BIO-INSPIRED MULTIPLE SCALES PLACE RECOGNITION FOR ELECTRIC SUBSTATIONS

Gang Wen[1, 4] , Fangrong Zhou[1, 4] , Hui Zhang[3] , Hao Pan[1, 4] , Jun Cao[1, 4] , Zhenyu Gao[3] ,
Yadong Liu[2] , Zhen Sun[2, *], Ling Pei[2]

[1]Joint Laboratory of Electric Power Remote Sensing Technology, Kunming, China - (1192381484, 42783590)@qq.com
[2]School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China - (zhensun, ling.pei)@sjtu.edu.cn
[3]Yunnan Power Grid Co., Kunming, China
[4]Yunnan Power Grid Co., Ltd. Institute of Electric Power Science, Kunming, China

Commission III, WG III/1

KEY WORDS: Electric facilities, Place recognition, Bio-inspired perception, Satellite remote images.

ABSTRACT:

We could get many helpful information and results from satellite remote sensing images and aerial images, including disaster monitoring, grid hidden danger identification, and electricity consumption management. In the recent years, novel computer vision and deep neural network have got a lot of attention in many fields because of mimicking mammalian cognitive mechanism as much as possible. With the in-depth of mammalian cognitive and motor mechanisms research, people trend to adopt these reliable and efficient methods for power grid management and maintenance.
For utilizing computing resources and improving analysing efficiency flexibly, we propose an assessing and verification framework based on bio-inspired perception and understanding, which summarizes the most appropriate image scale in the electric facilities place recognition. The proposed framework consists of different scenes aerial images datasets, several electric facilities place recognition methods, and credible evaluating methods mimicking mammals. Firstly, we gather satellite remote images and aerial images of sufficient electric power facilities in the United States via Google Earth and other public datasets. Then, several typical place recognition methods are adopted to testing recognition ability of multi-scale perception results, like SAD, NetVLAD, and GIST descriptor. To get more reliable result, multi-units and multi-scenes experiments are implemented roundly. After all experiments and evaluations, we could get that the most appropriate image scale is 1000m size and the highest recognition accuracy of electric power facilities location is 500m. Conclusion in our article shows the recommended perception form and scale closest to human cognition in the power grid management and maintenance.

1. INTRODUCTION

Large-scale scene monitoring is one of the fields, where computers would play a significant role. Power grid monitoring is a typical application on the background of large-scale scenes. Satellite remote sensing images and aerial images from high definition cameras are being used more and more widely with the development of aviation industry and unmanned aerial vehicle(UAV), which are the main sources of large-scale monitoring images.

Both place recognition and target recognition based on satellite remote sensing images need various images of different scales. The resolution ratio of analyzed images is higher, analysis results are more abundant. However, place recognition of electric facilities like substations does not need the highest resolution ratio. For instance, human being relies on rough visual input or other sensory signals when distinguishes electrical targets from various and complicated background. This objective law could be conformed to our intuition. As we know, brain can deal with information with high efficiency and low energy.

In this paper, we study how to implement electric facilities management and localization efficiently, which contains the listed parts as follows. Figure 1 shows the basic diagram. Diverse features and descriptors would be extracted from diverse satellite images. The goal of electrical place recognition is identifying the accurate location of substations. A key measure in the process of recognition is the speed and precision of results in the end. To get more reliable assessing results, this paper implements multi-unit contrast experiments with diverse bio-inspired algorithms of place recognition. As a result, the location of substations could be recognized with image sizes of high performance cost ratio.

Formally, our method contains typical algorithms of image matching and place recognition, which could simulate the recognition process of mammals like human beings. We apply Sum of Absolute Differences(SAD), new generalized Vector of Locally Aggregated Descriptors(NetVLAD) and the gist descriptor to complete our experimental design. The designed experiments are based on diverse dimensions representation of the real world scenes through self-supervised interaction methods. The validation scenarios come from a part of all substations in the U.S. We make a dataset that embodies satellite images of different scales in the static scenes and flying along the fixed trajectory. In our approach, matching results of place recognition are quickly computed with amount images.

In summary, we perform our experiments using objective statistical analysis and scale selection method. Algorithms aim to distinguish different substations for given object category, which use only visual observations. We show that scale 9 out-

* Corresponding author

performs in all three methods and recognition with SAD would get higher precision instead scale 1 being lowest. Moreover, time consumption of gist descriptor is the lowest and that of SAD is between other two methods. Additionally, the result of these experiments guides us to choose more appropriate image scale.

