Learning Depth

Fabian Sinz, Matthias O. Franz

15th December 2003

Abstract

The depth of a point in space can be estimated by observing its image position from two different viewpoints. The classical approach to stereo vision calculates depth from the two projection equations which together form a stereocamera model. An unavoidable preparatory work for this solution is a calibration procedure, i.e., estimating the external (position and orientation) and internal (focal length, lens distortions etc.) parameters of each camera from a set of points with known spatial position and their corresponding image positions. This is normally done by iteratively linearizing the single camera models and reestimating their parameters according to the error on the known datapoints. The advantage of the classical method is the maximal usage of prior knowledge about the underlying physical processes and the explicit estimation of meaningful model parameters such as focal length or camera position in space. However, the approach neglects the nonlinear nature of the problem such that the results critically depend on the choice of the initial values for the parameters.

In this study, we approach the depth estimation problem from a different point of view by applying generic machine learning algorithms to learn the mapping from image coordinates to spatial position. These algorithms do not require any domain knowledge and are able to learn nonlinear functions by mapping the inputs into a higher-dimensional space. Compared to classical calibration, machine learning methods give a direct solution to the depth estimation problem which means that the values of the stereocamera parameters cannot be extracted from the learned mapping.

On the poster, we compare the performance of classical camera calibration to that of different machine learning algorithms such as *kernel ridge regression*, *gaussian processes* and *support vector regression*. Our results indicate that generic learning approaches can lead to higher depth accuracies than classical calibration although no domain knowledge is used.