Urban Data Collectors: A Pragmatic Approach to Leveraging Urban Sensing

Rostand Costa
Digital Video Applications Lab - LAVID
Federal University of Paraíba - UFPB
João Pessoa, Brazil
rostand@lavid.ufpb.br

Raphael Resende
Graduate Studies Programme in Informatics - PPGI
Federal University of Paraíba - UFPB
João Pessoa, Brazil
raphael.resende@hotmail.com

Stenio Fernandes
Center of Informatics - CIn
Federal University of Pernambuco - UFPE
Recife, Brazil
sflf@cin.ufpe.br

Abstract— **This paper proposes the adoption of a more pragmatic approach to urban sensing to leverage the initial offering of open databases within the context of smart cities. In this sense, the use of mobility agents already available to provide mobility to traditional sensors can enable a preliminary instrumentation of cities more quickly and possibly at a lower cost. As proof of concept, we investigated the use of metropolitan bus equipped with sensors such as urban data collectors. We believe that this strategy can help overcoming many challenges of traditional approaches for instrumentation of cities, such as power supply, maintenance, operation, and protection of a large amount of static sensors. To assess the feasibility of sampling by collectors, we conducted experiments comparing averages using frequent readings vs averages using intermittent readings collected by sensors of a climatological micro station.**

Smart cities; IoT; wireless sensor networks; mobile sensing; urban instrumentatio.

I. INTRODUCTION

Ubiquitous smart environments equipped with low cost and easy implementation sensors and sensor networks open up a number of possibilities for large scale urban monitoring. From the point of view of deploying services and applications for smart cities, for example, it is consensus the extreme importance of building open repositories of sensor data, collected in real time from different sources [1]. Such repositories could support innovative applications for several scenarios, such as environmental monitoring, security, health, education, and urban mobility. Continuous data flow is able to populate such repositories for strategic real-time analysis or subsequent mining of large volumes of correlated data. [2].

However, data collection of urban monitoring of a large number of sensors is a challenge due to current technological limitations, in particular in the distribution and maintenance of sensors in large metropolitan areas. Initiatives to integrate emerging technologies such as *Internet of Things* (IoT), *Wireless Sensor Network* (WSN) and *Mobile Ad Hoc Network* (MANET) [3][4], have great potential to offer support for a large-scale environment for urban data sensing. This can lead to novel platforms for real-time analysis applications in strategic scenarios, defined by municipal administrators in several areas of interest [5][6].

There exist perceptions that the use of mobile sensors for data collection may be a key component to overcome the difficulties of deploying large-scale sensing in cities. [7]. In such an approach, the sensors collect data while moving, and distribute them at some specific collection points (e.g., in infrastructured WSNs) or between their peers (e.g., in *ad hoc* WSNs) [8]. Alternatively, mobile sensors can be directly connected in a distribution network with comprehensive coverage, such as the ones based on WiMax [9], cellular [10] (e.g. GPRS or HSDPA) or satellite technologies. Researches in this field envisage that using mobile sensors can significantly reduce the need for distribution of static sensors in order to cover most city areas. [7].

We advocate the use of existing infrastructure to provide power, shelter, and mobility for mobile sensors. For example, the deployment of specific sensing devices in the metropolitan public transport fleets is a turnkey solution. This is because, in the case of public transportation, buses, trams, and light rail trains often circulate on the same route. Therefore, this approach guarantees a periodic gathering of samples in a number of spots in a large area. It also allows an initial wider city coverage. Thus, the temporal and spatial granularity of the data to be collected can be directly related to the distribution strategy of mobile sensors in the urban transportation mesh.

As a preliminary feasibility study, we perform an experimental analysis to assess possible losses of precision of both frequent readings and intermittent readings, realized by sensors of environmental metrics. Using a micro weather station [11], we assessed the difference of the air average composition (CO and NO2), temperature, light intensity, sound levels, and humidity. Data were obtained using different strategies of aggregation in order to evaluate if the use of less frequent readings could be a viable strategy.

Initial results indicate that the use of more sporadic sampling of environmental measures do not imply significant losses in accuracy and may represent a promising alternative for the collection of urban data.