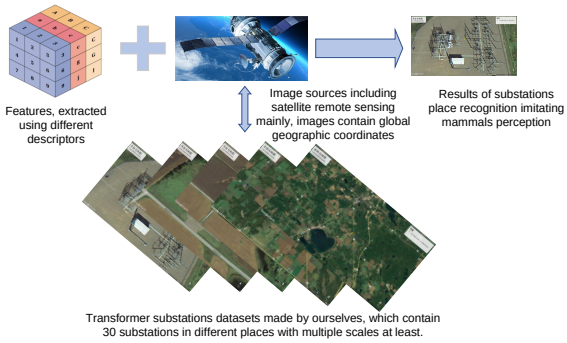


Figure 1. Diagram of Multiple Scales Substation Recognition

2. RELATED WORK

2.1 Algorithms for Electric Facilities Recognition

Failure to deal with the failure of overhead transmission lines in time will bring huge losses to the people. In order to minimize the economic losses caused by circuit problems, the power company will regularly inspect the power grid and make necessary planned repairs or replacements. The scale of the power grid is gradually expanding, the number of overhead transmission lines is increasing year by year, and the environment for towers and power lines is harsh. Most of the power lines are distributed in the wilderness. Due to natural erosion and man-made damage, a series of problems such as aging lines, damaged insulators, and loose anti-vibration hammers appear. If the maintenance is not timely repaired, the entire power system will lose a lot of electric energy, causing losses to the power company. Due to the vast territory of our country, most of the high-voltage transmission lines are distributed in hills, swamps, fields and other areas, and the terrain is complex. The traditional manual overhead transmission line inspection is not only labor-intensive, low in work efficiency, but also has personal safety problems. Usually maintenance personnel It is necessary to cross complex terrains such as mountains, rivers, and fields. More importantly, manual power line inspection will affect the daily operation and maintenance of the line.

In power inspection image processing, target extraction is the key. The complexity of the background and the variety of changes make the difference between the image background and the target very small, so the extraction of target features and the removal of image background are the bottleneck problems in the power inspection image processing.

The methods of target detection mainly include image segmentation, traditional machine learning and deep learning methods with a very high recognition rate. The discriminative model and the generative model are the classic target tracking methods. According to the characteristics of the cross structure of the transmission tower, the linear shape, and the insulator hanging on the top of the tower, this paper adopts the traditional object detection method to identify and locate the classic iron tower, concrete tower and steel pipe tower, that is, the ORB feature, HOG feature and LBP of the transmission tower are extracted. The fusion feature of the feature, using the XGBoot classifier to identify, target tracking can improve the detection efficiency and accuracy, so on the basis of identifying and locating the transmission tower, the KCF algorithm is used to track the transmission tower to realize the identification and tracking of the transmission tower. The convolutional neural network can automatically select excellent target features and use the improved YOLOv3 to identify insulators, that is, replace the Darknet-53 feature extraction convolutional network of YOLOv3 with MobileNets. According to the linear characteristics of the transmission line, the LSD algorithm is used to extract the straight line in the image, and then through two straight line screening, the least squares method is used to fit the line to obtain the transmission line target, and then the lossless Karman filter is used to realize the power line tracking. If the above targets are lost, the system will search and identify the target again, and then carry out target tracking to realize the long-term automatic identification and tracking function of the target.

The linear feature of the original image is enhanced by the improved Marr-Hildley edge detection algorithm, the straight lines and curves are extracted from the image by the Hough transform, and finally the power lines are identified by the morphological method in image processing. The above method only realizes the function of extracting and identifying power lines, but it needs to be further improved. In 2012, Zhang Jingjing et al. used Hough transform to extract power line segments, used K-means in Hough space to cluster and filter straight lines, and used Kalman filter to track power lines in Hough space(Zhang et al., 2012). In 2016, Chen et al. developed an improved clustering radon transform for extracting straight lines from satellite images(Chen et al., 2016).

2.2 Place Recognition Algorithm

For extraction and identification of insulators with high efficiency, domestic and foreign scholars have carried out various researches on the identification and positioning of insulators using visual algorithms. The operation speed of the satellite and air vehicle is very fast. During the inspection process, the geographic location is changing all the time. Therefore, the image background is also complex and changeable. Extracting insulator information from the complex background is the key to further discovering insulator defects.

Place recognition is the key element of agent motion and management. The key to Visual Place Recognition lies in the representation of easily distinguishable location image features. Early visual location recognition research generally used traditional manual local features (such as SIFT(Lowe, 1999), SURF(Bay et al., 2006), ORB(Rublee et al., 2011), etc.) as the feature representation of the location image, such as DBoW2(Zhang et al., 2019) using ORB features and bag-of-words model for location recognition. Or adopt or design a global feature vector to represent the current position image, such as HOG(Mizuno et al., 2012), Gist(Arandjelovic et al., 2016) and other features.

