Ground Camera Images and UAV 3D Model Registration for Outdoor Augmented Reality

Weiquan Liu[1]*, Cheng Wang[1]†, Yu Zang[1], Shang-Hong Lai[2], Dongdong Weng[3], Xuesheng Bian[1], Xiuhong Lin[1], Xuelun Shen[1], Jonathan Li[1,4]

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

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

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

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

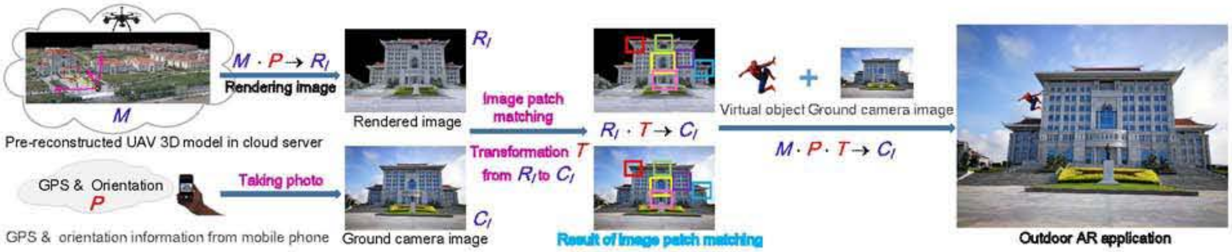

Figure 1: Overview of the proposed virtual-real registration approach for outdoor AR application. First, using position information acquired by mobile phone to render corresponding synthetic image from pre-constructed UAV 3D model; second, matching ground camera image and UAV 3D model rendered image to find transformation relationship between the cross-domain images; finally, registering virtual object to the ground camera image by the inferred spatial relationship of UAV 3D model and ground camera image.

ABSTRACT

This paper presents a novel virtual-real registration approach for augmented reality (AR) in large-scale outdoor environments. Essentially, it is a pose estimation for the mobile camera images (ground camera images) in 3D model recovered by Unmanned Aerial Vehicle (UAV) image sequence via Structure-From-Motion (SFM) technology. The approach considers to indirectly establish the spatial relationship between 2D and 3D space by inferring the transformation relationship between the ground camera images and the UAV 3D model rendered images. Specifically, the proposed approach can overcome the positioning errors, which are deterioration and drift in the GPS, and deviation of orientation. The experimental results demonstrate the possibility of the proposed virtual-real registration approach, and show that the approach is robust, efficient and intuitive for AR in large-scale outdoor environments.

Keywords: Virtual-real registration, outdoor AR, cross-domain image matching.

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

1 INTRODUCTION

The virtual-real registration is a hot topic for AR applications, for that the performance of the AR is strongly depended on the accuracy of the virtual-real registration. To date, most of AR applications are applied for the indoor environments, while few of them pay attention to outdoor environments. Due to the huge range of uncontrolled factors and complexity of the large scale data in the outdoor environments, such as the range of scenes become large-scale, hundreds of objects, dramatic changes of illumination, or the geo-registration for AR driving in outdoor environments [1].

For AR methods based on visual fiducial markers, it is difficult to cover such an uncontrolled outdoor environment with pre-placed markers. Besides, AR application with multiple sensors suffer some problems, like deterioration in the GPS precision, deviation in the gyroscope sensors, drift in the output of inertial and magnetic sensors. Such factors sometimes lead to unsatisfactory outdoor AR results.

In this paper, we propose a novel virtual-real registration approach for AR in large-scale outdoor environments (the details are in Section 2). The goal of the approach is to establish the spatial relationship between the UAV 3D model and the ground camera images, by utilizing the matching relationship of the UAV 3D model rendered images and ground camera images (we call them cross-domain images).

However, it is challenging to match ground camera images and UAV 3D model rendered images, because of different image domain. Besides, rendered images from UAV 3D model are inevitably contaminated with huge distortion, blurred resolution, and obstructions (bottom in Figure 2), which go beyond the reach of handcrafted descriptors, e.g. SIFT, BRIEF, BRISK, ORB, etc. Thus, we propose to use deep neural network to match the different domain images, using image patch (blob) to replace traditional keypoints (learning invariant feature descriptors for the cross-domain image patches).

Especially, the siamese and triplet networks are the recent mainstream architectures in image patch matching and feature descriptors learning. H-Net and H-Net++ [2] are the latest work for cross-domain image patch matching, but they are time consuming and the retrieval accuracy by the feature descriptors is not enough. In this paper, we use the HardNet [3] to learn the invariant feature descriptors for the ground camera images and UAV 3D model rendered images.

Our main contributions are listed as follows. a) Designing a so-

---

*e-mail: wqliu1026@163.com

†Corresponding author. e-mail: cwang@xmu.edu.cn

2019 IEEE Conference on Virtual Reality and
3D User Interfaces
23-27 March, Osaka, Japan
978-1-7281-1377-7/19/$31.00 ©2019 IEEE

Figure 2: Matching cross-domain image patch pairs. Top: ground camera image patches; Bottom: UAV 3D model rendered image patches.

lution which indirectly establishes the spatial relationship between ground camera image and UAV 3D model. b) Image patch feature descriptor learned by deep neural network is used instead of handcraft keypoint for the cross-domain image matching. c) The proposed approach is possible, lightweight and intuitive.

