Unlocking Troves of Data

Feature Extraction From Historical Print Maps Through an Object-Based Classification Approach

Capstone Peer Review Presentation

Delphine Khanna - GEOG 596A - May 6, 2019

Outline

- Background
- Goals
- Project Methodology
- Proposed Timeline
- Anticipated Results / Outcome
- Further developments

Background

1. Historical Geospatial Data

Historical geospatial data?

- GIS analysts are used to accessing geospatial datasets from multiple sources
- However, most of that data is relatively recent, from the last 20 to 30 years
- What happens when someone needs access to older historical data?

Why historical data?

There might be different reasons for needing to use geospatial data from previous historical periods:

- Researcher studying a past phenomenon, or interested in the evolution of specific geographic features over time (Madry, 2006)
- Increasingly, as input for computational models and machine learning algorithms (Bidney & Piekielek, 2018)

Historical print maps

- Historical print maps are a great source of data about the past:
  - Related to topography, land cover, urban development, human demographics, environmental dynamics, and more
  - They represent facts from the 18th to 20th century, or even earlier
- However, that data is locked on paper and cannot be leveraged easily in GIS systems

Background

2. Feature Extraction

Feature extraction from historical print maps

- Traditionally: manual feature extraction
  - Digitize the features of interest using digitization tablets or heads-up digitization techniques (Bolstad, 2016)
  - It works, but very expensive and slow

- Automated or semi-automated feature extraction
  - Research in that field for the last 15 years
  - Different approaches have been proposed

Different methods of automation

- First: scan, georeference, crop, pre-process

- Pixel-level classification
- Template-based
- Machine-learning
- Object-Based Image Analysis (OBIA)

Different methods of automation

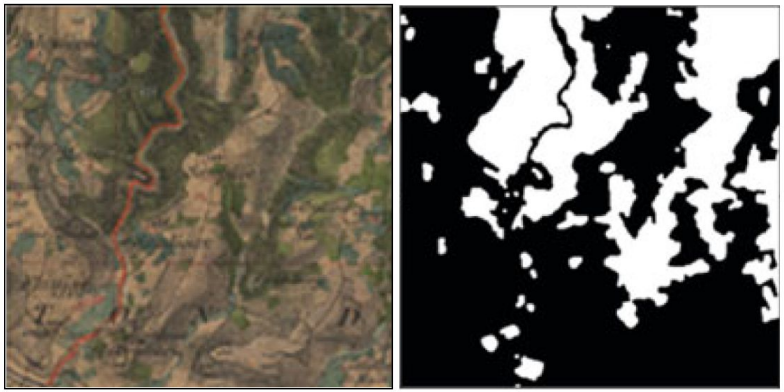
- **First: scan, georeference, crop, pre-process**

- Pixel-level classification
- Template-based
- Machine-learning
- Object-Based Image Analysis (OBIA)

Different methods of automation

- First: scan, georeference, crop, pre-process

- **Pixel-level classification**
- Template-based
- Machine-learning
- Object-Based Image Analysis (OBIA)

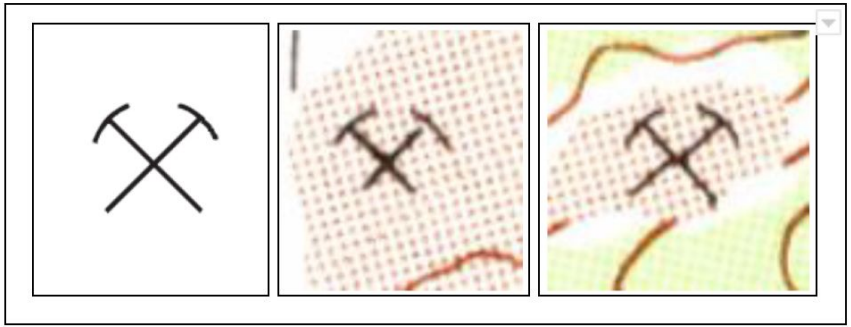


Extracting forested areas from a 19th century map (Herrault et al, 2013).

Good for hue-based areal symbologies.

Different methods of automation

- First: scan, georeference, crop, pre-process
- Pixel-level classification
- **Template-based**
- Machine-learning
- Object-Based Image Analysis (OBIA)



Quarry symbol template & occurrences on a USGS map (Chen, 2015).

Good for individual symbols.

Different methods of automation

- First: scan, georeference, crop, pre-process
- Pixel-level classification
- Template-based
- **Machine-learning**
- Object-Based Image Analysis (OBIA)

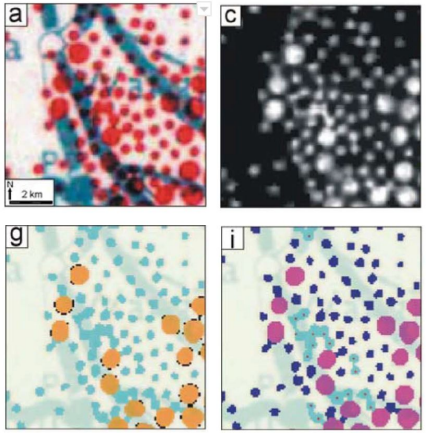
Locating individual buildings on a 1966 USGS map with a Convolutional Neural Network (Uhl et al., 2017).

Good for when very large training samples can be secured.

Different methods of automation

- First: scan, georeference, crop, pre-process

- Pixel-level classification
- Template-based
- Machine-learning
- **Object-Based Image Analysis (OBIA)**



Extracting dots symbolizing population numbers from a 1962 Kenya map (Kerle & de Leeuw, 2009).

Relatively versatile.
Does not require training samples.
Selected for this project.

Background

3. Soviet Military Maps

Large map series

- Automating feature extraction is particularly interesting with large map series
  - Putting in place the workflow is time consuming
  - So, ideally you want to use the workflow many times

- Soviet military maps
  - One fascinating example of large map series
  - My project will focus on some of those maps

1936 Soviet map representing Eastern Azerbaijan at 1:50,000.

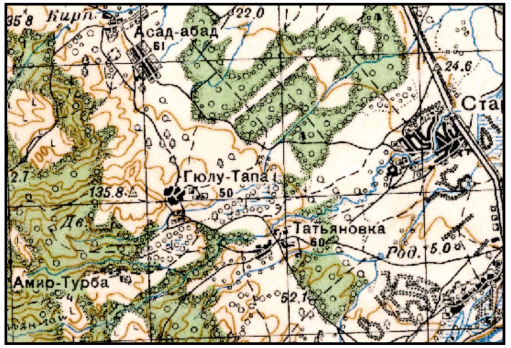
Soviet military maps

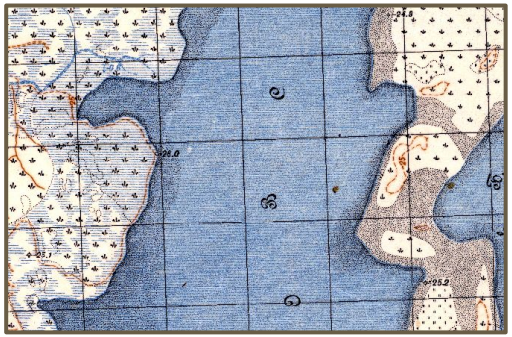
- Military maps were produced in the Soviet Union for most of the 20th century
  - Most were highly classified, and unknown to the public
  - The Soviet army first focused on mapping Russia and countries surrounding it
  - Later on, after World War II went on to map most of the terrestrial earth surface
  - Several series exist at different scales from 1:1,000,000 to 1:10,000
  - Large series using the same symbology consistently
  - (Davies & Kent, 2017)

Details of Washington, D.C. (1975)
At 1:25,000.