The remainder of the paper is organized as follows. Section II provides further detail on our proposed approach. Section III describes he methodology and the objectives of our preliminary experimental analysis. Section IV provides the analyses and the obtained results and Section V brings our final considerations.

II. A PRAGMATIC APPROACH FOR LARGE SCALE SENSING

Open data platforms are a key component to the context of Smart Cities. Among other advantages, an open platform helps:

i) Improving the scalability aspects in the collection and distribution of data;
ii) Transforming, aggregating, and pre-processing such data to facilitate strategic studies of urban management; and
iii) Developing a service layer to support the creation of a number of urban data applications.

We advocate the adoption of a more pragmatic approach to the development of large-scale urban sensing platforms. Considering different types of data, measurement, and implementation techniques, our intention is to explore strategies that involve less cost and less time for the provision of urban data on open platforms.

In this sense, public transportation systems (e.g., buses, trams, and taxis cabs) could be used as mobility agents in order to provide an alternative to collecting data in motion. A clear reason to use such an approach is because, in the case of public transportation systems, buses often circulate in the same route, generating several periodic samples of a large area. In addition, there would be negligible costs to turn the buses into mobile agents.

We call Urban Data Collectors the combination of mobility agent and traditional sensors, in well-defined routes derived from the public transportation system. Bus stations and stops, in turn, can host sink nodes, thus emulating an infrastructured WSN and enabling the delivery of samples collected at each bus round.

The potential for urban information that can be monitored and added to the city's open data repository is quite wide. The sensing may include GPS, microphone, gyroscope, accelerometer, detector of carbon dioxide (CO2), temperature, humidity, luminosity, and the like. Such viability is because deployment issues (e.g., power supply, shelter, and safety of the devices) are strongly minimized by allowing the sensors to be installed inside the bus.

As an incentive for active participation for large-scale deployment of mobile sensors, the partners involved can be directly benefited, since specific applications can be made available to them.

In summary, we believe that a more pragmatic approach is able to consolidate part of the scientific and technological advances in the areas of wireless communications, ubiquitous and pervasive computing, wireless sensor networks, big data and cloud computing, among others, in platforms that can support the emergence and development of strategic services for smart cities.

In the following sections, we present how we designed and carried out an initial experiment to investigate an important aspect of our approach: the feasibility of data collection using a system of noncontinuous sampling.

III. EXPERIMENTAL DESIGN

In order to make a preliminary assessment of the feasibility of the proposed approach, we designed and performed a preliminary experimental study. The main objective of our experiment was to determine potential loss of precision in the computation of physical environmental quantities as the sampling interval increases.

For the data measurements, we used a meteorological micro station manufactured by Acrobotics Industries [7], which is part of an open-source kit developed at Fab Lab Barcelona for monitoring the environment [12]. The hardware consists of two printed-circuit boards, namely: 1) an interchangeable daughterboard or shield and 2) an Arduino-compatible, data-processing board. The shield is called *Ambient Board* and it has six sensors: air composition (CO, NO2), temperature, light intensity, sound levels, and humidity.

A. Metodology

We collected data from all physical quantities at each minute, which is the lowest time interval supported by such meteorological micro station. The Ambient Board was kept in a stationary state, in a same indoor environment, during the sampling period. The data collected by the micro station were then retrieved and consolidated into averages for each hour and for each of the observed environmental physical quantities. To simulate several different frequencies of collection, we aggregated the data and calculate the averages at each hour into four groups:

- $\overline{1M}$**:** Sampling every ONE minute;
- $\overline{5M}$**:** Sampling every FIVE minutes;
- $\overline{10M}$**:** Sampling every TEN minutes;
- $\overline{30M}$**:** Sampling every THIRTY minutes.

Using the average $\overline{1M}$ as the baseline, the percentage change with respect to the equivalent average $\overline{5M}$, $\overline{10M}$, and $\overline{30M}$ for each hour was calculated.