In recent years, with the development of deep learning in the field of computer vision, great success has been achieved in image classification, face recognition, target detection, target tracking and other fields(LeCun et al., 2015). The reason why

deep learning can achieve such achievements depends on the powerful features of deep learning Representation ability, deep learning features have better robustness than traditional handcrafted features. Therefore, applying deep learning to position recognition tasks has become a feasible direction, which can solve the insufficient performance of traditional manual features when faced with large environmental changes. At the same time, how to design a visual position recognition algorithm based on deep learning with excellent performance can be used in Meet the requirements of the accuracy of location recognition while High real-time performance is an urgent problem to be solved.

2.3 Bio-inspired perception and behaviors

As early as 1948, Tolman et al. proposed the concept of cognitive maps, which believed that animals must have the ability to obtain spatial information, which provides navigation information for its space activities(Tolman, 1948). Recent research has also revealed that the human brain has an organizing framework, which generates and utilizes cognitive maps.(Epstein et al., 2017).

The important function of cognitive map is to establish a grid map model(Stensola et al., 2012) for spatial information to realize the coding expression of inter-information. Another important function of cognitive maps is to use landmarks to relocate. The study found that the hippocampus(Douglas, 1967) know the map snapshot, it also stores the landmarks in the experience map to provide guidance when entering the same area again.

The mechanics of mammalian spatial cognition have a lot of potential for informing new algorithms that can help mobile robots navigate better. In this investigation, the cognitive model proposed by Zeng et al.(Zeng and Si, 2017) uses grid cells and head direction cells to describe the allocentric position and orientation within the environment. When a plane explores a new environment using cameras, this article investigates how a compact cognitive map encapsulates the surroundings efficiently and stores information sparsely.

With the advancement of technology and the development of research, various improved or newly designed handcrafts continue to appear. The artificial feature method is used for the position recognition task, which improves the accuracy and efficiency of the position recognition algorithm, but there are still some limitations. The current mainstream location recognition algorithms based on traditional handcrafted features have achieved good performance in indoor and outdoor situations with small environmental changes, but when faced with large environmental changes, the algorithm performance is average. Therefore, location recognition technology for large environmental changes has become the mainstream of current research.

3. METHODS

3.1 Architecture

Remote sensing is a non-contact technology for acquiring information on the earth's surface. By detecting and recording the electromagnetic wave radiation information of the ground target, processing, analyzing and applying it, the position, nature, attribute and changing law of the ground target can be determined. The remote sensing process is the process of acquiring, processing and applying images of ground targets, which is essentially the process of collecting, processing and applying the electromagnetic wave information of ground targets.

As shown in the Figure 2, we put forward to an recognizing and assessing architecture for multi-scale images of electric substations.

3.2 Describing satellite remote sensing image of substation

3.2.1 GIST Descriptor GIST descriptors are mainly used for scene recognition and were proposed by Antonio Torralba of MIT. The recognition and classification of scenes by GIST does not require image segmentation and local feature extraction. Compared with local features, this feature is a more "macro" feature description method, ignoring the local features of the picture.

We noticed that: most cities look like the sky and the ground are tightly connected by building facades; most highways look like a large surface extruded skyline filled with concavities (vehicles); and forest scenes will Included in a closed environment with vertical structures as backgrounds (trees) and connected to a textured horizontal surface (grass). The spatial envelope can characterize this information to a certain extent.There are five ways to describe the spatial envelope: naturalness, openness, roughness, expansion, ruggedness. Calculation method of GIST512 is listed as below:

- 32 Gabor filtering is convolved in 4 scales and 8 directions, and 32 feature maps are obtained with the same size as the input image;
- Divide each feature map into 4*4=16 areas, and calculate the mean value in each area;
- The regional mean number (16) of a feature-map is multiplied by the number of feature-maps (32) to obtain 512-dimensional GIST features.

GIST of different dimensions is characterized by the number of Gabor filters, specifically the number of filter directions and the number of scales.

3.2.2 SAD Algorithm Commonly used region-based local matching criteria mainly include the absolute value of the corresponding pixel difference (SAD, Sum of Absolute Differences) in the image sequence. The SAD algorithm is one of the simplest matching algorithms, which is expressed by the formula as (1)

$$SAD(u,v) = \arg\min SUM|Left(u,v) - Right(u,v)| \quad (1)$$

This method is to define a window D with the source matching point of the left eye image as the center, the size of which is $(2m+1)(2n+1)$, count the sum of the gray values of the window, and then gradually calculate in the right eye image. The difference between the gray-scale sum of the left and right windows, and the center pixel of the area with the smallest difference finally found is the matching point. Basic process is shown as below

- Construct a small window in the same manner as the convolution kernel.

