Visible Light Positioning with TinyML: Improving Data Quality and Reducing Data Collection Effort

Jakub A. Trzykowski [1] Ran Zhu [1] Qing Wang [1]

[1]EEMCS, Delft University of Technology, The Netherlands

1. Introduction

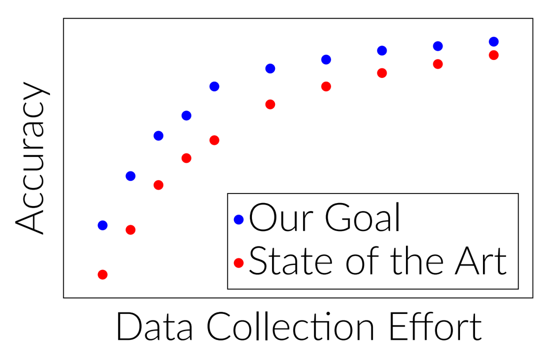


Figure 1. Our goal is to lower data collection effort while maintaining accuracy.

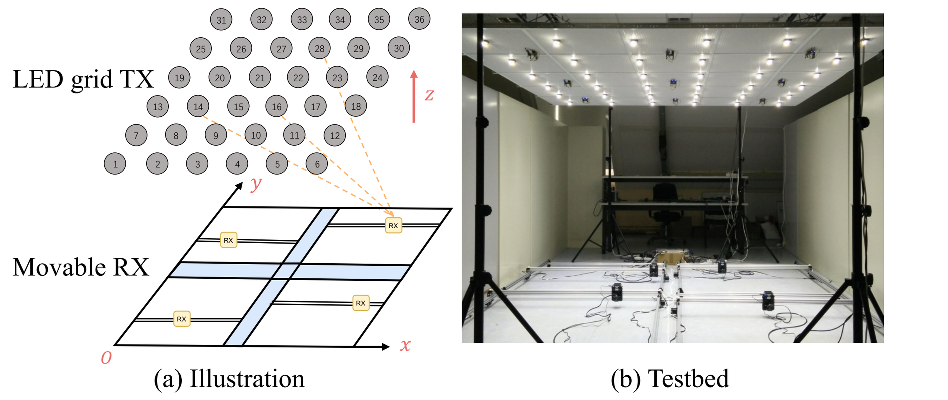


Figure 2. DenseVLC [1] Testbed used to collect the RSS data.

Research questions:

- How can the data cleaning and augmentation pipeline [3] be further improved to increase the accuracy and lower data collection effort?
- How do spatially irregular data acquisition strategies compare to collecting data in a rigid grid?

2. Improving Data Cleaning

Observation: The majority of noise are zero measurements.
Improvement: Revised RSS continuity scoring with brightness boosting, to improve retention of bright measurements.

