Vision based terrain reconstruction for planet rover using a special binocular bundle adjustment*

Min-yi SHEN1,2, Zhi-yu XUANG†1,2, Ji-lin LIU1,2

(1) Institute of Information and Communication Engineering, Zhejiang University, Hangzhou 310027, China
(2) Zhejiang Provincial Key Laboratory of Information Network Technology, Zhejiang University, Hangzhou 310027, China
†E-mail: xiangzy@zju.edu.cn
Received Nov. 14, 2007; revision accepted Feb. 18, 2008

Abstract: This paper presents a pure vision based technique for 3D reconstruction of planet terrain. The reconstruction accuracy depends ultimately on an optimization technique known as ‘bundle adjustment’. In vision techniques, the translation is only known up to a scale factor, and a single scale factor is assumed for the whole sequence of images if only one camera is used. If an extra camera is available, stereo vision based reconstruction can be obtained by binocular views. If the baseline of the stereo setup is known, the scale factor problem is solved. We found that direct application of classical bundle adjustment on the constraints inherent between the binocular views has not been tested. Our method incorporated this constraint into the conventional bundle adjustment method. This special binocular bundle adjustment has been performed on image sequences similar to planet terrain circumstances. Experimental results show that our special method enhances not only the localization accuracy, but also the terrain mapping quality.

Key words: 3D reconstruction, Binocular bundle adjustment (BBA), Scale-invariant feature transform (SIFT), Re-projection error, RANSAC
doi:10.1631/jzus.A0720057

INTRODUCTION

In planetary exploration missions, high-precision landing-site topographic information is crucial for engineering operations and the achievement of scientific goals. In particular, large-scale landing-site mapping will be extremely important for current and future landing missions such as the Mars Exploration Rovers (Di et al., 2004). Several systems have been proposed (Li et al., 2002; Di et al., 2005; Labrie and Hebert, 2007) for the Mars Exploration Rover Mission. In their implementation, the initial location and heading information of each rover was provided by the telemetry data acquired by onboard sensors. The onboard navigation system consists mainly of an IMU, an odometer, and some solar imaging cameras. Local rover localizations on the landing site were based on the onboard navigation sensors. The rover automatically estimates its position using wheel odometry and IMU data. A visual odometry experiment will improve localization accuracy by overcoming problems associated with wheel odometry such as slippage and low accuracy. Finally bundle adjustment (BA) was applied and constraints were imposed on the projection model or movement of the camera between images, resulting in high precision landing site topographic mapping products.

BA was originally used in photogrammetry (Wolf and DeWitt, 2000). It is basically a steepest-descent algorithm that searches for an optimal model by minimizing the error between the observed 2D feature points and the re-projected feature points from the reconstructed model. Most of the 3D reconstruction systems available today use a turntable to rotate an object to capture its images from different...
view points (Wong and Chang, 2004; Labrie and Hebert, 2007). For these systems, the object is constrained to rotate around a fixed axis. By incorporating this constraint into the traditional BA method, a more accurate reconstruction model can be obtained.

This paper presents a pure vision based terrain reconstruction for a planet rover using SIFT (scale-invariant feature transform) features and a special binocular bundle adjustment (BBA). In contrast to the work in (Li et al., 2004), which still needs manual help for feature matching, a SIFT feature has the characteristics of scale and rotation invariant and it guarantees a full automatic match between the projected features produced by the same 3D point. We initialized the localization process using stereo vision on some stereo images to obtain the structure and motion without INS (inertial navigation system) data. As shown in Fig.1, after merging local 3D structure models into a common coordinate system, a final BBA can be applied to the whole sequence. This paper aims to study how reconstruction accuracy could be improved by employing constraints inherent between the binocular views in an efficient way.

The remainder of this paper is structured as follows. First, we briefly describe the detection and tracking of features in Section 2. Then we show how to initialize the structure and motion without INS data in Section 3. In Section 4 we present the special BBA based pose estimation, using both photometric and geometric constraints. Section 5 shows how to merge all local 3D models to obtain a final 3D model and describes the experimental results of this work. Finally, we conclude the paper in Section 6.

FEATURE TRACKING

We use the SIFT features. This approach has been named the ‘scale-invariant feature transform’ (Lowe, 2004) as it transforms image data into scale-invariant coordinates relative to local features. The SIFT is invariant to translation, scaling (Lowe, 1999), and rotation. It is also partially invariant to illumination variations as well as affine for 3D projection. These features locate interest points at maxima/minima of a difference of Gaussian function in scale space. Each interest point has an associated orientation, which is the peak of a histogram of local orientations. The resulting feature descriptor, which captures the orientation information of the local image region, contains 128 elements. The features are highly distinctive, in the sense that a single feature can be correctly matched with high probability against a large database of features. All the feature points can be automatically selected.

The best candidate match for each keypoint is found by identifying its nearest neighbor in the database of keypoints from training images. The nearest neighbor is defined as the keypoint with minimum Euclidean distance for the invariant descriptor vector. We perform feature space outlier rejection to remove incorrect matches. It has been found that comparing the distance of a potential match to the distance of the best incorrect match is an effective strategy for outlier rejection (Brown et al., 2005). Only the verified candidates are accepted as inliers.

Suppose that the match distance of the second closest neighbor is $d_{sc}$, as it is the best matching outlier. In order to verify a match, we compare the match distance of a potentially correct match $d_c$ to the outlier distance, accepting the match if

$$d_c < r_d \cdot d_{sc}. \quad (1)$$

Typically the distance ratio $r_d$ is set as 0.8.

Fig.2 shows an example of matched SIFT features in frames.