A New Navigation System for Unmanned Aerial Vehicles in Global Positioning System-Denied Environments Based On Image Registration with Mutual Information and Deep Learning

Çağla Şahin, *TOBB University of Economics and Technology, Turkish Aerospace Industries*
İmam Şamil Yetik, *TOBB University of Economics and Technology*

BIOGRAPHY (IES)

Çağla Şahin is graduated from Bilkent University, Electrical and Electronics Engineering and she has been pursuing her M.S degree at Electrical and Electronics Engineering at TOBB University. She has been working at Turkish Aerospace Industries (which is one of the R&D companies in Turkey) as an avionic system design engineer since 2016.

İmam Şamil Yetik is an electrical engineer with a concentration on biomedical signal and image processing with the following educational background and experience: BS graduate of Electrical and Electronics Engineering at Bogazici University, MS graduate of Electrical and Electronics Engineering at Bilkent University, Turkey, PhD graduate of Electrical and Computer Engineering at University of Illinois at Chicago, Postdoc at Biomedical Engineering, University of California at Davis. He has joined the Department of Electrical and Electronics Engineering at TOBB ETÜ as an Assoc. Prof. in 2013.

ABSTRACT

In this paper, we develop an alternative navigation system for Unmanned Aerial Vehicle (UAV) in Global Positioning Systems (GPS)-denied environment. We use two image inputs, one is acquired with an on-board camera placed on the UAV (which is the large-area image) and the other is from satellite images (which is small known image) with GPS information. We use a convolutional neural network (CNN) architecture based on Oxford's Visual Geometry Group network (VGG-16) and utilize normalized variant mutual information between these two images to obtain position of the UAV. Satellite images are labelled and given to the UAV. When GPS information is lost, our algorithm starts to function and images from UAV camera are searched whether satellite image is seen by cameras on UAV image or not. If the UAV is in that area, our algorithm finds the GPS information from satellite image data.

INTRODUCTION

Inertial Navigation Systems (INS) and GPS provide latitude, longitude, altitude and velocity information of UAV. In GPS-denied environments we have to rely on information from INS only. However, INS alone cannot be used since it is affected by the drift of UAV, and accumulating errors from the sensors producing unusable information. GPS vulnerability is an important problem that needs to be solved for UAVs [1]. To remedy this problem and maintain a safe flight in the GPS-denied environments, there are alternative approaches such as geo-referenced navigation and vision-based navigation to correct errors stemming from INS.

https://doi.org/10.33012/2020.17203

Terrain navigation methods focus on terrain model and aircraft knows its position by using terrain model and radar altimeter sensors. These methods with terrain models are especially employed in very high speed aircrafts such as jet fighters. It cannot be used for low speed aircrafts, since altitude alteration is not suitable for precise ground matching [2]. There is also another alternative navigation system which utilizes cross correlation between two image sets. Both cross correlation and mutual information have positive and negative sides for different use cases [3]. Cross correlation is a measure for similarity of two images, however, it is affected by intensity. For example, when UAV flies on the same area in different season, cross correlation cannot detect the area because of the intensity value.

This paper proposes a novel navigation method based on registration with mutual information and deep learning. Our method uses two types of images sets: one image set is acquired with an on-board camera placed on the UAV and the other image set comes from satellite images with the exact GPS information. When there is jamming or loss in GPS information producing a GPS-denied environment, the images on the satellite image dataset are searched in the images acquired by the on-board camera within a certain neighborhood of the last tracked position of the UAV.

We obtain one image set from on-board camera of UAV in [4], the images are acquired in Cedaredge, Colorado. The second dataset is from Google Maps [5], we find Cedaredge, Colorado and save the places which are shown in drone dataset as satellite images. We also label the dataset whether the view of small image which is satellite image is shown in drone data images or not. When we label the images, we know the correct information whether the satellite image is viewed in drone dataset before the proposed solution.

