

생성적 적대 신경망을 이용한 행성의 장거리 2차원 깊이 광역 위치 추정 방법

Planetary Long-Range Deep 2D Global Localization Using Generative Adversarial Network

아하메드 엠.나기브¹ · 투안 아인 뉴엔² · 나임 울 이슬람³ · 김재웅⁴ · 이석한[†]
Ahmed M.Naguib¹, Tuan Anh Nguyen², Naeem Ul Islam³, Jaewoong Kim⁴,
Sukhan Lee[†]

Abstract: Planetary global localization is necessary for long-range rover missions in which communication with command center operator is throttled due to the long distance. There has been number of researches that address this problem by exploiting and matching rover surroundings with global digital elevation maps (DEM). Using conventional methods for matching, however, is challenging due to artifacts in both DEM rendered images, and/or rover 2D images caused by DEM low resolution, rover image illumination variations and small terrain features. In this work, we use train CNN discriminator to match rover 2D image with DEM rendered images using conditional Generative Adversarial Network architecture (cGAN). We then use this discriminator to search an uncertainty bound given by visual odometry (VO) error bound to estimate rover optimal location and orientation. We demonstrate our network capability to learn to translate rover image into DEM simulated image and match them using Devon Island dataset. The experimental results show that our proposed approach achieves ~74% mean average precision.

Keywords: Global Localization System, Conditional Generative Adversarial Network

1. Introduction

Autonomous navigation is a necessity for planetary rover to be able to traverse long-range distances. Although there are numerous localization techniques developed to assist rover autonomous navigation, such as Visual Odometry (VO) and wheel odometry, they often suffer from growing error due to the lack of absolute reference. Filtering algorithms and bundle adjustments can effectively reduce the growing error, but they cannot totally eliminate it. Thus, for a long-range mission, a global localization algorithm is needed in which the location and orientation of the rover is estimated with respect to global absolute reference such as planetary inertial frame, Universal Transverse Mercator (UTM) frame, Topocentric frame, ... etc. There are many research approaches that address the problem of planetary global localization. Multi-frame Odometry-compensated

Received : Jan. 15. 2018; Revised : Feb. 11. 2018; Accepted : Feb. 13. 2018

※ This research was supported in part by the "Space Initiative Program" of National Research Foundation (NRF) of Korea (NRF-2013M1A3A3A02042335), sponsored by the Korean Ministry of Science, ICT and Planning (MSIP), in part, by the "3D Visual Recognition Project" of Korea Evaluation Institute of Industrial Technology (KEIT) (2015-10060160), in part, by the "Robot Industry Fusion Core Technology Development Project" of KEIT (R0004590), and in part by, the KIAT under the Robot industry fusion core technology development project (R0004590) supervised by the KEIT. This research was also supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (2017R1A6A3A11036554).

※ The original ideas presented in this paper are suggested by Prof. Sukhan Lee

1~4. Intelligent Systems Research Institute, School of Information and Communication Engineering Sungkyunkwan University, Suwon, Korea (ahmed.m.naguib@gmail.com, tuanh2193@gmail.com, naeem.rhamza@gmail.com, create@skku.edu)

† Corresponding author: Intelligent Systems Research Institute, School of Information and Communication Engineering Sungkyunkwan University, Suwon, Korea (lsh1@skku.edu)

Global Alignment (MOGA) ^[1] uses LIDAR data and match it to Digital Elevation Map (DEM). LIDAR is often too heavy and power demanding to deploy on a rover. 3D stereo reconstruction, on the other hand, typically generates reliable point cloud up to 40m from the rover. With DEM resolution of about 2m per pixel, the accuracy global localization by matching 3D features from a single frame becomes limited. In order to accurately localize rover from a single frame, 2D images based matching is needed, since mountains, craters, skylines, and other 2D features are not mostly limited by distance.

Stein et al. ^[2] matches skyline in rover 2D image with that rendered from DEM to estimate rover global pose. Similarly, Visual Position Estimator for Rovers (VIPER) ^[3, 4] matches skyline in rover 2D image with that rendered from DEM using local geometrical shape features such as mountain peaks and shapes. Li Wei, et al. ^[5] solves the same problem as VIPER. They used more robust skyline detection by major line detection ^[6] method. Strong skyline was used with a more robust Bayesian Network for matching. Both approaches, however, suffers from low DEM resolution as well as artifacts in both DEM rendered images, and/or rover 2D images due to illumination variations and small terrain features as shown in [fig. 1]. Yicong Tian, et al. ^[7] shows that deep CNN outperforms traditional approaches in matching local and global images under variations. They use Faster R-CNN to match buildings in street-level image and geo-tagged tilted aerial photos for the purpose of geo-localization in urban environment. Similarly, Lin et al. ^[8] use CNN pretrained on ImageNet ^[11] and Places ^[12] to match google street view with geo-tagged tilted aerial photos. Workman et al. ^[9], on the other hand, use CNN pretrained on Places ^[12] to match street-level image with ortho-maps for estimating optimal

geo-location.