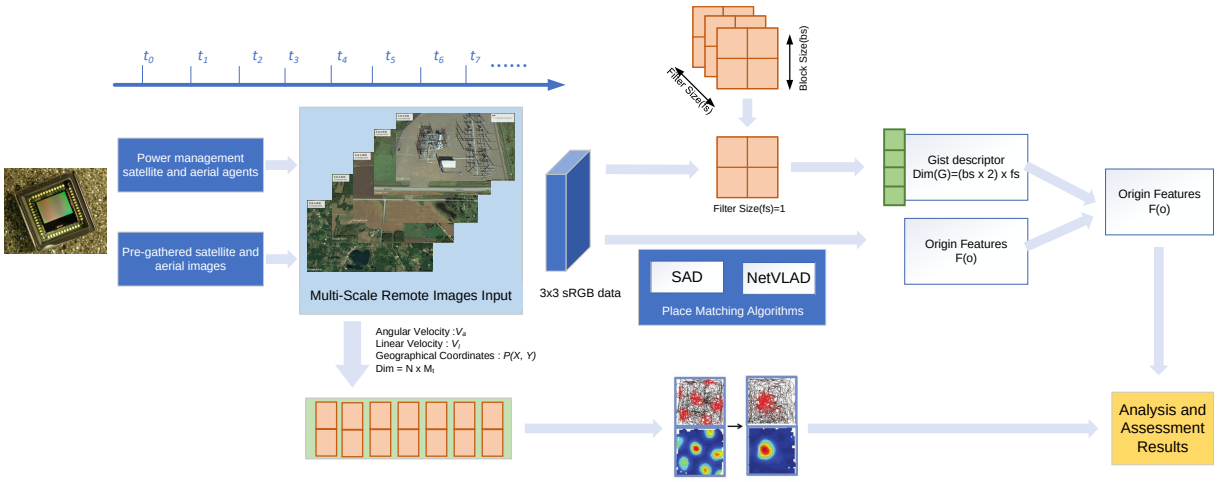
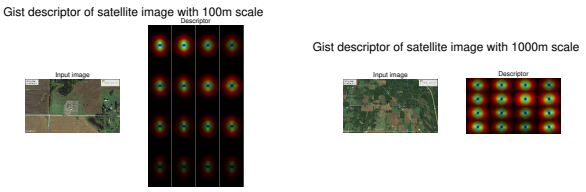


Figure 2. Architecture of proposed methods



(a) Image with 100m scale (b) Image with 1000m scale

Figure 3. The result of gist descriptor with different scale images

- select all pixels in the coverage area of window after cover the image on the left with the chosen window.
- Perform the same operation as in the previous step on the right image.
- Perform subtraction on the left and right coverage area to get the total of all pixel point differences.
- Move the window of the image on the right and repeat the actions of step 3 and 4. Notablely, it would jump out the search field if the window is out of the range.
- For determining the best matching pixel block in the left image, the lowest SAD value indicates required window in this range.

When the template size is determined, the SAD algorithm is the fastest.

3.2.3 NetVLAD Feature Detector NetVLAD is an improved version of VLAD. Usually, multiple local features of an image are obtained by traditional methods such as SIFT. Vector of Locally Aggregated Descriptors (VLAD) is one of the methods of compressing several local features into a global feature of a certain size. Through clustering, the feature dimensionality reduction is realized. VLAD is widely used on image retrieval tasks.

However, it is not feasible to directly embed VLAD layers in trainable CNN architectures, because $\alpha_k(x_i)$ in VLAD is discontinuous, so it is not feasible for $\alpha_k(x_i)$ modified, resulting

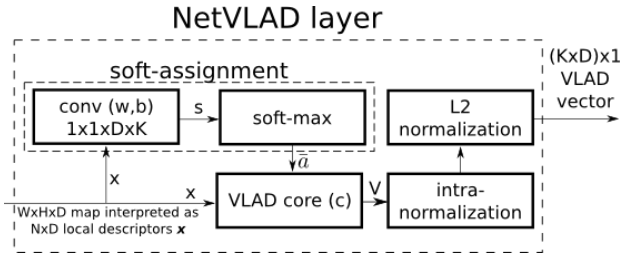


Figure 4. Architecture of NetVLAD Network

in NetVLAD. Replace the indicator function $\alpha_k(x_i)$ in VLAD with the derivable $\bar{\alpha}_k(x_i)$

$$\bar{\alpha}_k(x_i) = \frac{e^{-\alpha\|x_i - c_k\|^2}}{\sum_{k'} e^{-\alpha\|x_i - c'_k\|^2}} = \frac{e^{-\alpha\|x_i\|^2 - 2\alpha c_k^T x_i + \|c_k\|^2}}{\sum_{k'} e^{-\alpha\|x_i\|^2 - 2\alpha c_{k'}^T x_i + \|c_{k'}\|^2}} \tag{2}$$

