Compression of spatio-temporally correlated and massive georeferenced data

Antoine Crinière (1), Aline Roumy (2), Thomas Maugey (2), Michel Kieffer (3), and Jean Dumoulin (4)

(1) INRIA, I4S, Rennes, France (antoine.criniere@inria.fr), (2) INRIA, SIROCCO, Rennes, France, (3) L2S, Centrale Supelec, Paris, France, (4) IFSTTAR, COSYS, SII, Bougenais, France

During the past decades, the concept of smart-sensing in smart-territories has experienced a sustained growth thanks to the large scale deployment of sensor networks. This leads to problems related to the acquisition, management, and storage of a huge volume of data generated by these sensors. Usually, data generated by sensors measuring physical quantities related to their environment are largely correlated. This correlation may be spatial: two closely located sensors are likely to generate closely related data. Depending on the sampling rate, data generated by a given sensor may also be temporally correlated [1].

Many research works have considered the source coding problem of data generated by a network of sensors [2,3]. Most of these papers focus on coding efficiency. The way source coding has to be achieved when a random part of the data has to be retrieved has been only recently considered [4]. The INTERCOM project* aims to provide only the requested sources without having to transmit and decode the whole encoded data base although no re-encoding is needed. It proposes an interactive way to access multiple subsets of spatio-temporally correlated georeferenced data [5] while ensuring a convenient transmission rate and storage cost. The main challenge for interactive compression of multisources is then to provide an efficient exploitation of the temporal and spatial correlations while offering the possibility to extract the requested data only.

In this communication, we first analyze different compression schemes that are able to exploit not only the temporal but also the spatial correlation between the data. A special focus is made on a scheme where some sensors are used as references to predict the remaining sources. Then, an adaptation of the scheme is proposed to offer interactivity and free selection of some sources by a client.

Six months of weather parameter measurements have been used to generate our data set. For instance, temperature data have been recorded for almost 600 sensors distributed all over the European continent and gathered thanks to the METAR network.

* (http://www.intercom.cominlabs.ueb.eu/)

[1] D. Pan et al. Lossless differential compression of weather radar data in universal format using motion estimation and compensation. 26th Conference on IIPS, 2010.
[2] F. Marcelloni and M. Vecchio, An efficient lossless compression algorithm for tiny nodes of monitoring wireless sensor networks, Computer Journal, vol. 52, no. 8, pp. 969–987, 2009.
[3] T. Srisooksai et al. Practical data compression in wireless sensor networks: A survey, Journal of Network and Computer Applications (35) 2012
[4] Dupraz, E., Maugey, T., Roumy, A., and Kieffer, M. Rate-storage regions for massive random access. CoRR, abs/1612.07163, 2016
[5] Crinière, A., Roumy, A., Maugey, T., Kieffer, M., and Dumoulin, J. Sélection optimale de capteurs de référence pour le stockage de données spatialement corrélées. In GRETSI 2017, Juan-les-Pins, France 2017.