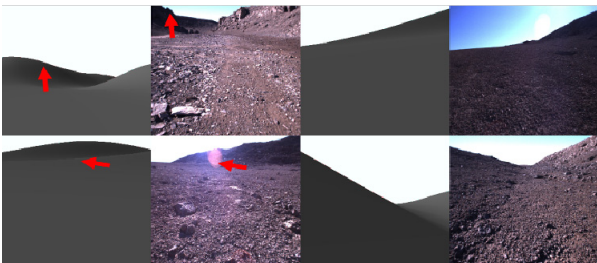
In this paper, we propose a novel Planetary Long-Range Deep 2D Global Localization Using Generative Adversarial Network. Our objective is to search a given space defined by the error bound of VO to find rover location. We divide the 6DOF space into 6D grid cells, each with specific pose for rover camera. For each cell, we simulate a virtual camera with a cell's pose and render virtual 2D image. While simulated image does not reflect local features that are too small to be seen by DEM, the general underlying terrain and skyline are mostly preserved. We match this simulated image from the cell with true image captured by rover camera. In this case, we use initial matching by extracted 2D major lines ^[6] in both images. We also obtain matching score for each cell and choose the optimal cell as rover 6D Pose.

2. Long-Range Deep 2D Global Localization Network

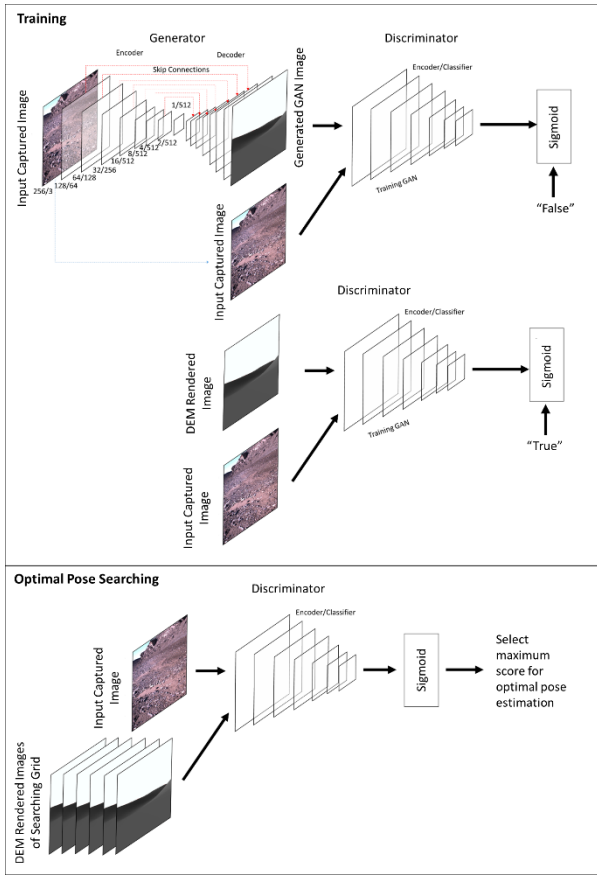
Uncertainties in captured rover images due to local geometrical shapes that are too small to appear in DEM, and illumination variations may lead to difficulties in matching images using conventional approaches. To overcome this challenge, we propose Long-Range Deep 2D Global Localization. We use conditional generative adversarial network (cGAN) to generate “fake” DEM-simulated image corresponding to a captured rover image. We then use this “fake” pair against “true” pair of rover image and “true” DEM-simulated images to train a CNN discriminator to distinguish between the fake and true pairs, effectively matching the true correspondences.

2.1 Generator

Similar to image-to-image translation ^[10], Generator (G) of cGAN uses “U-Net” to generate new images, this is a 15-layers encoder-decoder with skip connections between mirrored layers in the encoder and decoder stacks. Input image is scaled to 256x256x3 and output image is 256x256. Each convolutional layer of the encoder uses stride of 2 to cut the spatial resolution by half. The resolution and number of filters per layer are as follows: {128/64, 64/128, 32/256, 16/512, 8/512, 4/512, 2/512, 1/512}. The decoder is a mirrored structure of the encoder with concatenated input of skip connections from each corresponding layer. We used leaky ReLU activation for all the layers and batch



[Fig. 1] Examples of artifacts that pose a challenge in matching rover images with DEM rendered images. Top Left and Bottom Right: example of rocks in rover image too small to appear in DEM. Bottom Left and Top Right: examples of illumination variations such as lens flare and shallow dynamic range



[Fig. 2] Architecture of Long-Range Deep 2D Global Localization network. Top: Training phase. Bottom: Classification phase

normalization for all but the first layer. The structure of the generator is shown in [fig. 2].

