Runtime-based approaches to investigate and optimise data quality of crowd-sourced observations for environmental monitoring

Scope: Europe has invested a significant amount of resources in infrastructure and services to achieve an autonomous, accurate and multi-level Earth observation capacity. Initiatives such as Copernicus or similar (INSPIRE Infrastructure for Spatial Information in Europe, UrbanAtlas) have been seminal in providing a mapping of spatial planning, forest/agricultural areas, wetlands or artificial surfaces; yet, clearly the burden of investing in additional equipment or of maintaining the current infrastructure is unsustainable. Innovative ways of complementing the in-situ infrastructure with citizen-sourced data from thousands of mobile sensors at a lower cost are thus being sought. Whilst citizen participation in the environmental policy making process is still in its infancy, there are signs of a growing interest. This research is in the context of the Scent project (https://scent-project.eu/), funded by the EC H2020 Programme. In Scent, through a constellation of smart collaborative technologies, citizens will become the 'eyes' of the authorities and policy makers and will monitor land-cover/use changes through everyday activities. In this way, the costly in-situ infrastructure will be augmented with a people-generated and people-centric web of observations.

However and despite recognising the power and benefits of citizen science data, the data quality of observations is still ambiguous; innovative and at the same time reliable methods are being investigated for determining the data quality of observations or inversely, the data uncertainties. There are two categories of methods/ways this problem can be approached: (i) design-based approaches and (ii) runtime approaches. This thesis will concentrate on the second category. There are many research studies on the way that the data collected from sensors (including smartphone cameras) can be evaluated and how we can flag as invalid the measurements that are indeed a mistake and not a result of a phenomenon that was unexpected for the time/place. Some state of the art methods include the application of an ELO rating scheme[1], fuzzy logic[2] or others, yet the problem is very often application-driven. In the context of this thesis, observations from the following (not exclusive) tasks/actions from citizens will be collected:

o smartphone images for several land-cover/use types within the Kifisos river area; annotation of images according to an image taxonomy/classification; images to measure water level from a pre-installed measuring tape placed in the river bank

o smartphone videos in order to measure the flow of the water in the Kifisos river according to given guidelines

o soil moisture measurements with a portable sensor and through a Scent-based smartphone app

o reporting of events that are relevant to the environmental problem at hand through a Scent-based smartphone app

In the context of thesis, field test experiments will be organised with a team of volunteers contributing to the task execution for testing the methods and approaches that the candidate will investigate.

Prerequisite Knowledge:

[1] https://en.wikipedia.org/wiki/Elo_rating_system

[2] Timms GP, de Souza PA, Reznik L, Smith DV. Automated Data Quality Assessment of Marine Sensors. *Sensors (Basel, Switzerland)*. 2011;11(10):9589-9602. doi:10.3390/s111009589.

- Theoretical background on statistical analysis and/or basic data analytics/machine learning background
- Some experience with programming languages (Python, C or Java).

Knowledge to be acquired from the thesis:

- Knowledge of the state of the art in citizen science and crowd-sensing
- Hands-on experience with designing large-scale demonstrations of crowd-sourcing campaigns

Supervisors:

Νικόλαος Ουζούνογλου (nuzu@cc.ece.ntua.gr), Άγγελος Αμδίτης (angelos@esd.ece.ntua.gr)

Further info:

Αθανασία Τσέρτου (atsertou@iccs.gr), Μαρία Κρομμύδα (maria.krommyda@iccs.gr)