Multi-sensor Data Analysis for Aerial Image Semantic Segmentation and Vectorization

Vladimir A. Knyaz[1,2], Vladimir V. Kniaz[1,2], Sergey Yu. Zheltov[2], Kirill S. Petrov[2]

[1] Moscow Institute of Physics and Technology (MIPT), Moscow, Russia - (kniaz.vv, kniaz.va)@mipt.ru
[2] State Research Institute of Aviation Systems (GosNIIAS), Moscow, Russia - zhl@gosniias.ru

Key Words: Multi-sensor data analysis, image vectorization, semantic segmentation, maps updating, convolutional neural networks.

Abstract

One of the urgent and constantly in demand problems is updating maps. Maps, representing geo-information in vector form, have undoubted advantages in compactness and ”readability” compared to aerial photographs. The issue of maps actuality is critically important for rational urban planning, precision farming, the relevance of the cadastre and other geospatial applications. Various sources of data are used for maps updating, with aerial imagery being the main and rich source of information. Automatic processing of aerial photographs makes it possible to efficiently extract vector information, providing operational monitoring and accounting for changes that have appeared. The presented study addresses the problem of multi sensor information fusion in order to obtain accurate vector information. We use aerial images as a main data source and additionally the data of laser scanning and ground survey to increase performance of automatic image semantic segmentation and vectorization. The proposed framework is demonstrated on the task of forest monitoring.

1. Introduction

The actuality of maps is very important factor, that is necessary for the most of geospatial applications for correct decisions making. Maps representing geo-information in vector form have undoubted advantages in compactness and ”readability” compared to aerial photographs. The issue of maps actuality is critically important for sound urban planning, precision farming, the relevance of the cadastre and other geospatial applications. Various sources of data are used for maps updating, with aerial imagery being the main and rich source of information. Automatic processing of aerial photographs makes it possible to efficiently extract vector information, providing operational monitoring and accounting for changes that have appeared.

The scientific community pays great attention to the problem of image vectorization due to the high demand for a reliable and effective method of extracting vector information from remote sensing data. The first photogrammetric procedures for processing of stereo pairs of aerial photographs utilized complex photogrammetric devices such as stereo comparators and stereo plotters. Such processing required a large number of manual operations from the user. Advances in computers and digital image processing have given a powerful impetus to research in the field of automatic photogrammetric image processing methods, in particular for image segmentation and vectorization (Gruen and Li, 1995, Gruen and Li, 1997).

Nowadays, with significant advances in computers power and possibilities for acquiring and storing a huge amount of data, machine learning based methods demonstrate the state-of-the-arts results in this problem as in pixel-wise segmentation and vectorization. Data-driven methods exploit the power of hidden information that could be retrieved by analysis of huge amount of data. So, the representative and diverse task-oriented dataset is necessary part of machine learning study. A number of datasets exists (Bastani and Madden, 2021, Mohanty, 2018, Bayanlou and Khoshboresh-Masouleh, 2021), that are available for scientific community interested in semantic segmentation and image vectorization problems.

The presented study addresses the problem of multi sensor information fusion in order to obtain accurate vector information. We use aerial images as a main data source and additionally the data of laser scanning and ground survey to increase performance of automatic image vectorization. The proposed framework is demonstrated on the task of forest monitoring.

The main contributions of the study are the following:

- the framework for multi sensor data analysis for accurate vectorization of aerial images
- case study of the proposed technique in task of forest monitoring
- the estimate of improvement of vectorization quality provided by the proposed technique

2. Related work

The problem of image vectorization attracts attention of scientific groups in photogrammetry and computer vision for a number of decades. A notable progress in methods of semantic image segmentation and vectorization have been reached with appearing digital images and developing method of digital image processing. The study (Gruen and Li, 1995, Gruen and Li, 1997) proposed active contour models (Snakes) as a task of least squares minimization and then extended it for integrating multiple images for 3D linear features extraction. The concept of Least Squares B-spline Snakes improved performance of active contour models by implementing: internal quality control through computation of the covariance matrix of the estimated parameters; the exploitation of any a priori known geometric and photometric information; and the simultaneous use of any number of images through the integration of camera models. Compared to the two-image approach the proposed multi-image mode allows controlling blunders.