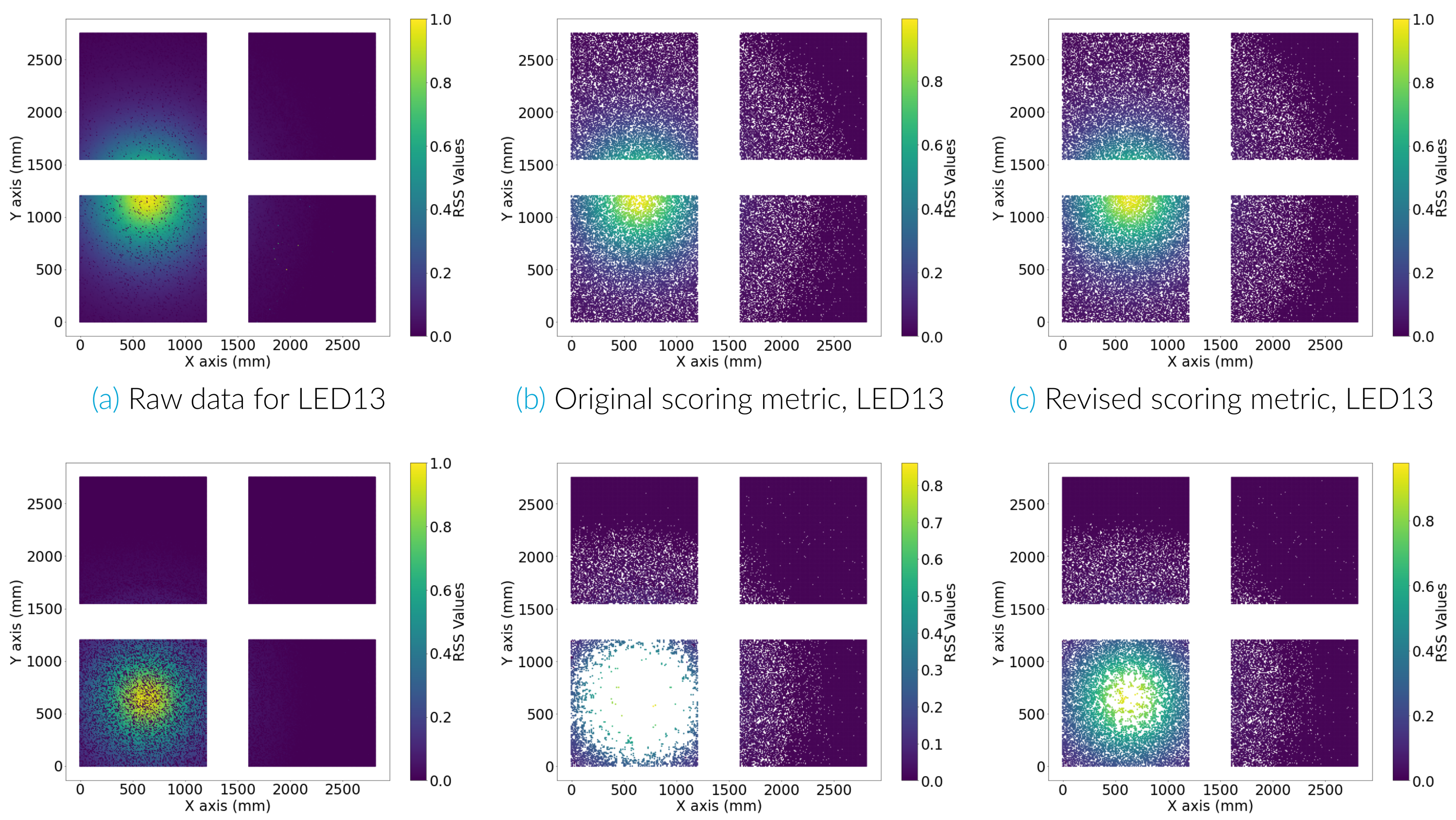
Original

$$S_{x,y} = \left| I_{x,y} - \frac{1}{|\mathbf{I}_N|} \sum_{I \in \mathbf{I}_N} I \right|$$

Revised

$$\widetilde{\mathbf{I}_N} := \frac{1}{|\mathbf{I}_N|} \sum_{I \in \mathbf{I}_N} I$$

$$S_{x,y} = \begin{cases} (I_{x,y} - \widetilde{\mathbf{I}_N}) \cdot r, & \text{if } I_{x,y} \geq \widetilde{\mathbf{I}_N} \\ (\widetilde{\mathbf{I}_N} - I_{x,y}), & \text{otherwise,} \end{cases}$$



