The HKUST Frog Pond - A Case Study of Sensory Data Analysis

Wenwei Xue, Bingsheng He, Hejun Wu, and Qiong Luo

Department of Computer Science
The Hong Kong University of Science and Technology
Clear Water Bay, Kowloon
Hong Kong, China
{wwxue, saven, whjnn, luo}@cs.ust.hk

Abstract. Many sensor network applications are data-centric, and data analysis plays an important role in these applications. However, it is a challenging task to find out what specific problems and requirements sensory data analysis will face, because these applications are tightly embedded in the physical world and the sensory data reflect the physical phenomena being monitored. In this paper, we propose to use field studies as an alternative for identifying these problems and requirements. Specifically, we deployed an experimental sensor network for monitoring the frog pond in our university and analyzed the collected sensory data. We present our methodology of sensory data collection and analysis. We also discuss preliminary analytical results from the collected sensory data, together with our generalization for similar sensor network applications. We find that this case study helped us identify and understand several problems, either general or specific, in real-world sensor network application deployment and sensory data analysis.

1 Introduction

Sensor network applications pose a number of novel problems for networking ([4][9][10]) and data management ([7][15]). Nevertheless, many more problems and requirements in real-world sensor network applications are to be identified and understood, especially for sensory data analysis. Due to the tight integration of these applications with the physical world, field studies are effective, sometimes necessary, for identifying problems and requirements. In this paper, we present a case study of sensory data analysis for a small-scale real-world sensor network application. Our goal is to identify and understand problems and requirements specifically for sensory data analysis.

From our case study, we observe that most of the problems in our sensory data analysis rose because the sensor network application was deeply embedded in the physical environment and the sensory data reflected the physical phenomena under study. For instance, we find that even though there were inherent trends in the readings of individual sensors as well as strong correlations between readings of multiple sensors, outliers were common and the causes of some outliers were hard to determine.

The remainder of the paper is organized as follows. Section 2 introduces the deployment of the case study. Section 3 presents our preliminary analytical results of the collected sensory data and discusses the generalization of our experience. Section 4 compares related work and Section 5 concludes the paper.

2 Deployment of the Case Study

We started the case study around a frog pond on the HKUST campus in April 2004. The frog pond is located at the northeastern corner of the campus and is surrounded by two pagodas and various plants. Throughout the late spring, the frogs in the pond croak loudly all day long.

The smart sensor nodes we used in the case study were the Crossbow MICA2 motes [3]. Each mote consists of an Atmel Atmega128L low-power micro-controller running TinyOS [13] with a 900MHz radio channel. Mote 0 connects with a PC-grade base station through a PC interface card, and other motes each consists of a MICA2-compatible sensor board. The scale of the case study was small due to our resource limit.

We deployed a total of nine MICA2 motes in two groups, with each group in a pagoda around the frog pond. Fig. 1 shows the deployment of the two groups. The base station (Mote 0) of each group was connected to the serial port of a notebook through a MIB510CA interface board and a serial cable.
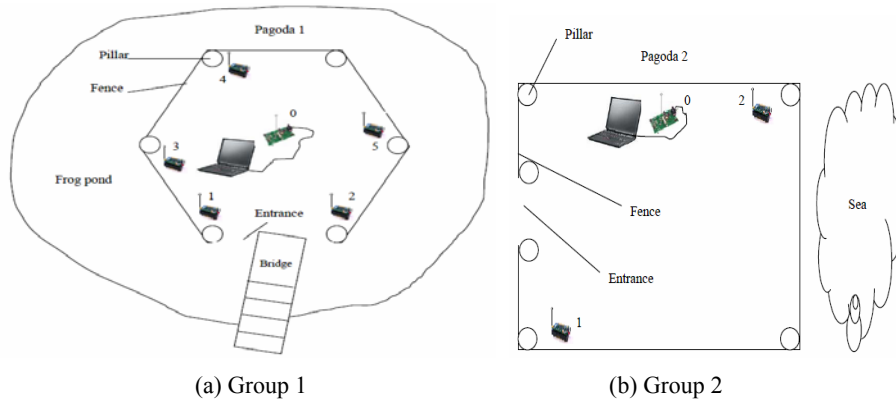

