Stereo Ground Truth with Error Bars

Daniel Kondermann[1], **Rahul Nair[1]***, Stefan Meister[1], Wolfgang Mischler[1], Burkhard Güssefeld[1], Katrin Honauer[1], Sabine Hofmann[2], Claus Brenner[2] and Bernd Jähne[1]

[1] *Heidelberg Collaboratory for Image Processing at IWR, Ruprecht-Karls-Universität Heidelberg*

[2] *Institute of Cartography and Geoinformatics, Leibniz Universität Hannover*

* *Corresponding author: rahul.nair@iwr.uni-heidelberg.de*

Keywords : Benchmarking, Reference Data, Stereo, Flow, Uncertainty Analysis, Performance Metrics

Figure 1. **Dataset acquisition**. Reference Data (in this case stereo) generated using auxiliary modalities such as LIDAR (right two panels) are subject to measurement uncertainty. [1]

Abstract

Performance analysis with reference data plays a vital role in research and development of machine vision systems used in automotive assistance, industrial quality control and the like.

Only by analyzing a method and its results we can be able to identify beneficial properties thereof and then use these findings to improve models and inference. One of the key performance analysis techniques employed is the **comparison of algorithm output with reference data** of higher accuracy, colloquially called ground truth.

The talk will focus on two aspects of this process in context of low-level vision (e.g. stereo, flow): **reference data generation** and **performance metrics** that allow us to compare different algorithms.

An aspect often neglected when discussing ground truth based on auxiliary modalities (e.g. LIDAR) is that the data generated is a result of measurement. Due to the nature of any measurement process, this kind of data is always subject to measurement uncertainty [1]. With vision methods improving, the error of the reference data may not always remain smaller than that of the vision method itself. Therefore, it has to be properly incorporated into the analysis. The first part of the talk will discuss how uncertainty can be quantified (Figure 2) in context of dataset generation (Figure 1).

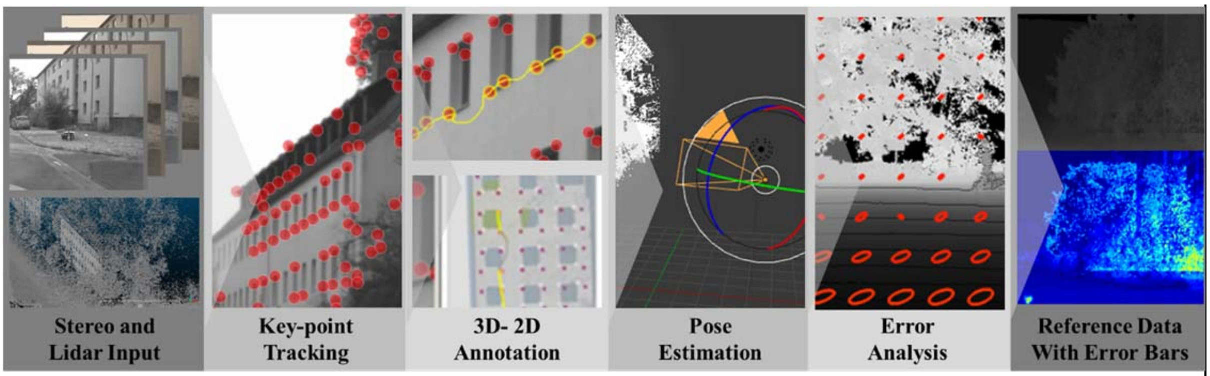

Figure 2. Reference Data Generation: The measurement uncertainty of the input data must be taken into consideration during data processing (left four panels) and subsequent performance analysis. This can be done by means of covariance analysis and error propagation (right two panels). From [1]

Subsequently, the question remains how this information can be used to evaluate algorithm results. This will be the topic of the second part.

Performance Metrics are often used to reduce the rich data available to a single number to be able to rank algorithms. While we understand that such a reduction is required in some form, we believe that by reducing to a single number, we are losing additional information on algorithm characteristics. The talk will show some ways of benchmarking methods while retaining the multi-dimensional nature of algorithm characterization [2] (cf. Figure 3.)

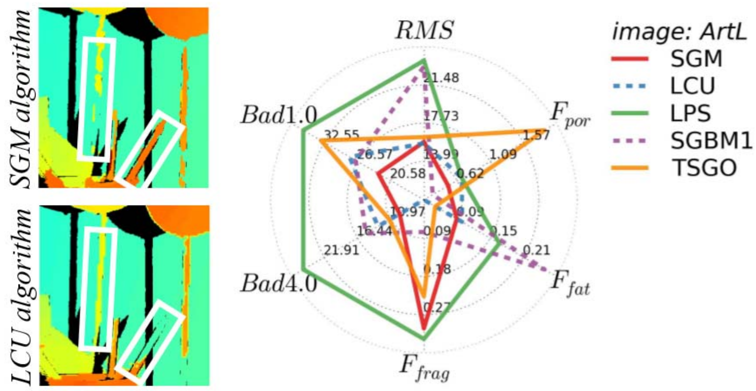

Figure 3: Multidimensional Performance Metrics. Reduction of the reference data to a single number during benchmarking leads to a loss of information. Using multiple complementary metrics (arms of the radar chart) allows researchers to better characterize methods and discover niche strengths of individual techniques (colored lines). .From [2]

[1] Kondermann, D., Nair, R., Meister, S., Mischler, W., Güssefeld, B., Honauer, K., Hofmann, S., Brenner, C. & Jähne, B. (2014). Stereo ground truth with error bars. In *Computer Vision--ACCV 2014* (pp. 595-610). Springer International Publishing.

[2] Honauer, K., Maier-Hein, L., and Kondermann, D. "The HCI Stereo Metrics: Geometry-Aware Performance Analysis of Stereo Algorithms." *Proceedings of the IEEE International Conference on Computer Vision*. 2015.