Reduce the $e^{-\alpha\|x_i\|^2}$ in the numerator and denominator, and then let $w_k = 2\alpha c_k$ and $b_k = -\alpha\|c_k\|^2$, you can get $\bar{\alpha}_k(x_i)$ final expression for

$$\bar{\alpha}_k(x_i) = \frac{e^{w_k^T x_i + b_k}}{\sum_{k'} e^{w_{k'}^T x_i + b_{k'}}} \tag{3}$$

$\bar{\alpha}_k(x_i)$ represents the weight that assigns x_i to the cluster center c_k, ranging from 0 to between 1. The larger the $\bar{\alpha}_k(x_i)$ is, the closer the distance between x_i and c_k is. α is a constant, when $\alpha \to +\infty$, the value of $\bar{\alpha}_k(x_i)$ is infinitely close to 1 or 0. Finally get the expression of $V(j,k)$

$$V(j,k) = \sum_{i=1}^{N} \frac{e^{w_k^T x_i + b_k}}{\sum_{k'} e^{w_{k'}^T x_i + b_{k'}}} (x_i(j) - c_k(j)) \tag{4}$$

Where w_k, b_k, c_k are all learnable parameters.

3.3 Assessment Criteria

To begin, we pre-define the key variable notations in order to properly perform the metric assessment. Variable q_s^i is the ex-

tracted features from remote sensing images acquired by satellite in real time, using method above. Variable r_s^i is the precomputed features from the datasets made by ourselves. The subscript s denotes the ID of different scales. The superscript i indicates the ID of the different substation coordinates in the datasets. We suppose that the dataset contains N different substations in the Eq. (6), M in the Eq. (5) is total number of remote sensing images of substations, and each substation has S remote sensing images of different scales.

On the whole, this research conducts a realistic experimental approach for reliably evaluating the objective performance of brain-like cognitive recognition at various scales. For the substation location recognition scenario in this paper, we produce a dataset of satellite remote sensing images at different altitudes of different substations. Each remote sensing image has a unique location label. According to the current standard picture size, each remote sensing image is a RGB image with a consistent resolution of 1080p, i.e. the image size is 1920×1080. The satellite is programmed to fly above the substation in an overhead perspective, and the experiment is carried out. As long as the images are identified, the geographic coordinates of selected substations are displayed.

Subsequently, a size mixing strategy could be used in the process of identifying substation geographic coordinates. For matching and recognizing images of any scale, satellite remote sensing photographs are captured and processed via strategy above. After the satellite acquires the photographs, it compares and searches the collection for substation images of various scales. Calculating the Euclidean distance between feature vectors based on the characteristics retrieved from the preceding section is the most basic matching method. The Euclidean distance equals 0 in the event of perfect agreement. Euclidean distance $d_s(q, r)$ shown in the Eq. (5) is a common method to calculate feature similarity. Notations q and r are the corresponding simplified representation of q_s^i and r_s^i.

$$d_s(q, r) = \sqrt{\sum_{i=1}^{M} (q_s^i - r_s^i)^2}, \forall S \tag{5}$$

As the basic expression for calculating the Euclidean distance, it is evident that Eq. (5) is rudimentary and inaccurate to assess the similarity between two features. Therefore, we try to generate feature vectors under different scale conditions, normalize them to obtain the similarity of real-time images under different scale data set conditions, e.g., Euclidean distance. Then it is useful to sum up all the Euclidean distances and normalize them to get the similarity measurement of this photograph after matching.

Firstly, we apply the equation to calculate the local average of K different regions of the real-time remote sensing image. The division of regions is determined according to the division of image scales, and the K values are set as 10, 25, 50, 100, 200, 400, and N in this paper. After the local average is calculated, it is normalized as shown in the Eq. (6)

$$\mu_s(l) = \frac{\sum_{i=1+K(l-1)}^{K*l} d_s(i)}{K} \quad 1 \le l \le \frac{N}{K} \tag{6}$$

Then we can get the similarity results of the acquired images in a single scale using the local standard deviation and local mean. We sum up the similarity measurements at multiple scales using Eq. (7) to get the consistent optimal matching results in our datasets. Eq. (7) is the mathematical formulas, where $\mu_s(l)$ is the mean value and $\sigma_s(l)$ is the standard deviation value.

$$d_{\text{all}} = \sum_{n=1}^{S} \frac{d_s(i) - \mu_s(l)}{\sigma_s(l)}, \quad \forall i, \quad \forall s \tag{7}$$