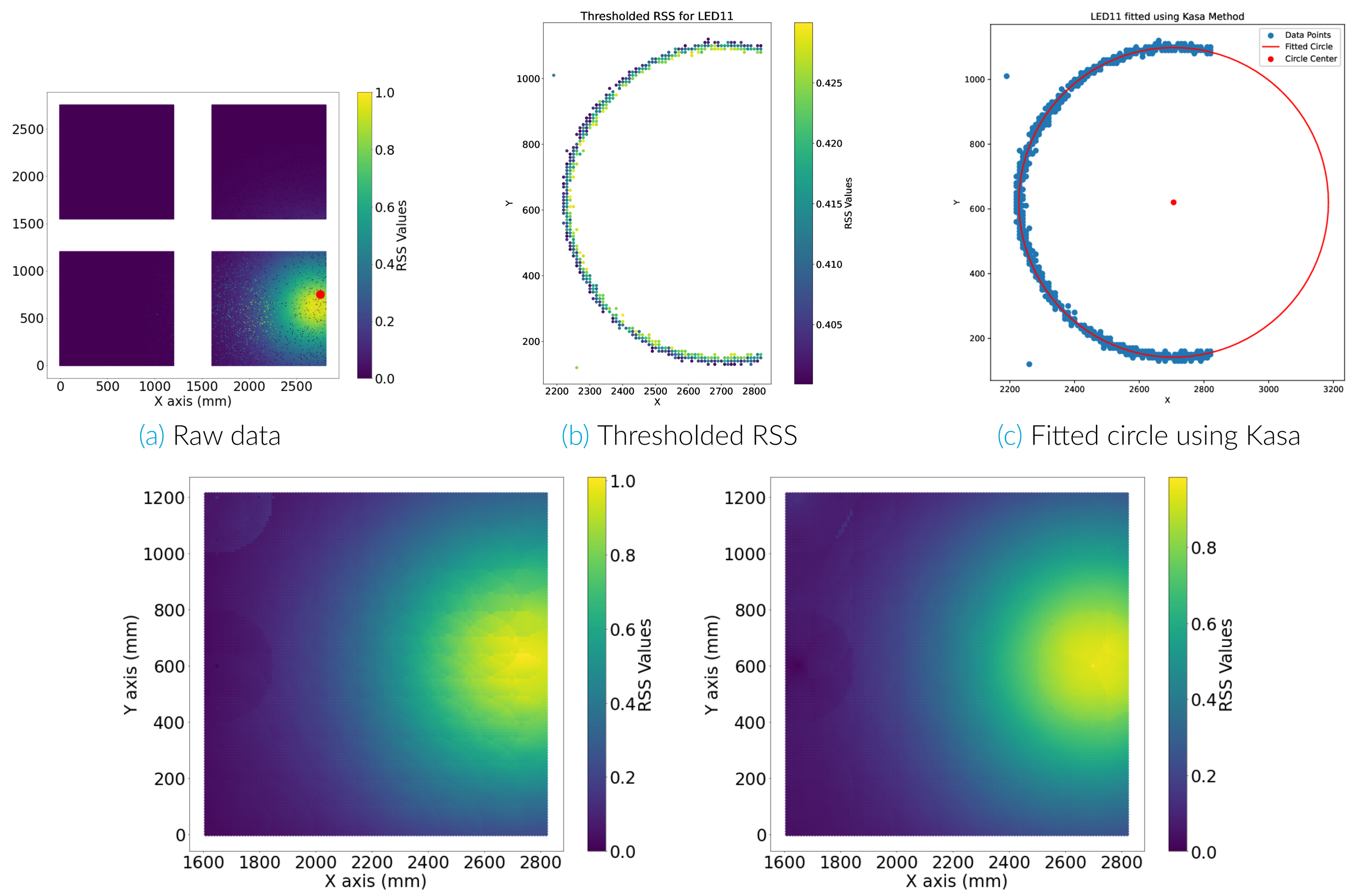
(a) Raw data for LED13 (b) Original scoring metric, LED13 (c) Revised scoring metric, LED13

(d) Raw data for LED7 (e) Original scoring metric, LED7 (f) Revised scoring metric, LED7

Figure 3. Comparison between raw noisy measurements and clean ones with different cleaning methods. Our method retains many more valuable samples even when a lot of noise is present.

