Correction of Mobile Positioning and Direction via CNNs Based on Street View Images

Lei-Ju Chu and Wei-Ming Chen*
National Ilan University, Taiwan

We present an application of convolutional neural networks (CNNs) in street view images recognition to help users correct positioning errors and find the right direction under GPS positioning. We create a CNN model for each GPS coordinate point, and the datasets for each model are street view images of eight directions of each point and image samples of the surrounding neighboring and other points. The output labels of each model are eight directions and others. Therefore, we could use a CNN model to rough identify whether an image belongs to the assigned coordinate point and to know which possible direction it is facing. In addition, in order to improve the efficiency of matching possible points, we propose a spiral search neighbors approach. Finally, we use speeded up robust features (SURF) for more precise positioning.

I. INTRODUCTION

In outdoor environments, smart phones with GPS [1], [2] capability can track the location of the device and navigate to the destinations that device users want to go. However, sometimes there are two annoying problems with mobile navigation applications, positioning error and wrong direction.

This paper presents a street view images recognition system which uses a spiral search and convolutional neural networks to narrow down the positioning region and then uses SURF [3] to help correct positioning errors and find the right direction.

Our proposed spiral search is used to determine the next possible search point if the previous point is not the target. In 2D image processing, CNN [4], [5] is usually used for pattern recognition tasks and is good for identifying the objects in the image, with the abilities of feature extraction and self-learning. Although CNN may need a long time to train a model, it is fast when predicting the result of an input. So we could use the CNNs to assist quickly searching a possible position region and a possible direction. However, in the image size and the focal length, the photo the user took is usually different from the image that is in the rough direction of the possible point based on CNNs. Hence, we use SURF to match the user photo with the image of the point we find for obtaining more precise positioning.

II. PROPOSED METHOD

A. Datasets

We pre-prepared many street view images for training CNN models and extracting SURF key-points to assist the follow-up precise positioning, and each image is a part of a panorama in a GPS coordinate and represents a direction.

As Fig. 1 shows, we define each CNN model is responsible for identifying an assigned GPS coordinate point, and the distance of neighbor points is about the same. We train each CNN model via the street view images of eight principal directions of the assigned point to find the right direction. The eight directions are north (N), northeast (NE), east (E), southeast (SE), south (S), southwest (SW), west (W), and northwest (NW), respectively.

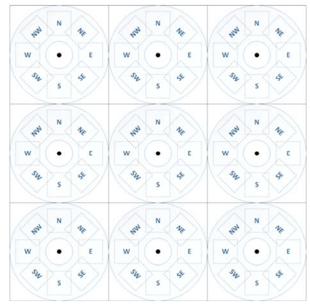

Fig. 1. Neighboring GPS coordinate points with street view image regions of eight directions

We increase trained images of each direction of each point through the technique of image augmentation because each direction of each point only has one original image is not enough to train each CNN model. Image augmentation transformations we do such as rotation, scale, shift, contrast, trapezoidal, etc. In addition, we sample these images in other points to train each CNN model to identify others; moreover, we input samples of surrounding eight GPS points to train each model not to be confused with the non-target GPS point.

B. Convolutional Neural Networks

Convolutional neural network consists of two major stages: training and prediction. As above-mentioned, we train numerous small CNN models, and each CNN model is trained and predicts for the assigned GPS coordinate point. The input datasets of each CNN model are images of eight directions and other neighbors, and there are nine output labels for identifying whether it is the right coordinate and which direction it is facing. Our CNN architecture for each point is shown in Fig. 2.

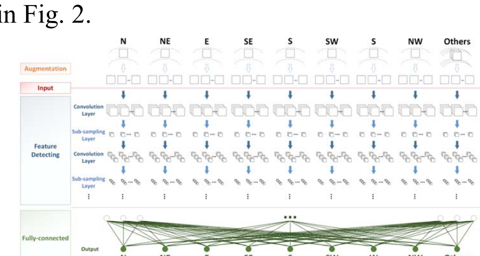

Fig. 2. CNN architecture for each GPS coordinate point

C. Spiral Search

We proposed an approach of finding out neighboring regions based on the concept of cellular networks to improve the efficiency of the matching processes. We use squares rather than hexagons, because we consider that the square is closer to the streets of cities rather than the hexagon, and many cities plan their streets to form grids and most roads are straight. We define many square areas, each for a GPS coordinate point only and form a large square network.

