Image Inpainting of Multi-Spectral Image with Laser Lines Based on Generative Adversarial Network

L Lu¹, G Zhong¹ and J Dong*¹²³
¹College of Information Science and Engineering, Ocean University of China, Qingdao, China
²Institute for Advanced Ocean Study, Ocean University of China, Qingdao, China
³Frontiers Science Center for Deep Ocean Multispheres and Earth System, Ocean University of China, Qingdao, China
*dongjunyu@ouc.edu.cn

Abstract. This paper presents a Generative Adversarial Network based on image in-painting, which can reconstruct the shape using a multi-spectral image with a laser line. One of the difficulties in multi-spectral photometric stereo is to extract the laser line, because the required illumination for multi-spectral photometric stereo, e.g. the red, green, and blue lights, may pollute the colour of the laser line. In this paper, we present a method, which uses the Generative Adversarial Network based on image in-painting, to separate a multi-spectral image with a laser line into a clean laser image and an uncorrupted multi-spectral image without the laser line, to reconstruct the shape using a multi-spectral image with a laser line. To make the proposed method applicable to real-world objects, a rendered image dataset obtained using the rendering models in ShapeNet has been used for training the network, and the evaluation shows the superiority of the proposed approach over several previous methods, on both rendered images and real-world images.

1. Introduction

There have been many algorithms that can perform high-precision 3D reconstruction of the target, including laser scanning technology [1], photometric stereo [2], structure from motion [3], multi-view stereo [4], etc. There is a recent trend that attempts to combine deep learning with photometric stereo. However, many issues exist when such 3D reconstruction methods are used in an uncontrolled environment, e.g. underwater 3D imaging: (1) Image acquisition is difficult, while deep learning usually requires a large amount of data for training; (2) The images required by photometric stereo algorithms should be obtained when the camera and the target are relatively still, while it is difficult to achieve in an underwater environment; (3) The result obtained by the photometric stereo algorithm does not have scale accuracy, that is, only the shape is relatively accurate, and the height information is not accurate.

In order to solve the first issue, we can use the rendered images as the training set [5]. For the second issue, we could use a multi-spectral photometric stereo algorithm, which only needs a single-color image ([6], [7], [8], [9]). Finally, for the third issue, we could add a laser line on the RGB image, and correct the result of the multi-spectral photometric stereo algorithm through the height information of the pixels on the laser line ([10], [11]).
However, in the process of 3D reconstruction using multi-spectral images with laser lines, there are the following problems: (1) The laser line will pollute the imaging result of the color light source, that is, the area covered by the laser light source will lose pixel information; (2) The color light source will influence the result of laser line extraction.

Inspired by the strong ability of Generation Adversarial Networks to model the data distribution, in this study, we propose a GAN based on image in-painting to solve the above issues. The proposed network can separate the multi-spectral image with a laser line into a clean laser image and an uncorrupted multi-spectral image without the laser line, which accordingly produce accurate 3D reconstruction.

The main contributions of this article are summarized as follows: (1) Through the improvement of the network proposed by Isola [12], we propose a Generative Adversarial Network based on image in-painting to realize the effective estimation of the pixel values at the locations covered by the laser line in the multi-spectral image; (2) The proposed network can extract the laser line in the multi-spectral image.

The remaining of this article is organized as follows: Section 2 introduces the related work; Section 3 introduces the network structure, related parameters and training process of the Generative Adversarial Network proposed in this article; Section 4 introduces the rendered image dataset used in this article and the results of the rendered images and the real images; Section 5 concludes this paper.

2. Related work
In this section, we review some traditional methods on laser line extraction algorithms in section 2.1, and image in-painting algorithms based on deep learning in section 2.2.

2.1. Laser line extraction algorithms
The laser scanning method is the earliest proposed and fully studied 3D reconstruction method, and even commercially available [13], but most of these devices or algorithms can only process the laser line extraction problem when the laser is the only light source or the ambient light is very weak relative to the laser brightness. In terms of algorithmic development in this field, there are many kinds of research studies on laser line thinning. Ta et al. [14] proposed a laser line detection method, using the color space to enhance the laser signal and reduce the discrimination effect of white ambient light. Song et al. [15] proposed a hybrid laser image point extraction algorithm using SVD decomposition in the HSV space of the image. Pavel et al. [16] proposed an algorithm based on color segmentation to extract the laser line.