Nowadays, with significant advances in computers power and possibilities for acquiring and storing a huge amount of data,

machine learning based methods demonstrate the state-of-the-arts results in this problem as in pixel-wise segmentation and vectorization.

The study (Pu, 2021) provides a sight on various optical and light detection and ranging (LiDAR) sensors. It analyses current various techniques and methods for data classification and identify limitations and recommend future directions. They main conclusions of the study are the following: a large group of studies on the topic were using high-resolution satellite, airborne multi-/hyperspectral imagery, and airborne LiDAR data; a trend of “multiple” method development for the topic was observed; machine learning methods including deep learning models were demonstrated to be significant in improving classification accuracy; unmanned aerial vehicle- based sensors have caught the interest of researchers and practitioners for the topic-related research and applications.

The approach, named PolyMapper (Li et al., 2018), skips conventional pixel-wise segmentation of images and tries directly to forecast the object vector representation. PolyMapper directly retreive the topological map from overhead images as collections of building footprints and road networks. Evaluation of the proposed technique on both existing and self-collected large-scale datasets demonstrated that proposed learnable model can predict polygons of building footprints and road networks very close to the structure of existing online maps. The developed model work in a fully automated mode. Quantitative and qualitative evaluation demonstrated the state-of-the-art level of performance.

Several studies have been addressed to the problem of image vectiorization in computer graphics context, where image vectorization remains a major challenge (Ma et al., 2022). Layer-wise Image Vectorization (LIVE) technique, proposed in (Ma et al., 2022), simultaneously converts raster images to vector graphics and maintains image topology. LIVE generates compact vector forms with semantic layer-wise structures. By adding progressively new bezier paths and optimizing these paths with the layer-wise framework governed by newly designed loss functions, LIVE presents plausible vectorized forms that outperforms the methods-analogues.

The specially collected dataset MUNO21 (Bastani and Madden, 2021) is designed for solving the map updating task. The main goal of creating this dataset is to solve practical map updating problem, specifically, updating an existing map by adding, removing, and shifting roads, without introducing errors in parts of the existing map that remain up-to-date. The evaluation of several state-of-the-art road extraction methods on MUNO21 showed that further improvements in accuracy is necessary for automatic map updating.

The study (de Castro et al., 2021) analyses a set of techniques for UAVs in vegetation monitoring, which applied to diverse agricultural and forestry scenarios. Three general categories are highlighted: sensors used for surveys and applying vegetation indices for classification, technological goals pursued, and precision farming and precision forestry applications. UAV flight operations, spatial resolution requirements, and computation and data analytics are consudered, along with the ability of UAVs for characterizing relevant vegetation features. The authors analyse UAV-based technological solutions for a better use of agricultural and forestry resources and more efficient production with relevant economic and environmental benefits.

Semantic segmentation and image vectorization techniques are also in demand for woodland aerial imagery analysis. With increasing availability of unmanned aerial vehicles for acquiring various kind of remote sensing data, such methods very important for precision forestry. They are applied for palm tree inventories, continuous monitoring, vulnerability assessments, environmental control, and long-term management. The study (Gibril et al., 2023) addresses the reliability and the efficiency of various deep vision transformers in extracting date palm trees from multiscale and multisource VHSR images. The evaluation of various vision transformers, such as Segformer, Segmenter, UperNet-Swin transformer, the dense prediction transformer, with various levels of model complexity, has shown that transformers models demonstrate the state-of-the-art performance on woodland imagery. Also the current state of multimodal data processing (Knyaz, 2019) allows to hypothesize that multi sensor data fusion can improve the performance in tasks of segmentation and vectiorization

The comprehensive review (Li et al., 2019) summarises recent remote sensing applications in urban forestry from the perspective of three distinctive themes: multi-source, multi-temporal and multi-scale inputs. It reviews how different sources of remotely sensed data offer a fast, replicable and scalable way to quantify urban forest dynamics at varying spatiotemporal scales on a case-by-case basis. Combined optical imagery and LiDAR data results as the most promising among multi-source inputs; in addition, future efforts should focus on enhancing data processing efficiency. For long-term multi-temporal inputs, in the event satellite imagery is the only available data source, future work should improve haze-/cloud-removal techniques for enhancing image quality.