2 METHOD

The schematic pipeline of the proposed virtual-real registration approach for outdoor AR is shown in Figure 1. Details are as follows:

Step 1: Using UAV to obtain the aerial image sequence of the large-scale outdoor scene, and using SFM algorithm to reconstruct the outdoor scene as an UAV 3D model through the aerial images, meanwhile, deploying this UAV 3D model to the cloud server. Step 2: Using mobile phone to capture a ground camera image which contains GPS and orientation information. Step 3: Rendering a synthetic image from the UAV 3D model by the above positioning information in Step 2. Step 4: Matching the UAV 3D model rendered image and ground camera image by HardNet. Step 5: Calculating the spatial relationship from UAV 3D model to ground camera image.

In formula, the spatial relationship between the UAV 3D model and the ground camera image can be defined as:

As shown in Figure 1, denoting the UAV 3D model as M, ground camera image as C_I, UAV 3D model rendered image as R_I, and the projection matrix from UAV 3D model to the rendered image is P. Here, P is obtained from the positioning information (GPS and orientation) by the mobile phone. First, the relationship between UAV 3D model rendered image and UAV 3D model is $M \cdot P \rightarrow R_I$, second, assuming the transformation matrix for R_I to C_I is T, that is $R_I \cdot T \rightarrow C_I$. Thus, the spatial relationship for UAV 3D model to ground camera image is as follow:

$$M \cdot P \cdot T \rightarrow C_I \tag{1}$$

Obviously, the above derivations verify that the core essential of the proposed virtual-real registration approach is the matching problem of UAV 3D model rendered images and ground camera images.

The strategy for matching ground camera images and UAV 3D model rendered images is as follows. First, randomly selecting 2,000 points from the two images; second, extracting the feature descriptors for these image patches by the trained network; third, performing Nearest Neighbor Search (NNS) to retrieve the feature descriptors, only the TOP1 retrieval results and the *cosine* similarity greater than 0.9 are retained; finally, using RANSAC to filter the significant mismatched results.

3 DATASET

In order to demonstrate the capability of the proposed virtual-real registration approach of AR in outdoor environments, a 3-square-kilometer campus is employed as the testing site. Where more than 100 buildings are involved, and more than 30,000 aerial images of the campus are captured by the UAV, then UAV 3D model is reconstructed by the SFM algorithm (as the pre-constructed UAV 3D model shown in Figure 1). In addition, 10,000+ ground camera images are captured by mobile phone, and the corresponding 10,000+ rendered images are generated in the UAV 3D model.

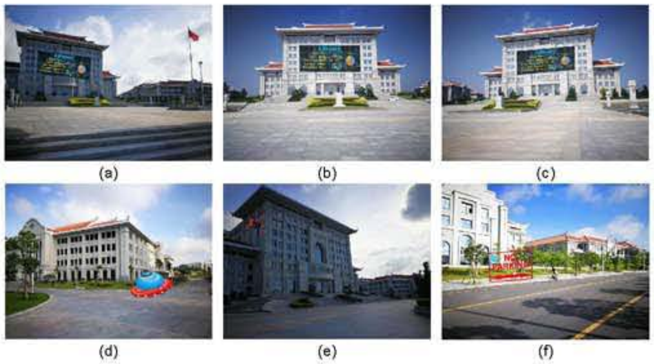

Figure 3: AR applications by our virtual-real registration approach.

As the size of image patches for the input of neural network should be carefully designed, the size of the cross-domain image patches is collected between in the neighborhood of 250*250 and 500*500 pixels. Finally, we collect 200,000+ labeled matching cross-domain image patch pairs from the 10,000+ corresponding ground camera images and UAV 3D model rendered images. Some examples are shown in Figure 2.

4 EXPERIMENTS

For testing the proposed AR virtual-real registration approach, several representative scenes on the campus are selected. All of the experimental AR application results are presented in Figure 3. The first row in Figure 3 shows that the virtual label *Library real-time information* is registered to library building. The carton UFO, carton spider-man and NO PARKING sign are registered into the outdoor environments in Figure 3 (d)-(f) respectively.

The above experiments demonstrate the capability of the proposed virtual-real registration approach of AR in outdoor environments. However, there are still some limitations, e.g. the virtual objects with slight deformation after registration and still with a gap to the targets.

5 CONCLUSION

In this paper, we propose a novel, mobile devices based, portable, lightweight, efficient, intuitive virtual-real registration approach for AR in large-scale outdoor environments. We indirectly infer the spatial relationship from 3D space to 2D space by using the matching relationship between ground camera images and UAV 3D model rendered images. The application results demonstrate that the capability of the proposed virtual-real registration approach of AR in large-scale outdoor environments. In the future work, we will explore the influence of illumination, weather and season differences on our approach, and explore to combine multiple sensors to improve the accuracy for the virtual-real registration in outdoor AR applications.

ACKNOWLEDGMENTS

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

REFERENCES

[1] H.-P. Chiu, V. Murali, R. Villamil, et al. Augmented reality driving using semantic geo-registration. In *IEEE VR*, pp. 423–430, 2018.
[2] W. Liu, X. Shen, C. Wang, Z. Zhang, C. Wen, and J. Li. H-net: Neural network for cross-domain image patch matching. In *International Joint Conference on Artificial Intelligence*, pp. 856–863, 2018.
[3] A. Mishchuk, D. Mishkin, F. Radenovic, et al. Working hard to know your neighbor's margins: Local descriptor learning loss. In *Advances in Neural Information Processing Systems*, pp. 4826–4837, 2017.