TLD Based Visual Target Tracking for Planetary Rover Exploration

Chongyang Zhang, Kaichang Di, Zhaoqin Liu, and Wenhui Wan

Institute of Remote Sensing Applications, Chinese Academy of Sciences
P.O. Box 9718, Datun Road, Chaoyang District, Beijing 100101, P.R. China
(zhangchy,kcdi,liuzq,whwan)@irsa.ac.cn

Abstract. Visual target tracking is one of the key technologies to implement full automatic exploration for a planetary rover and improve exploration efficiency. A novel visual tracking system is developed based on the Tracking-Learning-Detection (TLD) algorithm in combination with stereo image matching to achieve 3D tracking of a science target. Experimental results using stereo image sequences demonstrate the excellent performance of TLD tracking and the overall effectiveness of the 3D tracking.

Keywords: visual target tracking, planetary rover exploration, TLD.

1 Introduction

During the Mars Exploration Rover (MER) and the Mars Science Laboratory (MSL)missions, as shown in Fig.1, it usually takes a minimum of 3 cycles(3 sols, a sol is a Marian day,24 h 39 min 35 s) for a mars rover to approach a designated science target and place instruments for in situ exploration[1]. In the ongoing and future planetary rover exploration missions, to command the rover to approach a science target more efficiently, e.g., within a single command cycle, the rover should be capable of reliably tracking the target tens of meters away, locking it while traversing the rough terrain and avoiding obstacles. Visual target tracking, along with camera handoff, is one of the key technologies to accomplish single cycle target approach and instrument placement, which will significantly improve the efficiency of rover exploration.

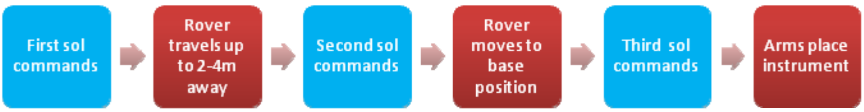


Fig. 1. 3-cycles of Mars rover target approach and instrument placement

Severalvisual tracking methods for planetary rovers have been researched and havebeen demonstrated on the Marsokhod rover at Ames Research Center[2]and on Rocky 7 Rover at the Jet Propulsion Laboratory[3]. TheMarsokhod tracker used the sign of the difference of Gaussian(SDOG) to match the target templates to new images. The Rocky 7 tracker used three-dimensional information from stereo images combined

T. Tan et al. (Eds.): IGTA 2013, CCIS 363, pp. 29–35, 2013.

with intensity information. Affine tracker was proposed for tracking at multiple image resolutions[4],but extensive tests showed that itwas unreliable when image changes between frames were large and prior knowledge of the pose was not supplied[5]. The final technology infused into MER Flight Software was Normalized Cross Correlation (NCC) matching with template image magnification and roll compensation[6].

To reduce mission risk, the flight project operation team tends to choose existing mature technologies rather than new approaches. Due to the limited computing resource on the rover, the tracking algorithm executed onboard was simple[6]. However, for academic research on earth, it is possible to explore more sophisticated method to achieve better performance, which may be applicable in the future rover missions.

Tracking-Learning-Detection(TLD) methodis a novel algorithm that integrates adaptive tracking with online learning of the object-specific detector[7]. It achieves real-time target tracking and is resistant to occlusions and appearance changes. When the designated target goes out of the camera view and comes backagain, TLD is able to detect and relock the target.

In this paper, we investigate the TLD method and combine it with stereo image matching for 3D tracking of science target on a planetary rover. Stereo image sequences are used to test the effectiveness of the combined target tracking method, which will be helpful for target approach within one command cycle in future planetary rover missions.

2 Tracking-Learning-Detection

TLD decompose the long-term tracking task into three components: tracking, learning and detection. Each of the three components deals with different aspect of the problem; the components are running in parallel and are combined in a synergetic manner to suppress their drawbacks[7].

The tracking component is based on Median-Flow tracker. The target is designated by a bounding box then flow motion between consecutive frames is estimated using Lucas-Kanade method[8]. The tracker estimates the displacements and scales of a number of points within the object-defined box, estimates their reliability and votes with 50% of the most reliable displacements for the object motion using median[9].

