Preprocessing of SAR Image for Building Detection

Egor Busko
FAMCS of
Belarusian State University
Minsk, Belarus
busko.egor@gmail.com

Alexander Nedzved
FAMCS of
Belarusian State University
Minsk, Belarus
NedzvedA@tut.by
ORCID 0000-0001-6367-5900

Shiping Ye
Zhejiang Shuren University
Hangzhou, China
zjsruysp@163.com
ORCID 0000-0002-9771-7168

***Abstract.* In this paper, we propose high-resolution SARS image processing algorithm to improve quality of the YOLO v4 algorithm by special processing and alignment of different polarization channels.**

***Keywords:* SAR image, buildings detection, image processing, high-resolution SARS images**

I. INTRODUCTION

Detection and highlighting buildings in satellite images is very useful for many applications. It's helpful in building maps, creating a territory building plan, finding malicious and illegally constructed objects, etc. The main difficulty of the SAR image analysis task lies in the large number of different structures recognition. The task is complicated cause of the various shape, color characteristics and size of the objects.

The basis of SAR image establishment is the reflection features of the scattering of the emitted radar signal by various surface types [5]. The total intensity of signal reflection (pixelgray value) is affected by the characteristics of flatness and regular properties of objects. As a result of processing, the "raw" SAR data is converted into a gray image. The gray value in a pixel in SAR images, which is not affected by lighting, chemical composition (except salt and ice) and temperature of objects, depends on three factors: the SAR system, the SAR processing, and the properties of object. So, an object can be classified by properties of it, such as geometry, dielectric properties, and so on. Volumetric objects (for example, vegetation or other index decoration) correspond to an average level of gray value and texture, surfaces (for example, a calm water surface) shifts brightness to a dark value and buildings to be a bright value. The dielectric properties of the material affect the intensity of the reflected signal. The difference in the coefficient values for different materials makes it possible to identify by SAR.

There have been developed many algorithms for recognizing buildings in optical satellite images. Most of those algorithms are based on the analysis of object shape, texture, shadow, s boundary, etc. [1, 2]. Recently, neural network methods have been used for buildings extraction in optical satellite images [3]. The training dataset consists of satellite original images and their masks. They are binary images or contours of regions, where the pixel value corresponds to classes of objects. But there are only a few neural networks that can get quality results, for example, Fully convolutional network (FCN) [8], Mask R-CNN [9], CNNs [10]. Such methods have been successfully employed in computer vision and remote sensing fields for optical image classification, but few applied to SAR images.

The preparing of mask or region plays a very important role in quality of detection by neural network method. The size of mask is depending on size of shadow. Therefore, each building should be covered by a mask that changes according to patterns of shadow.

In this paper we propose preprocessing of SAR images to detect buildings for analysis by neural networks, where the training areas for detection are selected for each building whose shadow is detected by using the shadow shape. It allows us to improve recognition of regions of buildings The experiments were performed on set of high-resolution SARS images.

II. BUILDINGS REPRESENTATIONS IN SAR IMAGES

Buildings in SAR images have their own characteristics that allow to be detected. Each type of building has its own characteristics of shadow. So, the geometry of shadow is pattern for buildings recognition.

The basic feature for building detection in SAR image is shadow. The structure of shadow depends on features of buildings. The shadow, which is formed by location of building details with conductive properties angle and speed of satellite motion reflections of waves from corners of construction element of building.

The paper [5] proposed that SAR building shadow is described by geometrical point of view, then defined an evaluation function implementing the ratio of exponentially weighted averages (ROEWA) which is used for the matching between the predicted structure and the observed SAR image.

The projection of the three-dimensional building into the two-dimensional slant image plane influence shadow and produces effects such as layover, shadow and foreshortening. In addition, there are specular

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

reflections for close building for urban regions. The paper [5] proposes dense alignment for the buildings images along the radar look of sight (RLOS) for compensation the multi-path reflection effects [6] like as in Fig 1.