IV. RESULTS AND ANALYSIS

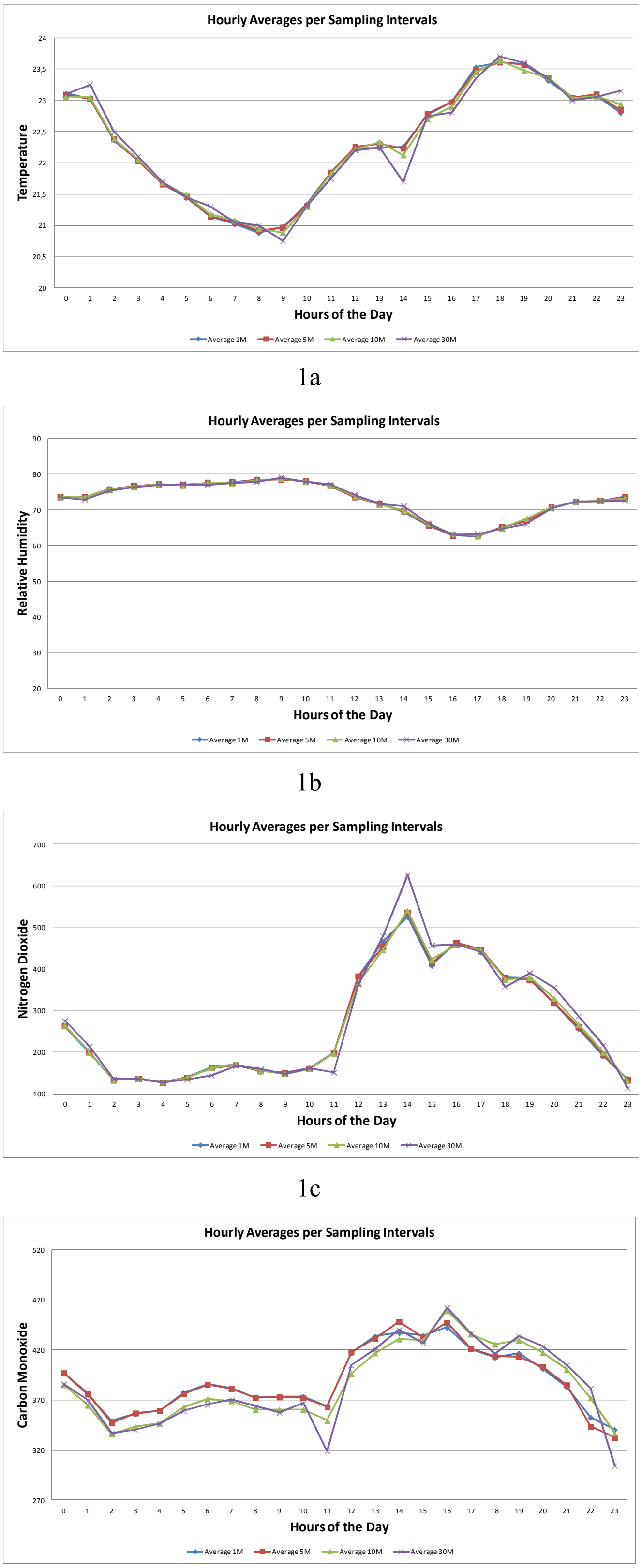
The results presented in this section refer to data collected by a meteorological micro station installed in João Pessoa, Brazil, during two months of 2014 (June and July), reaching nearly half a million individual samples.

In this study, we calculated the averages using groups described in section 3. The various averages obtained for each hour of the day were superimposed to allow a better view on how the behavior of the quantity curve on the day was captured by each grouping strategy.

As can be seen, on this particular day there was no significant variation of hourly averages obtained of temperature, relative humidity and nitrogen dioxide, whose graphs show almost coincident curves (Figures 1a, 1b and 1c).

This overlap is not as evident as in the case of carbon monoxide (Figure 1d), which presents a wider difference between the averages, especially in the case of $\overline{10\text{M}}$ and $\overline{30\text{M}}$. However, a condition to be observed is that even with a slight increase in the value of the obtained averages, the general trend of the curve of quantity remained the same.