3. Materials and Methods

3.1 Sensors used for the study

We use materials of forest area survey for developing and testing the proposed framework. An unmanned aerial vehicle (UAV) equipped with a set of multimodal sensors was used for the surveying. Three sensors acquired information during surveying flights. The equipment of the UAV includes:

- color (RGB) camera SONY DSC-RX1R;
- multi spectral camera MicaSense RedEdge-M;
- laser scanner AGM-MS3.

Images of the sensors are presented in Figure 1.

(a) laser scanner AGM-MS3 (b) multi spectral camera MicaSense RedEdge-M (c) color camera SONY DSC-RX1R

Figure 1. Sensors used for the surveying

Color infra-red (CIR) images are acquired by MicaSense RedEdge-M multi spectral camera. It acquires images with

photometric resolution (image depth) of 16 bits of five spectral bands: Blue, Green, Red, RedEdge (RE) and Near Infra Red (NIR). Main technical characteristics of the MicaSense RedEdge-M camera are presented in Table 1.

Parameter		Value				
Band		Blue	Green	Red	RE	NIR
Wave length	nm	475	560	668	717	840
FWHM	nm	20	20	10	10	40
Image depth	bit	16	16	16	16	16
Image width	pix	1280	1280	1280	1280	1280
Image height	pix	960	960	960	960	960

Table 1. Technical characteristics of MicaSense RedEdge-M multi spectral camera.

Color (RGB) images are acquired by SONY DSC-RX1R camera. It acquires high resolution color images with photometric resolution of 8 bits. These high resolution images are used for generating dense digital elevation model and high quality orthophoto. Main technical characteristics of the SONY DSC-RX1R camera are presented in Table 2.

Parameter		Value
Wave length	nm	535…670
Bits Per Sample		8
Image width	pix	6000
Image height	pix	4000

Table 2. Technical characteristics of SONY DSC-RX1R camera.

Laser scanner AGM-MS3 provides range data of high accuracy and high spatial resolution with high acquisition rate up to 604 kHz. Main technical characteristics of the AGM-MS3 laser scanner are presented in Table 3.

Parameter		Value
Data acquisition rate	kHz	up to 640
Range	m	up to 300
FOV	grad	up to 360
Deflection unit rotation speed	rps	up to 20
Spatial accuracy	mm	30…50
Range accuracy	Mm	30
Dimensions	mm	124x124x113
Weight	kg	1.1
Power consumption		12 V, 1.3 A
Operating temperature	^oC	-20 ... +55

Table 3. Technical characteristics of AGM laser scaner.

3.2 Multi-Sensor Data

Multi sensor data acquired during the survey includes six sets of images: RGB, images in Blue, Green, Red, RedEdge and Near Infra Red bands and range data.

Samples of the acquired data for the same surveyed area are shown in Figure 2.

For tree detection, classification and vectorization current standard processing procedure includes several phases, that are presented as Algorithm 1.

