2. Automated Streetlight Census from Street-Level Imagery: Integrating Computer Vision, Depth Estimation and Geospatial Processing

Leonardo Bridi[1], Eleonora Andreotti[1]⋆, Davide Magurno[2], Andrea Lugli[2], and Maurizio Napolitano[1]

[1] Digital Commons Lab, Fondazione Bruno Kessler, Via Sommarive 18, 38123 Povo, Trento, Italy.

[2] Sis-Ter Srl SB, Via Emilia 69, 40026 Imola (BO), Italy.

Public lighting is one of the most pervasive urban infrastructures, playing a crucial role in urban management, yet it remains largely undocumented in digital form: no extended catalogues of street poles and lamps currently exist, and their mapping and maintenance still rely on time-consuming and costly field surveys, dependent on manual data collection, integration of heterogeneous sources (GIS, street-view imagery, satellite data) and extensive post-processing. This work presents StreetLightScanner, an automated pipeline for detecting, classifying and geolocating streetlight infrastructure from street-level images collected with a smartphone or a camera-car. Unlike traditional photogrammetric approaches, focused on rigorous geometric reconstruction through calibrated metric cameras and bundle adjustment, the proposed framework adopts a geometry-relaxed paradigm, achieving operational robustness through redundancy, semantic segmentation and probabilistic aggregation across heterogeneous observations, and thus reducing the dependence on controlled acquisition geometry and explicit camera calibration. Methodologically, the system is built around a mask-aware classification core: streetlight components are detected and segmented with SAM3, exploiting its open-vocabulary, text-driven Promptable Concept Segmentation, and the resulting masks are propagated into a custom ConvNeXt V2 classifier implementing masked pooling. This design preserves the structural cues of thin, elongated poles and suppresses the background clutter that degrades naive crop-and-classify pipelines. Two dedicated classifiers, for pole geometry and luminaire typology, are trained on a custom-annotated dataset and matched to the categories of the target municipal lighting register. Around this core, a modular architecture adds AI-based monocular depth estimation and GPS-based geometric modelling, supporting three operational modes - single-image campaigns, crowdsourced street-view imagery and full Structure-from-Motion reconstruction. Redundant observations are consolidated through a statistical ⋆Corresponding author. eandreotti@fbk.eu L. Bridi et al. consensus mechanism and exported as GIS-compatible assets. Validated on a real-world case study in Mattarello (Trento, Italy), the system substantially reduced field-survey effort while detecting streetlights at ∼ 93% F1 and geolocating them within ∼1 m (median) after a lightweight GT-anchored correction. Future developments will address dynamic acquisition scenarios and additional classes of urban assets.