Satellite Image-based UAV Localization using Siamese Neural Network

Seong-Ha Ahn
Department of Electronic Engineering, Pusan National University,
2, Busandaehak-ro 63beon-gil Geumjeong-gu
Busan, 46241, Republic of Korea

Ho-Sun Kang
Department of Electronic Engineering, Pusan National University,
2, Busandaehak-ro 63beon-gil Geumjeong-gu
Busan, 46241, Republic of Korea

Jang-Myung Lee
Department of Electronic Engineering, Pusan National University,
2, Busandaehak-ro 63beon-gil Geumjeong-gu
Busan, 46241, Republic of Korea

E-mail: seongha7379@pusan.ac.kr, hosun7379@pusan.ac.kr, jmlee@pusan.ac.kr
www.pusan.ac.kr

Abstract

We present a method for UAV localization using pre-existing satellite images. The use of Unmanned Aerial Vehicles (UAVs) has rapidly increased in several applications such as surveillance, search, and defense. When in GPS-denied situations, however, the onboard GPS signal may be noisy or inaccurate. The proposed method is based on a Siamese Neural Network that contains two instances of the same neural architecture and weights. Siamese Neural Network learns the similarity metric so that can recognize the same place from two raw images. Convolutional Neural Network is used as a backbone in Siamese Neural Network to overcome variation due to differences such as perspective, shadow angle, and presence of vehicles. We describe UAV localization pipeline and a dataset for training and testing our networks. Finally, the performance of the proposed method was shown in accuracy.

Keywords: visual localization, uav, satellite image, siamese neural network, image retrieval

1. Introduction

Unmanned Aerial Vehicles (UAVs) are aircraft that remotely controlled or fly autonomously with an onboard manipulation system. Nowadays, the use of UAVs has rapidly increased in several applications such as surveillance, search, and defense [1]. Localization is an essential task to utilize UAVs in such tasks. Currently, the onboard navigation system is mostly relied on Global Positioning System (GPS). When in GPS-denied situations, however, the onboard GPS signal may be noisy or inaccurate. Visual localization, which is the problem of estimating a camera pose, can be a useful alternative in GPS-denied environments. Image retrieval-based approach is one of the common visual localization methods [2]. In image retrieval tasks for UAVs visual localization, satellite imagery can be utilized as a geodatabase. In recent year, deep neural networks have

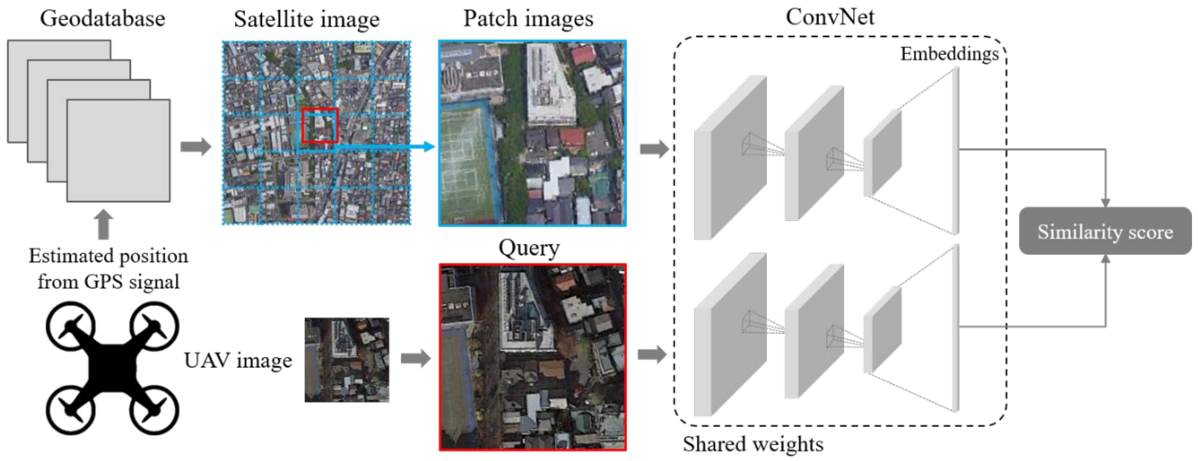

Fig. 1. Image retrieval pipeline of visual localization for UAVs. georeferenced satellite imagery is divided into patches for search and UAVs image is used as query image. The closest patch image will be retrieved by comparing similarity of every patch and query image pairs.

been successfully applied in various image processing and analysis tasks. We apply deep neural architecture in the image retrieval system to analyze images efficiently. In this paper, we propose a satellite image-based UAVs visual localization method with image retrieval approach. In section 2, we describe the deep neural architecture, called Siamese Neural Networks that learns metric for similarity. Dataset preparation for training neural architecture will be covered in section 3. We train and evaluate the proposed method in section 4. Finally, we summarize our study in section 5.