First, the method we proposed finds the closest GPS coordinate point in our defined square network based on a rough GPS coordinate uploaded by the user, and then checks the coverage area of the rough point to determine the order of the spiral search. If the CNN model of the closest point predicts the input image does not match, and then we match the next possible point according to the order of the spiral search until we find out. We propose four spiral search orders: top-left, top-right, bottom-left, and bottom-right, as shown in Fig. 3. In the first step, we choose the opposite y-axis direction of the majority of the coverage area, choose the top by default. In the second step, we choose the opposite x-axis direction of the majority of the coverage area, choose the right by default. We will search the next in a clockwise or counter-clockwise direction based on the above steps until we find.

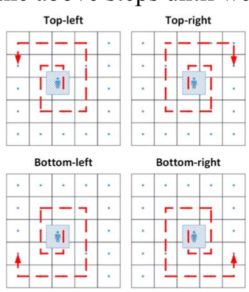

Fig. 3. Spiral search

D. *Speeded Up Robust Features*

According to the above methods, we can find a possible coordinate and its rough direction. Because of the focal length is different between the user image and the pre-prepared image, the coordinate and the direction are not accurate enough. For this reason, we proposed to use speeded up robust features (SURF) to assist positioning. Fig. 4 shows an example of extracting key-points with SURF, and the building in the picture is at Eluanbi Lighthouse, Taiwan.

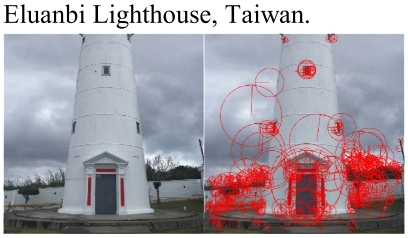
Fig. 4. Example of extracting key-points with SURF

We saved key-points and local features of each pre-prepared street view image into our image feature databases through SURF in advance. We also extract key-points and local features of the user-uploaded image and then we could do feature matching with the user-uploaded image and the image of the possible direction of the possible point based on CNNs, and then according to the corresponding key-points we will know their variances in translation, rotation, and scale. Therefore, we could find the more precise position and direction in accordance with these variances.

III. Experiment

Due to lack of human resources and time, we are unable to collect street view images from around the world. We only collected images of limited areas to verify our ideas. We took a street view panorama for each GPS coordinate point, and these points were about 3 meters apart. Each panorama was cropped into eight square images to representing eight directions, and each was extracted key-points via SURF and augmented to 50 images for training the CNN model of the assigned point. Fig. 5 presents an example of a panorama we took in the GPS coordinates of 25°05'23.4"N 121°28'19.9"E.


Fig. 5. Example of street view panorama

Even though the panorama is taken on a rainy day and the lightness is dark, our proposed system is able to recognize the possible target GPS coordinate via CNNs and the spiral search to improve the accuracy down to the 3 meter range. Furthermore, through SURF matching we could get an average accuracy of less than 1 meter even in areas where GPS signal is weak, and we could get the more precise direction.

IV. Conclusion

This paper has presented an application that applies CNNs on correction of mobile positioning and direction, and the experiment result is good in the limited region. We believe that it could be extends to assist autonomous cars to drive themselves if we could collect more street view images in the whole world and the points are the smaller distance apart.

References

[1] U.S. Coast Guard, Navstar GPS User Equipment: Introduction, September 1996.
[2] A. El-Rabbany, Introduction to GPS: The Global Positioning System, Boston: Artech House, 2002.
[3] H. Bay, A. Ess, T. Tuytelaars and L. Van Gool, "Speeded-Up Robust Features (SURF)," Computer Vision and Image Understanding, pp. 346-359, 6 2008.
[4] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard and L. D. Jackel, "Handwritten Digit Recognition with a Back-Propagation Network," Advances in neural information processing systems, pp. 396-404, 1990.
[5] M. Matsugu, K. Mori, Y. Mitari and Y. Kaneda, "Subject independent facial expression recognition with robust face detection using a convolutional neural network," Neural Networks, vol. 16, pp. 555-559, 2003.