ASSESSING THE CHALLENGES OF ENVIRONMENTAL SIGNAL PROCESSING THROUGH THE SENSORSCOPE PROJECT

Guillermo Barrenetxea, François Ingelrest, Yue M. Lu and Martin Vetterli

LCAV, I&C School, Swiss Federal Institute of Technology Lausanne (EPFL), Switzerland
{Guillermo.Barrenetxea, Francois.Ingelrest, Yue.Lu, Martin.Vetterli}@epfl.ch

ABSTRACT

SensorScope is a collaborative project between network, signal processing, and environmental researchers that aims at providing a cheap and out-of-the-box environmental monitoring system based on a wireless sensor network. It has been successfully used in a number of deployments to gather hundreds of megabytes of environmental data. With data gathering techniques well mastered, the efficient processing of the huge amounts of the acquired information to allow for useful exploitation has become an increasingly important issue. In this paper, we present a number of challenging and relevant signal processing tasks that arise from the SensorScope project. We believe the resolution of these problems will benefit from a better understanding of the underlying physical processes. We show an example to demonstrate how physical correlations between different sensing modalities can help reduce the sampling rate.

Index Terms— Wireless sensor networks, environmental monitoring, environmental signal processing, sampling, new challenges

1. INTRODUCTION

A *Wireless Sensor Network* (WSN) [1] is a self-organized multi-hop wireless network, composed of small communicating devices called *sensor motes* which are generally deployed over an area of interest. Thanks to embedded or external sensors, these motes are able to gather various information about their environment, such as wind speed, air temperature, or soil humidity. WSNs usually operate in an *n-to-1* communication paradigm, in which collected data is forwarded to a base station (*sink*). The sink is then in charge of sending this data to a server where it is stored and further processed.

Environmental monitoring is one of the most important applications of WSNs. Current data collection techniques are indeed rather limited and make use of very expensive sensing stations, leading to a lack of appropriate observations. Thanks to a WSN architecture, it is possible to deploy a large quantity of cheap stations to get dense spatial and temporal measurements that can be used to create prediction models of potentially dangerous climate changes.

In this paper, we first describe the SensorScope project, which is a WSN-based system with applications in environmental monitoring. Section 2 summarizes the development of the system, including its construction, networking issues, and field experiments. During the last 15 months, the SensorScope system has been successfully deployed at several locations, generating a huge amount of environmental data. With the data gathering technique well mastered, now the question becomes: how to efficiently process the data from sensor network measurements to allow for useful exploitation? In the second part of the paper, we present a list of challenging and relevant signal processing problems that arise from our SensorScope project. We believe the resolution of these problems will benefit from a better understanding of the underlying physical processes. In Section 4, we show an example to demonstrate how physical correlations between different sensing modalities can help reduce the sampling rate.

2. THE SENSORSCOPE PROJECT

Although environmental monitoring is getting a lot of attention nowadays, only few efforts have been put into improving current data collection techniques. Most measuring campaigns are indeed based on lofty sensing stations, and their effectiveness is limited due to numerous drawbacks:

- **High cost:** They are expensive (*e.g.*, €60 000 — $ 86 000 — for a high-precision station), allowing only for the deployment of a few stations.

- **Reduced flexibility:** They are large and heavy, preventing them from being easily moved to different locations on the monitored area.

- **Inefficient data storage:** They commonly use embedded data loggers that require manual downloading, resulting in limited storage and no immediate feedback.

Measure	Sensor	Range	
Air humidity	Sensirion SHT75	$0-100\,\%$	$\pm\,2\,\%$
Air temperature	Sensirion SHT75	$-20-60°C$	$\pm\,0.3°C$
Precipitation	Davis Rain Collector	$0-\infty\,mm$	$\pm\,10\%$
Soil moisture	ECH20 EC-5	$0-55\,\%$	$\pm\,10\%$
Solar radiation	Davis	$0-1500\,W/m^2$	$\pm\,5\%$
Surface temperature	Zytemp TN901	$-20-70°C$	$\pm\,0.6°C$
Water content	Irrometer Watermark	$-200-0\,kPa$	unknown
Wind direction	Davis Anemometer	$0-360°$	$\pm\,7°$
Wind speed	Davis Anemometer	$1.5-79\,m/s$	$\pm\,5\%$

Table 1. Environmental values gathered by a SensorScope station.