2. Image Retrieval using Siamese Neural Networks

2.1. *Siamese neural networks*

Siamese neural networks (SNN) is a neural architecture that has two networks which share its weights. Siamese neural networks employ a unique structure to learn a similarity metric between two input [3]. To extract features effectively from input images, Convolutional Neural Networks (CNN) is used in feature extraction layers of siamese neural networks. Figure 1 shows the image retrieval pipeline using CNN-based siamese neural networks. The proposed method uses semi-supervision of GPS system. The georeferenced satellite image is selected according to the noisy GPS signal from database. Satellite image is divided into few numbers of patches. The network takes two input: patch of satellite image and UAVs image. Input images are projected onto the low dimensional space which is called embedding space. By training siamese neural networks with UAVs and satellite patch imagery, the networks are trained to extract discriminative features to measure similarity. It helps the networks to recognize if two input images are taken from the same place or not. Every patch images and UAVs image are compared by measuring the distance in the embedding space.

2.2. *Triplet loss*

One of the options for training a siamese neural networks to learn similarity metric is to use triplet loss. The input data are projected into the embedding space via the networks. When Anchor, Positive and Negative images are input, triplet loss is calculated as the euclidean distance between three embedded points. (see Figure 2) The loss function can be described as follows:

$$L(A,P,N) = \max\left(\left|\left|f(A) - f(P)\right|\right|^2 - \left|\left|f(A) - f(N)\right|\right|^2 + \alpha, 0\right) \quad (1)$$

Where a is a margin between positive and negative pairs, and f is an embedding. In our implementation, we set a as 1.

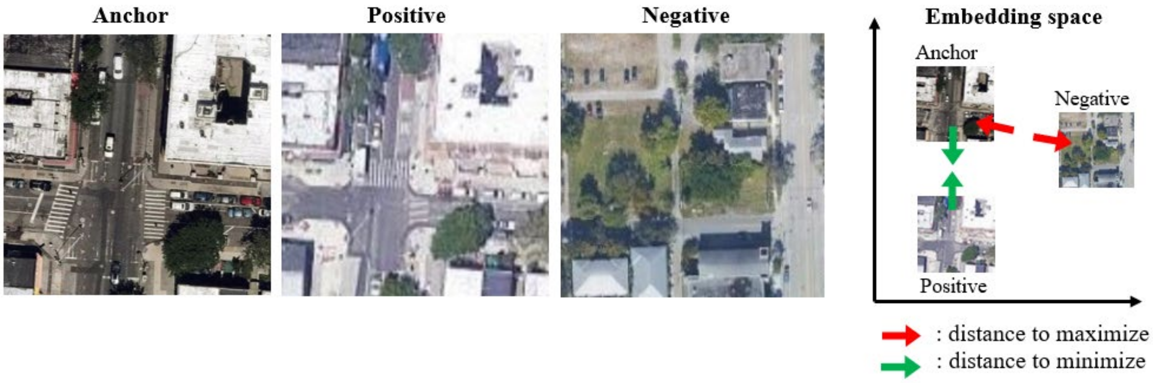

Fig. 2. The distance from the Anchor to Positive is minimized and the distance from the Anchor to Negative is maximized.

3. Dataset

To simulate UAVs and satellite imagery, we used two satellite imagery sources. As shown in Figure 3, the difference in characteristics of satellite imagery sources causes various visual representations.

3.1. *Database properties*

We build two databases: UAVs image database and satellite imagery database. These databases are consist of total 88 correspondence images. Database image pairs from 3 different cities in New York and Florida. Each image covers a 0.16km^2 area of land. UAVs database image size with 0.125m resolution is 3200*3200 pixels, satellite image size with 0.5m resolution is 800*800 pixels. In our experiments, we divide database images into 16 patches per each. For training and validating, we generate approximately 30k and 0.5k, 0.5k image pairs respectively with the augmented images generated by rotating and adding a random value to channels.

3.2. *Anchor, Positive and Negative image*

As described section 2, triplet loss is calculated by measuring euclidian distance between Anchor, Positive and Negative images embedded in embedding space. For positive images, we create 4 shifted images from the original patch. The pixel range for the shift is limited to ±25 percent of total pixels in the image. The Anchor and Positive image pairs are created by combining 8 images. The Negative images are sampled in the same database image except the same patch and the other images.

