2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)

Learning to Match 2D Images and 3D LiDAR Point Clouds for Outdoor Augmented Reality

Weiquan Liu[1]*, Baiqi Lai[1]*, Cheng Wang[1]†, Xuesheng Bian[1], Wentao Yang[1], Yan Xia[2], Xiuhong Lin[1], Shang-Hong Lai[3], Dongdong Weng[4], Jonathan Li[5]

[1] Fujian Key Laboratory of Sensing and Computing for Smart City, School of Informatics, Xiamen University, Xiamen, China

[2] Photogrammetry and Remote Sensing, Technical University of Munich, Munich, Germany

[3] Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan

[4] Beijing Engineering Research Center of Mixed Reality and Advanced Display, School of Optics and Photonics, Beijing Institute of Technology, Beijing, China

[5] Department of Geography and Environmental Management, University of Waterloo, Waterloo, Canada

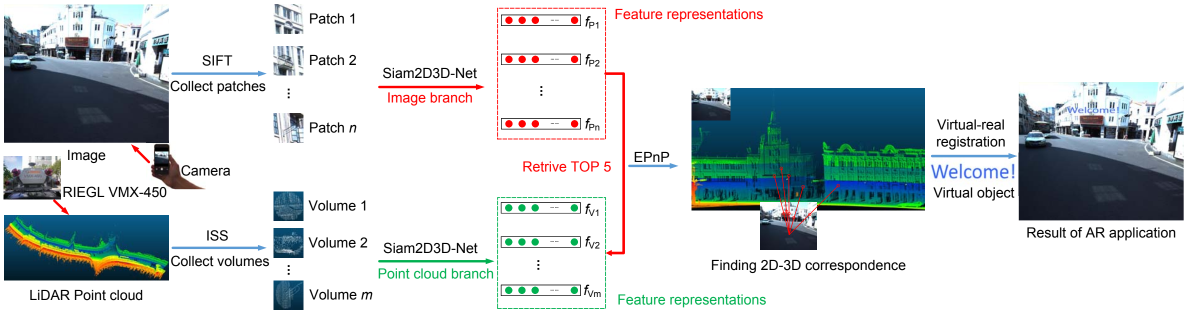


Figure 1: Overview of the proposed outdoor AR virtual-real registration approach. 1) Using MLS system to scan outdoor 3D LiDAR point cloud as the 3D map, and capturing images by camera; 2) Using the keypoint detector of SIFT and ISS to extract the keypoints in 2D image and 3D LiDAR point cloud, respectively; 3) Using the above keypoints to collect the image patches and point cloud volumes; 4) Using proposed Siam2D3D-Net to extract the feature representations of image patches and point cloud volumes; 5) Retrieving the TOP 5 results from each image patch feature representation to the database of point cloud volume feature representations; 6) Using EPnP and RANSAC algorithm to find the 2D-3D keypoint correspondences to calculate the camera pose; 7) Using the camera pose to compute the transformation matrix from 3D space to 2D space so that achieve virtual-real registration.

Abstract

Large-scale Light Detection and Ranging (LiDAR) point clouds provide basic 3D information support for Augmented Reality (AR) in outdoor environments. Especially, matching 2D images across to 3D LiDAR point clouds can establish the spatial relationship of 2D and 3D space, which is a solution for the virtual-real registration of AR. This paper first provides a precise 2D-3D patch-volume dataset, which contains paired matching 2D image patches and 3D LiDAR point cloud volumes, by using the Mobile Laser Scanning (MLS) data from the urban scene. Second, we propose an end-to-end network, Siam2D3D-Net, to jointly learn local feature representations for 2D image patches and 3D LiDAR point cloud volumes. Experimental results indicate the proposed Siam2D3D-Net can match and establish 2D-3D correspondences from the query 2D image to the 3D LiDAR point cloud reference map. Finally, an application is used to evaluate the possibility of the proposed virtual-real registration of AR in outdoor environments.

Keywords: Outdoor AR, virtual-real registration, 2D-3D feature representation, cross-domain data matching.

Index Terms: Human-centered computing—Visualization—Visualization techniques—Augmented reality

1 Introduction

With the development of various sensors and cross-domain data processing methods, AR applications are gradually being applied in outdoor environments [3]. Due to the complexity of outdoor environment, achieving accurate virtual-real registration is challenging.