As a result, there is a lack of appropriately dense spatial and temporal observations, which prevents environmental researchers from providing accurate climate models. From this perspective, it is highly relevant to rely on a WSN, which allows to gather data at the required density. Moreover, thanks to wireless communications, measures may be automatically transmitted to a database server, allowing for real-time (*e.g.*, storms) as well as long-term (*e.g.*, ice melting) monitoring. SensorScope provides such a new generation of environmental monitoring system centered around a WSN.

2.1. The Sensing Stations

Stations include a TinyNode sensor mote[1] and 7 external sensors that gather a total of 9 different environmental measures. The set of these sensors, provided in Table 1, has been carefully chosen with the help of environmental researchers to especially target hydrology and micro-climatology domains. To ensure the quality of the measurements, all sensors are calibrated before deployment. In a first step, they are tested in the laboratory, and in a second step their readings are compared to reference outdoor sensors over several days. We require the correlation coefficient obtained for the measured values to be higher than 0.98. The price of a station, including all sensors, is around € 900 ($ 1280).

To get a sufficient autonomy for long-term outdoor operations, stations also include an energy management module based on solar energy. It is composed of a solar panel and two batteries: the first one is primarily used to power the station and is charged directly thanks to the solar power, while the second one is used as a backup buffer in case of a failure of the primary battery. Through the various campaigns (see Table 2), we proved that this system is reliable. These deployments indeed lasted up to 6 months during which all batteries were always fully charged, even in case of cloudy weather. This theoretically makes the batteries' recharge cycle-count the only limiting factor for long-term deployments.

2.2. Networking

Using a multi-hop WSN makes it possible for SensorScope to gather data over a wide area with only one sink and to arbitrarily modify the monitored area by moving/adding/removing stations whenever needed. Since wireless stations are monitoring their network neighborhood, these changes are quickly and automatically incorporated without the need to reconfigure the network. A station may also fail without impacting on data gathering: if it was indeed part of a route to the sink, a new route will automatically be created and used to replace the deprecated one.

Besides delivering gathered data to the sink, the network is also responsible for time synchronization. To allow for useful exploitation of data, each measure must indeed be timestamped, and since the stations are subject to a substantial *time drift* (crystals have a correct but not very high precision), it is needed to regularly synchronize the stations. In SensorScope, the current time is simply regularly propagated from the sink to the network by mean of multi-hopping. The WSN used is also very energy-efficient and is able to turn off the radio (which is the biggest energy consumer in the system) most of the time, without impacting on data gathering. Thanks to this mechanism and the solar energy system, stations should theoretically never undergo a power outage.

2.3. Deployments

Over the last 15 months, we have run 6 outdoor deployments, ranging in size from 6 to 97 stations, from the EPFL campus to high mountains. Table 2 provides details about them. During these campaigns, we gathered hundreds of megabytes of environmental data which are freely available for download on our website[2].

One of the most challenging deployments occurred on the rock glacier located at 2 500 m (8 200 ft) on Le Génépi, in Switzerland. This site was chosen for always being the source of dangerous mud streams during intense rain periods, which caused several victims in the last decade. The authorities in charge did not have any measures of rain in that site, and asked us to deploy SensorScope there. Gathered data during this campaign allowed us to model a particular micro-climate that plays an important part in the mud streams, and that will help in flood monitoring and prediction on the site.

3. CHALLENGES IN SIGNAL PROCESSING

We can see from Table 2 that the deployments of the SensorScope project have generated a large amount of environmental data. How to efficiently process the acquired data to allow for useful exploitation has now become an increasingly important issue. In this section, we present a list of challeng-

---

[1] http://www.tinynode.com

[2] http://sensorscope.epfl.ch

Place	Dates	Size	Data	Characteristics
Campus of EPFL	2006-11-01 – 2007-05-15	97 stations	17 GB	Urban environment
Plaine Morte	2007-03-12 – 2007-03-16	13 stations	32 MB	Alpine glacier
Campus of EPFL	2007-07-18 – n/a	10 stations	48 MB	Urban rooftops
Morges	2007-08-03 – 2007-09-04	6 stations	32 MB	Border of a river
Le Génépi	2007-08-27 – 2007-11-05	16 stations	247 MB	Rocky glacier in high mountain
Grand St Bernard	2007-09-13 – 2007-10-26	23 stations	303 MB	Pass in high mountain