Tracking failures are automatically detected by calculating Forward-Backward errors. First the tracker produces a trajectory by tracking the points forward, and then the points located in the last frame are tracked back, generating abackward tracking trajectory. If the inconsistency between the two trajectories is larger than a threshold, the forward tracking result is considered incorrect[10].

The main purpose of the learning component is to update the object detector in runtime after initialized in the first frame. A novel learning paradigm called P-N learning, or alternatively explained as growing and pruning, was proposed to solve the problem[11]. P-expert(growing event) discovers the object and thus increases generalization of the object detector. More positive samples are added to the online model represented by a set of 15*15 intensity normalized patches. N-expert(pruning event) generates negative training examples. Its goal is to discover clutter in the background that the detector should remove from the model. The P-N learning establishes a positive-negative feedback system, which maintains the tracking system stable and reliable.

The task of detection component is to localize patches contained in the online model and to efficiently adjust its decision boundary by the growing and pruning events[9]. The object detector is based on 2bit Binary Patterns(2bitBP), which measure gradient orientation in a certain area and output four possible codes. The detection classifier has the form of a randomized forest. The forest consists of several trees. Each tree is built from one group of features. If the patch reaches the end of the tree, it is considered positive; otherwise if it can't reach the end, the patch is treated as negative. Evaluation of an unknown patch by the tree is very efficient. The final decision is obtained by the majority vote. Even though the classification result for a single tree may not be often correct, the final decision made by the whole forest is quite reliable.

Given an image sequence and an initial bounding box of the target defined in the first frame, the TLD procedure can be briefly described as follows.

TLD.Init()
  Read the first frame;
  buildGrid(): generate all possible scales and shifts of the initial bounding box with the following parameters: scales step=1.2 and horizontal/vertical step= 10% of width/height;
  getOverlappingBoxes():find the boxes that havemore than 60% overlap with theinitial box andgenerate positive samples, others are considered as negative samples;
  classify():describe the features based on 2bitBP, and classify thesamplesusing randomized forest.
TLD.processFrame()
for each frame
  lucasTracker():track points using Lucas-Kanade method and eliminate errors;
  bbPredict(): predict the bounding box based on the correctly tracked points;
  tld.detection(): classify the object and background using the former trained classifier, update the classifier and detector;
  tld.learning(): generate new positive and negative samples, update online model of the object.
end

Through tracking, learning and detection, the TLD method achieves state-of-the-art performancefor tracking the object in a long term, with the object appearance changes and moves in and out of the camera view. It has a promising application to the visual target tracking for a planetary rover. More details about the TLD method could be found in[7,9-12].

3 TLD-Based Rover Tracking Framework

Due to the remarkable performance of the TLD tracking method, we apply it to the visual target tracking system for a planetary rover. Fig. 2 shows the framework of the 3D tracking system. For a stereo pair of cameras equipped on a rover, first a target is designated in the left camera view, and then the feature points within the target bounding box are tracked based on TLD.

For those points that are tracked correctly in the left image, corresponding points on the right image are matched through NCC simultaneously.To improve the speed and reliability of the NCC stereo matching, epipolar constraint of the stereo camera is applied. From the corresponding images points, 3D positions of the points on the target are calculated through space intersection. During the rover driving, according to the position and orientation between the stereo camera and the target, camera pointing or camera handoff may be necessary to keep the target in the camera views or transfer the target to another stereo camera, e.g., from mast-based navigation camera to body-based hazard avoidance camera.

