SEAMLESS MOSAICING OF VERY HIGH RESOLUTION SATELLITE DATA AT CONTINENTAL SCALE: A CASE-STUDY FOR BIG DATA SCIENCE FROM SPACE

Pierre Soille

European Commission, Joint Research Centre (JRC)
Institute for the Protection and Security of the Citizen (IPSC)
Global Security and Crisis Management Unit, Via E. Fermi, 2749, I-21027 Ispra (Va), Italy

ABSTRACT

The seamless mosaicing of massive very high resolution imagery addresses several aspects related to big data from space. Data volume is directly proportional to the size the input data, i.e., order of several TeraPixels for a continent. Data velocity derives from the fact that the input data is delivered over several years to meet maximum cloud contamination constraints with the considered satellites. Data variety results from the need to collect and integrate various ancillary data for cloud detection, land/sea mask delineation, and adaptive colour balancing. This paper details how these 3 aspects of big data are handled and illustrates them for the creation of a seamless pan-European mosaic from 2.5m imagery (Land Monitoring/Urban Atlas Copernicus CORE_03 data set).

1. INTRODUCTION

The generation of seamless multiresolution mosaics of very high resolution (VHR) data at continental scale offers a wide variety of big data challenges. First, the shear amount of input data need to be addressed effectively and efficiently. In particular, the desired results should be obtained with as little as possible computations. Second, the method need to cope with data updates without requiring the reprocessing of the whole input data set. Third, it needs to take into account the integration of various additional data needed for achieving better results (e.g., cloud free reference, land/sea mask, reference for colour balancing). All these aspects are detailed in separate sections hereafter and are illustrated for the generation of a seamless pan-European mosaic from uncalibrated pansharpened SPOT5 and SPOT6 imagery.

This paper is organised as follows. Section 2 describes the principle of the underlying methodology for creating a seamless mosaic. Section 3 shows how the data volume matching 2.5m continental imagery is addressed. Data velocity considerations are detailed in Sec. 4. The integration of various additional data sources is presented in 5. Before concluding, results are given in Sec. 6.

2. SEAMLESS MOSAICING METHODOLOGY

The mosaicing methodology is based on the morphological compositing method described in [1]. In the simplest case of two overlapping image, morphological compositing is obtained by a seeded region growing procedure initiated by labelling the regions of no-overlap with the index of their corresponding image. The labelled regions are then propagated within the domain of overlap with a propagation driven by the image content. In practice, the propagation fronts originating from the two domains with no-overlap meet automatically at the boundary of salient image objects such as river or lake boundaries, field or road boundaries, etc. By doing so, the visual detection of the boundary, i.e., *seamline*, along which one switches from one image to the other in the resulting mosaic is hardly detectable even without applying image feathering. The method extends directly to any number of input images with an arbitrary number of overlaps while miminising the appearance of undesirable objects such as clouds in the output mosaic [1].

3. ADDRESSING DATA VOLUME

Data volume should not be simply addressed by providing the necessary capacity in terms of computing power and I/0 data rates. Ideally, the task at hand should be fulfilled with as few resources as possible. This is required by efficiency considerations that should be taken into account not only from purely economical but also from environmental considerations (large scale data centres have a non-neglectable impact on carbon emissions).

Efficient handling of large image data volumes need to take into account by considering adpative, scalable, modular, and distributed computing schemes. For the generation of VHR seamless mosaics at continental scale, this was achieved as follows:

- Adaptivity through multiresolution: the input image was downsampled twice with a factor 10 so that the resulting data consists of a pyramid with 3 levels (2.5m, 25m, and 250m). All subsequent computations are

doi: 10.2788/1823

performed at most appropriate resolution level with a speed factor of up to 10,000 when considering the coarsest level instead of the finest.