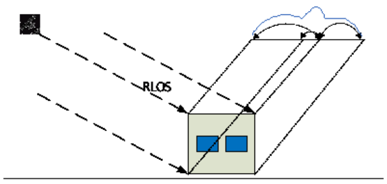


Fig. 1. Geometry scheme of SAR image generation for one floor building with flat roof

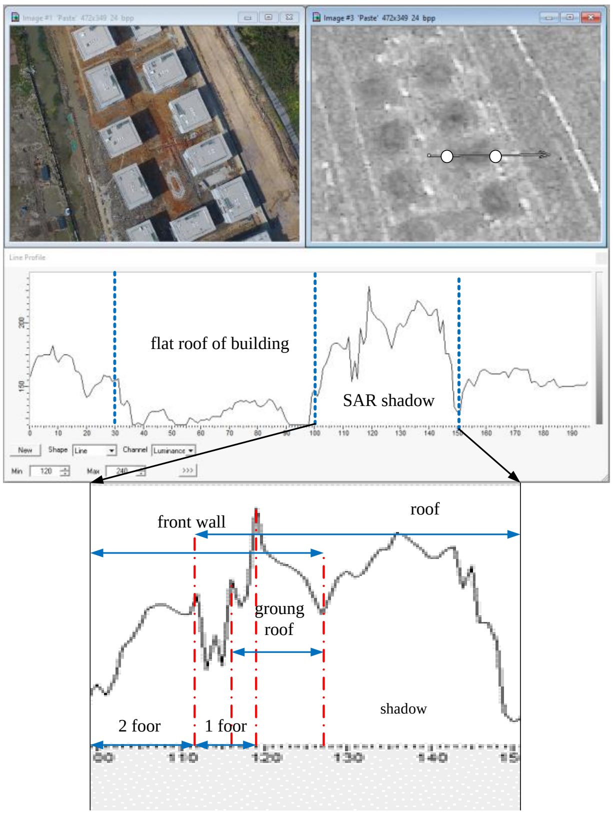


Fig. 2. Sample of profile of building on SAR image with shadow description

Profiles of shape of building shadow in the SAR image are divided into two basic types: the flat roof and the gable roof building. For example, the single bounce generated by the isolated flat roof of one floor building reflections from the ground, front wall and roof. The structure of one store building with flat roof is not complex. For multistory building, brightness profile has more complex structure, like as Fig. 2.

For urban buildings shadow has regular character where each local maximum corresponds to one floor.

Azimuthal and lateral resolution has different concepts. In this case, there is different brightness representation, but the structural elements of the SAR shadow are preserved.

III. Preparing of Dataset for Teaching Procedure

A. Data creation

Common SAR image processing consists of seven basic steps: creation, calibration, multilooking, selection of region of interest, indexing, segmentation and classification [7].

The first step is data creation (Fig 3). Imperially, it was found that the double polarization of VV+VH [4] allows to obtain more accurate results than the double polarization of HH+HV. Therefore, we use only combination VV+VH. In this step two branchs of processing are created for every type of SAR polarization.

Fig. 3. Sample of contrasting of a SAR satellite image

B. SAR Image Calibration

The second step is calibration. The calibration radiometrically corrects the SAR image that the pixel values are changed to the backscattering value of the radar beam from the reflecting surface. It is automatically determined based on satellite image metadata. Calibration radiometrically corrects a SAR image so that the pixel values truly represent the radar backscatter of the reflecting surface. The calibration corrections are realized by the SNAP software that automatically determine what corrections need to be applied to the image. Calibration is essential for quantitative estimation of SAR images.

Multilooking is used to produce a product with a conditional pixel size of the image. Its accumulation is formed by averaging the pixel resolution in range and azimuth, increasing radiometric resolution, but deteriorating spatial resolution. As a result, the image has an approximate square pixel size. It corresponds to conversion from inclined range to ground range.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

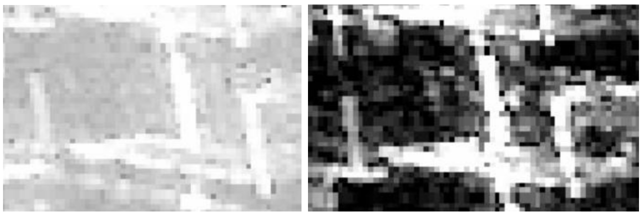
The multilooking is an optional step. It is not required when the image is adjusted for terrain. But we do it in common processing scheme.