Soviet military maps

- After the collapse of the Soviet Union in the early 1990's:
  - More and more of those maps started appearing in the West
  - Thousands of map sheets have already surfaced, and it is suspected that there exist many more that are still unknown to us (Watt, 2005)
  - That cartographic body represents a wealth of valuable historical information
  - Known for their remarkable level of details





Details of Azerbaijan (1936) at 1:50,000.

Goals

Goals

- Build a workflow to extract features semi-automatically
- From Soviet military maps dating from the 1930’s for a region of the Belarus Republic
- Using the Object-Based Image Analysis (OBIA) approach

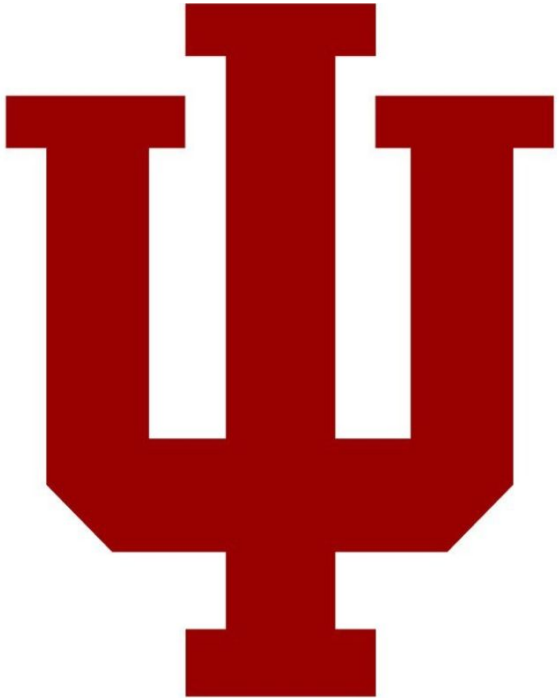
- Strive to create a workflow as streamlined as possible
- Assess the potential for real-life use
- Tackle some complex symbology types
- Help assess the usefulness of OBIA for this type of analysis

Project Methodology

1. Securing the Data

Indiana University's Collection

- Indiana University Bloomington Library
- Started digitizing their collection of Soviet military maps thanks to a grant (Crowe, 2018)
- High resolution GeoTIFFs
  - Not yet public
  - But they gave me access to them
- No complete series
- All files are georeferenced and cropped

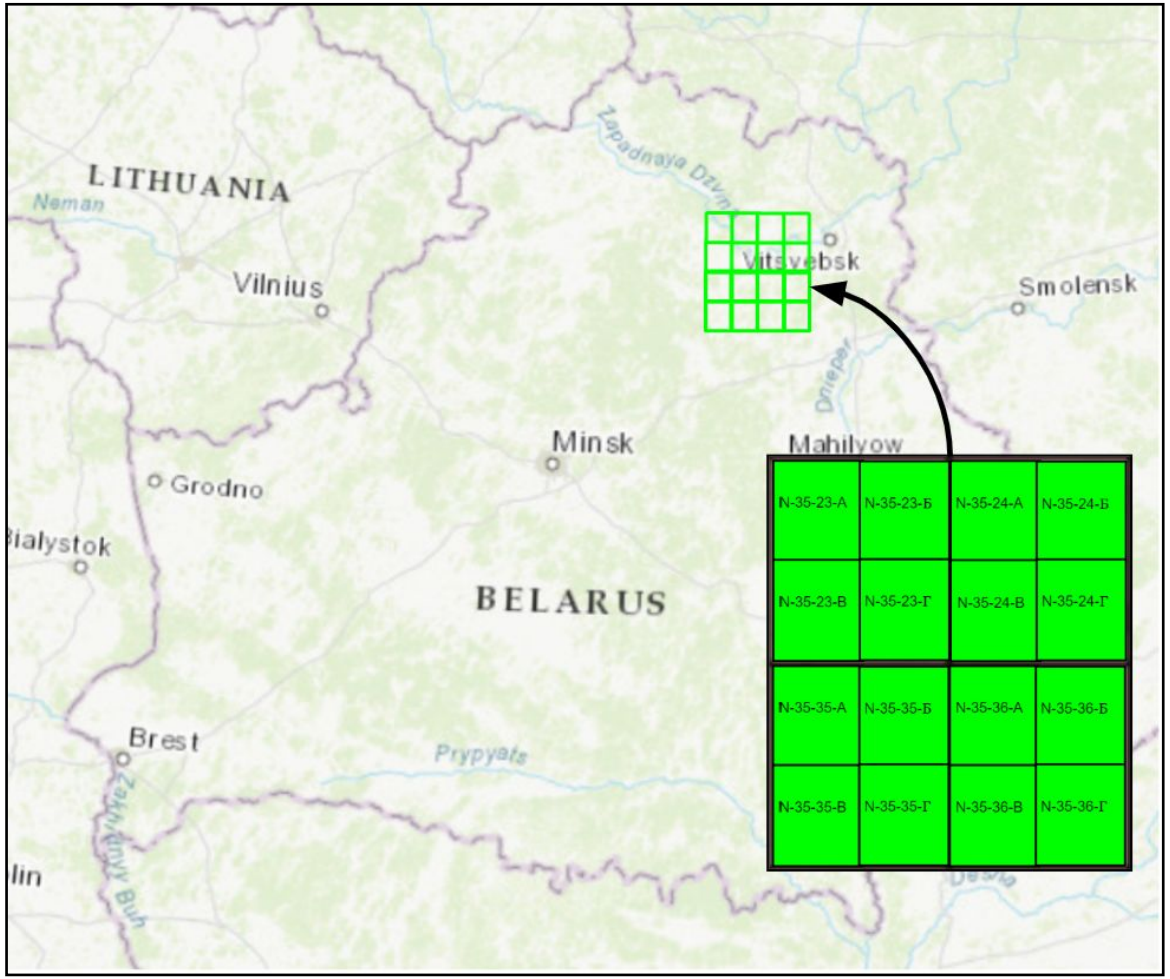


IU Logo

Area of Interest

- **Type:** 16 topographic 1:50,000 map sheets
- **Location:** region East of Viciebsk, Belarus Republic.
- **Years:** 1936 & 1937
- **Coordinate system:** Pulkovo 1942 3 Degree GK Zone 16
- **Datum:** Pulkovo 1942
- **Unit:** meter
- **IDs:**

N-35-23-А, N-35-23-Б, N-35-23-В, N-35-23-Г, N-35-24-А, N-35-24-Б, N-35-24-В, N-35-24-Г, N-35-35-А, N-35-35-Б, N-35-35-В, N-35-35-Г, N-35-36-А, N-35-36-Б, N-35-36-В, N-35-36-Г.