(a) Group 1   (b) Group 2

Fig. 1. Deployment of two groups of Motes

Group 1 was deployed in the pagoda that is surrounded by the frog pond. Each of Motes 1-5 was attached with a MTS310CA sensor board, which includes a temperature sensor, a light sensor, a microphone, a 2-axis accelerometer and a 2-axis magnetometer. We installed TinyDB [12] on the motes and used the TinyDB GUI to collect sensor readings and to log the readings to a text file.

Group 2 was deployed in the pagoda that is near the frog pond and overlooks the sea. Its Motes 1-2 used the MTS420CA weather sensor boards. This type of sensor board consists of a humidity and temperature sensor, a barometric pressure and tem-

perature sensor, an ambient light sensor, a 2-axis accelerometer and a GPS module. We configured the Xlisten program and the corresponding on-mote module XSensorMTS400 downloaded from the TinyOS SourceForge CVS [14] for logging the sensory data from this type of sensors.

It was cloudy with intermittent rain on the day of our data collection. We collected one-day data in four two-hour periods of the day: 6:30 – 8:30 (morning), 12:30 – 14:30 (noon), 17:30 – 19:30 (dusk), and 22:00 – 24:00 (night). We set the sample period to be 30 seconds. At the end of the eight-hour data collection, we logged thousands of sensor readings per group.

We used homegrown programs to pre-process the collected sensory data before conducting further analysis. First, we converted the sensor readings from raw ADC counts generated by the sensor boards to more human-friendly engineering units (e.g., Celsius degrees for temperature) using conversion formulas provided by Crossbow. Next, we parsed and imported the sensor readings (both ADC counts and engineering units) into a Microsoft Access database. Finally, we used SQL queries and Microsoft Excel Charts to perform preliminary data analysis.

3   Sensory Data Analysis

In this section, we analyze the sensory data we collected in the experiment. From a database perspective, our focus of the analysis is on identifying and understanding the problems and requirements that are specific for sensory data. Apparently, the analytical results we present here are preliminary. However, the methodology and insights gained from these initial analytical results are valuable for more advanced analysis and are otherwise unavailable or less convincing without the case study.

3.1   Trends in Readings of Individual Sensors

We first give an analysis of individual sensor readings using some typical examples.
Light. Fig.2 shows the light readings of Mote 1 and Mote 5 in Group 1. We pick these two motes because in our deployment they were the nearest (Mote 1) and farthest (Mote 5) motes from the Group 1 base station. Each point in the figure corresponds to one light reading. In the morning, the light readings kept on increasing due to the sunrise. At noon, the light readings were at the highest for the day and slightly decreased past noon. At dusk, the light readings decreased sharply due to the sunset and then jumped up to a certain level because the lamps around the pagoda were lit. The light readings at night remained almost constant with the lamplight.

The two motes in Fig. 2 had similar readings. The other three motes of Group 1 also had similar readings to these two. This similarity was because the area of a pagoda is small and thus the motes in this group were located near to one another. The proximity of motes also made readings of other sensors (e.g., humidity, temperature) of a group similar.

However, for Group 2, the readings of the ambient light sensors remained a constant value of 131.448624 Lux due to a bug in the XSensorMTS400 program that we

