AER1515 – Assignment 01

This project implements feature point detection and its matching between stereo pair images from KITTI dataset. For a given input RGB image from left camera, the features which are described to be an image region that is salient, local, repeatable, compact and efficient, are identified and studied by visual inspection for unreliability on matching. As the features are detected in its corresponding right camera image as well, we match the feature with a brute force matcher. For each match pair identified, the disparity between the x values are calculated, and they are converted into depth values using the predefined functions and calibration details provided. The calculated depth values are then documented as a text file and attached for evaluation on test images. A RANSAC based outlier rejection method has been implemented on the match pairs, and promising results have been noted. The method has been fine tuned on train images by evaluating the results for corresponding changes to its confidence and reprojection threshold values.

Feature Detection:

ORB (Oriented FAST and Rotated BRIEF) feature detector has been implemented for detecting features and descriptors on the given image. It is a combination of features from FAST keypoint detection and descriptors from Binary Robust Independent Elementary Features (BRIEF) method. It is inevitable to note the presence of few (to many) unreliable keypoints being detected by the algorithm. For example, in images 2, 4 and 5 on the training set, it can be noted that the keypoints identified on the branches of trees are not reliable. They are an effect of illumination changes and are more prone to external disturbances. So, these keypoints are failed to be matched between the descriptor pair. Images 1 and 3 from training are also interesting examples for the effect of illumination on keypoint identification, where in image 3, almost all keypoints detected on the gradient between the bright and shadow regions of the van, are eliminated as outliers through our RANSAC approach.

Feature Matching:

Brute Force (BF) based feature matching is implemented. A descriptor of one feature of the left image is chosen initially and it is evaluated with all features of the right image, by calculating the binary hamming distance function between them. By setting ‘`crossCheck = True`’, features from the right images are also evaluated with left images and this OpenCV implementation is robust by filtering most of the unreliable keypoints during its operation. The selected matches are visualized on the image pairs and the count of the matches returned are also displayed. Given the images are rectified, then only the horizontal matches are displayed by equating their corresponding y values.

For each chosen matches, its disparity is calculated and used to determine depth by utilising the provided calibration data and functions. The depth map for chosen matches on the test images are attached in the requested format as a part of submission.

Outlier Rejection:

Detecting and removing outliers are of paramount importance of the vision task, as it may lead to many false positive results. RANSAC (RAndom SAmple Consensus) is applied as the outlier rejection method, by exploiting its implementation on OpenCV's `findHomography` built in function. Given the co-ordinates of matches in the corresponding left and right keypoints array, RANSAC can be applied between them to return a homography matrix and a status array that holds the value of either '0' or '1' which determines whether the particular match has been identified as an inlier by the RANSAC algorithm. Given only matching keypoints with status '1' (i.e., an inlier), outliers are rejected from the initially detected and matched features. An interesting observation would be to match only the outliers identified by RANSAC, between the image pairs. This would provide more details and insights about unreliable keypoints which were detected as our initial step. This will be a future work considering the scope of the assignment.

RANSAC is fine-tuned on the training dataset, by changing the values of the maximum number of iterations, the confidence score and the threshold value of the reprojection error (C). Considering 2000 maximum iterations with OpenCV implementation, 1600 has been set. The confidence score is set to be 0.99 and ransacReprojThreshold is set 7 after many iterations of evaluation on the training data, which is the maximum allowed reprojection error to treat a point pair as inlier. This can be written as,

$$\|\texttt{dstPoints}_i - \texttt{convertPointsHomogeneous}(\texttt{H} * \texttt{srcPoints}_i)\|_2 > \texttt{ransacReprojThreshold}$$

where H is the determined Homography matrix.

References:

[1]. Tutorial Image Feature Extraction and Matching, Kaggle kernel notebook, https://www.kaggle.com/wesamelshamy/tutorial-image-feature-extraction-and-matching.

[2]. Introduction to Image Understanding, class notes, Prof. Sanja Fidler, UTM.

[3]. Generating Dense Disparity Maps using ORB Descriptors, 2016, https://sourishghosh.com/2016/dense-disparity-maps-orb-descriptors/

[4]. Camera Calibration and 3D Reconstruction, OpenCV document, https://docs.opencv.org/master/d9/d0c/group__calib3d.html#gafd3ef89257e27d5235f4467cbb1b6a63