Area of Interest

Each raster:

- 18 by 15 km
- 1 file ≅ 770 Mb

Total of the 16 rasters:

- 72 by 60 km
- ≅ 12.3 Gb

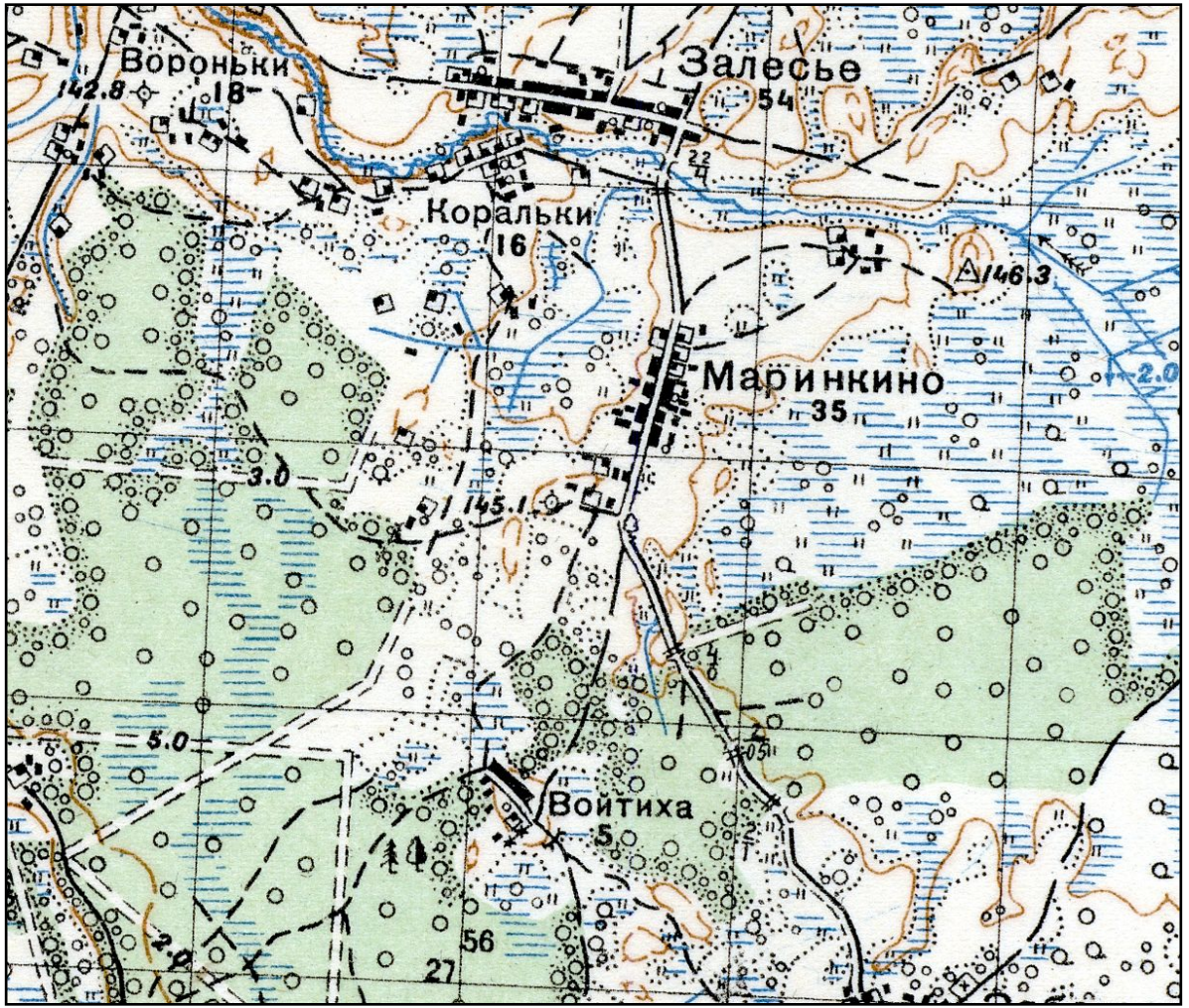
3 color bands:

- Red, Green, Blue

The 8 rasters for the bottom half of the AOI.

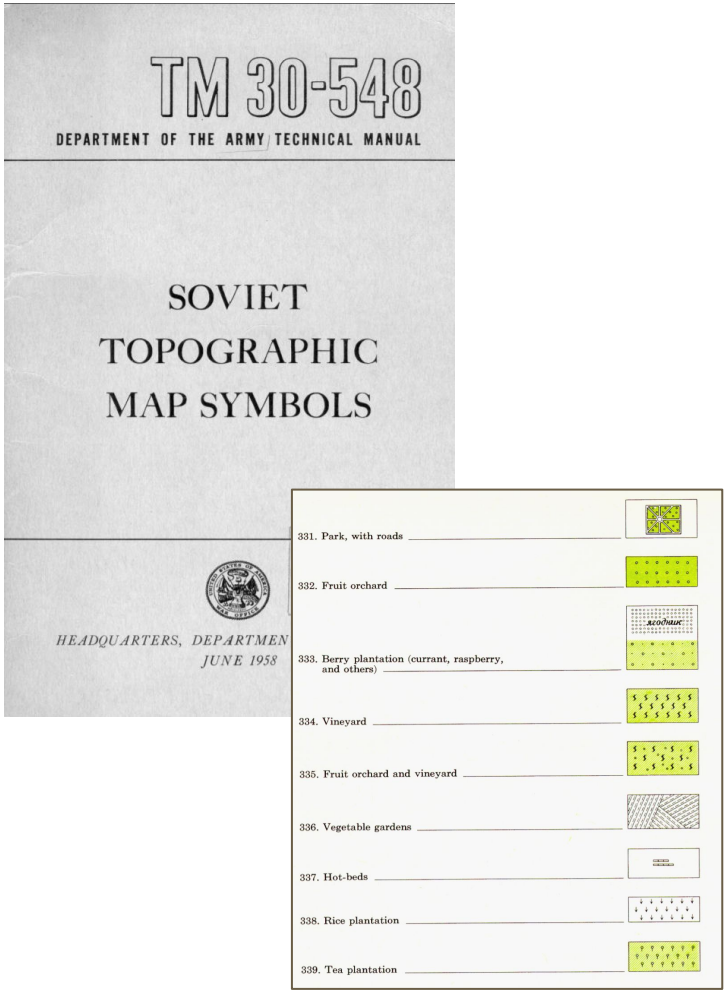
Area of Interest

Detail of the AOI: surroundings of Zales'ye, Belarus



Symbolization guide

- No legends on the map sheets
- The map users relied on a separate guide book
- Symbolization guide published by the U.S. Army in 1958 (Dept. of the Army Headquarters, 1958)
  - Found on the UC Berkeley Library's website
  - Seems to describe my maps perfectly



