Title goes here

Anonymous CVPR submission

Paper ID Occular Tension

Abstract

Talk about overall problem of vision-based SLAM. Then give 1 sentence summary of each paper. Then tell that we've implemented one and show our results.

1. Introduction

Introduction goes here.

2. Analysis of Robust Large Scale Monocular Visual SLAM

2.1. Problem Statement

This paper focuses on the problem of Visual Simultaneous Localization And Mapping (VSLAM) which estimates a 3D model or map of the environment in which a camera moves along a trajectory in [?]. Moreover, the paper focuses on using calibrated monocular cameras and making the algorithm robust, accurate, and scalable. Monocular VSLAM comes with the added challenge of not being able to observe the scale of the scene of the environment. In order to overcome this, loop closures (which occur when the camera returns to a previously observed location) need to be detected. This is an issue in large environments where many scenes look alike, and results in an erroneous 3D model if loop closures are not detected properly. Thus, the authors of the paper intend to focus not only on the general problem of monocular VSLAM but also tackle a key subproblem of dealing with loop closure.

2.2. Innovative Contribution

To solve the problem of monocular VSLAM, the authors propose a framework consisting of three parts:

1. A Structure from Motion algorithm based on the *Known Rotation Problem* [?] is used to estimate submaps which are parts of the camera trajectory and the unknown environment [?],
2. A loopy belief propagation algorithm is used to efficiently aligns many submaps based on a graph of relative 3D similarities to produce a global map that is consistent up to a scale factor.
3. An outlier removal algorithm that detects and removes outliers in the relative 3D similarity graph is used to reject wrong loop closures.

2.3. Proposed Method

The paper proposes a four-part framework to implement the innovations that solve monocular VSLAM: keyframe selection, submap reconstruction, pairwise similarity estimation, and large scale relative similarity averaging.

2.3.1 Keyframe Selection

From the set of all frames in the video captured by the monocular camera, only a subset is used as camera motion input for VSLAM to give it a reasonable computational complexity. For each frame, Harris Points of Interest (PoI) are detected and tracked using a Lucas-Kanade tracker. When the Euclidean distance between the PoI of the current frame and previously selected keyframe is greater than a specified threshold, which is usually 5% of the image width, the frame is selected as a keypoint.

2.3.2 Submap Reconstruction

2.4. Experimental Evaluation

2.5. Subsequent Conclusions