However, the premise of the above algorithm for laser line detection is that the color of the background and the laser line is significantly different, such as natural light ([14], [15]), or dark illumination environments [16], which is not applicable to the extraction of laser lines in trichromatic laser images.

2.2 Image in-painting algorithms based on deep learning
There are many image in-painting algorithms based on deep learning. Such as, Hong et al. [17] designed a learnable fusion block, which can predict the alpha composition map and achieve smooth transition. Yu et al. [18] proposed a novel method using a contextual attention layer to achieve image in-painting. Huang et al. [19] proposed a method using mid-level structural cues for image in-painting. Jimmy et al. [20] proposed a novel CNN architecture, which enables conventional CNN to learn translation variants of irregular spatial data, and achieves superior performance on image in-painting and super-resolution.

Although image in-painting algorithms based on deep learning have been greatly developed, these algorithms cannot directly deal with the problem in this article. Cause they are usually suitable for specific images (such as face images) and images under natural light. There is not an image in-painting algorithm for multispectral images. In addition, most of these existing algorithms need to
provide a mask map, or calculate the mask map through interactive operations, but in this article, the laser image (i.e., the mask map) is unknown.

3. Method
In this section, we will introduce the details of the algorithm we proposed, which mainly include the architecture of the network, the parameter setting and training information of the Generative Adversarial Network we proposed. We first present the architecture of our network in Section 3.1, and detail the parameter settings, loss function and training process in Section 3.2.

3.1. Architecture
We designed an improved Generative Adversarial Network base on the network proposed by Isola [12]. The architecture of the proposed network contains twenty-one layers with different weights, where the generator part is composed of an encoder-decoder network composed of eight encoders and eight decoders, and the discriminator is composed of five convolutional networks. We did not make any changes to the discriminant network, but only improved the generator network.

The structure of the proposed network is shown in Figure 1.

The input values of all layers should be activated before performing other operations, the activate function for all the generator layers is the Leaky ReLU non-linearity function, whose negative slope is 0.2, while the activate function of de-convolution layers is ReLU. In decoder_8 to decoder_3, a dropout strategy is adopted, and the parameter is 0.5. All of these outputs value of the coder-encoder layers in the generator should be batch normalized before output. And in the final layer of decoder, an activation function 'tanh' is adopted.

The details of the generator are shown in Table 1.

![Figure 1. The architecture of the proposed network.](image-url)
The architecture of encoder_1 to encoder_3

The architecture of encoder_4 to encoder_8

The architecture of decoder_8 to decoder_2

The architecture of decoder_1

Figure 2. The details of the generator, in which (a) is the architecture of encoder_1 to encoder_3, (b) is the architecture of encoder_4 to encoder_8, (c) is the architecture of decoder_8 to decoder_2, and (d) is The architecture of decoder_1.

Table 1. Details of our proposed network.

| Name            | Input_channels | Output_channels |
|-----------------|----------------|-----------------|
| Encoder_1       | 3              | 64              |
| Encoder_2       | 64             | 128             |
| Encoder_3       | 128            | 256             |
| Encoder_4       | 256            | 512             |
| Encoder_5~8     | 512            | 512             |
| Decoder_8~5     | 512            | 512             |
| Decoder_4       | 512            | 256             |
| Decoder_3       | 256            | 128             |
| Decoder_2       | 128            | 64              |
| Decoder_1       | 64             | 1               |

3.2. Training

The loss function was based on conventional l2 residual minimization and that with l1 residual minimization, and their sum was optimized using the Adam Optimizer [21]. We use a Gaussian distribution with a mean of 0 and a standard deviation of 0.02 to initialize the network weights, and the learning rate during training is 0.0005.