Algorithm 1: Remote sensing data processing

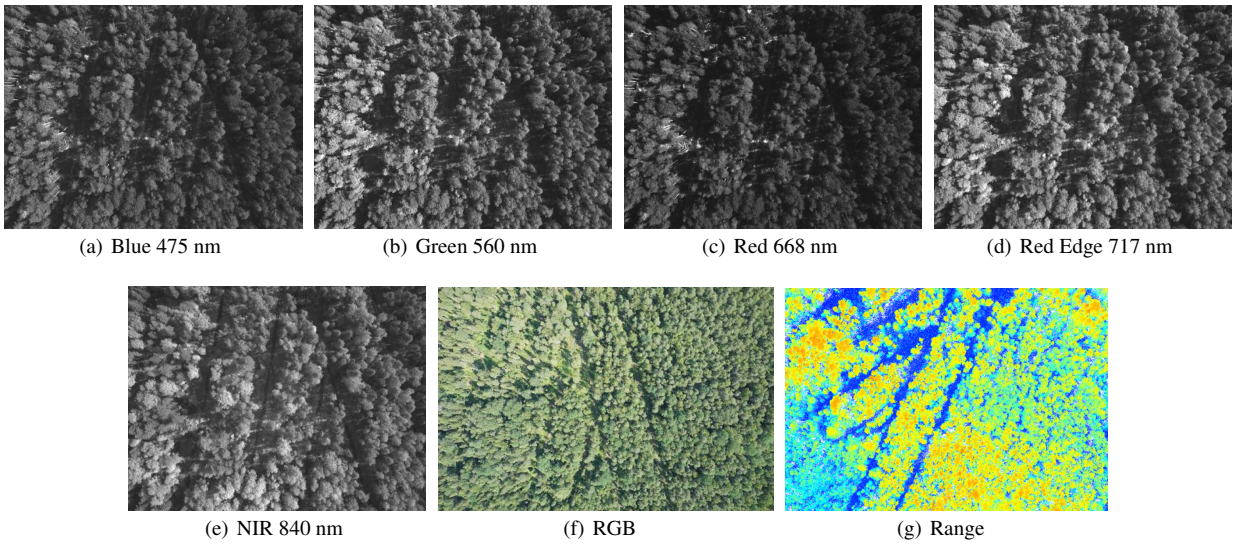
Input:
Set of color (RGB) images $\{I_c^j\}, j = 1, \ldots, N_c$ of the area
Set of multi band (CIR) images $\{I_{mb}^j\}, j = 1, \ldots, N_{mb}$ of the area
Range data R of the same area
Output:
Digital elevation model H_e
Color orthophoto O_c
Multi band CIR rthophoto O_{mb}

Processing procedure ;
Procedure `Pre-processing():`
  `/* Color image pre-processing */`
  for $I_c^j \in I_c$ **do**
    Correct distortion(I_c^j))
  `/* Multi-band image pre-processing */`
  for $I_{mb}^j \in I_{mb}$ **do**
    CorrectDistortion(I_{mb}^j))
    CalibrateRadiometry(I_{mb}^j))
    NormalizePixelValue(I_{mb}^j))
  return;
Photogrammetric image processing ;
Procedure `Photogrammetric processing(t,s):`
  `/* Photogrammetric image processing */`
  for $I_c^j \in I_c$ **do**
    FindDescriptors(I_c^j))
    MatchDescriptors(I_c^j))
    OrientImage(I_c^j))
  GenerateSparsePointCloud($\{I_c\}$))
  GenerateDensePointCloud($\{I_c\}$))
  H_e = GenerateDEM($\{I_c\}$))
  O_c = GenerateRGBOrthophoto($\{I_c\}$))
  O_{mb} = GenerateCIROrthophoto($\{I_{mb}\}$))
  return H_e, O_c, O_{mb};

Multi band images were processed by standard procedure, that includes the following phases:

3.2.1 Radiometric calibration. Radiometric calibration is performed for accounting gain and exposure of the camera, directional parameters (the position of the sensor and the position of the sun), irradiance parameters (exploiting special tools such as light sensors or reflectance panels). Using this data, raw digital array (raw image) are converted into sensor reflectance (or irradiance), and then to reflectance values of imaged surface.

Radiometric calibration allow to calculate absolute spectral radiance values basing on the raw pixel values of an image. Absolute spectral radiance is measured in $W/(m^2 \cdot sr \cdot nm)$. Such factors as sensor gain and exposure settings, sensor sensitivity and black-level, and lens vignette effects are compensated by radiometric calibration, thus providing accurate data for radiometric analysis of images in order to obtain adequate information about surface reflectance needed for the task of tree classification.