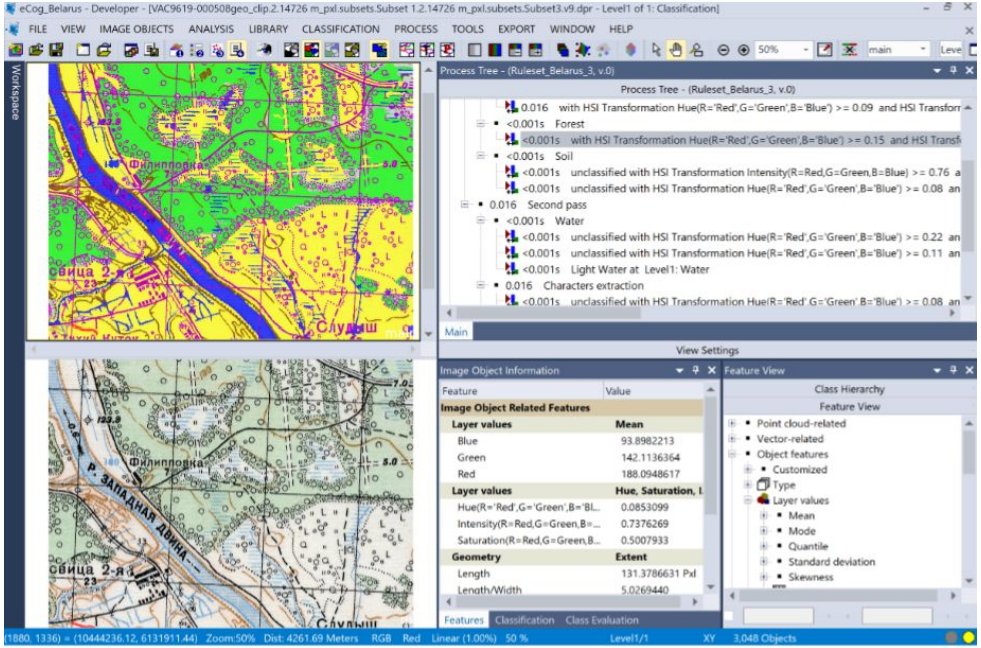
Project Methodology

2. Workflow

OBIA approach

- Overall approach selected: Object-Based Image Analysis (OBIA)

- Main software applications:
  - eCognition
  - ArcGIS Pro



eCognition interface

Workflow

AOI raster dataset (lower half)

1. **Acquire raster files**
2. Prepare AOI raster
   - Color calibration (?)
   - Create mosaic raster
   - Downsampling (?)
3. Define subsets
4. Develop rule set on subsets
   - Preprocessing steps
   - Segmentation
   - Classification
5. Run ruleset on entire AOI Raster
6. Generate shapefile output
7. Accuracy assessment
8. Production potential assessment
9. Creation of illustrative maps

Legend

- ArcGIS Pro
- eCognition
- Photoshop
- Internet / File system

Workflow

AOI raster dataset (lower half)

1. Acquire raster files

2. Prepare AOI raster

- Color calibration (?)
- Create mosaic raster
- Downsampling (?)

3. Define subsets

4. Develop rule set on subsets

- Preprocessing steps
- Segmentation
- Classification

5. Run ruleset on entire AOI Raster

6. Generate shapefile output

7. Accuracy assessment

8. Production potential assessment

9. Creation of illustrative maps

Legend

- ArcGIS Pro
- eCognition
- Photoshop
- Internet / File system

Workflow

AOI raster dataset (lower half)

1. Acquire raster files
2. Prepare AOI raster
   - **Color calibration (?)**
   - Create mosaic raster
   - Downsampling (?)
3. Define subsets
4. Develop rule set on subsets
   - Preprocessing steps
   - Segmentation
   - Classification
5. Run ruleset on entire AOI Raster
6. Generate shapefile output
7. Accuracy assessment
8. Production potential assessment
9. Creation of illustrative maps

Legend

- ArcGIS Pro
- eCognition
- Photoshop
- Internet / File system

Workflow

AOI raster dataset (lower half)

1. Acquire raster files
2. Prepare AOI raster
   - Color calibration (?)
   - **Create mosaic raster**
   - Downsampling (?)
3. Define subsets
4. Develop rule set on subsets
   - Preprocessing steps
   - Segmentation
   - Classification
5. Run ruleset on entire AOI Raster
6. Generate shapefile output
7. Accuracy assessment
8. Production potential assessment
9. Creation of illustrative maps

Legend

- ArcGIS Pro
- eCognition
- Photoshop
- Internet / File system

Workflow

AOI raster dataset (lower half)

1. Acquire raster files
2. Prepare AOI raster
   - Color calibration (?)
   - Create mosaic raster
   - **Downsampling (?)**
3. Define subsets
4. Develop rule set on subsets
   - Preprocessing steps
   - Segmentation
   - Classification
5. Run ruleset on entire AOI Raster
6. Generate shapefile output
7. Accuracy assessment
8. Production potential assessment
9. Creation of illustrative maps

Legend

- ArcGIS Pro
- eCognition
- Photoshop
- Internet / File system

Workflow

AOI raster dataset (lower half)

1. Acquire raster files

2. Prepare AOI raster

- Color calibration (?)
- Create mosaic raster
- Downsampling (?)

3. Define subsets

4. Develop rule set on subsets

- Preprocessing steps
- Segmentation
- Classification

5. Run ruleset on entire AOI Raster

6. Generate shapefile output

7. Accuracy assessment

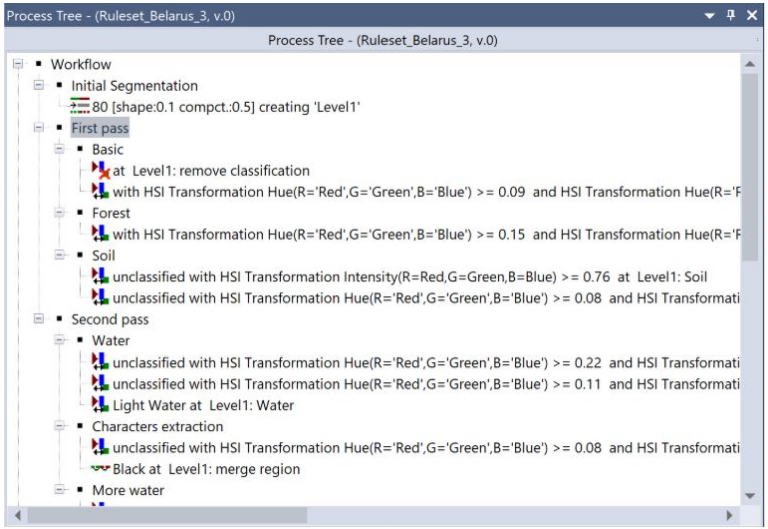
8. Production potential assessment

9. Creation of illustrative maps

Legend

- ArcGIS Pro
- eCognition
- Photoshop
- Internet / File system

Workflow



Rule set example

1. Acquire raster files

2. Prepare AOI raster

- Color calibration (?)
- Create mosaic raster
- Downsampling (?)