We perform feature extraction for two image sets using a convolutional neural network architecture based on Oxford's VGG network (VGG-16), which is trained on the 1000-labels ImageNet dataset [6]. These features first enter the binary classifier feed-forward neural network and the algorithm decides whether the target image exists in the main image acquired by UAV. If the output is positive, i.e., the probability of presence is greater than a certain threshold, another neural network predict the location of the target image in the main image by estimating four parameters of the boundary, which are (x_c, y_c, h, w). These parameters represent the center point in x- and y- directions, height and width, respectively. In addition, we find mutual information between these two images for varying rotation angles and predict the location of the target image based on mutual information.

We approach the problem as a regression. To train the model, we use backpropagation algorithm where the loss is Euclidean loss. When the images are matched, the GPS information for the UAV is directly obtained from corresponding image in the satellite dataset. We emphasize that our architecture does not only contain an end-to-end convolutional neural network, but also uses intermediate features from mutual information. Our method is highly adaptive due to inherent characteristics of neural networks. In addition, our method is robust to varying environmental conditions, since the mutual information uses joint density of two images unlike many other template matching navigation solutions using intensity values.

PROBLEM DESCRIPTION

In this paper, we develop an alternative navigation system for UAVs. There is an on-board camera of UAV and it acquires pictures of UAV's route. When the aircraft arrives in a non-GPS area and generates a GPS invalid alert, our algorithm starts to function. Flight path of the UAV is assumed to be known in advance and satellite images of the flight path are given in advance on the UAV as a small satellite image dataset.

In this work, first problem is to know whether the image from UAV camera exists in the satellite image set or not. In order to solve this problem, a multilayer neural network with a binary output is used. We train the system before UAV flies on this area. We take sample video from UAV on this area and we take satellite images from these area with known GPS information. After we train the system, whenever UAV flies on the area and GPS invalid alert is taken, this algorithm starts and decides whether the satellite image with exact GPS information is in the image from the camera on UAV. In this framework, we provide only the GPS information belonging to the satellite images, i.e., without altitude information.

Second problem is scaling. When the algorithm decides the image from satellite dataset is present in the image from the camera of UAV, satellite image is searched in image from UAV. The size of the object in satellite image is different from object size from image of on-board camera. To solve this problem, we first train a neural network, whose inputs are the features of two images extracted by the VGG-16 network and predict coordinate of objects from the satellite image in the image from camera with a multilayer feed-forward neural network. However, neural networks mainly are vulnerable to small changes, which human eyes cannot even notice [7]. For example, addition of a small Gaussian noise may cause a neural network classifier to misclassify with a high confidence. Therefore, we do not directly use these predictions for the coordinates. Instead, we predict the size of the target image on the image acquired by the camera of UAV, we then resize the satellite image size with image object size from camera, then apply the normalized variant mutual information between resized satellite image and on-board camera image to make our system more robust. To provide the boundaries predicted by the neural network as a prior information to the mutual information block, we multiply the mutual information values corresponding the predicted boundaries with a small coefficient $c > 1$, e.g., $c = 1.2$. Note that setting $c = 1$ means no information is passed to the mutual information from the neural network other than size of the object. Furthermore, setting c to a large value may cause output of mutual information block to converge to the output of neural network directly, which is an undesired case. For this purpose, we set c to a value slightly greater than 1 as stated above.