Finally, the optimal matching result of the current remote sensing image in the dataset, i.e., the identified coordinate position of the substation, may be determined by finding the minimal value of the normalized joint similarity. The normalized result of the joint similarity takes 0 as the average value and takes 1 as the standard deviation. The calculation formula for matching the optimal substation location is shown in the Eq. (8). $\mathrm{M}(i)$ denotes the distance value of the best matching result. $\bar{d}_{\text{all}}$ and $\sigma\left(d_{\text{all}}\right)$ denote the mean and variance of the similarity distance, respectively.

$$\mathrm{M}(i) = \arg\min_{i \in N} \quad \frac{d_{\text{all}} - \bar{d}_{\text{all}}}{\sigma\left(d_{\text{all}}\right)} \tag{8}$$

As a consequence, the joint similarity is utilized as the basis for measuring the matching accuracy in practice. A value that is far in the negative direction from zero indicates a better matching estimation result. It is necessary to prevent some missing points from appearing while determining shared similarity within a particular range. The best matching estimation results are found in the feature results of datasets. We artificially set a certain threshold T. The matching results are valid when the joint similarity is less than the threshold T.

3.4 Statistical analysis and Scale selection

In the regular process of evaluating the accuracy of location recognition results, it is necessary to generate a precision-recall curve, which is strongly correlated with the above-mentioned threshold T. Precision refers to how many of the samples predicted to be positive are truly positive, and recall refers to how many of the positive examples in the sample are predicted correctly. The area under curve(AUC) is considered as the probability that the model will rank a random positive category sample above a random negative category sample. Generally, the value of AUC is between 0.5 and 1, the closer to 1, the higher the prediction accuracy.

This paper uses the student t test for statistical analysis. Student t can be used to assess whether there are statistical differences between multiple groups of experiments at different scales when multiple groups are conducted. Eq. (9) can be used to calculate the statistical test value t for comparing two groups of AUC, where $s_{(\cdot)}^2$ denotes the square deviation of the elements in the two AUC sets, respectively, and n is the sequence length at the time of testing the flight dynamics data, which is consistent at different scales in this paper.

$$t = \left(\overline{\mathrm{AUC}_2} - \overline{\mathrm{AUC}_1}\right) \cdot \sqrt{\frac{n}{s_1^2 + s_2^2}} \tag{9}$$

After the calculation was completed, the student test t calculated in Eq. (9) was checked against the t-values in the table to

correspond with a degree of freedom of $2n-2$ and a significance of $p < 0.05$. According to the method, we use the alternative hypothesis rather than invalid null hypothesis, when the value of calculated student t is greater than that in the lookup table.

A phenomenon is applied in this article that the field of grid cell tends to growth in multiples of $\sqrt{2}$ for each level lattice cell. Stensola and colleagues found this feature(Stensola et al., 2012). We followed this rule to determine the height corresponding to the different scales of the distribution.

On the self-made dynamic flight dataset, We choose the minimum size when the satellite height from the ground is 20 m. According to the existing basis, we determine 9 different image scales based on different satellite image heights, and the height from the ground used in the dataset in this paper is approximately as follows: 100 m, 200 m, 500 m, 1000 m, 2000 m, 5000 m, 10000 m and 20000 m. The minimum scale height is selected mainly based on the ability to completely cover the substation facilities and brain-like cognitive experience.

On the self-made static top-view dataset, we choose a similar approach to the flight dynamic data, where the growth rate of adjacent scale images is kept approximately in the ratio of $\sqrt{2}$. It should be noted that scaling at different spatial scales is based on different ground elevations at one view point. Then, the height sets of satellite images at different scales in the dataset are as follows: 20 m, 60 m, 100 m, 200 m, 500 m, 1000 m, 2000 m and 5000 m.

4. EXPERIMENTS

4.1 Collection of Satellite remote images dataset

To accomplish proposed idea sufficiently, we design two kinds of ways to acquire datasets based on Google Earth mannually.

4.1.1 Dynamic flight dataset In this way, we plan a flight route based on Google Earth, which overflew one substation known specific geographic coordinate. The flight path length is $9.84km$ and view of flight angle is looking down although these parameters could be configured according to designed experiments. We collect flight data at different altitude from the ground, which contain $100m$, $200m$, $500m$, $1000m$, $2000m$, $5000m$, $10000m$, and $20000m$. Multiple flight altitudes help us to verify recognition results variety of multiple scale satellite remote sensing images. We could find out detailed flight process via saved $*.kml$ file. The path diagram has been shown in the figure 5 and the yellow line display the trajectory.

