Doctoral Dissertation
Doctoral Program in Computer and Control Engineering (37^{th}cycle)

Towards Scalable and Robust Visual Geo-localization for Real World Applications

By

Gabriele Berton
******

Supervisor(s):
Prof. Barbara Caputo, Supervisor
Prof. Carlo Masone, Co-Supervisor

Prof. Margarita Chli, ETH Zurich
Prof. Javier Civera, University of Zaragoza
Prof. Sebastian Scherer, Carnegie Mellon University
Prof. Giorgos Tolias, Czech Technical University

Politecnico di Torino
2025

Abstract

Visual Geo-localization is a fundamental computer vision task that aims to determine the geographical location of a query image by matching it against a database of geo-tagged reference images. This thesis advances the field by addressing key challenges within the two sub-tasks of Visual Place Recognition, which aims to localize images within a large area, and Astronaut Photography Localization, where the goal is to obtain precise localization of a photo taken from space.

Visual Place Recognition (VPR) has the goal of localizing photos within a large map, through means of image retrieval, by retrieving query's most similar image from a database of geo-tagged images; VPR finds applications in robotics, augmented reality, autonomous cars, and image verification at large, *e.g.*journalists verifying a photo's location. The first work of this thesis is a comprehensive benchmark of existing methods, to better understand which are the limitations of existing models, and which are the open challenges that prevent VPR models from being deployed at large. After finding that scalability presents still a major issue in VPR pipelines from the literature, the next chapter introduce CosPlace, a training technique that frames the training as classification, eliminating the need for expensive mining and reducing the training complexity from quadratic to linear: CosPlace allows to train models on previously unfeasibly large datasets, such as the San Francisco eXtra Large dataset of 41M images, leading to robust representations that achieve impressive results even with small descriptors. Noting that a main limitation of CosPlace is robustness to viewpoint invariance, we then present EigenPlaces, a training technique that explicitly imbues robustness to viewpoint into the model by carefully selecting ideal training samples.

Given the lessons learned from VPR, the second half of the thesis focuses on the problem of Astronaut Photography Localization (APL), which seeks to provide location data for photos taken by astronauts from the International Space Station.

This problem finds applications in disaster management and climate change research, and it has been historically tackled manually: since the launch of the International Space Station scientists at NASA have already localized over 300k images. With the aim to automate this task, we present EarthLoc, which looks to understand the feasibility of such problem through computer vision means: this is done with a deep learning model, trained on a large corpus of satellite imagery, that retrieves the most similar satellite image from a given area, and adopts targeted techniques to create ideal training batches to increase robustness to the domain gap. As EarthLoc shows that the problem can be tackled with modern computer vision, the next chapter presents AstroLoc, which achieves hugely better results by making use of the 300k manually annotated images. AstroLoc relies on a combination of two task-specific losses, and uses both astronaut photographs and satellite imagery for training: its robustness is such that it even works well on 50 years old imagery from the Space Shuttle era, showcasing robustness to high color jittering and long-term temporal changes. Retrieval techniques require post-processing for confidence estimation and refinement: we therefore present EarthMatch, an iterative matching pipeline that can be built on top of existing retrieval methods, and allows us to filter out bad predictions, as well as to provide precise pixel-wise GPS location for the astronaut photos.

Collectively, this work provides scalable, efficient, and robust solutions for real-world localization problems, advancing the state-of-the-art in both urban and space-based geo-localization. Our contributions open new possibilities for autonomous systems, environmental monitoring, and large-scale visual understanding of our planet.