적대적 생성 신경망을 이용한 단일 RGB 기반 깊이 이미지 생성

푸옹[1], 호서가[1], 서가가[1], 허언민[1], 한우철[1], 조경은[1]*
[1]동국대학교 멀티미디어공학과
*e-mail : cke@dongguk.edu(교신저자)

Depth image generation from single RGB image using Generative adversarial network

Phuong Minh Chu[1], Shujia Hao[1], Jiajia Xu[1], Yanmin He[1], WooChul Han[1], Kyungeun Cho[1]*
[1]Department of Multimedia Engineering, Dongguk University-Seoul

Abstract

Depth images are important for many applications such as 3D scene reconstruction. This paper proposes a depth image generation method from a single RGB image. The input of the proposed method is a 2D image captured by a conventional camera, which is generally less expensive. Next, a generative adversarial network is employed to generate a depth image estimation from the input image. Experimental results verify that the proposed method provides high-quality depth images from single 2D images. In addition, the proposed depth image generation method does not require high computational resources. We can use a general PC with a moderate-capacity graphics processing unit for handling the calculations.

1. Introduction

Generative adversarial networks (GANs) [1,2] are being extensively studied in deep learning. A GAN model consists of generator and discriminator. The generative network creates candidates that are evaluated by the discriminative network. In addition, both networks compete to raise the error rate of each other. Both generator and discriminator are trained until we obtain the optimal parameters. Finally, the generator can generate high-quality synthetic candidates. In this study, we aim to develop a method that automatically generates a depth image from a single RGB image based on GANs. We used Kinect datasets in experiment for evaluating the proposed method.

2. Related work

Nowadays, a variety of deep neural networks [3- 5] to create depth images from single RGB images are presented. Eigen et al. [3] used a multi-scale deep network for depth map prediction from a single image. In [4], the authors proposed a depth extraction method from video using non-parametric sampling. Besides, Liu et al.[5] employed a deep convolutional neural network to estimate depth images from single RGB images. These methods have limitation of the quality. Therefore, we proposed a new method based on GAN. In experiment, we evaluated our method using the NYU-Depth V2 dataset [6].

3. Proposed methodology

We customized pix2pixHD [2] model to satisfy two goals. First goal is to generate high-quality synthetic depth images from RGB ones. Second goal is to minimize the hardware requirements that the proposed method can run on a limited GPU. The proposed model contains two networks as shown in Figure 1. Generator G learns to map RGB image X and random noise vector Z into depth image Y, G: $\{X, Z\} \rightarrow Y$. G is trained to create depth images that cannot be detected from the ground truth using adversarial discriminator D. On the other hand, D is trained to distinguish the synthetic depth images produced by generator G.

In our GAN model, we employed a single global generator having same structure with generator in [2]. Moreover, the discriminator D is decomposed into two sub-discriminators, D_1 and D_2. D_1 works with the full-resolution of depth images, whereas D_2 works with half-scale of the depth images. Therefore, D_1 processes a global view of the synthetic depth image to guide G to produce globally consistent images, whereas D_2 directs G to generate sharp and accurate images. Moreover, D_2 helps G avoiding repeated patterns on the

synthetic depth images. Both D_1 and D_2 have the same structure employed in [2].

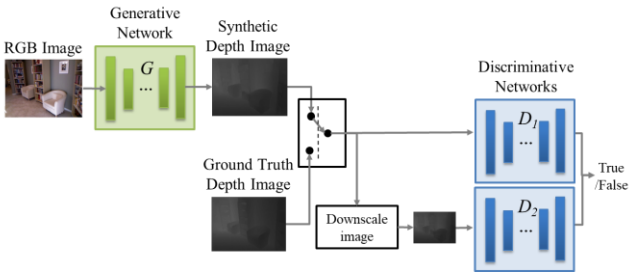


Figure 1. Framework of depth image generation from single RGB image

4. Experiment

For the experiments, we employed the NYU-Depth V2 dataset [6] obtained from Kinect sensors. The NYU dataset consists of 1449 pairs of RGB and depth images, from which 795 pairs were employed for training and 654 pairs for testing. The qualitative results are demonstrated in Table 1. For quantitative evaluation, we used ground-truth depth images. The average relative error (REL) and root-mean-square error (RMSE) are 0.168 and 0.538, respectively. In addition, we only required common PC with Nvidia GTX 970 graphic card (4 GB) to run the experiment.

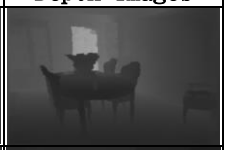
RGB Images	Generated Depth Images	Ground Truth Depth Images
		
		
		

Table 1. Results of proposed method

5. Conclusion

In this paper, we propose a GAN-based method to transform a single RGB image into the corresponding depth image because GANs are flexible and powerful tools for generating synthetic data. Experimental results indicate that the proposed method can effectively create high-quality depth images. The comparison of the proposed method to existing methods shows that our method achieves the best performance. In addition, our method does not have high PC hardware requirements including high-end GPUs. In future work, we will modify the GAN structure to improve depth estimation quality.

Acknowledgments

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government (MSIP) (2018R1A2B2007934).

Reference

[1] P. Isola, Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-Image Translation with Conditional Adversarial Networks” in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5967-5976, Honolulu, Hawaii, USA, 21-26 July 2017.

[2] T. C. Wang, M. Y. Liu, J. Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, “High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs”, arXiv preprint arXiv:1711.11585, pp. 1-14, 2017.

[3] D. Eigen, C. Puhrsch, and R. Fergus, “Depth map prediction from a single image using a multi-scale deep network”, in Neural Information Processing Systems 27 (NIPS 2014), pp. 1-9, Montréal, Canada, 8-13 December 2014.

[4] K. Karsch, C. Liu, and S. B. Kang, “Depthtransfer: Depth extraction from video using non-parametric sampling”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 11, pp. 2144-2158, 2014.

[5] F. Liu, C. Shen, and G. Lin, “Deep convolutional neural fields for depth estimation from a single image”, in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5126-5170, Boston, MA, USA, 7-12 June 2015.

[6] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, “Indoor segmentation and support inference from rgbd images”, in 12th European Conference on Computer Vision, pp. 746-760, Florence, Italy, 7-13 October 2012.