4. Experiments

In this section, we will evaluate the experimental results of our network, including both rendered and real-world images. All of these results were acquired with an Nvidia GT730 graphics card.
4.1. Dataset

4.1.1. Rendered image dataset. We use the Blender Python API to render a total of 13977 models of five categories in ShapeNet, such as bus, car, airplane, ship and train, into multispectral images with laser lines, multispectral images without laser lines, and laser images. The settings were as follows: the resolution of the rendered images was set to 256 * 256; the origin of the coordinate system was set to the center of these models; the lamp type was set to "spotlight", and their beam were set to 60 degrees, the initial position were set to be at 0 degrees, 120 degrees and 240-degree, the optical axis points to the origin and can move along the ring as a whole at 15 degrees per time; 501 red spotlights were set to simulating the effect of a laser illumination, while their beam was set to 1 degree, the position moved along a straight line, and the color was set to (1,0,0) or (0,1,0), representing red or green laser, respectively. Each model would be rendered to 24 multi-spectral images without laser lines, 24 multi-spectral images with laser lines, and one laser image. Partly of these rendered images are shown in Figure 3.

![Figure 3](image)

4.1.2. Real-world images. We use four objects to create a collection of real-world images at different angles, including multi-spectral images with laser lines, multi-spectral images without laser lines and the pure laser images. Partly of these real-world images are shown in Figure 4.

![Figure 4](image)

4.2. Laser line extraction results

We input the multi-spectral image with a laser line into the proposed GAN network to obtain an intermediate result image. After making a difference between it and the input image, we then obtain the desired laser image through threshold segmentation, median filtering, dilation and corrosion. The threshold is set to 0.35 in the experiment.

4.2.1 Extraction results of laser lines in rendered images. First, we input the rendered image into our proposed network for training 100 times, which takes about 5 hours with a Nvidia GT730 graphics card. The training process is shown in Figure 5, and the results of laser line extraction of these rendered images are shown in Figure 6.
Figure 5. Rendered image training process. (a) is the training process of red laser, and (b) is the training process of green laser.

Figure 6. Processing results of rendered images. (a) are the input multispectral images with laser lines, (b) are the ground truth, (c) are the output results of our network, (d) are the ground truth of the laser images, (e) are the predicted laser images, (f) are the laser prediction error images.

4.2.2. Extraction results of laser lines in real-world images. The laser line extraction results using real images are shown in Figure 7.

Figure 7. Real-world images processing results. (a) are the input multi-spectral images with laser lines, (b) are the ground truth, (c) are the output results of our network, (d) are the ground truth of the laser images, (e) are the predicted laser images, (f) are the laser prediction error images.
4.3 Analysis and discussion
Comparing the results of our network with the results of the algorithm proposed by Criminisi [22][23] and Lu [24], we further verify the advantage of the proposed network. These comparison results are shown in Figure 8 and Figure 9, where all the analysis results are compared in the laser line area obtained in the previous step.

Figure 8. Compares the results of our network with Criminisi’s algorithm [22][23]. (a) are the multi-spectral images with laser lines, (b) are the ground truth, (c) are the output results of our network, (d) are the MSE of our result, (e) are the results of Criminisi’s algorithm, and (f) are the MSE of Criminisi’s algorithm.

Figure 9. Compares the results of our network with Lu’s algorithm [24]. (a) are the multi-spectral images with laser lines, (b) are the ground truth, (c) are the output results of our network, (d) are the MSE of our result, (e) are the results of Lu’s algorithm[24], and (f) are the MSE of Lu’s algorithm[24].

5. Conclusion
In this work, we have studied the three-dimensional reconstruction algorithm of multi-spectral images with laser lines. The Generative Adversarial Network based on image in-painting has generated laser images and multi-spectral images without laser lines, and realized the three-dimensional reconstruction based on multi-spectral images. Experiments show that the proposed network can effectively extract the laser lines in the multi-spectral images with laser lines, and at the same time, the pixel values of the area covered by the laser lines are also effectively estimated.
Acknowledgments
This work was supported by the Natural Science Foundation of Shandong Province of China (ZR2018ZB0852).
Authors have no conflict of interest to declare.