3. Define subsets

4. Develop rule set on subsets

- Preprocessing steps
- Segmentation
- Classification

5. Run ruleset on entire AOI Raster

6. Generate shapefile output

7. Accuracy assessment

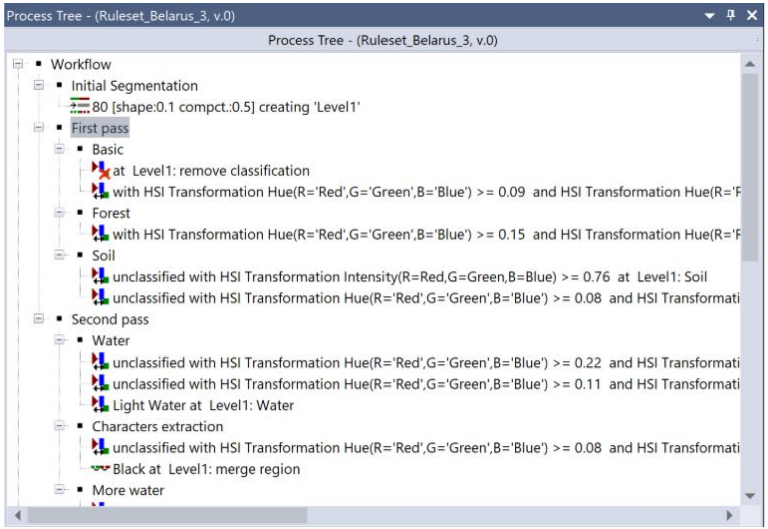
8. Production potential assessment

9. Creation of illustrative maps

Legend

- ArcGIS Pro
- eCognition
- Photoshop
- Internet / File system

Workflow



Rule set example

1. Acquire raster files

2. Prepare AOI raster

- Color calibration (?)
- Create mosaic raster
- Downsampling (?)

3. Define subsets

4. Develop rule set on subsets

Preprocessing steps

Segmentation

Classification

5. Run ruleset on entire AOI Raster

6. Generate shapefile output

7. Accuracy assessment

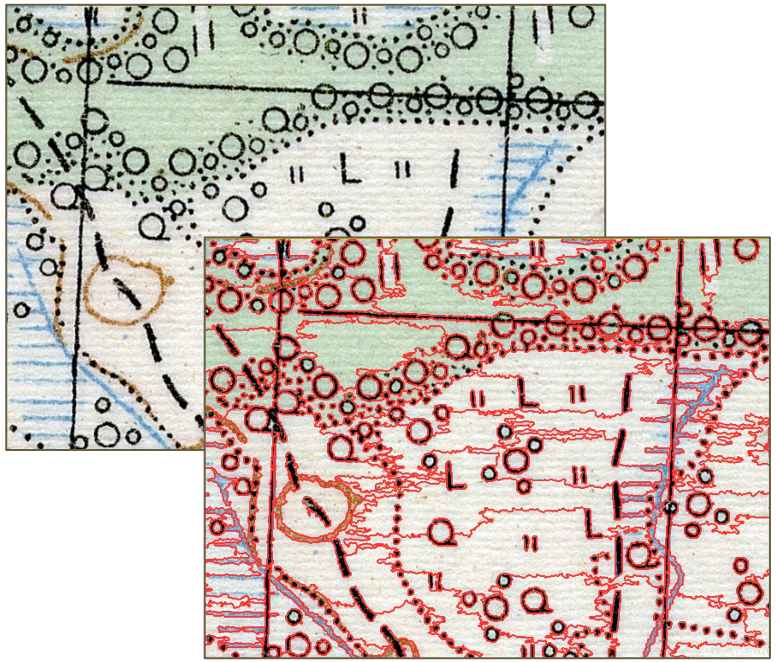
8. Production potential assessment

9. Creation of illustrative maps

Legend

- ArcGIS Pro
- eCognition
- Photoshop
- Internet / File system

Workflow



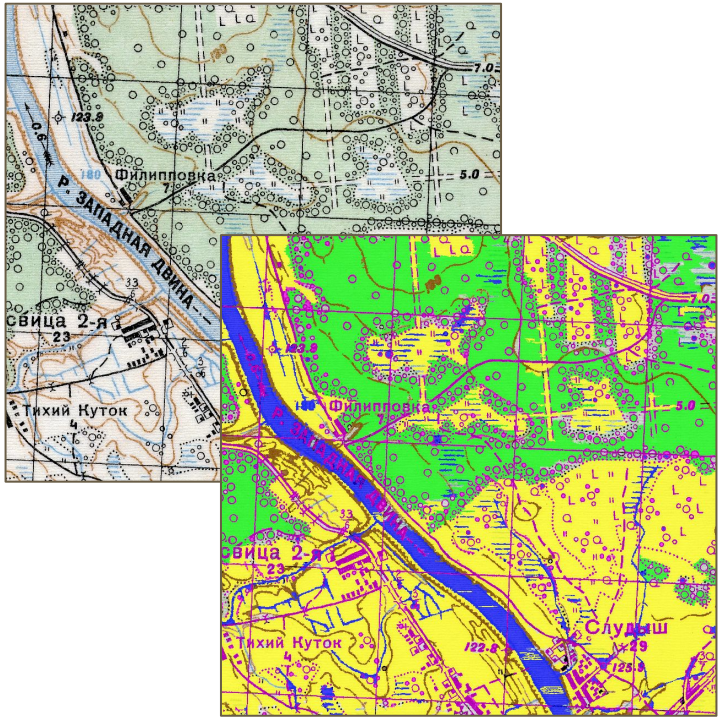
Segmentation example

1. Acquire raster files
2. Prepare AOI raster
   - Color calibration (?)
   - Create mosaic raster
   - Downsampling (?)
3. Define subsets
4. Develop rule set on subsets
   - Preprocessing steps
   - **Segmentation**
   - Classification
5. Run ruleset on entire AOI Raster
6. Generate shapefile output
7. Accuracy assessment
8. Production potential assessment
9. Creation of illustrative maps