Table 2. All SensorScope deployments, performed since the beginning of the project.
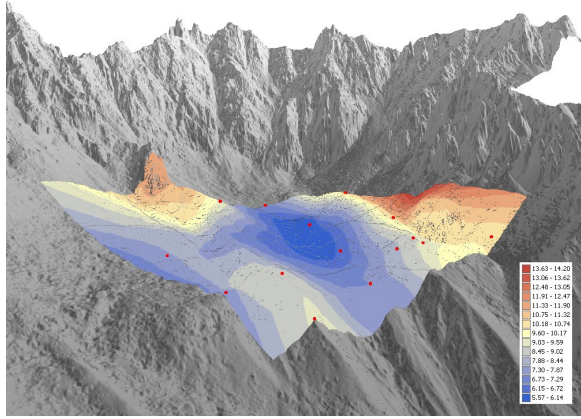

Fig. 1. Data sampling and interpolation example during the Genepi deployment: air temperature snapshot over the glacier.

ing and relevant signal processing tasks that arise from the SensorScope project for environment monitoring.

3.1. Sampling and Faithful Reconstruction of the Physical Phenomena

The sensor network for environmental monitoring can be viewed as a spatial-temporal sampling device for some physical phenomenon of interest (*e.g.*, air temperature or wind speed). While the physical process itself is a function $f(\boldsymbol{x}, t)$ of the continuous spatial variables $\boldsymbol{x} \in \mathbb{R}^3$ and the time variable t, the measurements we get from the sensor network are only samples ("snapshots") of the continuous phenomenon at discrete spatial locations $\{\boldsymbol{x}_k\}_{k=1}^K$ and time instances $\{mT\}_{m \in \mathbb{Z}}$ (see Fig. 1). An important question is, of course, if the samples $f(\boldsymbol{x}_k, mT)$ are a faithful representation of the original function $f(\boldsymbol{x}, t)$. If so, how can we reconstruct $f(\boldsymbol{x}, t)$ from $f(\boldsymbol{x}_k, mT)$? If not, what approximations can we get based on the samples $f(\boldsymbol{x}_k, mT)$.

As a fundamental and ubiquitous issue in signal processing, the sampling and reconstruction process has been extensively studied in the past, with prominent results represented by the Kotelnikov-Shannon-Whittaker sampling theorem and its various extensions (see, for example, [2] and the references therein). In the classical setting, the signals to be sampled are assumed to be bandlimited in the frequency domain. For signals that are not strictly bandlimited (as is often the case in practice), one usually needs to apply a lowpass antialiasing filter on the signals before sampling. This practice leads to reconstructed signals that are the optimal L^2 approximations of the original signals.

However, the distributed setting of the sensor networks poses new difficulties that are not well-handled by the classical sampling approach. Though we can easily apply an antialiasing filter along the temporal dimension of the signals, it is physically infeasible for us to apply a spatial filter before sampling, since the sensors can only observe the local value of the physical phenomenon. This lack of spatial filtering determines that spatial aliasing will be a key issue in the sampling process of the sensor network. How to recover or get a good approximation of the original signals in the presence of spatial aliasing becomes a challenging sampling problem that needs to be carefully addressed.

3.2. Data Integrity in the Sensor Network Measurement

An attractive feature of employing a WSN (such as the SensorScope) in environmental monitoring applications is that we can deploy a large number of inexpensive and lightweight sensing stations in the area of interest and conduct unprecedently dense observations of the physical processes. However, as a price to pay for using simple sensing devices, the measurements we get are not as reliable and accurate as what we would get from the traditional expensive and heavy whether stations. Consequently, it is desirable and sometimes necessary to preprocess and improve the quality of the raw data from the sensor networks before we present them to environmental researchers. In particular, the related problems we need to address include the following:

- **Sensor calibration**: Although all sensors have been pre-calibrated before deployment, their readings are still subject to drift during long-term operations. Note that manual calibrations often become impractical if the deployment is in remote areas. In these cases, how to perform automatic sensor calibration from the gath-

ered data becomes a challenging problem. As shown in [3], the feasibility of automatic calibration relies on a good understanding of the correlations between different measurements.

- **Outlier detection**: Due to various reasons, some sensors may fail; and worse yet, the failed sensors might not always be able to report this situation to the central station. It is therefore important to detect outliers in the sensor measurement data in order to improve the robustness of the overall system.