References
[1] T. Liu, N. Wang, Q. Fu, Y. Zhang and M. Wang, "Research on 3D Reconstruction Method Based on Laser Rotation Scanning," 2019 IEEE International Conference on Mechatronics and Automation (ICMA), Tianjin, China, 2019, pp. 1600-1604.
[2] R. J. Woodham, "Photometric method for determining surface orientation from multiple images", Optical engineering, vol. 19, no. 1, pp. 191 139–191 139, 1980.
[3] L. Torresani, A. Hertzmann and C. Bregler, "Nonrigid Structure-from-Motion: Estimating Shape and Motion with Hierarchical Priors," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 5, pp. 878-892, May 2008.
[4] X. Chen, Q. Wu and S. Wang, "Research on 3D Reconstruction Based on Multiple Views," 2018 13th International Conference on Computer Science & Education (ICCSE), Colombo, 2018, pp. 1-5.
[5] L. Lu, L. Qi, Y. Luo, H. Jiao, J. Dong, "Three-Dimensional Reconstruction from Single Image Base on Combination of CNN and Multi-Spectral Photometric Stereo". Sensors 2018, 18, 764.
[6] Christensen, P.H., Shapiro, L.G. "Three-dimensional shape from color photometric stereo". Int J Comput Vision 13, 213–227 (1994).
[7] Ikeda and Y. Duan, "Color Photometric Stereometry for Albedo and Shape Reconstruction," 2008 IEEE Workshop on Applications of Computer Vision, Copper Mountain, CO, 2008, pp. 1-6.
[8] Ikeda, "Color photometric stereo for directional diffuse object," 2009 Workshop on Applications of Computer Vision (WACV), Snowbird, UT, 2009, pp. 1-6.
[9] R. Anderson, B. Stenger and R. Cipolla, "Color photometric stereo for multicolored surfaces," 2011 International Conference on Computer Vision, Barcelona, 2011, pp. 2182-2189.
[10] H. Fan, L. Qi, N. Wang, J. Dong, Y. Chen and H. Yu "Deviation correction method for close-range photometric stereo with nonuniform illumination," Optical Engineering 56(10), 103102 (6 October 2017).
[11] H. Fan, L. Qi, J. Dong and H. Yu, "Refractive laser triangulation and photometric stereo in underwater environment", Optical Engineering, 2017, vol. 56, no. 11.
[12] P. Isola, J. Zhu, T. Zhou and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 5967-5976.
[13] https://www.2grobotics.com/
[14] H. N. Ta, D. Kim and S. Lee, "A novel laser line detection algorithm for robot application," 2011 11th International Conference on Control, Automation and Systems, Gyeonggi-do, 2011, pp. 361-365.
[15] W. Song, X. Hu, J. Fu, Q. Zhou, T. Zhou and P. Si, "The method of hybrid-laser image spot extracts based on HSV space SVD for power transmission line detection", 2016 IEEE International Conference on Information and Automation (ICIA), Ningbo, 2016, pp. 1361-1364.
[16] Chmelar, Pavel & Dobrovolny, Martin. (2013). "The Laser Line Detection for Autonomous Mapping Based on Color Segmentation".
[17] X. Hong, P. Xiong, R. Ji and H. Fan, "Deep Fusion Network for Image Completion", https://arxiv.org/abs/1904.08060.
[18] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, T. S. Huang, "Generative Image In-painting with Contextual Attention", 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, Utah, 2018, pp. 5505-5514.
[19] J. Huang, S. B. Kang, N. Ahuja, and J. Kopf, "Image Completion using Planar Structure Guidance", ACM Transactions on Graphics (Proceedings of SIGGRAPH 2014), 33(4), 2014.
[20] J. SJ. Ren, L. Xu, Q. Yan and W. Sun, "Shepard Convolutional Neural Networks", Twenty-ninth Conference on Neural Information Processing Systems (NeurIPS 2015).
[21] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization ", 2014.
[22] A. Criminisi, P. Pérez, K. Toyama. "Region filling and object removal by exemplar-based image inpainting" [J]. IEEE Transactions on image processing, 2004, 13(9): 1200-1212.
[23] A. Criminisi, P. Perez, K. Toyama. "Object removal by exemplar-based inpainting" [C]//Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on. IEEE, 2003, 2: II-II.
[24] C. Lu, J. Feng, Z. Lin, S. Yan. "Exact Low Tubal Rank Tensor Recovery from Gaussian Measurements". International Joint Conference on Artificial Intelligence (IJCAI). 2018.