

Short Term Scientific Mission Report

- COST Action IC0903 -

Beneficiary (visiting scientist): Dr Milan Mirkovic, Faculty of Technical Sciences (FTS), Serbia

Host: Dr Gennady Andrienko, Fraunhofer IAIS

Period: From November 18th to November 30th

Place: Sankt Augustin, Germany

Purpose of the STSM

Main goal of the visit was for the visiting scientist to gain insight into methods and tools employed by the host group for extracting knowledge from geo-referenced data. Over the last decade, host group led by Dr Gennady and Dr Natalia Andrienko has made significant achievements in devising methods for spatial and temporal analysis of data that has geographical location associated with it. As a part of those efforts, the group has developed a software suite (Visual Analytics System for Spatial and Temporal Data) that enables interactive exploration of the data using aforementioned methods, and provides a set of tools for data transformation and analysis in this context. On the other hand, visiting scientist's group has developed a method for retrieving geo-referenced data associated to YouTube videos, a source that has not been explored yet by the host group.

In specific, objectives set for this visit were:

- To share knowledge of the techniques employed in research at FTS with the host group at the IAIS.
- For visiting scientist to get better insight into the visualization algorithms and tools that the host group has developed and is using.
- To provide a new set of data for the host group (and wider scientific community) retrieved from a service not so commonly used in analysis of episodic movement data (i.e. YouTube).
- To formulate and examine hypotheses on what constitutes differences in the ways episodic movement data is generated in freely available Flickr versus YouTube data sets, and how this differences influence the visual analysis process, and the type of information that can consequently be extracted from such data sets.
- To identify missing coverage of retrieval, computational and visualization methods in the currently available software toolset that would be needed to conduct visual analysis according to these findings.
- To integrate corresponding extensions with the visual analytics framework, both during and as

mid- to long term collaborative effort beyond the immediate STSM.

- To discuss future collaborations and outline a joint research paper based on the analysis of tracks present in community-contributed multimedia data.

For the most part, these objectives were achieved during the visit and some additional ones were defined and will be pursued in the future.

Description of the work carried out during the STSM

I spent the first three days of the visit on getting introduced to cutting-edge methods currently used by the host group. As a part of this introduction process, I was setup with the most recent version of V-Analytics software, and a tutorial was given to me by Dr Natalia Andrienko on some of its most commonly used features. A compilation of relevant papers providing theoretical background necessary for being able to better exploit potential of the V-Analytics was made by Dr Georg Fuchs and presented to me.

On the fourth day, I gave an hour-long seminar about the methods employed by the FTS group to crawl the Youtube for meta-data. It was attended by members of Knowledge Discovery and Mobility Mining groups, and we had a discussion afterwards on how approach of the FTS group could be improved, with respect to existing limitations imposed by the YouTube service. Particular focus was placed on quantifying performance of the crawler, and on potential methods to further enhance the size of returned result-set. Last day of the first week was spent on assessing possible use-cases where geo-referenced data from different sources could be used to either extract new knowledge or enhance/enrich existing information by providing more details.

It was decided that it would be interesting to see if well-known events could be tracked (both in spatial and temporal dimensions) across different communities (i.e. Flickr, YouTube and Twitter), and to identify patterns pertinent to each community. In order to do that, we decided to focus on a recent catastrophic event (hurricane Sandy, that hit the east coast of the US) since it was bound to resonate through social media. Hence, the first two days of the final week were spent on retrieving data from different sources that is geographically associated to metropolitan area of New York city (USA) and on agreeing upon some common data formats that would be used for the exchange of the retrieved meta-data. Finally, the remainder of the STSM was spent on identifying patterns (spatial, temporal and even behavioral – when it comes to habits of social-media users) that were present in the data, and discussing possible ways to use the aggregated dataset to extract knowledge that would escape detection if data from different services were observed separately.

During my whole stay, the whole team (Dr Gennady Andrienko, Dr Natalia Andrienko and Dr Georg

Fuchs) provided valuable advice and guidance (especially when it came to using V-Analytics software).