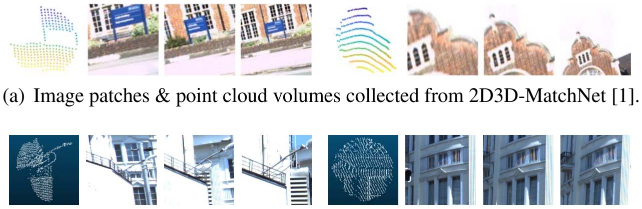
3D LiDAR point cloud data, an novelty remote sensing data, provides basic 3D information support for outdoor AR with its advantage of high precision and fast acquisition. Specifically, if the 2D images and the 3D LiDAR point clouds can be matched, the spatial relationship of 2D and 3D space will be established, which is a solution to achieve the virtual-real registration of AR. However, the great difference between the data modal of the 2D images and 3D LiDAR point clouds makes it challenging to match them directly.

To date, deep learning is a strategy for matching 2D images and 3D LiDAR point clouds, but there is lacking of high-quality dataset to train the deep neural network. 2D3D-MatchNet [1] is the first to use deep learning method to establish 2D-3D correspondences for 2D images and 3D point cloud. Meanwhile, 2D3D-MatchNet created a dataset of 2D image patch to 3D point cloud volume correspondences, however, the density of the 3D point cloud is too sparse to represent the details and contour information of the objects (as shown in Figure 2(a)). Thus, the dataset of 2D3D-MatchNet is hard to be used for outdoor AR.

In this paper, our goal is to learn the local invariant feature representations of 2D image patches and 3D LiDAR point cloud volumes, which can be used to match the 2D images and 3D LiDAR point clouds. The main contributions are listed as follows: a) We create a high-quality 2D-3D patch-volume dataset; b) We propose a neural network, Siam2D3D-Net, to jointly learn the local feature represen-

*indicates equal contribution. e-mail:wqliu1026@163.com
†Corresponding author. e-mail: cwang@xmu.edu.cn

978-1-7281-6532-5/20/$31.00 ©2020 IEEE
DOI 10.1109/VRW50115.2020.00178



(a) Image patches & point cloud volumes collected from 2D3D-MatchNet [1].

(b) The high-quality 2D-3D patch-volume dataset collected in this paper.

Figure 2: The 2D-3D dataset of 2D3D-MatchNet and ours.

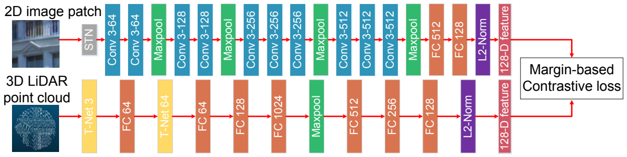


Figure 3: The architecture of Siam2D3D-Net.

tations for 2D image patches and 3D LiDAR point cloud volumes; c) We design an adaptive strategy to obtain an adaptive margin of margin-based contrastive loss to optimize Siam2D3D-Net.

2 Method

The pipeline of the proposed AR virtual-real registration approach in outdoor environments is shown and described in Figure 1.

Dataset. 1) Using the RIEGL VMX-450 MLS system to scan about 15 km street scene. The point density of acquired 3D LiDAR points is about 4000 points/m^2. The resolution of the camera image is 2452×2056 pixels. 2) Based on the data collection method of 2D3D-MatchNet, to avoid false voting, constraining the area of keypoint voting is located within a radius of 50 meters from the image GPS position. Totally, we collect 13,884 pairs of matching 2D-3D patch-volume, several examples are shown in Figure 2(b).

Network. Our proposed Siam2D3D-Net (Figure 3) consists of two branch with not shared parameters. One is the image branch, which is constructed with STN (Spatial Transformer Network) module and modified VGG network, to learn the image feature representations; the other one is point cloud branch, which is modified PointNet, to learn the LiDAR point cloud feature representations. The inputs of Siam2D3D-Net are paired 2D image patches and 3D LiDAR point cloud volumes. In detail, the size of the 2D image patches are resized to 128×128 pixels, and are normalized by Gaussian. The 3D LiDAR point cloud volumes are padded or down-sampled to 1024-points, and then achieve normalization based on the center point of the point cloud volume. The outputs of Siam2D3D-Net are 128-dimensional feature vectors.