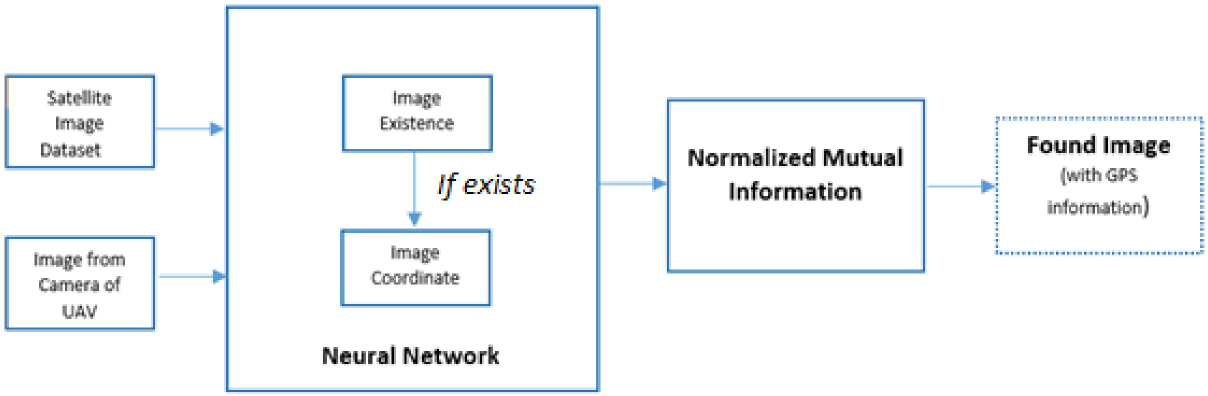

Figure 1: Diagram of proposed navigation system algorithm

NAVIGATION SYSTEM BASED ON NEURAL NETWORK

We use a convolutional neural network architecture based on Oxford's VGG network (VGG-16), trained on the 1000-labels ImageNet dataset. The VGG-16 network provides feature extraction of an image as shown in the following figure:

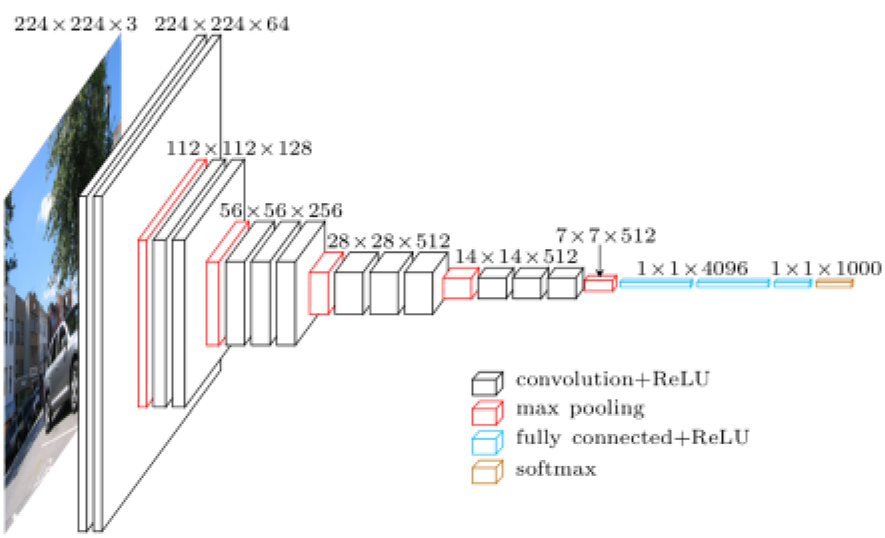

Figure 2: Architecture of VGG-16 [8]

Input of first convolution layer is a fixed 224x224x3 size RGB image. After convolution layers with ReLu activation function and max pooling, we obtain a feature map whose size is 7x7x512. VGG-16 architecture continues with fully connected network and ReLu activation function in order to classify the images in terms of type [6]. However, in our system, we do not continue to use the fully connected network which is used in VGG-16, we design different fully connected layers to detect the existence of the image and predict the boundaries of the target image.

Because of the fact that VGG-16 network only takes 224x224 size images, we fix size of images before giving them in the CNN, i.e., resize images to 224x224. We give a satellite image and an image from camera to VGG-16 network and take 25088 features for each image.

In order to decide whether the satellite image exists in the image from UAV camera, these features enter the final feed-forward neural networks. We use two layer feed-forward neural networks for existence of image. As an activation function of the first neural network, we use ReLu and as a second neural network activation function, we use sigmoid function. Mathematically, ReLu function is defined as $y = \max(0, x)$ and sigmoid function is defined as $y = \frac{1}{1+e^{-x}}$. Because sigmoid function fixes the value in [0-1], we take the value between 0 and 1 while using this neural network, which presents the probability an image exists in the other image. We decide whether the satellite image is shown in image on UAV by the output of sigmoid function. To train this classifier network, we use backpropogation algorithm with cross entropy loss.