Speckle reduction remove specific noise that is caused by random constructive and destructive interference during construction of the image. The resulting image of a particular pixel is obtained by adding a set of values from antenna sensors. Speckle noise is represented as graininess caused by chaotic alternation of light and dark pixels. The presence of it makes difficult to analyze radar images. Speckle filters are applied to SAR images to reduce the amount of speckle at the cost of blurred features.

The order of recording the return signal values in radar sensing of the earth's surface depends on the direction of motion of the satellite and the direction of sight, as a result of which the original images are not always correctly oriented relative to the cardinal directions. Terrain Correction is decoding of the image by correcting SAR geometric distortions. It includes geocoding and orthocorrection.

Geocoding is coordinate reference of the original or converted radar image without removing distortions for the relief. Such transformation converts an image from Slant Range or Ground Range Geometry into a Map Coordinate System.

Orthocorrection includes not only the coordinate reference, but also the elimination of distortions associated with the terrain, which uses a digital elevation model. As a rule, geocoding and orthocorrection of radar images is performed by orbital data. Such transformation is used a Digital Elevation Model (DEM) to correct for inherent SAR geometry effects on images such as shadow, foreshortening, layover. For terrain correction use range-Doppler algorithm.



a) b)

Fig. 4. Image of the building: a) before contrasting, b) after contrasting

Then SAR image usually contrasted for best representation. This operation corresponds to expression of intensity like as logarithmic transformation. It is possible spend analysis and comparation SAR images after such transformation because all distortions should be corrected (Fig. 4).

C. Selecting data samples by region of interest

The ROI selection is important step that depends on basic task. It is necessary doing because usually SAR satellite image has very big size and it is not possible to download it memory of computer. We use traditional way where the image is cut into tiles.

We use the method of cutting the image into sections. For realization of some functions, we use software QGIS [11]. The subdivided areas represent neighborhoods of houses with small buildings. We also left one skyscraper in one image.

After we have sliced our large satellite image into many small images, we need to label them. We will train the YOLO network to recognize 3 classes of houses: small cottages (private_building), multistory houses (multistory_building) and tall skyscrapers (skyscraper).

Using the tool for graphic annotation of images LabelImg [12] we mark our images with classes (Fig. 5).

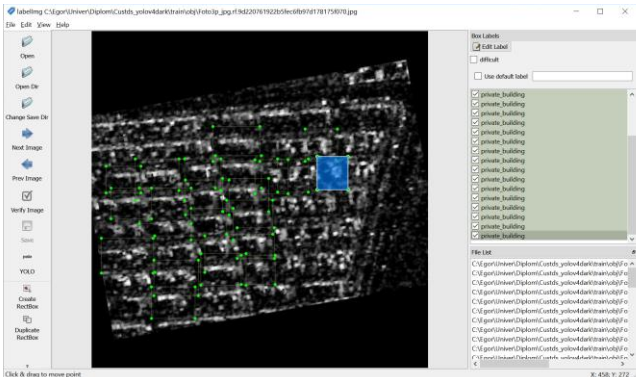


Fig. 5. Example of image marking

With the help of this utility we select our objects with rectangles. Each shape contains information about the location of an object in the image.

After labeling, each new image is supported by an annotation txt file in the same directory and with the same name, which includes the object's number and coordinates of the object in this image, for each object on a new line: <object-class> <x> <y> < width> <height>. The first field object-class is an integer representing the class of the object. It ranges from 0 to (number of classes - 1). In our current case since we have 3 classes from 0 to 2. This annotation was classified into the following classes:

- multistory_building;
- private_building;
- skyscraper.