Legend

- ArcGIS Pro
- eCognition
- Photoshop
- Internet / File system

Workflow



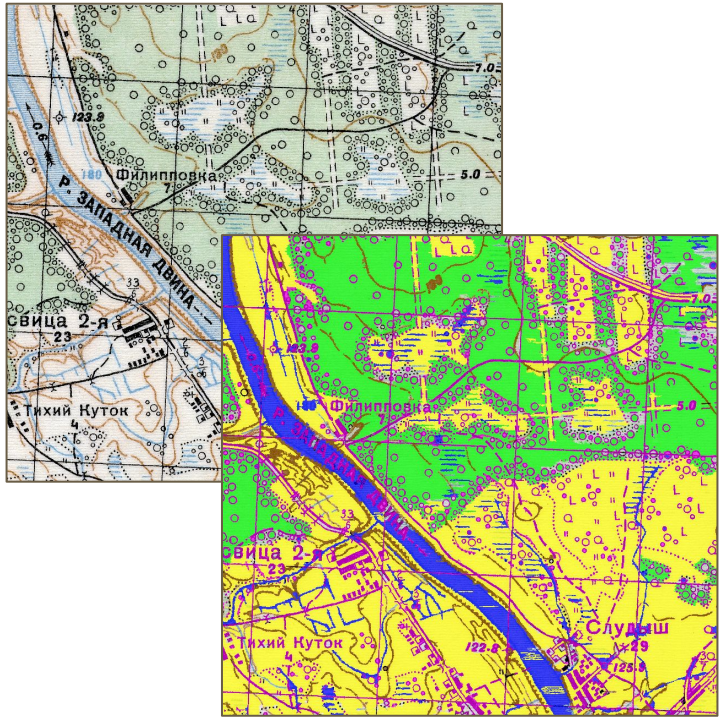
Classification example

1. Acquire raster files
2. Prepare AOI raster
   - Color calibration (?)
   - Create mosaic raster
   - Downsampling (?)
3. Define subsets
4. Develop rule set on subsets
   - Preprocessing steps
   - Segmentation
   - **Classification**
5. Run ruleset on entire AOI Raster
6. Generate shapefile output
7. Accuracy assessment
8. Production potential assessment
9. Creation of illustrative maps

Legend

- ArcGIS Pro
- eCognition
- Photoshop
- Internet / File system

Workflow



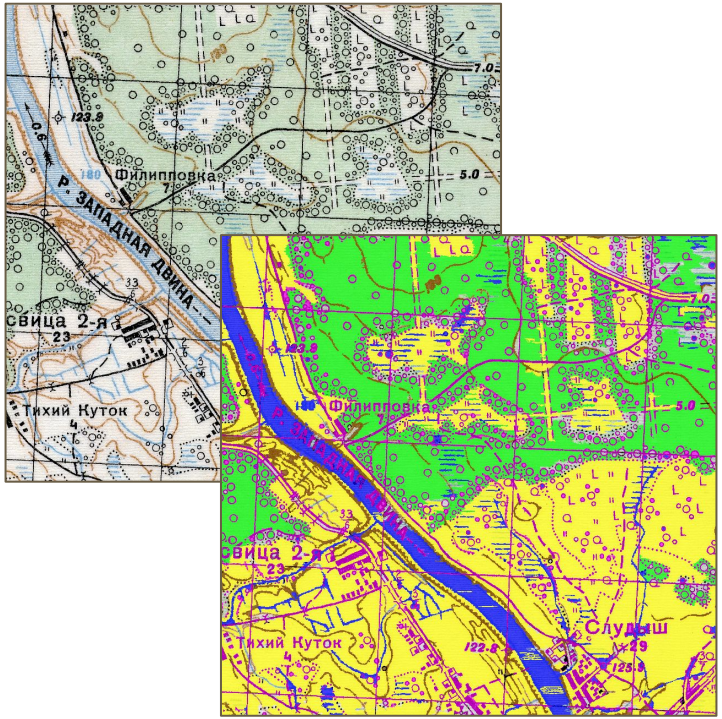
Classification example

1. Acquire raster files
2. Prepare AOI raster
   - Color calibration (?)
   - Create mosaic raster
   - Downsampling (?)
3. Define subsets
4. Develop rule set on subsets
   - Preprocessing steps
   - Segmentation
   - Classification
5. **Run ruleset on entire AOI Raster**
6. Generate shapefile output
7. Accuracy assessment
8. Production potential assessment
9. Creation of illustrative maps

Legend

- ArcGIS Pro
- eCognition
- Photoshop
- Internet / File system

Workflow



Classification example

1. Acquire raster files

2. Prepare AOI raster

- Color calibration (?)
- Create mosaic raster
- Downsampling (?)

3. Define subsets

4. Develop rule set on subsets

- Preprocessing steps
- Segmentation
- Classification

5. Run ruleset on entire AOI Raster

6. Generate shapefile output

7. Accuracy assessment

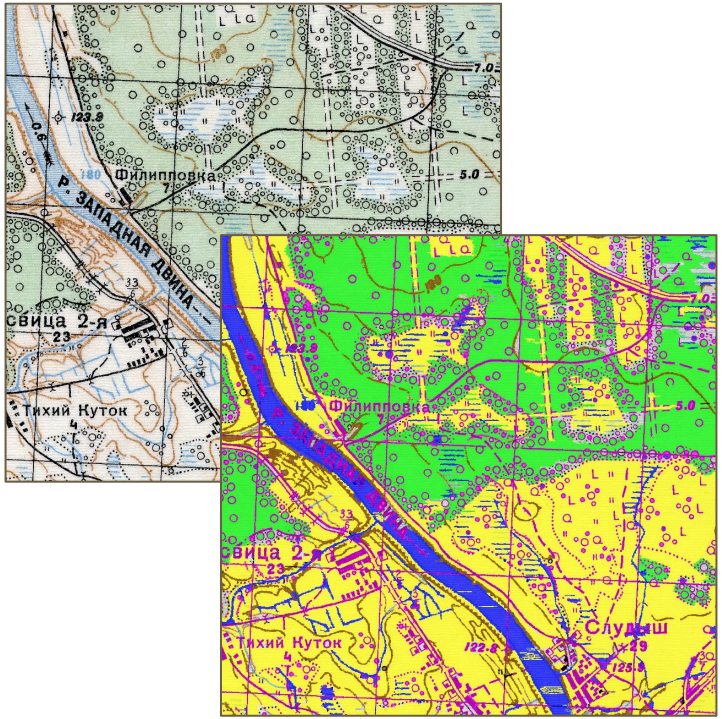
8. Production potential assessment

9. Creation of illustrative maps

Legend

- ArcGIS Pro
- eCognition
- Photoshop
- Internet / File system

Workflow



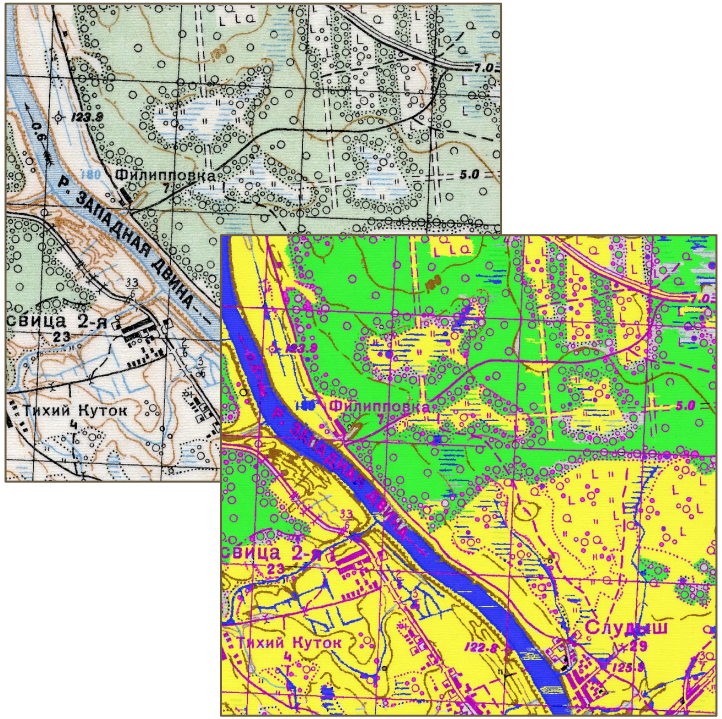
Classification example

1. Acquire raster files
2. Prepare AOI raster
   - Color calibration (?)
   - Create mosaic raster
   - Downsampling (?)
3. Define subsets
4. Develop rule set on subsets
   - Preprocessing steps
   - Segmentation
   - Classification
5. Run ruleset on entire AOI Raster
6. Generate shapefile output
7. **Accuracy assessment**
8. Production potential assessment
9. Creation of illustrative maps