used. As a result, a comparison of light readings between the two groups is not done. We are developing our own data logging programs for future usage.
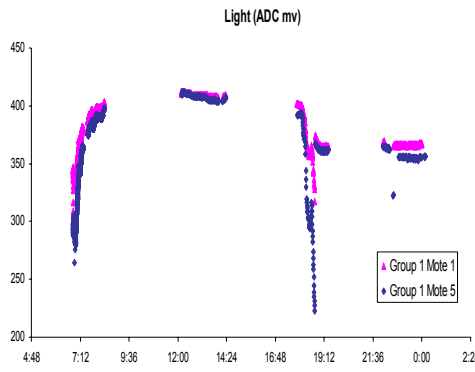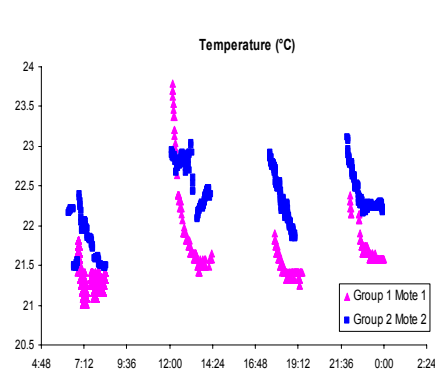

Fig. 2. Light readings of Group 1

Fig. 3. Temperature readings of two groups

Temperature. Fig. 3 shows the temperature readings of Group 1 Mote 1 and Group 2 Mote 2. We put the readings from different groups in one figure for comparison purpose. Again, temperature readings from motes of the same group were similar due to their proximity.

The temperature readings of Group 1 motes varied from 21 to 24°C, whereas those of Group 2 motes varied from 21 to 23°C. The temperature measured by Group 2 was often slightly higher than that measured by Group 1 (except around noontime), even though the two pagodas were within a distance of 20 meters from each other. We think there are two possible reasons: (1) the temperature sensors of the two groups are made by different companies and therefore differ in hardware characteristics, (2) the microclimates in the two pagodas differ due to their different geographical locations.

Humidity. Humidity sensors were available only in Group 2. The humidity readings of the two motes in Group 2 are illustrated in Fig. 4. Most of the time, the readings remained at the level of around 90%.

Note there were some abnormally high humidity readings (larger than 130%) of Mote 1 at the beginning of the morning period. These abnormal readings were because some rain drops splashed onto the Mote by accident. The water made the humidity sensor malfunction and return abnormally high readings. This kind of physical failure is not uncommon and is recoverable [11]. After being dried, the sensor returned to normal operation.

Noise. Microphone sensors were available only in Group 1. Fig. 5 shows the noise readings of Mote 1 and Mote 5 in Group 1. Unlike temperature, light, and humidity readings, which are more continuous, the noise readings are more discrete. The scattered data points in the noise readings in Fig. 5 usually suggest the actual, sudden changes (events) in the sound level in the environment. In comparison, those outlier points in temperature, light, and humidity readings in the previous figures were often due to errors.

From Fig. 5 we see that the frogs croaked most actively in the early morning and were most quiet around noontime. Also, some of the high data points in the figure were because people passing by were talking.

Correlation. Our analysis on correlation of sensor readings is limited and will be an important part of our future work. So far we have found that the temperature and humidity readings were inversely correlated and that the temperature and light readings were not correlated as found in other environments [5].
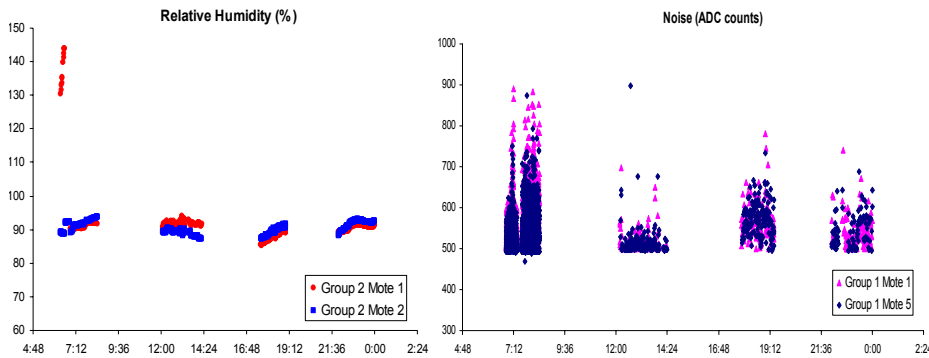