The second and third x and y records are the x and y coordinates of the center of the bounding box, normalized (divided) by the image width and height, respectively. The fourth and fifth records, width and height are respectively the width and height of the

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

bounding box, again normalized (divided) by the width and height of the image, respectively.

We do not have classes of wooden and metal houses, due to the lack of a training set. Such objects are detected by the brightness threshold. Wood and metal buildings can be identified using a uniform brightness control area with a high value for metal buildings and a low value for wood structures. For metal buildings, this is a small range of maximum brightness values of about 5% of the brightness histogram. For wooden buildings, this is a small range of maximum brightness values of about 10% of the brightness histogram.

IV. Building detection

A. Segmentation and classification

Two last steps of building detection are segmentation and classification. They are very complicated procedures that have many realizations.

The segmentation is the process of partitioning a SAR image into multiple regions (connected sets of pixels, that correspond to objects). The goal of segmentation is to simplify the representation of an image into more meaningful to analyze. The classification is defining visual content to segmented regions. It is final step for detection building on SAR image. These two steps can be combined through the use of a CNN.

For our research we use images from RadarSAT-2 satellite. The resolution of such images about 1.7 meter per pixels. Type of such pixels is float.

B. Selection of buildings by CNN

Many detection systems repurpose classifiers or localizers to perform detection by using R-CNN, VGGNet, ResNet, Inception, and so on. They use the model only for images with selected scales in separate location. As rule, such algorithms use images with very high resolution.

YOLO based on a single neural network to the full image. This network divides the image into regions and predicts bounding boxes and probabilities for each region. Such box can include parts of different class of building with SAR shadow. These bounding boxes are weighted by the predicted probabilities. The basic method for detection of buildings we define as YOLO V4. It is multi-object detector. In our solution three types of images (VV, VH and RVI) are merged by concatenation. A layer that concatenates two inputs is along a specified axis, which corresponds to the concatenate layer. The inputs must be of the same shape except for the concatenation axis. Concatenation takes as input a list of tensors, all the same shape except for the concatenation axis, and returns a single tensor, the concatenation of all inputs. It converted multiple inputs into same shape layer. It is possible to realize by adding additional Dense layer to inputs of VV image and to result of such merging. The same shape outputs are used for every concatenation layer and for input. The basic architecture YOLO is shown at Fig. 6.

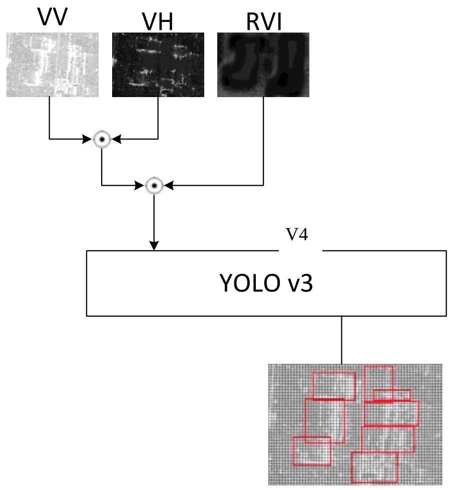


Fig. 6. Modification of YOLO v4 network for VV, VH and RVI inputs

C. Formation of a dataset for the YOLOv4 network

We will use the tool for systematization, preparation and improvement of training data Roboflow [13]. For the YOLO network, it is desirable to use images whose sizes are multiples of 32. Therefore, this problem was solved by adding black stripes at the edges of the images. In addition, we have a fairly small set of images to train the network. To enlarge the data, we use the rotation of the images by 15, 165, 90, 180, 270 degrees, as well as the reflection vertically and horizontally. All our markings are automatically corrected according to the changes made. After that, we form a dataset, which will consist of a set of images for training the network (train), images for verification (validation) and images for testing (test) our trained network (Fig. 7). We assigned most of the images to the training set (90% percent) and 5% to the testing set.