Legend

- ArcGIS Pro
- eCognition
- Photoshop
- Internet / File system

Workflow



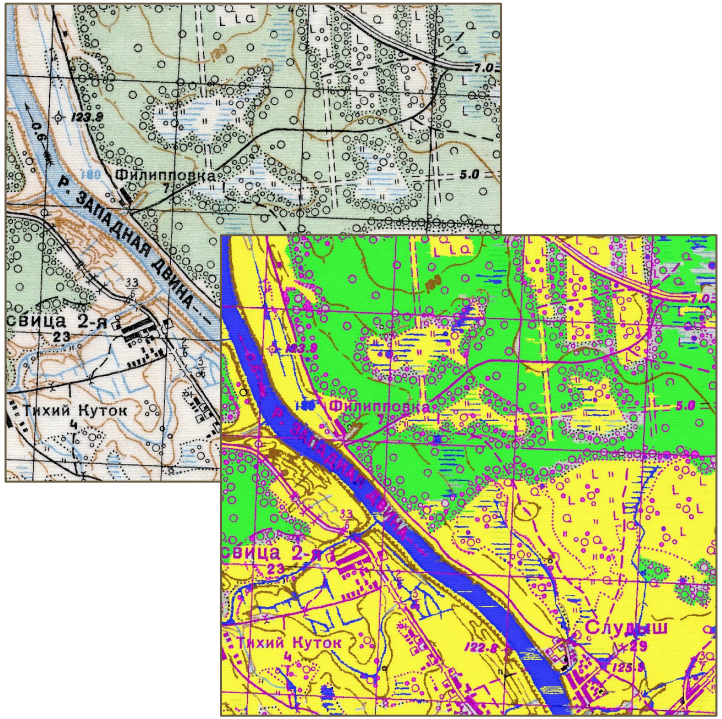
Classification example

1. Acquire raster files
2. Prepare AOI raster
   - Color calibration (?)
   - Create mosaic raster
   - Downsampling (?)
3. Define subsets
4. Develop rule set on subsets
   - Preprocessing steps
   - Segmentation
   - Classification
5. Run ruleset on entire AOI Raster
6. Generate shapefile output
7. Accuracy assessment
8. **Production potential assessment**
9. Creation of illustrative maps

Legend

- ArcGIS Pro
- eCognition
- Photoshop
- Internet / File system

Workflow



Classification example

1. Acquire raster files
2. Prepare AOI raster
   - Color calibration (?)
   - Create mosaic raster
   - Downsampling (?)
3. Define subsets
4. Develop rule set on subsets
   - Preprocessing steps
   - Segmentation
   - Classification
5. Run ruleset on entire AOI Raster
6. Generate shapefile output
7. Accuracy assessment
8. Production potential assessment
9. **Creation of illustrative maps**

Legend

- ArcGIS Pro
- eCognition
- Photoshop
- Internet / File system

Project Methodology

3. Features targeted

Land cover features

- Focus on extracting land-cover styles features, such as forested areas, barren land, built-up areas, and wetlands

- Belarus has many wetlands/swamp areas, so it will be an important aspect of the project.

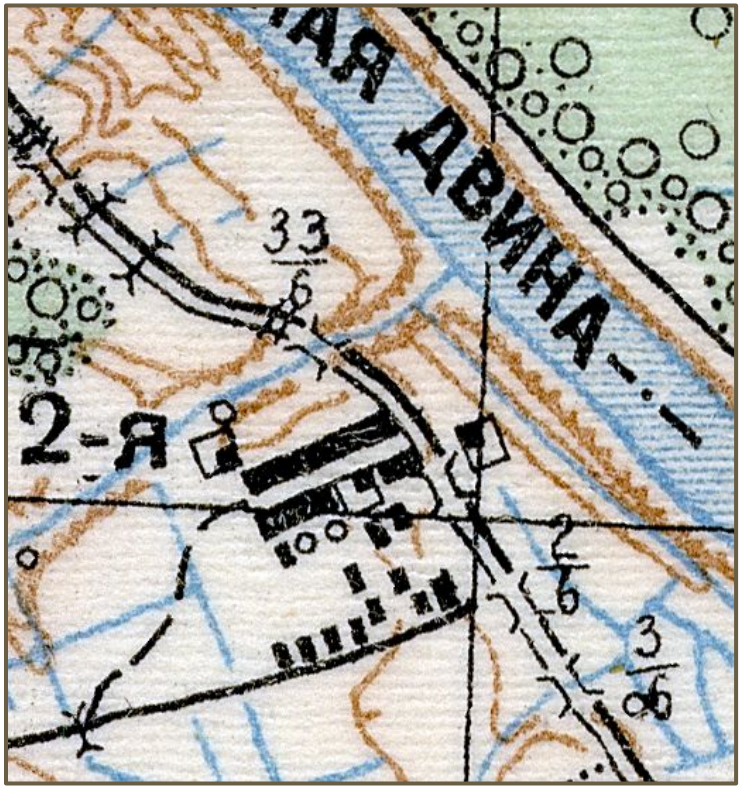
Forested wetland in Belarus (ramsar.org).

Challenging symbology

Built-up areas

- Same color as labels, roads, vegetation and tree symbols, etc.

- Variable shapes, which in many cases reproduce the actual shapes of the buildings



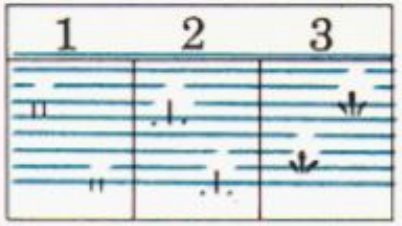
Example of built-up area.

Challenging symbology

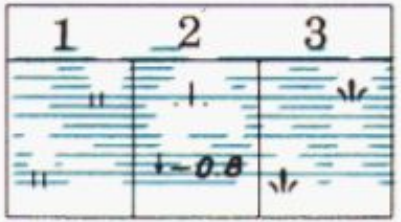
Wetlands / swamps - multi-layered symbology:

- Stripes: wetlands / swamp areas
- Symbols: 1. grass, 2. moss, 3. reeds or cane
- White areas: A. impassable, B. hard to pass, C. passable (military concept of “going”)
- (Dept. of the Army Headquarters, 1958)

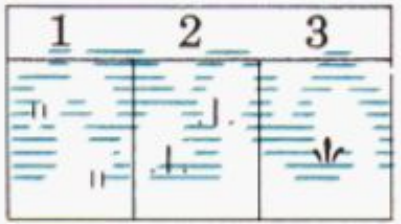
A.



