Stereo Image Based Depth Reconstruction for Crate-Detection

Feature Matching and Interpolation

Graduate

Joel Stohler

David Peter

Introduction: Crates used to transport drinks and vegetables are often distributed to retailers and then collected for reuse. At the collection points, their deposit value must be determined. For this, crates are classified by computer vision. A depth image of the crate stack is estimated to identify the region of interest of the front face, for a reliable and fast recognition. Crate fronts can be planar, ribbed, sliced or indented, and are typically unicolor, making depth estimation challenging due to repetitiveness or rarity of visual features. Traditional depth estimation methods rely on pattern projectors, which are costly and require specialised equipment. This work investigates an alternative to projector-based methods by relying on the advanced matching capabilities of a CNN matchers, like LightGlue. The performance is evaluated with respect to traditional matching algorithms. In this work, improvements and challenges regarding accuracy and efficiency are highlighted.

Approach: A pipeline has been developed to generate depth images based on stereo vision, where features are first extracted using DISK, SIFT and SuperPoint extractors. These features are then matched using brute force and LightGlue matching. Feature matches are indicated in Fig. 1. The disparity of the matched points is used to compute their depth. The error of the features depth derived from the ground truth can be seen in Fig. 2. The point wise depth has been interpolated to create a continuous depth image. To optimise the performance of the algorithms, an automatic evaluation pipeline for parameter tuning has been implemented.

Result: The result, shown in Fig. 3, indicates a sufficient accuracy to support identification of the front face of the first stack and to distinguish it form other depth regions. The technology works effectively without the need of pattern projectors. DISK was found to produce a significant number of strong features on box surfaces, while LightGlue improves the number of matched features on repetitive patterns.

Fig. 2: Color coded depth errors of the matched features derived from a ground truth. Displaying the match quality.
Own presentment

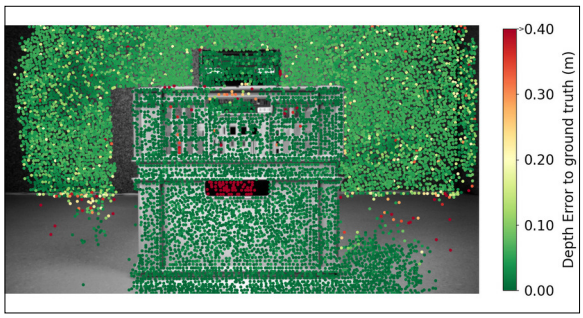


Fig. 3: Interpolated depth image of stacked crates in the foreground and a tower of crates in the background.
Own presentment

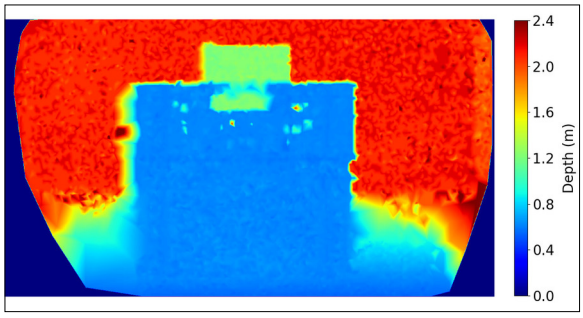
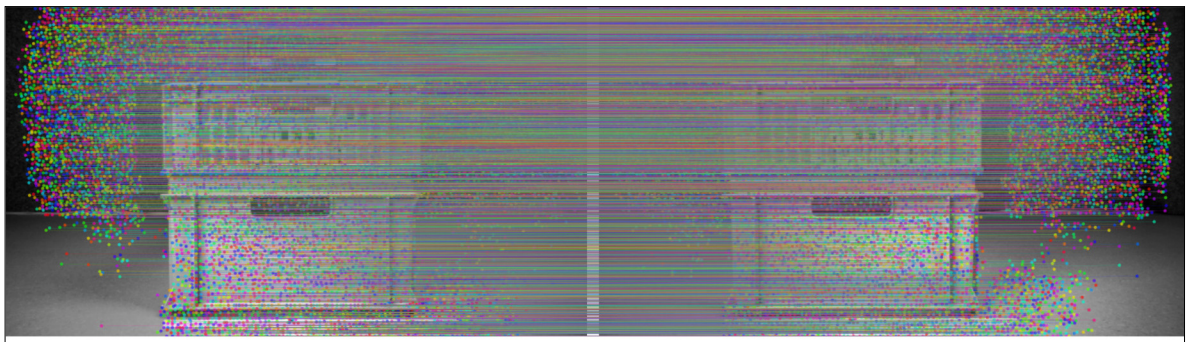


Fig. 1: Matching of features between a pair of stereo images, considering epipolar constraints.
Own presentment



Advisor
Prof. Dr. Martin Weisenhorn

Co-Examiner
Jonas Schmid, Stettbacher Signal Processing, Dübendorf, ZH

Subject Area
Image Processing and Computer Vision