2.2 Discriminator

Discriminator (D) uses a 7-layers CNN to discriminate real and fake images' pairs. Input pair are concatenated and each layer cuts the resolution by half. The resolution and number of filters per layer are as follows: {128/64, 64/128, 32/256, 16/512, 8/512, 4/512, 1/1}. Finally, a unit sigmoid is used to estimate matching probability between the paired input images. Similar to the generator, we used leaky ReLu activation for all the layers and batch normalization for all but the first layer. The structure of the generator is shown in [fig. 2].

2.3 Training Procedure

Let's assume we have a dataset of pair samples:

$\{x_i, y_i\}_{i=1}^N: x_i \in A, y_i \in B$ where A and B are rover and DEM simulated images respectively. For every mini-batch of paired samples, we use back propagation to minimize generator's, and maximize discriminator's log likelihood functions:

$$\begin{aligned} \mathcal{L}_{L1}(G) &= E_{x \sim P_{data}(x), z \sim P_z(z)} [\|y - G(x, y)\|_1] \\ \mathcal{L}_{cGAN}(G, D) &= E_{x, y \sim P_{data}(x, y)} [\log D(x, y)] \\ &\quad + E_{x \sim P_{data}(x), z \sim P_z(z)} \left[\log (1 - D(x, G(x, z))) \right] \end{aligned}$$

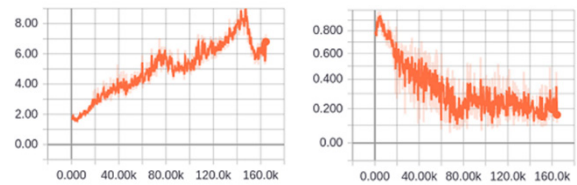
Where $\mathcal{L}_{L1}(G)$, $\mathcal{L}_{cGAN}(G, D)$ are log likelihood functions of generator and discriminator respectively, $G(x, y)$ is generated "fake" DEM image by generator, and $D(x, y)$ is discriminator output probability of matching x to y.

3. Experimental Results

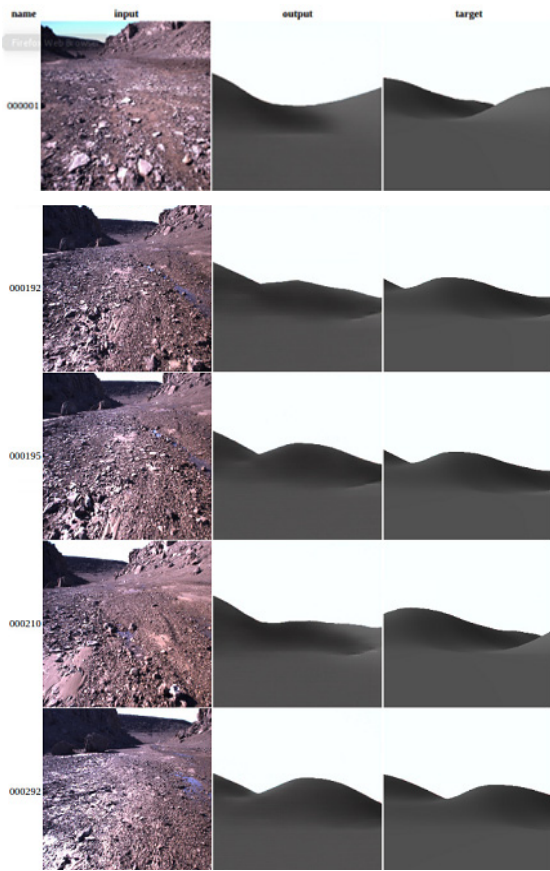
We used Devon Island dataset^[13] bundle 1 which includes 2056 images with ground truth location. For each image, we used OpenGL based renderer to generate DEM-simulated image at ground truth location. We divided the set into 80%/20% training and testing samples. We trained cGAN for 80 epochs on the training samples.

The evolution of the log likelihood function can be seen in [fig. 3] and the examples of generator output can be seen in [fig. 4].

We then assumed a VO error bound of 100m around each ground truth location. We divided the error bound into 100x100x180 Tx, Ty, Rz grid cells. For each cell, we rendered DEM-simulated images. We used discriminator to evaluate each cell and the best match is selected as optimal estimated global location of the rover. [Table 1] shows the accuracy of proposed algorithm as well as L1-loss between estimated location and ground truth.