- **Denoising**: Noise is inherent in any data acquisition process, and hence denoising is a key step in improving the quality of the sensor network measurements.

4. HOW PHYSICS MAY HELP — AN EXAMPLE

We believe the successful resolution of the signal processing tasks listed in Section 3 depends on, among other things, a better understanding of the underlying physical processes. Although each sensor has low precisions and might be unreliable, we have a large number of these sensors observing the same physical reality, and therefore their measurements should be correlated. In this section, we present a simple example to show how signal processing algorithms can benefit from exploiting the physical correlations between different sensor measurements.

We know from the discussion in Section 2 that each sensor station SensorScope is equipped with multiple sensors capable of measuring different physical quantities. Suppose we are measuring N different quantities[3] $f_1(t), f_2(t), \ldots, f_N(t)$ (*e.g.*, air temperature, moisture, etc.). Since these quantities are different manifestations of the same physical reality, they are often linked by some given physical law. In the simplest case, we assume these functions satisfy a system of linear differential equations with constant coefficients. In the frequency domain, this constraint can be written as

$$\boldsymbol{A}(\omega) \left(\widehat{f_1}(\omega), \widehat{f_2}(\omega), \ldots, \widehat{f_N}(\omega) \right)^T = \boldsymbol{0}, \qquad (1)$$

where $\widehat{f_i}(\omega)$ is the Fourier transform of the ith function $f_i(t)$, and $\boldsymbol{A}(\omega)$ is a known M-by-N matrix of functions of ω.

If we assume that each function is bandlimited to $[-\sigma, \sigma]$ for some $\sigma > 0$, but only take samples at a fraction of the corresponding Nyquist rate ($1/T_0 = \sigma/\pi$), can we still perfectly reconstruct the original signals? The following proposition gives a positive answer, in which the physical correlations given in (1) plays a key role.

Proposition 1 *The N functions $f_1(t), f_2(t), \ldots, f_N(t)$ can be perfect reconstructed from their sample values $\{f_i(nKT_0)\}$*

---
[3]For simplicity of notation, we only consider 1-D functions.

($1 \leq i \leq N, n \in \mathbb{Z}$) if and only if the following matrix

$$\boldsymbol{V}(\omega) \stackrel{def}{=} \begin{pmatrix} \boldsymbol{I}_N & \boldsymbol{I}_N & \cdots & \boldsymbol{I}_N \\ \boldsymbol{A}(\omega) & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}(\omega + c) & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{A}\left(\omega + (K-1)c\right) \end{pmatrix}$$

has full column rank for every $\omega \in [-\sigma, -\sigma + c]$, where $\boldsymbol{I}_N$ is the N-by-N identity matrix, K is some positive integer representing the undersampling factor, and $c = 2\sigma/K$.

Since $\boldsymbol{V}(\omega)$ is a matrix with $N + MK$ rows and NK columns, an immediate consequence of the full column rank condition in Proposition 1 is the following upper bound on the undersampling factor:

Corollary 1 *To perfectly reconstruct the N functions $f_1(t)$, $f_2(t)$, ..., $f_N(t)$, one must have*

$$K \leq \frac{N}{N - M}, \qquad (2)$$

where M is the number of rows of $\boldsymbol{A}(\omega)$.

Note that the upper bound in inequality (2) is an increasing function of M (*i.e.*, the number of equations provided by the physical model). This result is fairly intuitive, as it shows that the more prior information we have about the physical correlations between the signals, the less samples we need to take while still being able to fully reconstruct the original signals.

5. CONCLUSION

In this paper, we presented a list of challenging and relevant signal processing tasks that arise from the SensorScope project, which is a WSN-based system with applications in environmental monitoring. We believe that the successful resolution of these signal processing problems can greatly benefit from a better understanding of the underlying physical processes. One example is to exploit the correlation between different physical quantities measured by the sensor stations.

6. REFERENCES

[1] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey," *Computer Networks*, vol. 38, pp. 393–422, 2002.

[2] M. Unser, "Sampling - 50 years after Shannon," *Proc. IEEE*, vol. 88, no. 4, pp. 569–587, Apr. 2000.

[3] L. Balzano and R. Nowak, "Blind calibration of sensor networks," in *Proceedings of the International Conference on Information Processing in Sensor Networks (IPSN)*, Apr. 2007.