(a) Blue 475 nm (b) Green 560 nm (c) Red 668 nm (d) Red Edge 717 nm

(e) NIR 840 nm (f) RGB (g) Range

Figure 2. Samples of data used in the study. From top to bottom: RGB image, one channel of CIR image, laser scanning data

The spectral radiance L (in $\frac{W}{m^2 \cdot sr \cdot nm}$) for pixel with value p is:

$$L = V(x, y) \cdot \frac{a_1}{g} \cdot \frac{p - p_{BL}}{t_e + a_2 y - a_3 t_e y} \quad (1)$$

with
p – the normalized raw pixel value,
$p_B L$ – the normalized black level value
a_1, a_2, a_3 – the radiometric calibration coefficients
$V(x, y)$ – the vignette polynomial function for pixel (x, y).
t_e – the image exposure time
g – the sensor gain setting
x, y – the pixel column and row number, respectively

3.2.2 Pixel Value Normalization. Multi band images from MicaSense camera have 16-bit depth. As the radiometric model operates with normalized pixel values in the range 0 to 1, so raw digital pixel values are converted to normalized form by dividing the the value the pixel by 2^N ($N = 16$ – the image depth). Such kind of transformation is also applied the black level values.

For transformation of raw pixel values into reflectance, images of special calibrated reflectance panel taken before or after flight a used. For each calibrated reflectance panel the transfer function is given The transfer function gives the relation between the raw pixels of the panel image to units of absolute reflectance (a value between 0.0 and 1.0) in the range of 400 nm to 850 nm with the increment of 1 nm.

The average value of radiance for the pixels located inside the actual panel area of the image $avg(L_i)$ is used for determining the reflectance calibration factor for each band. For the $i - th$ band the transfer function of radiance to reflectance is:

$$F_i = \frac{\rho_i}{avg(L_i)} \quad (2)$$

with

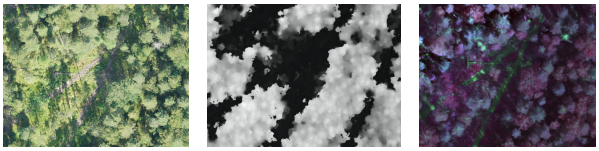
F_i – the reflectance calibration factor for band i,
ρ_i – the average reflectance of the calibrated reflectance panel for the $i - th$ band.

This factor is used for the $i - th$ band to convert all radiance values to reflectance. This same procedure is applied independently to each band for converting the images reflectance units.

3.3 Framework for Segmentation and Vectorization.

After pre-processing of the raw data acquired during the survey, images are processed by standard routine to generate photogrammetric products: digital elevation model (DEM) and orthophoto for testing area. These routine includes images orienting, feature detecting and solving the correspondence problem, sparse and dense point clouds generating, DEM recontructing, orthophoto producing. The resulting orthophoto is then used for semantic segmentation and vectorization.

The resulting products of photogrammetric processing (digital elevation model, RGB orthophoto, and CIR orthophoto) are shown in Figure 3.



(a) RGB orthophoto (b) digital elevation model (c) CIR orthophoto

Figure 3. Corresponding regions at RGB orthophoto, digital elevation model, CIR orthophoto

Due to the complexity of the problem of matching images of woodlands, which leads to a mismatch of the corresponding points, the output digital elevation model contains errors. These errors lead to significant distortion of the orthorectified image, and can dramatically affects the results of segmentation and decryption.

(a) height scale (b) Range data (c) digital elevation model (d) Composite image of (b) and (c)

Figure 4. Corresponding regions at range data, digital elevation model, and mixed image of half-transparent DEM and range image. (a) is the height scale, from 140 m (red) and 180 m (blue).

Figure 5 demonstrates a typical image distortion caused by errors in digital elevation model. Due to these errors it is not possible detect and separate several trees in the area A (shown by red rectangle).

To overcome the problem of orthophoto distortion caused by error in digital elevation model we use laser scanner data (Figure 2(g)) as additional information source. We use an image representation of the range data (Figure 4(b)), which allows us to perform cluster data analysis and detect individual trees in the forest area.

For data clusterization into individual trees in range image we use algorithm similar to the work (Kniaz et al., 2021), based on analysis of a set of images obtained by image binarization by sequentially decreasing threshold. The pixels of a range image are then aggregated in one cluster (individual tree) using gravitation model. An example of data clusterization is shown in Figure 5.