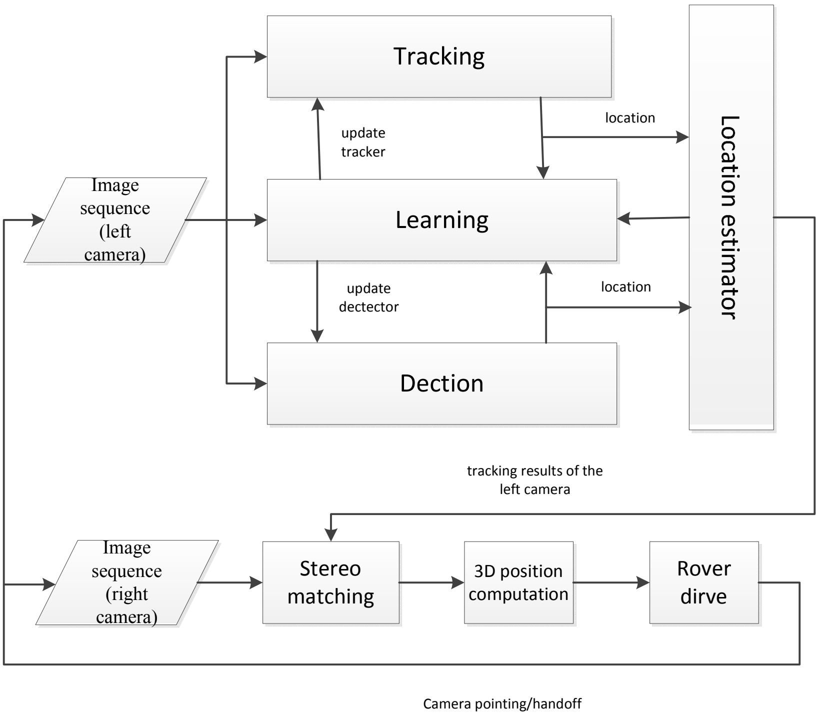


Fig. 2. The work flow of TLD-based 3D visual tracking system for rover exploration

4 Experimental Results

4.1 Experimental Data

The 3D tracking method has been tested with a number of image sequence datasets, two of which are reported here. Seq.1 is a 3D video captured by a Finepix Real 3D cameraat a deserted field near Thirteen Ming Tombs.Seq.2 was taken bya stereo camera mounted on a model roverat a field test in TianMo, a small desert near Beijing, whose terrestrial environment is similar to a planetary terrain(Fig.3).

For each test, the stereo camera started acquiringimage sequences from 10 to15 meters away to the target(rock);the tracking procedure was performed while the cameras approached the target at the speed of several centimeters per second. The tracking procedure ended when the distance between the camera and the target is 1m, i.e., within the workplace of the robotic arm. Both image sequences contain the situation that the target moves out of the camera view. For Seq.1, the stereo camera captured the image sequences at 20frames per second with an image size of 1240*720 pixels; for Seq.2, the rover stereo camera captured the image sequences at 2 frames per second with an image size of 1392*1040 pixels. The whole system is implementedinMatlabbased on the open source TLD code and runs on a regular PC, with Intel Core(TM)2 Duo CPU (3.00GHz) and 4.00GB RAM. The system achieved a real-time performance at 20 stereo frames per second.