The second network for navigation system is to predict four parameters, which are $(x_c, y_c, h, w,)$. These parameters represent the center point in x- and y- directions, height and width, respectively. We use three final feed-forward neural networks and as an activation function, we only use ReLu. To train this network, we use backpropogation algorithm with mean squared error loss. In both cases, Adam [9] is used as the optimizer with learning rate 0.001.

NAVIGATION SYSTEM BASED ON IMAGE REGISTRATION WITH MUTUAL INFORMATION

Mutual information is computed by using entropy of the signals. Entropy measures the disorder of the system, if the entropy is high, the disorder of the system is also increased. Shannon entropy is widely used in image registration studies [10], [11] and [12]. We denote Shannon entropy by $H(X)$ of random variable X and it is defined as:

$$H(X) = -\sum_{i=1}^{n} p_i \, log(p_i)$$

where p_i is the probability of the ith event. Entropy of the variable means the amount of information which is included in the variable. $P_{XY}(x, y)$ is joint probability distribution of X and Y, $P_X(x)$ is the probability distribution of X and $P_Y(y)$ is the probability distribution of Y. The relation between $P_{XY}(x, y)$, $P_X(x)$, $P_Y(y)$ are given in following equation:

$$P_X(x) = \sum_{y} P_{XY}(x, y)$$

$$P_Y(y) = \sum_{x} P_{XY}(x, y)$$

In order to find the similarity of two discrete variables X and Y, mutual information (MI) is computed, MI equation is given by the following equations [10], [11], [12].

$$MI(X, Y) = H(X) + H(Y) - H(X, Y)$$

$$MI(X, Y) = H(X) - H(X|\, Y)$$

$$MI(X, Y) = H(Y) - H(Y|\, X)$$

Where $H(X)$ and $H(Y)$ are marginal entropies, $H(X|\, Y)$ and $H(Y|X)$ are conditional entropies and $H(X, Y)$ is joint entropy.

$$H(X, Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} \log(p_{ij})$$

$$H(X|\, Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} \log(p_{i|j})$$

$$H(Y|\, X) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} \log(p_{j|i})$$

$$H(Y) - H(Y|\, X) = -\sum_{j=1}^{m} p_j \, log(p_j) + \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} \log(p_{j|i}) = \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} \log(p_{j|i}) - \sum_{j=1}^{m}\sum_{i=1}^{n} p_{ij} \log(p_j)$$

With Bayes theorem ($p_{ij} = p_i \, p_{j|i} = p_j \, p_{i|j}$), MI is calculated as the following equation [10]:

$$MI(X,Y) = H(X) - H(X|\,Y) = H(Y) - H(Y|\,X) = \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}\log\frac{p_{ij}}{p_i\, p_j}$$

Mutual information can take value 0 to infinity. When similarity is increased, the mutual information value also increases. In order to handle with larger mutual information values, it is needed to bound values, therefore, we use normalized variant on mutual information [13]

$$NVMI(X,Y) = \frac{MI(X,Y)}{H(X) +\ H(Y)}$$

We calculate normalized variant mutual information (NVMI) by shifting the resized satellite image on the image from on-board camera for different rotation angle (0, 90, 180, 270 degree) and we save each NVMI value in matrix with row, column and angle information. We choose the maximum NVMI information in NVMI matrix, and draw the expected image on the drone dataset image.

SIMULATIONS

In this section, we illustrate the performance of proposed navigation system in GPS-denied environment. As the UAV images we use the dronemapper data images [4] and for the target images with pre-known coordinates we use images from Google maps [5] at the same location. We work on a small datasets with 82 images in total. There are 90 annotated objects in these images, where the number of unique labels is 11. We split the annotated objects as %90 (81) and %10 (9) as the training and test sets, respectively. We repeat our experiments with 3 different random splits. To mitigate the small data problem, we use data augmentation. Here, we expand the training set by including the rotated version of training images, where the rotation angles are $\theta = [0, 90, 180, 270]$. An object in the image is counted as “found” if the Intersection of Union (IoU) score between the prediction the ground truth is greater than a certain threshold $\alpha = 0.20$.

Based on this setup, the accuracy of the algorithm on the test sets is %$\mathbf{88.89}$. The mean IoU score of the dataset is $\mathbf{0.702}$.