We process a subset of the images by the described above technique, and then use this subset as training data for `WireNet` neural network model (Kniaz et al., 2020). After training stage, we use `WireNet` for segmentation of the rest of the images. The results of segmentation demonstrated more than 18% improvement of average precision AP metric, used for evaluation, comparing with the standard segmentation procedure.

$$AP(S_{\text{gt}}, S_{\text{p}}) = \frac{1}{N_T \cdot N_O} \sum_{t=0}^{T_{\max}} \sum_{i=0}^{N_O} \mathbb{1}_{J(S_{\text{gt}}, S_{\text{p}}) > t}, \quad (3)$$

with

N_T is the power of the set of threshold values used;
N_O is the power of the set of detected objects;
$T_{\max}$ is the maximum of thresholds;
$\mathbb{1}_v$ is the indicator function;
S_{gt} is the ground truth segmentation;
S_{p} is the predicted segmentation;
$J(S_{\text{gt}}, S_{\text{p}})$ is the Intersection-over-Union value (or Jaccard index).

$$J(S_{\text{gt}}, S_{\text{p}}) = \frac{S_{\text{gt}} \cap S_{\text{p}}}{S_{\text{gt}} \cup S_{\text{p}}}. \quad (4)$$

At the next phase we use segmented images as mask ones for the classification of the detected trees. For this purpose, we overlay the mask images on the CIR orthophoto to select individual trees for classification. Then each tree inside the mask is classified by standard procedure using vegetation indexes.

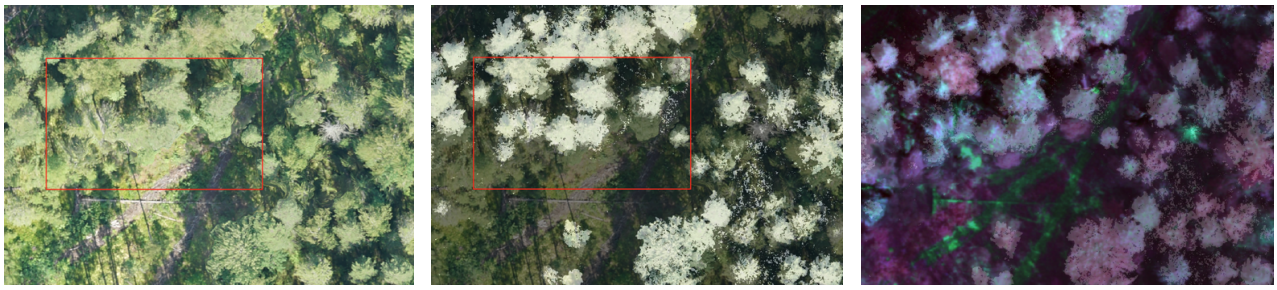
The whole framework can be presented as Algorithm 2.

Algorithm 2: Orthophoto segmentaion

Input:
Orthophoto O of the area
Range data **R** in image form I_r
Output:
Weights of the CNN model $\{W_i\}$

```
Data clusterization ;
Procedure Clusterization(R):
    R_t = CreateTrainingSubset(R)
    for r_j ∈ R_t do
        {C_t} = FindClusters(r_j))
        Aggregate({C_t})
        I_m^j = GenerateMaskImage({C_t})
    return {I_m};
Training ;
Procedure Training():
    for I_r^j ∈ I_r do
        I_s^j = FindAreas(I_m^j))
        Correct(I_s^j))
    TrainCNNModel({I_s}))
    return {W_i};
Inference ;
Procedure Segmentation(O):
    O_s = SegmentingOrthoPhoto(O, W_i)
    return O_s;
```

We estimated the performance of the proposed technique on the data, obtained during the forest area survey using various sensors. As a baseline we used the results of processing by standard processing routine. We applied average precision AP metric for estimating the performance of the proposed framework (Table 4), and evaluation results demonstrated up-to 18% progress in segmentation and vectorization accuracy comparing with the baseline.



(a) Area A (red rectangle) with indistinguishable image of separate trees (b) Semantic segmentation of the same area (c) CIR-based classification

Figure 5. Image distortion caused by errors in digital elevation model and semantic segmentation of this area

	$P_{[0.5]}$	$P_{[0.7]}$	$P_{[0.9]}$	AP
Standard procedure	0.663	0.576	0.381	0.540
Proposed Framework	0.817	0.743	0.625	0.728

Table 4. Performance evaluation

3.4 Conclusion

The framework for accurate semantic segmentation and vectorization is developed. It utilizes three type of data acquired during UAV-based aerial survey, specifically: RGB imagery, CIR imagery and range data acquired by laser scanner. To improve the performance of semantic segmentation range data is used for annotating training dataset in semi-automated mode.