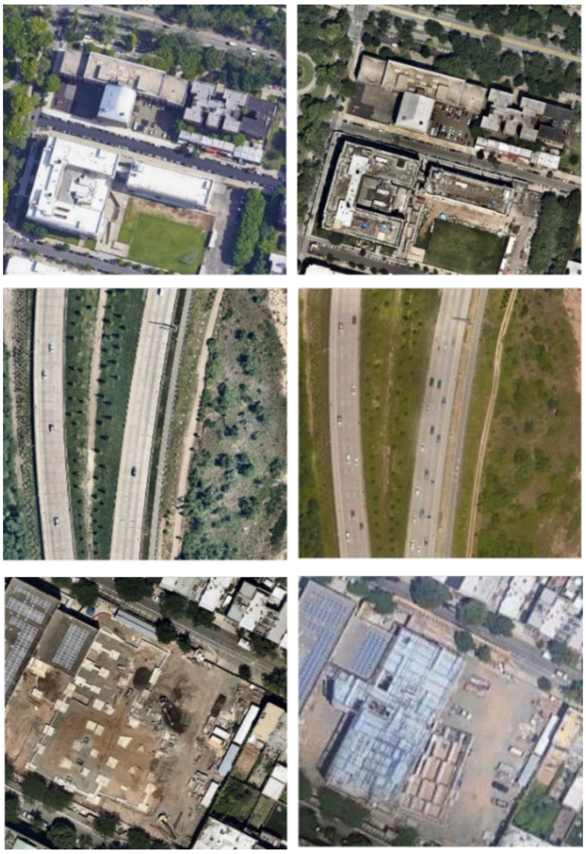

Fig 3. Various visual representation of the same area caused by different characteristic of satellite imagery sources.

4. Implementation

We implement the proposed system in Python and PyTorch. Experiments have been done using 2 RTX2080ti. Stochastic Gradient Descent (SGD) optimizer is employed to train the networks. Since the only image with the shortest Euclidean distance is assumed to be a *Positive*, we evaluate our methods by calculating top- k precision score. The proposed method achieves as Table 1. It retrieves correct images for most queries, however, it performed poorly in specific cases, such as when the area has too many structures. (see Figure 4) It shows that the dataset requires more images that contain various visual features for generalization.

Table 1. The experimental results on our test set.

Precision score

Top-k precision	score
k=1	0.65
k=3	0.74

5. Conclusion

We proposed and implemented an image retrieval system that queries UAV images and retrieves the most similar image from pre-existing satellite images using CNN-based Siamese neural networks. To simulate UAVs and satellite imagery, we created a dataset using two different satellite imagery sources. The proposed system performed decently on our test set, however, the performance should be improved in some specific cases. In order to deploy the system for practical usage, the more diverse dataset is necessary for generalization. Moreover, further research of the additional post-process is required to acquire global camera pose with high accuracy.

Acknowledgements

This research was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No.2019R1A2C2088859)

References

1. B. H. Y. Alsalam et al., "Autonomous UAV with vision based on-board decision making for remote sensing and precision agriculture," 2017 IEEE Aerospace Conference, Big Sky, MT, 2017, pp. 1-12, doi: 10.1109/AERO.2017.7943593.
2. Pion, Noé et al., (2020). Benchmarking Image Retrieval for Visual Localization.

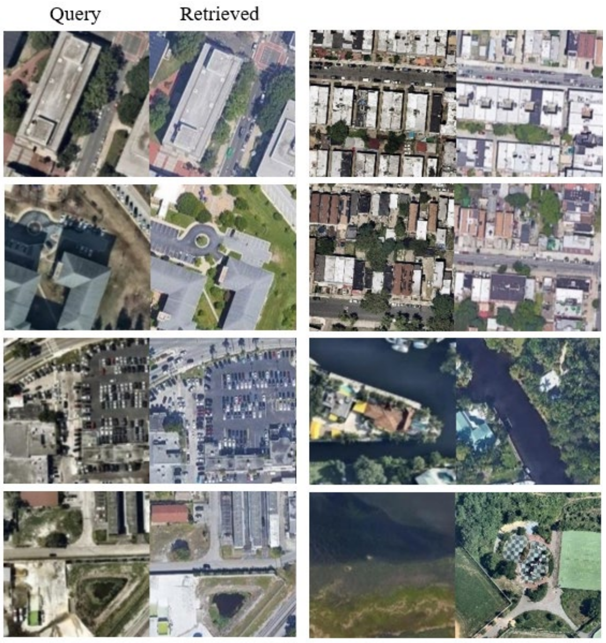

Fig 4. Example image pairs with good retrieval results (left column) and bad retrieval results(right column).

3. F. Schroff et al., "FaceNet: A unified embedding for face recognition and clustering," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015, pp. 815-823, doi: 10.1109/CVPR.2015.7298682.