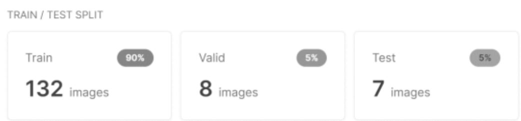


Fig. 7. Proportional partitioning of our dataset

You can also find out how many images of buildings of different classes appeared in our dataset (Fig. 8).

Fig. 8. The number of buildings of different classes in our set

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

Most of all there are houses of the private_building class in the SAR image, the skyscraper class turned out to be the rarest. Our datasets are now properly formatted for training and validation.

V. Conclusion

For the high-quality operation of the YOLO v4 algorithm, we optimize the data input through special processing and alignment of different polarization channels. The approach described in this paper has proven to be effective for detecting buildings. This approach can be used to detect any discrete objects. The key is to provide the correct training kit. This set must be balanced. The problems with our result are related to the lack of data for training. The dielectric properties of wooden structures form the low brightness of these objects. As a result, timber houses have a low detection rate. Complex buildings are often found not as one, but as several buildings. These problems are very difficult to solve. Thus, if the set is well balanced, then the neural network training result is better.

Acknowledgment

This work is supported by the National Foreign Experts' Program (G20200216025) and Public Welfare Technology Applied Research Program of Zhejiang Province under Grant (No.LGF18F030004, LGJ19F020002 and LGF19F020016).

References

[1] G. Cheng, J. Han, A survey on object detection in optical remote sensing images, ISPRS 117, 11–28 (2016).

[2] Salar Ghaffarian, Saman Ghaffarian "Automatic building detection based on supervised classification using high resolution google earth images" International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 2014, 40(3), pp. 101-106.

[3] X. Zhuo, F. Fraundorfer, F. Kurz, P. Reinartz, "Building Detection and Segmentation Using a CNN with Automatically Generated Training Data", IGARSS 2018 – 2018 IEEE Intern. Geoscience and Remote Sensing Symp., Valencia, 2018, pp. 3461-3464. doi: 10.1109/IGARSS.2018.8518521.

[4] Morton John Canty, Image Analysis, Classification and Change Detection in Remote Sensing, 3rd Edition 2014, ImprintCRC Press, 576 p.

[5] Z. Wang, L. Jiang, L. Lin, and W. Yu, "Building Height Estimation from High Resolution SAR Imagery via Model-Based Geometrical Structure Prediction", Progress In Electromagnetics Research M, vol. 41, 2015, pp. 11-24.

[6] A. Ferro, D. Brunner, L. Bruzzone, G. Lemoine, "On the relationship between double bounce and the orientation of buildings in VHR SAR images," IEEE Geosci. Remote Sens. Lett., vol. 8, no. 4, 2011, pp. 612–616.

[7] W. Liu, F. Yamazaki, " Building height detection from high-resolution TerraSAR-X imagery and GIS data," Proceedings of 2013 Joint Urban Remote Sensing Event, Sao Paulo, Brazil, CD-ROM, 2013, pp. 33-36.

[8] K. Bittner, S. Cui, P. Reinartz. "Building extraction from remote sensing data using fully convolutional networks." ISPRS – Intern. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII- 1/W1, 2017, pp. 481-486.

[9] Kang Zhao, Jungwon Kang, Jaewook Jung, et al. Building Extraction from Satellite Images Using Mask R-CNN with Building Boundary Regularization[C], 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, 2018.

[10] R. Hamaguchi, S. Hikosaka, Building Detection from Satellite Imagery using Ensemble of Size-Specific Detectors[C], 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, 2018.

[11] QGIS: Free open source geographic information system [Electronic resource], Documentation Center, https://qgis.org/ru/site/. – Date of access: 29.06.2021.

[12] LabelImg: graphical image annotation tool [Electronic resource], Documentation Center, https://tzutalin.github.io/labelImg/. – Date of access: 29.06.2021.

[13] Roboflow Modify Dataset [Electronic resource], Documentation Center, https://app.roboflow.com/datasets. – Date of access: 29.06.2021.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.