The image segmentation network model, trained on created dataset has demonstrated up-to 18% progress in segmentation and vectorization accuracy comparing with the standard segmentation procedure.

Aknowlegements

The research was carried out at the expense of a grant from the Russian Science Foundation No. 24-21-00269, https://rscf.ru/project/24-21-00269/

References

Bastani, F., Madden, S., 2021. Beyond Road Extraction: A Dataset for Map Update using Aerial Images. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 11885-11894. https://api.semanticscholar.org/CorpusID:238583806.

Bayanlou, M. R., Khoshboresh-Masouleh, M., 2021. Multi-task learning from fixed-wing UAV images for 2D/3D city modeling. *CoRR*, abs/2109.00918. https://arxiv.org/abs/2109.00918.

de Castro, A. I., Shi, Y., Maja, J. M., Peña, J. M., 2021. UAVs for Vegetation Monitoring: Overview and Recent Scientific Contributions. *Remote Sensing*, 13(11). https://www.mdpi.com/2072-4292/13/11/2139.

Gibril, M. B. A., Shafri, H. Z. M., Al-Ruzouq, R., Shanableh, A., Nahas, F., Al Mansoori, S., 2023. Large-Scale Date Palm Tree Segmentation from Multiscale UAV-Based and Aerial Images Using Deep Vision Transformers. *Drones*, 7(2). https://www.mdpi.com/2504-446X/7/2/93.

Gruen, A., Li, H., 1995. Linear Feature Extraction with 3-D LSB-Snakes. *ISPRS Journal of Photogrammetry and Remote Sensing*, 50(4), Road extraction from aerial and satellite images by dynamic programming.

Gruen, A., Li, H., 1997. Linear feature extraction with 3-d lsb-snakes. A. Gruen, E. P. Baltsavias, O. Henricsson (eds), *Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)*, Birkhäuser Basel, Basel, 287–298.

Kniaz, V. V., Grodzitskiy, L., Knyaz, V. A., 2021. DEEP LEARNING FOR CODED TARGET DETECTION. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIV-2/W1-2021, 125–130. https://isprs-archives.copernicus.org/articles/XLIV-2-W1-2021/125/2021/.

Kniaz, V. V., Zheltov, S. Y., Remondino, F., Knyaz, V. A., Bordodymov, A., Gruen, A., 2020. WIRE STRUCTURE IMAGE-BASED 3D RECONSTRUCTION AIDED BY DEEP LEARNING. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B2-2020, 435–441. https://isprs-archives.copernicus.org/articles/XLIII-B2-2020/435/2020/.

Knyaz, V., 2019. Multimodal data fusion for object recognition. *Proc. SPIE 11059, Multimodal Sensing: Technologies and Applications.*, 110590, 110590P. https://doi.org/10.1117/12.2526067.

Li, X., Chen, W. Y., Sanesi, G., Lafortezza, R., 2019. Remote Sensing in Urban Forestry: Recent Applications and Future Directions. *Remote Sensing*, 11(10). https://www.mdpi.com/2072-4292/11/10/1144.

Li, Z., Wegner, J. D., Lucchi, A., 2018. PolyMapper: Extracting City Maps using Polygons. *CoRR*, abs/1812.01497. http://arxiv.org/abs/1812.01497.

Ma, X., Zhou, Y., Xu, X., Sun, B., Filev, V., Orlov, N., Fu, Y., Shi, H., 2022. Towards layer-wise image vectorization.

Mohanty, S. P., 2018. CrowdAI dataset. https://www.crowdai.org/challenges/ mapping-challenge/dataset-files.

Pu, R., 2021. Mapping Tree Species Using Advanced Remote Sensing Technologies: A State-of-the-Art Review and Perspective. *Journal of Remote Sensing*, 2021. https://spj.science.org/doi/abs/10.34133/2021/9812624.