- Scalability: this is achieved by considering image processing algorithms having a linear complexity. For example, the region growing procedure at the basis of the mosaicing methodology is based on the a watershed segmentation algorithm satisfying this property;
- Modularity: the overall image processing chain is subdivided into modules that can be reused or relaunched separately in case of failure;
- Suitability for parallel distributed computing: whenever possible the tasks are performed in parallel, one for each image or set of independent images. A set of independent images is set of images that can be processed in parallel without generating any conflicting results. For the purpose of mosacing, two images are independent if and only if there is an empty intersection between (i) the data region of interest of the first image unioned with those of its intersecting images and (ii) the data region of interest of the second image unioned with those of its intersecting images.

4. ADDRESSING DATA VELOCITY

In many applications, the input data set is constantly augmented by new data. Efficiency constraints impose that not all the input data should be reprocessed for every single addition but should be restricted to the minimum amount of necessary data. For morphogical mosaicing, this is simply achieved by considering that the data region of interest of a given image is defined by its subdomain contributing to the mosaic calculated at the previous step. Consequently, only those images whose subdomains are interesecting the domains of the current set of additional images need to be processed.

5. ADRESSING DATA VARIETY

With the increasing availability of free and open geospatial data, the processing of a specific input data for a given application should exploit additional data sources suitable for enhancing the quality of the output data. In the case of the seamless mosaic, the following additional data sources where integrated in the processing chain:

- MODIS Terra 8-day composites to enable the accurate detection of clouds at the coarsest resolution;
- Landsat data to define reference from true colour image and achieve colour balancing at medium resolution (automatic learning);
- OpenStreeMap for defining a land/sea mask that is then conflated with the input imagery.

6. RESULTS

Figure 1 (left) shows the footprints of the input image data. It consists of (3.3 TeraPixels of pan-sharpened SPOT5 and SPOT6 imagery (Copernicus 2011-2013 CORE_03 coverage over the 39 European Environment Agency member and associate states) and (right) the resulting seamless mosaic (1.4 TeraPixels). It can be viewed at full resolution at `http://ghslsys.jrc.ec.europa.eu?shortcut=CopernicusSeamlines` and at `http://cidportal.jrc.ec.europa.eu/copernicus/services/webviewer/core003/` (seamline version). This data set is at the basis of the production of the Global Human Settlement Layer (GHSL) [2] for Europe.

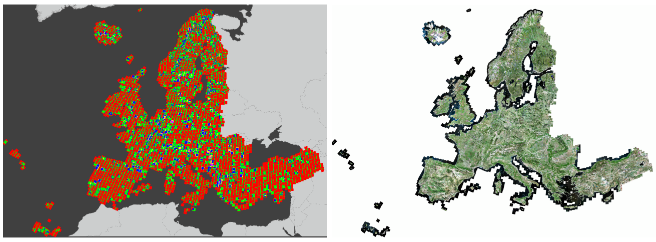

Fig. 1. Left: footprints of the 3,544 input VHR images with colour coding of the overlap degree (3.3 TPixels). Right: resulting seamless mosaic (1.4 TPixels).

7. CONCLUSION

This paper shows that big data science from space should not be solely based on generic data mining approaches for remote sensing expert knowledge need to be exploited to ensure efficient and effective processing. This has been illustrated for the generation of a pan-European VHR seamless mosaic. This type of product will become routine at medium resolution for the generation of global Sentinel-2A/B composites given the 5 day repeat time of these twin satellites. In this latter case, overall quality measures combining pixel quality measures and spatial coherence need to be maximised [3].

8. REFERENCES

[1] P. Soille, "Morphological image compositing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 5, pp. 673–683, May 2006. doi: 10.1109/TPAMI.2006.99

[2] M. Pesaresi and et al., "A global human settlement layer form optical HR/VHR RS data: concepts and first results," *Journal of Earth Observation Applications*, vol. 6, no. 5, pp. 2102–2131, 2013. doi: 10.1109/IGARSS.2010.5653039

[3] P. Soille, "Towards the compositing and mosaicing of Sentinel-2 images using quality and spatial coherence constraints," in *Proc. of Sentinel-2 for Science Workshop*. ESA, May 2014.