In Figure 3, there is a satellite view with GPS information and we label the target places in the satellite images.

Figure 3: Satellite view from Google Maps [5].

In Figure 4 and Figure 5, images are from on-board camera of UAV and red marked borders show the places which are from satellite images and have known GPS information in advance. Our algorithm finds these satellite images in the image from camera on UAV and matches the GPS information. Even though the images from on-board camera is taken at different angles 0, 90, 180, 270, the satellite image can be matched on the image from UAV. Figure 4 is for matching for satellite target 1 and Figure 5 is marked for satellite target 7.

Figure 4: Finding view of satellite image target 1 in the DJI 0001 image which is from on-board camera of UAV in four rotation angle.

Figure 5: Finding view of satellite image target 7 in the DJI 0012 image which is from on-board camera of UAV in four rotation angle.

In the figures, right image is viewed in satellite view from Google Map shown in Figure 3 and left image is from the camera of UAV. When UAV is flying on the known route, the satellite image view is matched with the image view from camera. Because GPS information from places shown in satellite image is known in advance, GPS information is known even in GPS-denied environment.

Figure 6: Left image (target 4 is boarded) which is from on-board camera and right image (target 4) which is from Google map shown in Figure 3.

Figure 7: Left image (target 7 is boarded) which is from on-board camera and right image (target 7) which is from Google map shown in Figure 3.

CONCLUSION

In this paper, we studied image registration based on deep neural networks and the mutual information on a small dataset. We used pre-trained VGG-16 network for feature extraction. Next, we trained a classifier and a predictor networks using these features, where the classifier network provides binary decisions about if a target image exists in the given image and the predictor network gives the candidate boundaries of the target object. The target image is resized based on these boundaries. Next, we created a mutual information map by shifting the resized target image on the UAV image and then locate the target where the mutual information is maximum. We demonstrate the performance of the proposed algorithm in the simulations.

REFERENCES

1. James, C., “Vulnerability assessment of the transportation infrastructure relying on the global positioning system,” *Volpe National Transportation System Center, August 200,* pp. 35
2. Conte, G., Doherty, P., “Vision-based Unmanned Aerial Vehicle Navigation Using Geo-referenced Information,” *EURASIP Journal on Advances in Signal Processing 2009, June 2009,* pp. 2
3. Ghannam, A., Abbott, A L., “Cross Correlation versus Mutual Information for Image Mosaicing”, *International Journal of Advanced Computer Science and Applications (IJACSA) 2013,* pp. 94-96
4. 4th Ave Reservoir Capacity Map – Cedaredge, Colorado, http://dronemapper.com/sample_data/, (accessed 2019)
5. Google Maps. "Cedaredge, Colorado"
6. Simonyan, K., Zisserman, A., “Very Deep Convolutional Networks for Large-Scale Image Registration,” *ICLR 2015, April 2015,* pp. 2-4
7. Goodfellow, I J., Shlens, J., Szegedy, C., “Explaining and Harnessing Adversial Examples,” *ArXiv, December 2014*
8. Garcia, A., Orts, S., Oprea, S O., Martinez, V V., “A Review on Deep Learning Techniques Applied to Semantic Segmentation,” *ArXiv, April 2017,* pp. 3
9. *Kingma, D P., Ba, J., “Adam: A method for Stochastic Optimization,” ArXiv, January 2017*
10. Bardera, A., Feixas, M., Boada, I., “Normalized similarity measures for medical image registration,” *SPIE 2004, Society of Photo-Optical Instrumentation Engineers, May 2004,* pp. 2-3
11. Dame, A., Marchand, E., “Using mutual information for appearance-based visual path following,” *HAL, December 2012,* pp. 3-4
12. Pluim, J P W., Maintz, J B A., Viergever M A., “Mutual-information-based registration of medical images: a survey,” IEEE Transactions on Medical Imaging, Vol. 22, No. 8, August 2003, pp. 986-990
13. Huang, S M., Huang, C C., Chou, C C., “Image registration among UAV image sequence and Google satellite image under quality mismatch,” *12th International Conference on ITS Telecommunications 2012, November 2012,* pp. 313