Description of the main results obtained

This short visit enabled both teams to gain insight into the research being conducted by the other team, which was one of the main goals of the mission. Visiting scientist was acquainted with the cutting-edge methods and tools developed and used by the host group, which should make future collaboration and identification of potential joint research topics much easier. Also, software (crawlers) for obtaining data from different sources (YouTube and Flickr respectively) was exchanged, effectively enabling both teams to build their own datasets. In addition, common data attributes were identified which will make future exchange of the datasets easier.

Experiment on the aggregated dataset showed that sudden (abnormal) events such as a natural disaster can easily be detected within all three observed communities. Even though one of them (Twitter) stands out so much in terms of response time (seconds and minutes versus hours and days when pictures/videos are in question) and sheer volume of data that it alone can be used to gain a fair amount of information regarding the observed event, it was our hypothesis that media-sharing sources (Flickr and Youtube) can be leveraged to gather more details on the event in question and/or to fine-tune its spatial and temporal boundaries. Our experiment to track hurricane Sandy through different social networks and services (both in space and in time) gave inconclusive results, in terms that we have succeeded in identifying the abnormal event (the hurricane) in all three services, but failed to detect the exact trajectory of it with the desired level of detail using only geo-referenced media data (which would enable us to document its progress and aftereffects, and identify potentially important sub-events such as local rescue efforts or particularly affected areas). This failure was mainly caused by the shortage of data, which in turn was caused by the constraints (and even changes) in the API of respective services that emerged during our efforts to create the dataset. We have already had some discussion on how to improve our data retrieval methods and expect to be able to create a more substantial dataset in near future. On the other hand, even our preliminary findings seem to be well aligned with other related research [1] when some temporal patterns are concerned, and we intend to try to identify what constitutes abnormal events in different social media and how other sources reflect them and react to them (e.g. a natural disaster caused a stir in all three mentioned services, but a concert or some other locally important event might cause abnormality to be detected in one service and go almost unnoticed in another). This should enable us to further explore the cause and effect of “offset” that is apparently present when abnormal events are observed through different social-media.

Immediate and tangible results of this STSM are twofold: one part is in the form of a dataset of

aggregated geo-referenced meta-data retrieved from Flickr and YouTube, covering the metropolitan area of New York city (USA) between October 20th and November 05th 2012, that will be made publicly available on the website of one or both of the groups (FTS/IAIS) after final data cleanup (right now it is available upon request). Even though this dataset did not prove sufficiently detailed for the needs of our experiment, it might prove valuable for other tasks (such as semantic analysis of titles/descriptions, individual movement analysis, etc.). The other part is a YouTube crawler (for retrieving publicly available, community contributed geo-referenced videos) that will be made freely available after some refinement and implementation of new features that should make data retrieval more efficient and more comprehensive.

Future collaboration with host institution

Joint efforts in immediate future will be concentrated on improving data retrieval methods and building a sufficiently large dataset to test our hypotheses on detecting/tracking abnormal events in different social media.

In addition to these, a few other topics have been discussed that are not directly related to the ongoing joint research but show a good potential for future collaboration (such as automatic identification of potentially hazardous crowd events, semantic analysis of meta-data associated to YouTube videos, improvement of Video Quality Assessment approaches by leveraging publicly-available meta-data, etc.)

Foreseen publications/articles resulting or to result from the STSM

Several conferences and journals have been considered for publication of results that ought to emerge from the joint research efforts of the IAIS and FTS groups. These results are expected to be twofold: (1) improved methods for retrieving metadata associated to publicly available multimedia content – which we plan to publish in journals or conferences related to computer science and information systems (such as <http://www.comsis.org/> or <http://www.telfor.rs/?lang=en>), and (2) comparison of differences between (and tracking of) abnormal events detected in different media-sharing services / social networks – which we plan to publish in related conferences (such as <http://www.mirlabs.net/cason12/> or <http://www.eurova.org/>).

References

[1] Chae, J., Thom, D., Bosch, H., Jang, Y., Maciejewski, R., Ebert, D. S., & Ertl, T. Spatiotemporal

Social Media Analytics for Abnormal Event Detection and Examination using Seasonal-Trend Decomposition, IEEE Conference on Visual Analytics Science and Technology 2012, October 14 - 19, Seattle, WA, USA