Fig. 3. Field test of visual tracking for target approach on a model rover

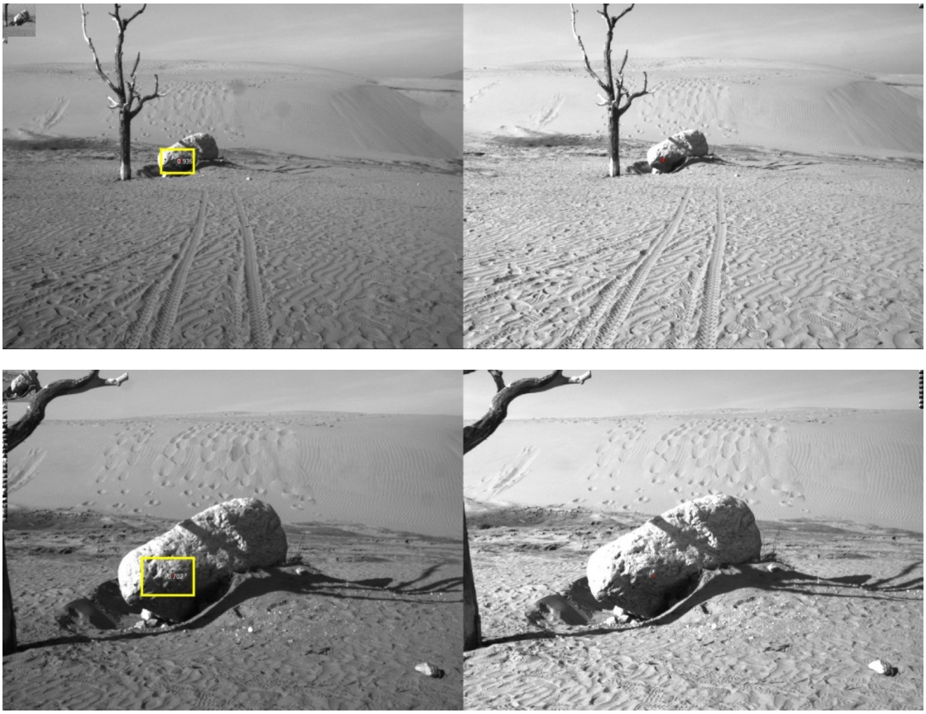
4.2 Results and Discussion

After a science target (rock) was manually defined by a bounding box on the first frame of the image sequences respectively, the TLD method successfully trackedand matched on all the frameswhere the target appear.If the target was out of the image, itwas detected correctly and being tracked again when the target reappeared in the frame.The tracking performance is listed in Table 1 and shown in Fig.5.The rectanglesin Fig.5 indicate the tracked target whose scale changes while approaching the target. The points on the selected target are those tracked correctly from the previous frame of the left camera image sequences and successfully matched on the frames of the right camera.Overall, the 3D tracking is stable and reliable.

Table 1. Tracking results of the two image sequences

	No. of frames	Out of camera views	Successfully tracked
Seq.1	760	17	743
Seq.2	470	8	462

(a) First and last stereo frames with tracked rock in Seq.1



(b) First and last stereo frames with tracked rock in Seq.2

Fig. 4. The tracking performances on the two datasets

5 Summary and Future Work

We combined the TLD method and NCC matching forvisual target tracking in planetary rover exploration. The target is tracked in the left camera image sequences and corresponding pointsarematched in the images of the right camera to achieve 3D tracking. Test results using 2 datasets demonstrated the excellent performance of TLD tracking and the overall effectiveness of the 3D tracking. In the future, we will further improve the 3D tracking method, e.g., by utilizing 3D information during TLD tracking. More tests will be performed to validate the tracking system so that the method could be used for single cycle target approach in future planetary rover exploration.

Acknowledgements. The research conducted in this paper was supported by National Basic Research Program of China (2012CB719902) and National Natural Science Foundation of China(41171355).

References

1. Bajracharya, M., Diaz-Calderon, A., Robinson, M., et al.: Target tracking, approach, and camera handoff for automated instrument placement. In: IEEE Aerospace Conference. IEEE Press, Montana (2005)
2. Wettergreen, D., Thomas, H., Bualat, M.: Initial results from vision-based control of the Ames Marsokhod rover. In: IEEE International Conference on Intelligent Robots and Systems, pp. 1377–1382. IEEE Press, Grenoble (1997)
3. Nesnas, I.A.D., Maimone, M.W., Das, H.: Autonomous vision-based manipulation from a rover platform. In: IEEE International Symposium on Computational Intelligence in Robotics and Automation. IEEE Press, Monterey (1999)
4. Nesnas, I., Bajracharya, M., Madison, R., Bandari, E., et al.: Visual target tracking for rover-based planetary exploration. In: IEEE Aerospace Conference. IEEE Press, Big Sky (2004)
5. Kim, W., Steele, R., Ansar, A., et al.: Rover-based Visual Target Tracking Validation and Mission Infusion. In: AIAA Conference on Space (2005)
6. Kim, W., Biesiadecki, J., Ali, S.: Visual target tracking on the Mars Exploration Rovers. In: International Symposium on Artificial Intelligence: Robotics and Automation in Space, Los Angeles, California (2006)
7. Kalal, Z., Matas, J., Mikolajczyk, K.: Tracking-Learning-Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 34(7), 1409–1422 (2012)
8. Lucas, B.D., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: International Joint Conference on Artificial Intelligence, pp. 674–679 (1981)
9. Kalal, Z., Matas, J., Mikolajczyk, K.: Online learning of robust object detectors during unstable tracking. In: IEEE 12th International Conference on Computer Vision. IEEE Press, Kyoto (2009)
10. Kalal, Z., Mikolajczyk, K., Matas, J.: Forward-Backward Error: Automatic Detection of Tracking Failures. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 23–26. IEEE Press, Istambul (2010)
11. Kalal, Z., Matas, J., Mikolajczyk, K.: P-N Learning: Bootstrapping Binary Classifiers by Structural Constraints. In: IEEE Conference on Computer Vision and Pattern Recognition. IEEE Press, San Francisco (2010)
12. Kalal, Z., Mikolajczyk, K., Matas, J.: Face-TLD: Tracking-Learning-Detection Applied to Faces. In: IEEE International Conference on Image Processing. IEEE Press (2010)