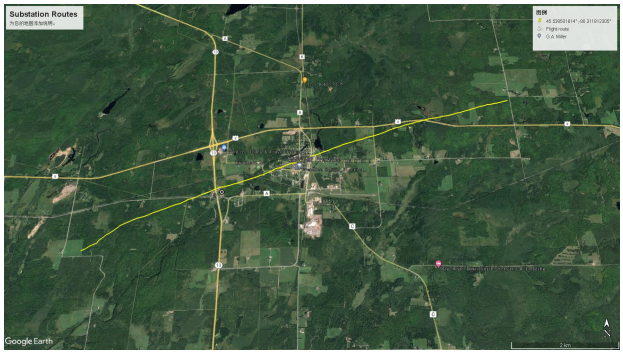
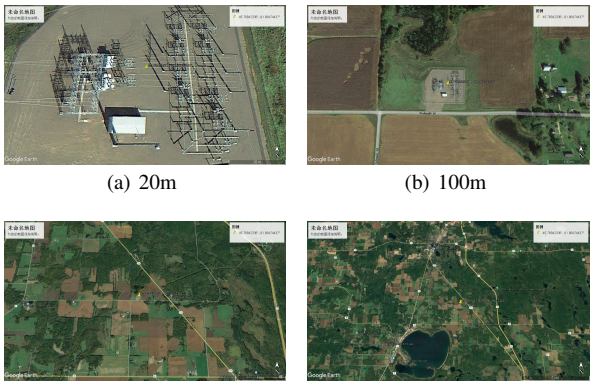


Figure 5. The flight trajectory

4.1.2 Static collected dataset Similarly, we gathered abundant static satellite remote sensing images of 30 substations in USA. The geographic coordinates are accurate through our tests. These images are collected at 5 different altitudes from the ground, which are $200m$, $500m$, $1000m$, $2000m$, and $5000m$. To simplify the influence of projection, these images are top view of the ground. Figure 6 shows two different scale images gathered.



(a) 20m (b) 100m

(c) 1000m (d) 5000m

Figure 6. Image samples in datasets made by ourselves

4.2 Experimental settings

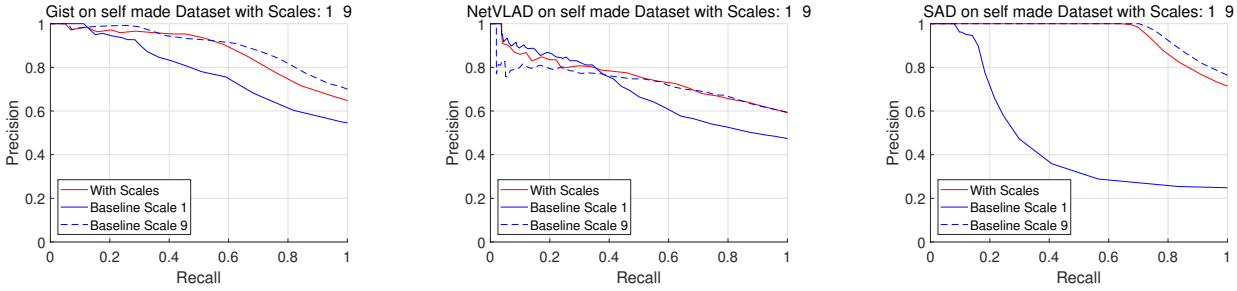
We design extensive experiments to obtain objective results of location recognition for substations. To avoid the error from imperfect processes and dataset sizes, we divided the self-made datasets into training-sets and test-sets rationally. Choosing different scales requires multi-group cyclic training and testing, depending on stable software and high-performance hardware platform.

The hardware and software platforms we used are described as follows. To run all programs on Windows 11 and Ubuntu 20.04.3 LTS, we used Python 3 (3.6.9 or 3.9.7 to run different programs), MATLAB R2020b, CUDA 11.0, and CUDNN 8.1.1. We trained and tested the location detectors of substations on an NVIDIA GeForce RTX 1080Ti GPU and an Intel Core i7-8700K with a clock speed of $3.70GHz$.

4.3 Performance of multiple scales images with different feature descriptors

We conducted adequate experiments to assess our proposed approaches to cover the shortage in previous study. To acquire the location recognition accuracy of substation satellite photos at various scales, accuracy-recall curves are produced for various scales. The GIST descriptor, SAD algorithm, and NetVLAD are evaluated primarily in terms of bio-inspired location recognition. The varied scales correlate to the height of the satellite from the ground where the photographs have been acquired, which is $20m$, $60m$, $80m$, $100m$, $200m$, $500m$, $1000m$, $2000m$, and $5000m$, respectively.

