{frequency}$$

Note that the weights for error evaluation (accuracy metrics) are increased compared to the PDR Challenge 2017.

5. Format of the result submission.

The file for result submission should include X coordinate in meters, Y coordinate in meters, and Unixtime in seconds. The coordinate system (origin, directions) are described in MAP folder. The trajectory file should be csv file and its format is [x(m),y(m),unixtime(s)].

If competitors have troubles for submitting the result data, we can provide methods for uploading data, such as PHP based uploading form for file submission.

Here is an example of the final submission trajectory.

[PDR_result_10_20180313].csv

#x,y,unixtime

9.235,20.345,1505693349.532

9.892,20.487,1505693349.993

10.023,20.003,1505693350.602

10.833,19.725,1505693351.004

11.432,19.700,1505693351.596

11.992,19.688,1505693351.999

12.681,19.692,1505693352.522