

Crowdsourcing ground-truth datasets for building change detection

Juste Raimbault,^{1,2,3,4} Julien Perret,^{1,5}

1. LaSTIG, Univ. Gustave Eiffel, IGN-ENSG
2. CASA, UCL
3. UPS CNRS 3611 ISC-PIF
4. UMR CNRS 8504 Géographie-cités
5. LaDéHiS, EHESS

An accurate quantification of past urban dynamics is key for designing future sustainable cities [1]. More particularly, dynamics of urban densification are a complex yet crucial issue, since they link to several conflicting sustainability goals [2] (e.g. increased access to public transport and amenities, and limited urban sprawl, but with negative externalities such as less urban green space or an increased urban heat island effect), and imply several stakeholders at multiple scales what makes these difficult to control through policies [3]. Since landowners are often implied in plot splitting or soft densification, an understanding of change dynamics is required at the building level. Raster data such as aerial imagery, on which machine learning methods provide a good performance for land-use change for example [4], is not sufficient for that purpose, and change detection must be performed on vector data of building geometries. To that purpose, geospatial vector data matching algorithms are an efficient tool [5], with the advantage of being robust to spatial noise in geometries and changes in data specifications. They however require some extensive parametrisation and provide varying performances [6]. An important issue for optimising this type of algorithms is the lack of ground-truth datasets with a large enough size and diversity. We propose in this contribution to tackle this issue, for the specific case of building change detection.

We therefore introduce a web application designed to annotate potential matching links between two vector datasets. Its main functionalities are that (i) it is fully based on git and javascript, requiring no database nor advanced web server deployment, and can be automatically deployed through github pages for example; (ii) it manages multi-user collaboration, with authentication linked to rights granted for the project git repository (implemented only for github for now); (iii) provides a dashboard with the user progress and the overall progress on all datasets displayed on maps; (iv) it allows collaborators annotating two vector datasets at two different dates, with aerial photographs being provided in the background for ground-truthing. The annotation is sequential on potential matching links (intersecting features between the two datasets), and the user must ground-truth first the type of link the algorithm should detect (no link, single, multiple), then the type of real world evolution (parametrised for building changes in our case), and can also highlight and comment data quality issues. The application is open source and available at <https://github.com/subdense/annotate>.

We show an example of the application in Figure 1. We currently deployed the application on samples within the functional urban areas of Strasbourg, France and Dortmund, Germany, as data quality issues are very different across countries (for example, French BDTopo had a specification change in 2015 leading to many geometry changes with polygon splits that do not correspond to real-world changes). We compare buildings between 2011 and 2021, for all buildings within 500m radius circles around 100 sampled points for each area. The annotation campaign is currently ongoing with several researchers experts on GIS data and urban densification. First results to optimise algorithms (in the sense of F-score for real-world change prediction) with partial ground-truth datasets were obtained for the same algorithms studied by [6] (the Multi-criteria matching algorithm and the Geometric Matching of Areas algorithm), and confirm previous results obtained with synthetic data, in particular that (i) each algorithm can be optimised to better perform in different urban form typologies (center, housing projects, suburbs); (ii) they can be combined into a multi-modelling approach for a better overall performance.

Our contribution thus provide a basis for a robust quantification of urban densification at the building level, and a generic web application to annotate vector datasets in a comparative way.

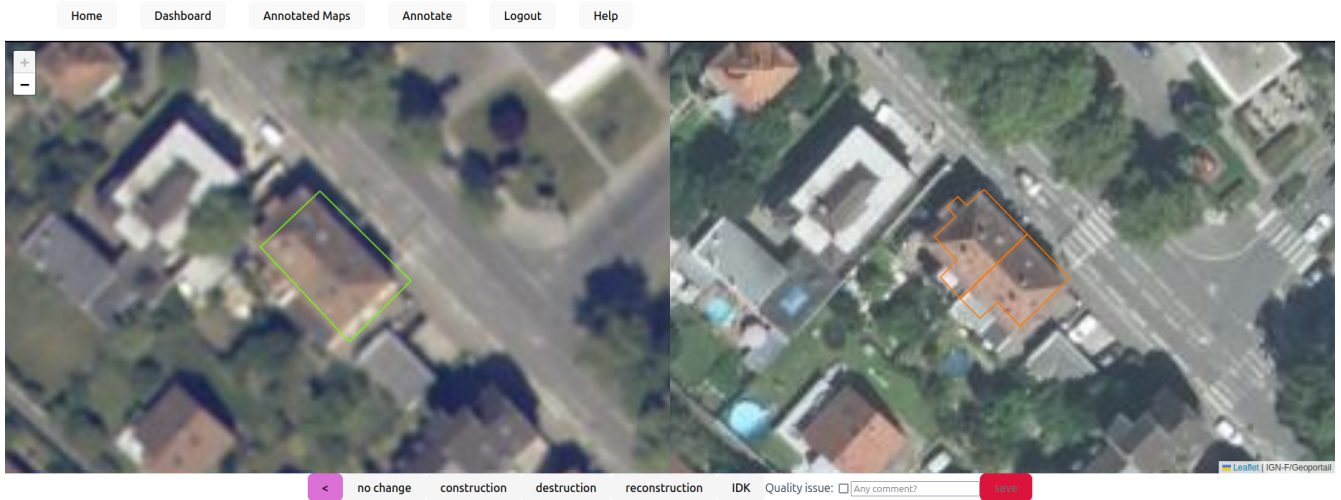


Figure 1: Screenshot of the annotation web application. On the top the user can access the different tabs, such as the dashboard with the annotation progress on the tasks the user was assigned, or global maps showing progress across all users and datasets in space. The main tab to annotate shows two aerial photographs of the same place at the two compared dates, with the vector features at each date surimposed. In this particular case, there seems to be no change in the real world, but the polygon data for buildings significantly changed. The user first provides the type of link, and then (shown in the screenshot) the type of real world building change (no change, construction, destruction, reconstruction, does not know), and can comment on a data quality issue. Note that aerial photographs are in some rare cases not enough to ground-truth change, thus the option to skip the step. Once clicking “save”, the next unit to annotate will be displayed. The application is currently parametrised for building change, but vector data sources, aerial imagery sources, and change labels can all be customised in a configuration file, and thus can be changed for other types of features, such as land plots or road networks.

- [1] Michael Batty. *Inventing future cities*. MIT press, 2018.
- [2] David Evers, Ivana KaturiĆ, and Ries van der Wouden. *Urbanization in Europe: Past Developments and Pathways to a Sustainable Future*. Springer Nature, 2024.
- [3] Mathias Jehling, Martin Schorcht, and Thomas Hartmann. Densification in suburban germany: approaching policy and space through concepts of justice. *Town Planning Review*, 91(3):217–237, 2020.
- [4] Lazhar Khelifi and Max Mignotte. Deep learning for change detection in remote sensing images: Comprehensive review and meta-analysis. *Ieee Access*, 8:126385–126400, 2020.
- [5] Emerson MA Xavier, Francisco J Ariza-López, and Manuel A Urena-Camara. A survey of measures and methods for matching geospatial vector datasets. *ACM Computing Surveys (CSUR)*, 49(2):1–34, 2016.
- [6] P. Guardiola, J. Raimbault, A. Olteanu-Raimond, and J. Perret. Optimising geospatial vector data matching algorithms for change detection with synthetic data. In *Conference on Complex Systems 2024, Exeter, September 2024*, 2024.