B.

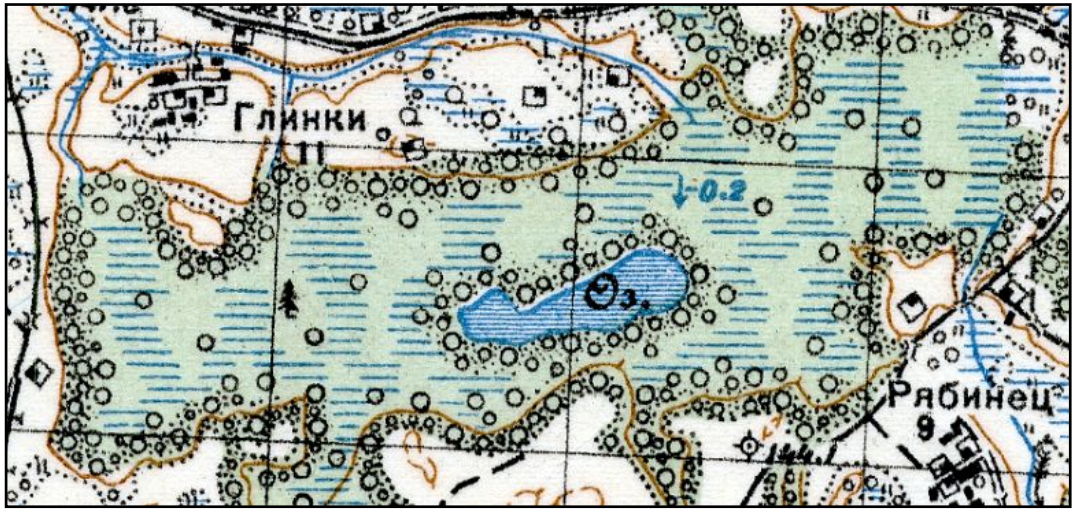


C.



Challenging symbology

Swampy forest!



Forested wetland in Belarus.

Textual elements

- No attempt to parse the textual elements, such as labels
- However, identify text elements as a separate class of eCognition objects

Example of textual elements

Project Timeline

Project Timeline

- May 2019
  - Peer review presentation.
  - Finalize the map set and create single raster
  - Start developing eCognition rule set, working on several raster subsets.

- June-July 2019
  - Develop any pre-processing steps that might be needed.
  - Estimate processing time, and proceed to downsampling the two rasters if necessary
  - Fully develop the eCognition rule set, still working on several raster subsets

- August 2019
  - Apply the rule set to the full raster.
  - Perform accuracy assessment

- September 2019
  - Assessment of potential for a production workflow.
  - Create illustrative maps.
  - Write final report.
  - Prepare conference presentation.

- October 2019
  - Present at conference.
  - Finalize final report.

Anticipated Results / Outcome

Project deliverables

- The eCognition rule set that I produce
- Any pre-processing script I create
- The final output shapefiles
- Several illustrative maps
- A report describing the process and discussing the results
- Presentation at a conference

Conference presentation

- Primary conference targeted: North American Cartographic Information Society (NACIS) 2019 Conference in Tacoma, WA, October 16-19

- I already submitted a proposal

Further Developments

Further developments

- Try out the workflow in a production context
- Apply the workflow to larger geographic areas as more map sheets from the same series become available
- I will share my results with the Indiana University Library staff

Thank you!

To my advisor Dr Nathan Piekielek

To Michelle Dalmau and Theresa Quill, Indiana University Bloomington

References (1)

Bidney, M., & Piekielek, N. (2018). In Defense of the Map Library. *Journal of Map & Geography Libraries*, *14*(1), 1–8. https://doi.org/10.1080/15420353.2018.1514856

Bolstad, P. (2016) *GIS Fundamentals: A first text on geographic information systems*, 5th edition. White Bear Lake, Minnesota: Eider Press.

Chen, Y. (2015). *Point Object Extraction From Scanned Topographic Maps.* (Master's Thesis, University of Wisconsin-Madison). Retrieved from https://minds.wisconsin.edu/bitstream/handle/1793/74932/Chen%20Yuying%202015.pdf;jsessionid=9C6044544506E8D8D32E14138420E117?sequence=1

Crowe, M. (2018, February 7). Digitizing captured Russian military maps. *Indiana University Bloomington Library - Communications*. Retrieved from https://libraries.indiana.edu/digitizing-captured-russian-military-maps

Davies, J. & Kent, A. J. (2017). The Red Atlas: How the Soviet Union Secretly Mapped the World. Chicago, IL: University of Chicago Press.

Department of the Army Headquarters (1958, June 12). Soviet Topographic Map Symbols; TM-30-548 Technical Manual. Retrieved from http://www.lib.berkeley.edu/EART/pdf/soviet.pdf

Herrault, P.-A., Sheeren, D., Fauvel, M., & Paegelow, M. (2013). Automatic Extraction of Forests from Historical Maps Based on Unsupervised Classification in the CIELab Color Space. In D. Vandenbroucke, B. Bucher, & J. Crompvoets (Eds.), *Geographic Information Science at the Heart of Europe* (pp. 95–112). https://doi.org/10.1007/978-3-319-00615-4_6

References (2)

Kerle, N., & Leeuw, J. de. (2009). Reviving Legacy Population Maps With Object-Oriented Image Processing Techniques. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(7), 2392–2402. https://doi.org/10.1109/TGRS.2008.2010853

Leyk, S., Boesch, R., & Weibel, R. (2006). Saliency and semantic processing: Extracting forest cover from historical topographic maps. *Pattern Recognition*, *39*, 953–968.

Madry, S. (2006). The Integration of Historical Cartographic Data within the GIS Environment. In S. N. Archer & K. M. Bartoy (Eds.), *Between Dirt and Discussion: Methods, Methodology, and Interpretation in Historical Archaeology* (pp. 33–60). https://doi.org/10.1007/978-0-387-34219-1_3

Uhl, J. H., Leyk, S., Yao-Yi Chiang, Weiwei Duan, & Knoblock, C. A. (2017). Extracting Human Settlement Footprint from Historical Topographic Map Series Using Context-Based Machine Learning. *8th International Conference of Pattern Recognition Systems (ICPRS 2017)*, 15 (6 .)-15 (6 .). https://doi.org/10.1049/cp.2017.0144

Watt, D. (2005). Soviet Military Mapping. *Sheetlines*, *74*. Retrieved from https://www.charlesclosesociety.org/files/Issue74page9.pdfQuill, T. (2015). Index to Soviet Military Topographic Maps at Indiana University Bloomington [Interactive map]. Retrieved from https://quill.carto.com/viz/084f1c22-e53f-11e4-9ce7-0e4fddd5de28/embed_map