The main experimental results of this paper are shown in Fig. 7. In the case of satellite photograph at various scales, the performance of the GIST descriptor and the NetVLAD network detector for location recognition is relatively consistent. As a contrast, the location recognition accuracy of SAD algorithm in the case of large scale decreases significantly, whose accuracy rate of recognition can reach about 25% at the lowest. In



(a) Gist (b) NetVLAD (c) SAD

Figure 7. Recognition performance for multiple scale images

the large scale case, the location recognition accuracy of GIST descriptor is higher than that of NetVLAD by about 8 percentage, which is consistent with the characteristics of GIST descriptor for large scale field scenes as discussed earlier. These two algorithms focus not only on the pixel features in the image but also on the global features after extraction. While the SAD algorithm, as a visual matching algorithm for small scale scenes, is more accurate in small scale scenes. In summary, the ideal scales for several methods in satellite image location recognition of substations in the field environment include: scale 8 ($2000m$ height) for GIST descriptor, scale 7 ($1000m$ height) for NetVLAD feature detector, and using scale 6 ($500m$ height) for SAD algorithm.

In terms of running time, the GIST descriptor is the most efficient one with $33.499s$ in our experiments. Nevertheless, the NetVLAD feature detector has the highest running time of $87.727s$ and the SAD algorithm takes $48.916s$.

5. CONCLUSION

In this study, we attempt to use brain-like methods to enhance efficiency when managing and maintaining a large number of substations. Existing studies are mostly based on rodents. Low power consumption and excellent resilience properties of animals when executing activities like navigation are shown by the grid cell theory. In order to establish a link with existing brain-like cognitive outcomes, our work designs and uses three different location recognition algorithms through numerous sets of replicated experiments on dynamic and static satellite remote sensing data of various scales produced by ourselves. Finally, the effects of various spatial visual information scales on target recognition, such as substations, are assessed. Conclusion of this article provides a quantitative foundation for effective and low-cost remote observation, administration, and repair of power infrastructure such as satellite substations and aircraft.

ACKNOWLEDGMENT

We acknowledge the effort from research and application of constructing satellite remote sensing technology power application comprehensive test site and environment wide area intelligent monitoring, which is numbered by YNKJXM20191246.

This work is also supported by the National Nature Science Foundation of China (NSFC) under Grant 61873163.

REFERENCES

Arandjelovic, R., Gronat, P., Torii, A., Pajdla, T., Sivic, J., 2016. Netvlad: Cnn architecture for weakly supervised place recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 5297–5307.

Bay, H., Tuytelaars, T., Gool, L. V., 2006. Surf: Speeded up robust features. *European conference on computer vision*, Springer, 404–417.

Chen, Y., Li, Y., Zhang, H., Tong, L., Cao, Y., Xue, Z., 2016. Automatic power line extraction from high resolution remote sensing imagery based on an improved radon transform. *Pattern Recognition*, 49, 174–186.

Douglas, R. J., 1967. The hippocampus and behavior. *Psychological bulletin*, 67(6), 416.

Epstein, R. A., Patai, E. Z., Julian, J. B., Spiers, H. J., 2017. The cognitive map in humans: spatial navigation and beyond. *Nature neuroscience*, 20(11), 1504–1513.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *nature*, 521(7553), 436–444.

Lowe, D. G., 1999. Object recognition from local scale-invariant features. *Proceedings of the seventh IEEE international conference on computer vision*, 2, Ieee, 1150–1157.

Mizuno, K., Terachi, Y., Takagi, K., Izumi, S., Kawaguchi, H., Yoshimoto, M., 2012. Architectural study of hog feature extraction processor for real-time object detection. *2012 IEEE Workshop on Signal Processing Systems*, IEEE, 197–202.

Rublee, E., Rabaud, V., Konolige, K., Bradski, G., 2011. Orb: An efficient alternative to sift or surf. *2011 International conference on computer vision*, Ieee, 2564–2571.

Stensola, H., Stensola, T., Solstad, T., Frøland, K., Moser, M.-B., Moser, E. I., 2012. The entorhinal grid map is discretized. *Nature*, 492(7427), 72–78.

Tolman, E. C., 1948. Cognitive maps in rats and men. *Psychological review*, 55(4), 189.

Zeng, T., Si, B., 2017. Cognitive mapping based on conjunctive representations of space and movement. *Frontiers in neurorobotics*, 11, 61.

Zhang, J., Liu, L., Wang, B., Chen, X., Wang, Q., Zheng, T., 2012. High speed automatic power line detection and tracking for a uav-based inspection. *2012 International Conference on Industrial Control and Electronics Engineering*, IEEE, 266–269.

Zhang, Q., Xu, G., Li, N., 2019. Improved slam closed-loop detection algorithm based on dbow2. *Journal of Physics: Conference Series*, 1345number 4, IOP Publishing, 042094.