3. Improving Data Augmentation

Observation: The imprecise LED positions make data augmentation less accurate.
Improvement: Employ robust circle fitting by Kasa [2] to accurately estimate the LED positions + IDW interpolation.



(a) Raw data (b) Thresholded RSS (c) Fitted circle using Kasa

(d) Augmented 16-to-1 cm with subpar LED position estimates (e) Augmented 16-to-1 cm with improved LED position estimates

Figure 4. The effects of inaccurate LED positions on the RSS data augmentation, along with methods that improve the accuracy.

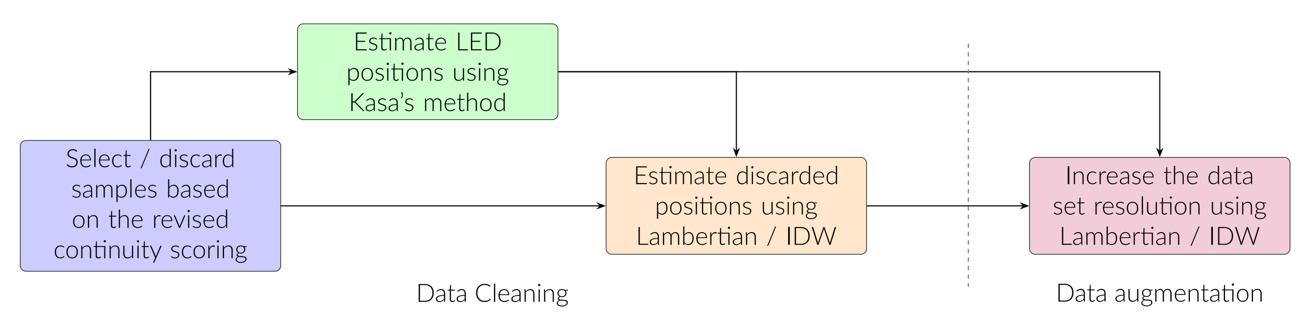
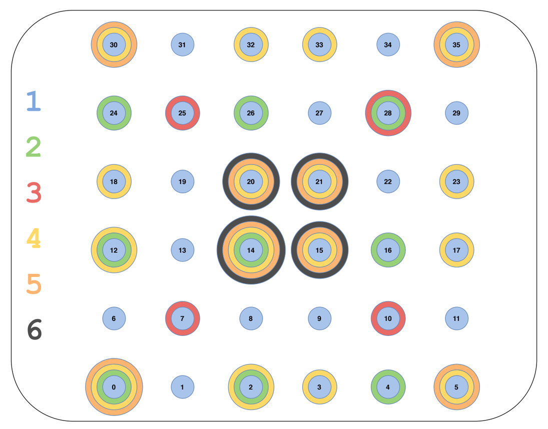


Figure 5. Summarized pipeline of data cleaning and augmentation. The estimated LED positions are used both to reconstruct missing points and increase the density of the dataset.