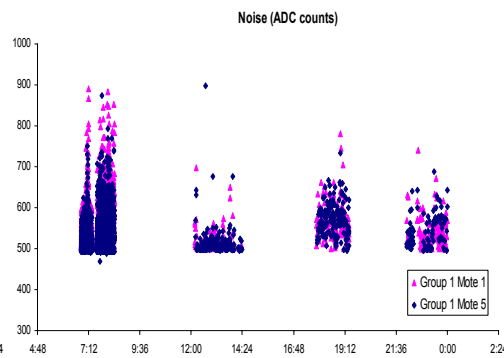**Fig. 4.** Humidity readings of Group 2

Fig. 5. Noise readings of Group 1

3.2 Discussion

Having presented our initial results with a specific sensor network application, we discuss how to generalize our findings to similar sensor network applications.

We start by summarizing the problems that we have encountered in the application deployment and data analysis. First, both hardware problems and software bugs are common in sensor network applications. The reasons include sensor networks being an emerging technology and the typical application environment being the physical world full of unpredictable events and changes. Second, data pre-processing and post-processing constitutes a large amount of work in order to facilitate sensory data analysis. This work mainly includes data cleaning and format conversion prior to analysis and visualization after analysis. Third, sensory data exhibits regularities as well as abnormalities, and the causes of outliers are hard to determine.

Based on this summary of experienced problems, we propose the following three requirements for a sensory data analyzer.

(1) The analyzer should have data acquisition functions that are fault-tolerant and adaptive, since the sensory data collection process determines the quality of sensory data. The fault-tolerance requirement is because hardware malfunctioning is common in field studies, as we have already experienced. It is thus desirable that a data collector is able to recover, to migrate the work from a failed node to a normal node, and to resume the work. The adaptivity requirement is to take advantage of the patterns and regularities captured in sensor readings. For instance, continuous quantities such as

temperature can be measured with a sampling frequency adapted to the changes in the temperature readings in order to improve power efficiency while keeping the quality of sensory data unaffected.

(2) The analyzer should have a set of basic functions for data pre-processing and post-processing operations. Data pre-processing is to further ensure the quality of data for analysis. Data post-processing is mainly for the presentation of analytical results. For example, the function convert() converts sensor readings from raw ADC counts to human-friendly engineering units, the function calibrate() performs hardware-specific calibration of the readings, and the function plot() plots data points and curves together with analytical summaries following user-defined criteria.

(3) As the core of the analyzer, the sensory data analysis functions include pattern and outlier detection, and correlation of multiple sensory attributes or multiple sensor nodes. We further discuss these two kinds of functions as follows.

First, detecting patterns and outliers in single-node single-attribute sensory data is the basic analytical operation. For instance, given the temperature readings of one sensor node, the basic analytical information about these readings must include a summary of the range, the trend, and the outliers of the data. As a result of measuring natural phenomena, sensory data has inherent patterns as well as outliers. Moreover, outliers sometimes are due to real events in the environments and sometimes due to system errors. It is necessary to pay special attention to outlier analysis.

Second, correlation analysis gives more insights into sensory data, especially because each sensor node has multiple sensory attributes and multiple sensor nodes work concurrently in a geographical region. The inherent correlations between natural phenomena as well as the temporal and spatial correlations of sensor nodes will be useful for both sensor network applications and system deployment. For example, when an application is detecting transient changes such as a sudden increase in the noise level, it can utilize the spatial correlation of a cluster of adjacent nodes to detect the noise with a high fidelity. In other words, if one sensor node detects a sudden increase of noise level, it might be a real event as well as a system error. But if multiple nearby nodes report the same event, the probability of a system error is much lower and that of a real event is much higher than reported by a single node.

In summary, we find several problems in sensory data analysis, ranging from hardware or software problems in the deployed applications to difficulties in producing meaningful analytical results out of sensory data. Correspondingly, we propose several requirements for sensory data analysis systems, including fault-tolerance and adaptivity of data collection, a set of data pre-processing and post-processing functions, and basic data analysis functions such as pattern, outlier, and correlation detection. Our ultimate goal is to build a general sensory data analysis system for various data-centric sensor network monitoring applications.