Loss function. Inspired by SiamAM-Net [2], we propose to use the margin-based contrasted loss with adaptive margin to optimize the Siam2D3D-Net, defined as follows:

$$L_{margin}(I_P,M_V,l)=\frac{1}{2}lD^2+\frac{1}{2}(1-l)\{\max(0,m-D)\}^2 \qquad (1)$$

$$\begin{cases} m=d+\ln(d+1) \\ d=\max\{\|f(P_k)-f(V_k)\|_2\cdot l\} \end{cases} \qquad (2)$$

where l is the label of 2D image patch I_P and 3D LiDAR point cloud volume M_V. If I_P and M_V are matched, $l=1$; otherwise, $l=0$. D is the feature distance between the learned features f_{I_P} and f_{M_V} of the input pairs, i.e. $D(f_{I_P},f_{M_V})=\|f_{I_P}-f_{M_V}\|_2$. $m>0$ is the adaptive margin. Denoting the data in a batch B is $\{P_k,V_k\}$, $k=1,2,\dots,K$, where B has $K/2$ matching pairs and $K/2$ non-matching pairs. d is the maximal distance of the matching pairs in batch B.

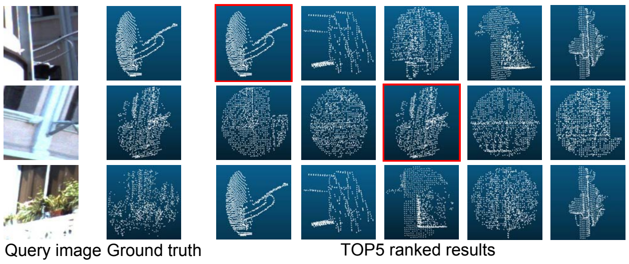


Figure 4: Several TOP 5 ranked results in testing data. 1st row: Correct TOP 1; 2nd row: Correct TOP 5; 3rd row: Wrond TOP 5.

Training strategy. We use 11,766 matching and 11,766 non-matching pairs of 2D image patches and 3D LiDAR point cloud volumes to train Siam2D3D-Net, which is implemented by the Pytorch framework and trained with a Nvidia 2080 Ti GPU. The image branch and point cloud branch are all optimized with Adam optimizer. The learning rate, initially at 6×10^{-5}. In addition, the parameters of the VGG network in the image branch are initialized with the VGG model which pre-trained on ImageNet.

3 Experiment

We use the rest 2,118 paired matching 2D image patches and 3D LiDAR point cloud volumes for testing. To demonstrate the performance of the learned feature representations by Siam2D3D-Net, several TOP 5 ranked results are shown in Figure 4, the correct retrieved results are labeled with red bounding boxes. In Figure 4, the 1st, 2nd and 3rd row show the correct TOP 1, correct TOP 5 and wrong TOP5 retrieved results, respectively. It can be viewed that high quality data are easier to retrieve the correct correspondence (1st row in Figure 4); with repetitive and similar structure will mislead the retrieval (2nd row in Figure 4); irregular and complex samples will increase the challenge (3rd row in Figure 4). Finally, to demonstrate the possibility of our proposed virtual-real registration of AR in an outdoor environment, we register a 'Welcome!' label to the building, as shown in the last column of Figure 1.

4 Conclusion

In this paper, we propose to match 2D images across 3D LiDAR point clouds to achieve virtual-real registration of AR in outdoor environments. We propose Siam2D3D-Net to jointly learn the feature representations of 2D image patches and 3D LiDAR point cloud volumes. In addition we create a high-quality 2D-3D patch-volume dataset from the data acquired by RIEGL VMX-450 MLS system. Experimental results demonstrate that the learned feature representations of 2D image and 3D LiDAR point cloud can be used to retrieve in Euclidean space. Finally, an application is used to demonstrate the possibility of the proposed virtual-real registration of AR in outdoor environments. In the future, we plan to explore more robust feature representations for 2D images and 3D LiDAR point clouds.

Acknowledgments

This work is supported by National Natural Science Foundation of China (No.U1605254, 41471379, 61371144).

References

[1] M. Feng, S. Hu, M. Ang, et al. 2d3d-matchnet: Learning to match keypoints across 2d image and 3d point cloud. In *IEEE ICRA*, 2019.

[2] W. Liu, C. Wang, X. Bian, et al. Learning to match ground camera image and uav 3-d model-rendered image based on siamese network with attention mechanism. *IEEE Geosci. Remote Sens. Lett.*, 2019.

[3] W. Liu, C. Wang, Y. Zang, et al. Ground camera images and uav 3d model registration for outdoor augmented reality. In *IEEE VR*, 2019.