4. Improved Pipeline Evaluation



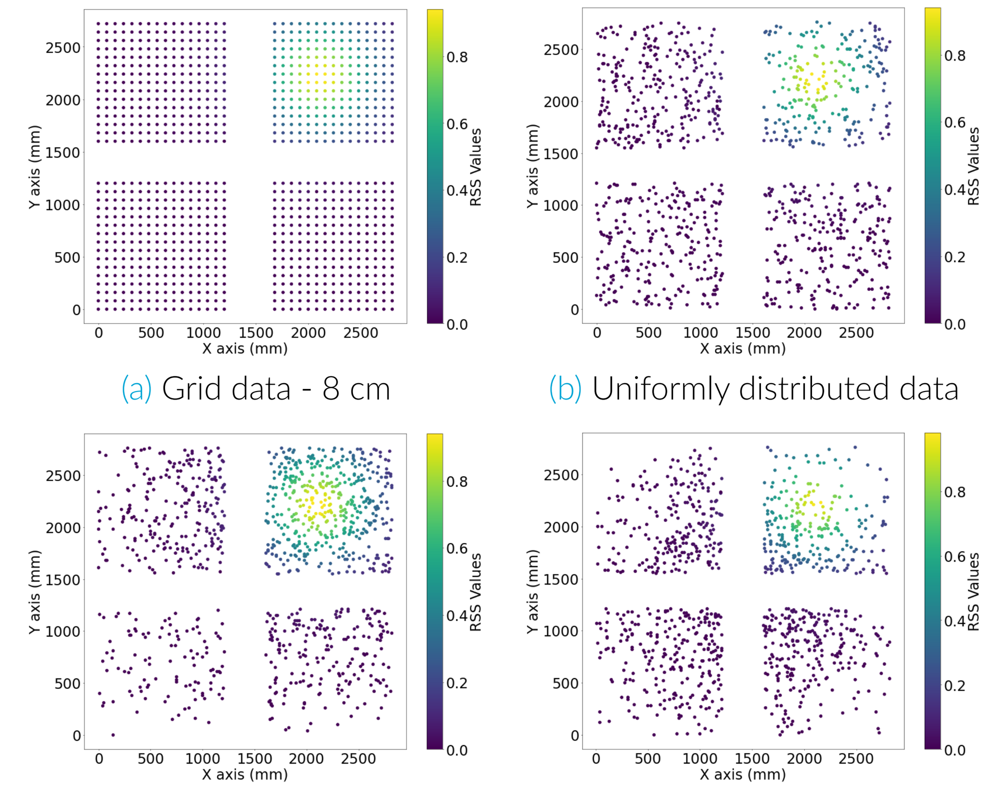
	Raw data	Clean data		Augmented data (from 16 cm)			
				Original		Revised	
Error against	Raw	Raw	Clean	Raw	Clean	Raw	Clean
Conf 1	1.64	6.06	0.79	7.86	1.70	6.68	1.38 ↓18.8%
Conf 2	4.55	13.52	2.91	14.39	4.66	14.01	4.09 ↓12.2%
Conf 3	40.53	49.65	24.20	51.49	27.13	51.20	26.93 ↓0.7%
Conf 4	2.55	7.61	1.51	9.84	3.33	8.71	3.08 ↓7.5%
Conf 5	6.61	11.32	5.93	15.96	9.23	16.30	9.44 ↑2.3%
Conf 6	12.28	17.26	10.17	25.14	16.08	25.21	15.31 ↓4.8%

Table 1. Average errors (in cm) between four model tested on six different LED topologies of varying sparsity (graphic on the left). Evaluated against raw and clean data.

Improvements in accuracy of reconstructing clean data by up to 20% compared to the original method.

5. Spatially Irregular Data Collection

Goal: Reconstruct the clean dataset from differently distributed samples of the RSS values – uniform, LED-centered normal, globally-centered normal.