4   Related Work

A number of sensor network projects have real-world deployment, including ALERT [1], GDI ([8][11]), PODS [2], Surveillance and NIMS [6].

ALERT (Automated Local Evaluation in Real Time) is a well-known, practical sensor network application [1]. It provides real time rainfall and water level information for forecast of flooding. ALERT mainly focuses on special-purpose sensory data statistics and uses them for prediction.

Both GDI ([8][11]) and PODS [2] deployed sensor networks in outdoor environments mainly for the purpose of system performance study. Specifically, GDI deployed a multi-tier sensor network for habit monitoring whereas PODS was deployed in Hawaii Volcanoes National Park.

Surveillance and NIMS [6] are two demonstrations. Surveillance built an energy-efficient surveillance system using a wireless sensor network, and NIMS focused on new, mobile sensing devices on a suspended infrastructure.

In comparison, our case study is at a smaller scale and a finer level, with a focus on identifying general problems and requirements for advanced sensory data analysis in real-world applications.

5   Conclusions

In this paper, we describe our case study of deploying a small-scale monitoring application at the frog pond in our university and analyzing the collected sensory data. Our goal is to identify the problems and requirements for sensory data analysis in real-world sensor network applications. We find that (1) data collection and logging functions need to be failure-aware and easy to resume, (2) data pre-processing such as format conversion and post-preprocessing such as visualization is necessary for sensory data analysis, and (3) essential sensory data analysis functions include pattern and outlier detection for readings of individual sensors and correlation detection for readings of multiple sensors.

Our future work includes designing and implementing advanced sensory data analysis tools and conducting larger-scale studies using these tools.

Acknowledgement

References

1.   ALERT Systems Organization Homepage. http://www.alertsystems.org.
2.   Edoardo Biagioni and Kent Bridges. The Application of Remote Sensor Technology to Assist the Recovery of Rare and Endangered Species. In Special issue on Distributed Sensor Networks for the International Journal of High Performance Computing Applications, Vol. 16, No. 3, August 2002.
3.   Crossbow, Inc. Wireless Sensor Network Products. http://www.xbow.com.

4. Deborah Estrin, Ramesh Govindan, John Heidemann, and Satish Kumar. Next Century Challenges: Scalable Coordination in Sensor Networks. MobiCOM 1999.
5. Wei Hong and Samuel Madden. Implementation and Research Issues in Query Processing for Wireless Sensor Networks. MDM 2004 Tutorial.
6. Intel Research – Exploratory Research – Sensor Networks. http://intel.com/research/ exploratory/wireless_sensors.htm.
7. Samuel Madden, Michael J. Franklin, Joseph M. Hellerstein, and Wei Hong. The Design of an Acquisitional Query Processor for Sensor Networks. SIGMOD Conference 2003.
8. Alan Mainwaring, Joseph Polastre, Robert Szewczyk, David Culler, and John Anderson. Wireless Sensor Networks for Habit Monitoring. WSNA 2002.
9. Loren Schwiebert, Sandeep K. S. Gupta, and Jennifer Weinmann. Research Challenges in Wireless Networks of Biomedical Sensors. MobiCOM 2001.
10. David C. Steere, Antonio Baptista, Dylan McNamee, Calton Pu, and Jonathan Walpole. Research Challenges in Environmental Observation and Forecasting Systems. MobiCOM 2000.
11. Robert Szewczyk, Joseph Polastre, Alan Mainwaring, and David Culler. Lessons from a Sensor Network Expedition. EWSN 2004.
12. TinyDB: A Declarative Database for Sensor Networks. http://telegraph.cs.berkeley.edu/ tinydb/.
13. TinyOS. http://www.tinyos.net.
14. SourceForge.net. CVS Repository: TinyOS. http://sourceforge.net/cvs/?group_id=28656.
15. Yong Yao and Johannes Gehrke. Query Processing for Sensor Networks. CIDR 2003.