[Fig. 3] Loss function (inverse of log likelihood function) evolution with training iterations. Left: Discriminator loss function. Right: Generator loss function



[Fig. 4] Examples of input rover images (left), generator's "fake" DEM simulated images (middle), and actual "true" DEM simulated images (right)

[Table 1] Performance of proposed algorithm on bundle 1 of Devon Island Dataset

	Training Data	Testing Data
Localization Accuracy	95.41%	73.92%
Mean L1-loss	8.89	35.23

4. Conclusion

In this paper, we proposed a novel algorithm for global localization of a planetary rover. We used cGAN to train a CNN discriminator to extract deep representations of rover 2D image and DEM rendered images; and use them for estimating a matching probability. To overcome challenging artifacts of rover image and/or DEM rendered images, we used a generator to generate "false" DEM simulated images and help train discriminator further. This approach elevates the requirement for massive training dataset which may not be attainable for

interplanetary environments. We divide pose searching space defined by VO error bound into a grid and query rover optimal pose by evaluating the discriminator at each cell using rover image and DEM rendered image at this cell perspective. Experimental results demonstrate the training performance, generated samples, and accuracy of our proposed approach.

References

- [1] P. J.F. Carle, P. T. Furgale, and T. D. Barfoot, "Long-range rover localization by matching LIDAR scans to orbital elevation maps," *Journal of Field Robotics*, vol. 27, no. 3, pp. 344-370, 2010.
- [2] F. Stein and G. Medioni, "Map-based localization using the panoramic horizon," *IEEE Transactions on Robotics and Automation*, vol. 11, no. 6, pp. 892-896, 1995.
- [3] F. Cozman and E. Krotkov, "Automatic mountain detection and pose estimation for teleoperation of lunar rovers," *International Conference on Robotics and Automation*, Albuquerque, NM, USA, pp. 2452-2457, 1997.
- [4] F. Cozman, E. Krotkov, and C. Guestrin, "Outdoor visual position estimation for planetary rovers," *Autonomous Robots*, vol. 9, no. 2, pp. 135-150, September, 2000.
- [5] L. Wei and S. Lee, "3D peak based long range rover localization," *2016 7th International Conference on Mechanical and Aerospace Engineering*, London, UK, pp. 600-604, 2016.
- [6] J. Kim and S. Lee, "Extracting major lines by recruiting zero-threshold canny edge links along sobel highlights," *IEEE Signal Processing Letters*, vol. 22, no. 10, pp. 1689-1692, October, 2015.
- [7] Y. Tian, C. Chen, and M. Shah, "Cross-View Image Matching for Geo-localization in Urban Environments," *arXiv:1703.07815*, 2017.
- [8] T.-Y. Lin, Y. Cui, S. Belongie, and J. Hays, "Learning deep representations for ground-to-aerial geolocation," *2015 IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, pp. 5007-5015, 2015.
- [9] Workman, Scott, Richard Souvenir, and Nathan Jacobs. "Wide-area image geolocation with aerial reference imagery." *Proceedings of the IEEE International Conference on Computer Vision*. P, 3961-3969, 2015.
- [10] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, pp. 5967-5976, 2017.
- [11] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211-252, Dec., 2015.

- [12] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva, "Learning deep features for scene recognition using places database," *Advances in Neural Information Processing Systems* 27, pp. 487-495, 2014.
- [13] P. Furgale, P. Carle, J. Enright, and T. D Barfoot, "The Devon Island rover navigation dataset," *The International Journal of Robotics Research*, vol. 31, no. 6, pp. 707-713, April, 2012.



Ahmed M. Naguib

2005 Electronics & Electrical
Communication Engineering, Cairo
University, Egypt (BSc.)
2014 Information and Communication
Engineering Sungkyunkwan
University (MSc.)

Area of Interest: Visual Recognition and Machine Learning



Tuan Anh Nguyen

2016 Information technology, PITT,
Vietnam (BSc.)

Area of Interest: Computer Vision, Deep Learning, Artificial Intelligence



Naeem UI Islam

2008 Electrical and Electronics Engineering,
UET, Pakistan (BSc.)
From 20013 Information and
Communication Engineering
Sungkyunkwan University (PhD.
Candidate)

Area of Interest: Machine Learning and Artificial Intelligence



Jaewoong Kim

2002 Information and Communication
Engineering Sungkyunkwan
University (BSc.)
2007 Information and Communication
Engineering Sungkyunkwan
University (MSc.)

2015 Information and Communication Engineering Sungkyunkwan
University (PhD. In Computer Science Engineering)

Area of Interest: Visual Recognition, Pattern Recognition, Image
Processing, and Machine Learning



Sukhan Lee

1972 Electrical Engineering, Seoul National
University (BSc.)
1974 Electrical Engineering, Seoul National
University (MSc.)
1982 Electrical Engineering, Purdue
University (Ph.D)

Area of Interest: Robotics, Computer Vision, Pattern Recognition,
Machine Learning, Field Robotics