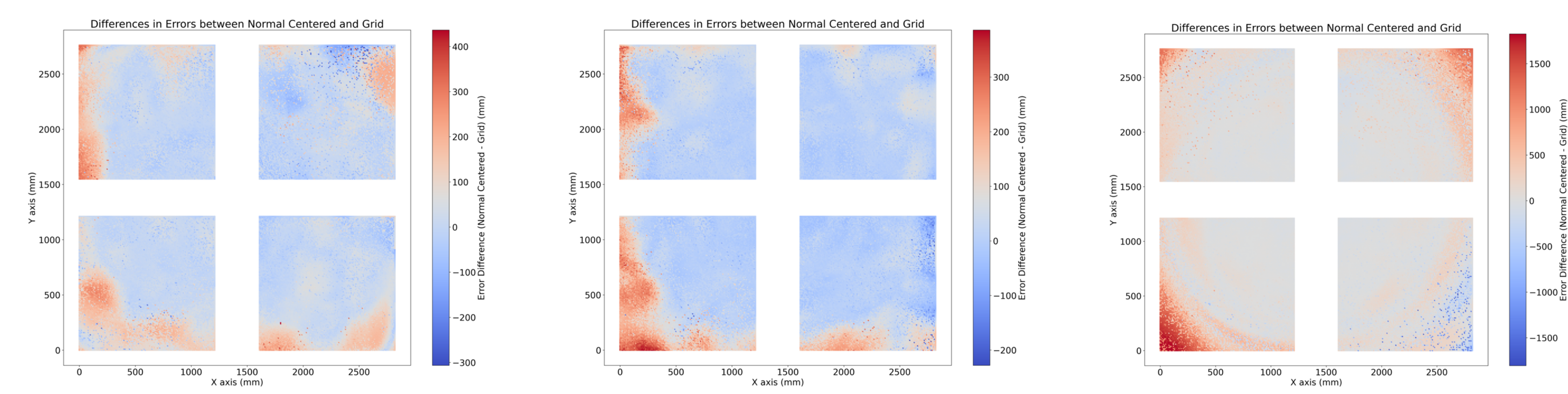
(a) Grid data - 8 cm (b) Uniformly distributed data

(c) Normally distributed, LED-centered data (d) Normally distributed, globally-centered data

Figure 6. Showcase of different sampling methods imitating different data collection strategies. All approaches sample around 1000 points.

	Grid	Uniform	Normal
~1000 samples			
Conf 1	1.16	1.17 ↑0.9%	1.13 ↓2.6%
Conf 2	3.65	3.90 ↑6.8%	3.76 ↑3%
Conf 3	26.12	25.60 ↓2%	25.64 ↓1.8%
Conf 4	2.18	2.57 ↑17.9%	2.48 ↑13.8%
Conf 5	8.62	8.21 ↓4.8%	8.38 ↓2.8%
Conf 6	15.47	15.70 ↑1.5%	15.01 ↓3%
~250 samples			
Conf 1	1.38	1.54 ↑11.6%	1.42 ↑2.9%
Conf 2	4.09	4.87 ↑19.1%	5.02 ↑22.7%
Conf 3	26.93	27.19 ↑1%	27.42 ↑1.8%
Conf 4	3.08	3.97 ↑28.9%	3.69 ↑19.8%
Conf 5	9.44	9.66 ↑2.3%	10.05 ↑6.5%
Conf 6	15.31	16.86 ↑10.1%	18.73 ↑22.3%

Table 2. Comparison of accuracies between models trained on augmented datasets constructed from structured, grid-like, from uniformly, and from LED-centered normally distributed samples.



(a) Conf 2 (b) Conf 4 (c) Conf 6

Figure 7. Comparison of error differences between rigid grid and globally centered normally distributed. Blue – normally distributed was better, red – grid was better.

Uniform sampling was inferior to grid in all cases, LED-centered normally distributed was considerably better with higher sample counts, while globally-centered normally distributed gave a local accuracy boost for denser LED configurations.

6. Conclusions

- The two-layer, 96-perceptron MLP performed comparably to a large 2.5k-neuron network used in [3].
- Improvements to the pipeline can lower the errors by around 20%.
- Need more robust interpolation methods to make the employment of alternative sampling strategies effective.
- More challenging datasets required to further evaluate and refine the methods.

References

[1] Jona Beysens, Ander Galisteo, Qing Wang, Diego Juara, Domenico Giustiniano, and Sofie Pollin. DenseVLC: A cell-free massive mimo system with distributed leds. In *Proceedings of the 14th International Conference on Emerging Networking EXperiments and Technologies*, pages 320–332, 2018.

[2] I. Kasa. A circle fitting procedure and its error analysis. IM-25(1):8–14.

[3] R. Zhu, M. Van Den Abeele, J. Beysens, J. Yang, and Q. Wang. Centimeter-level indoor visible light positioning. 62(3):48–53.

J.A.Trzykowski@student.tudelft.nl