Disparity Image Integration Method for 3D Mapping around Crane

Tomoki KOBAYASHI[*1], Junichi SUSAKI[1], Tetsuharu OBA[1], Yoshie ISHII[1], Marek OSOSINSKI[2]*

[1] Department of Civil and Earth Resources Engineering, Graduate School of Engineering, Kyoto University.
[2] Technical Research Institute, Tadano Ltd.

Recently, due to a severe shortage of manpower in construction sites, the automation of crane operation has been attracting attention. To realize this automation, it is essential to create an accurate 3D map around the crane. Considering the need to safely transport loads and the constantly changing conditions at construction sites, both accuracy and fast processing are important when generating a 3D map. Therefore, the objective of this research is to develop a method to generate 3D map in a highly accurate and semi-real-time way using video images taken from a monocular camera on the crane's boom. In this research, it is assumed that Controller Area Network (CAN) data, which includes information such as the swing degree of the crane's boom, is acquired by a sensor installed at the same location as the camera. Some previous studies have utilized Simultaneous Localization and Mapping (SLAM). However, SLAM performs depth estimation only for the feature points of keyframes, resulting in the issue of empty areas in the generated point cloud. In this study, we focus on the method of Semi-Global Matching (SGM) which generates disparity images between close frames. Disparity images refer to images that visualize the distance between objects and the camera based on the horizontal difference (disparity) of corresponding pixels detected in the rectified two images. In the disparity image, there may be pixels where disparity calculation fails due to occlusion or insufficient feature points. However, in this study, by referencing multiple aligned disparity images, multiple disparity candidate values can be prepared for each pixel, so we expected to obtain a dense 3D map compared to methods using SLAM. In our proposed method, we first extract frames from the input video images, apply SGM to the adjacent frames to generate multiple disparity images, and then align and integrate them to generate a final disparity image around the crane. We improved the accuracy by using CAN data. Specifically, we enhanced the accuracy of rectification by referring to the swing degree of the crane's boom between consecutive frames as the initial rotation angle for image rectification. We also detected and eliminated pairs with low rectification accuracy by comparing the matrices used for rectification with the boom swing degree data. Additionally, we achieved improved accuracy through projection transformation from central projection to orthographic projection, disparity alignment between disparity images, and integration by overlaying disparity images while taking the median value. As a result, we demonstrated the ability to generate dense a 3D map, which were difficult to obtain in previous research using SLAM, in semi-real-time processing. For example, the processing time is 45.7 seconds when applying to actual video images with a duration of 29.0 seconds. In the future, it will be necessary to apply our method to multiple types of video data accompanied by CAN data using actual crane machines or simulators in order to validate the applicability of the proposed method.

Keywords: photogrammetry, stereo matching, disparity image