A more accurate representation of the variation of average was calculated using the metrics described in Section 3. In summary, there were obtained the percentage change of average $\overline{5\text{M}}$, $\overline{10\text{M}}$ and $\overline{30\text{M}}$ in relation to the average $\overline{1\text{M}}$ of the temperature, humidity, carbon monoxide, and nitrogen dioxide for each hour in the same. In percentage terms, the variation of average $\overline{5\text{M}}$ and $\overline{10\text{M}}$ with respect to the average values $\overline{1\text{M}}$ for the considered day is relatively low, with differences in measurement accuracy of up to 1% in general, with peaks of up to 5% for some of the quantities in the case of the average $\overline{10\text{M}}$. This behavior also holds for the average $\overline{30\text{M}}$ in the case of quantities such as temperature and humidity. However, in the case of nitrogen dioxide and carbon monoxide, there are most significant peaks of variation of 10% and 25%, respectively.



1a

1b

1c

1d

Figure 1: *Hourly averages of temperature, humidity, carbon monoxide and nitrogen on a given day calculated with different sampling intervals.*

To map the behavior of the variation in a more comprehensive manner, the metrics of interest of all physical quantities considered were calculated for every hour of every day of the observation period and the results are summarized in Table 1. The values represent the mean total percentage changes of the metrics calculated throughout the observation period and were obtained with a confidence level of 95%. The confidence interval for each average is indicated in the table.

Table 1: Percentage average variation of the interest metrics calculated for the observation interval.

Quantity	**Metric**	**Average Percentage Variation**	**Confidence Interval (CI)**
Temperature	$T(d,h)\frac{5M}{1M}$	*0.22%*	± 0.00116123
	$T(d,h)\frac{10M}{1M}$	*0.34%*	± 0.0011931
	$T(d,h)\frac{30M}{1M}$	*0,59%*	± 0.0018738
Relative humidity	$RH(d,h)\frac{5M}{1M}$	*0.18%*	± 0.00073321
	$RH(d,h)\frac{10M}{1M}$	*0.29%*	± 0.0008929
	$RH(d,h)\frac{30M}{1M}$	*0.64%*	± 0.0023285
Carbon Monoxide	$CM(d,h)\frac{5M}{1M}$	*0.73%*	± 0.0023856
	$CM(d,h)\frac{10M}{1M}$	*2.44%*	± 0.0048109
	$CM(d,h)\frac{30M}{1M}$	*3.18%*	± 0.0077047
Nitrogen Dioxide	$ND(d,h)\frac{5M}{1M}$	*2.58%*	± 0.0180435
	$ND(d,h)\frac{10M}{1M}$	*3.88%*	± 0.0205051
	$ND(d,h)\frac{30M}{1M}$	*7.97%*	± 0.0255651
Noise Level	$AN(d,h)\frac{5M}{1M}$	*1.42%*	± 0.0045797
	$AN(d,h)\frac{10M}{1M}$	*2.38%*	± 0.0070871
	$AN(d,h)\frac{30M}{1M}$	*3.94%*	± 0.0165981
Luminosity Level	$AL(d,h)\frac{5M}{1M}$	*4.02%*	± 0.0278655
	$AL(d,h)\frac{10M}{1M}$	*6.05%*	± 0.0342902
	$AL(d,h)\frac{30M}{1M}$	*13.47%*	± 0.0468787

The percentage variation between the different strategies for grouping samples remained very low (below 1%) for all averages of temperature and relative humidity. For the case of carbon monoxide and noise level, the average variation is slightly larger but remaining below 4%. This maximum level of 4% of variation is also observed for average $\overline{5M}$ and $\overline{10M}$ of nitrogen dioxide and the average $\overline{5M}$ of luminosity level.

Such results, which are shown graphically in Figure 2, are encouraging and motivate us to continue investigating the proposed approach. Considering that a single sensor can make up to 60 readings in one hour and considering the use of the strategy of grouping (30M), it is possible to measure in 30 different geographical locations with same equipment, when moving, without significant loss of accuracy.

The use of low cost equipment along with an appropriate strategy for the distribution of sensors in public bus fleet may facilitate large-scale urban instrumentation.

We argue that fewer accurate (and maybe static) weather stations, which are more expensive, could be deployed along with the low cost ones, in order to monitor critical strategic areas.

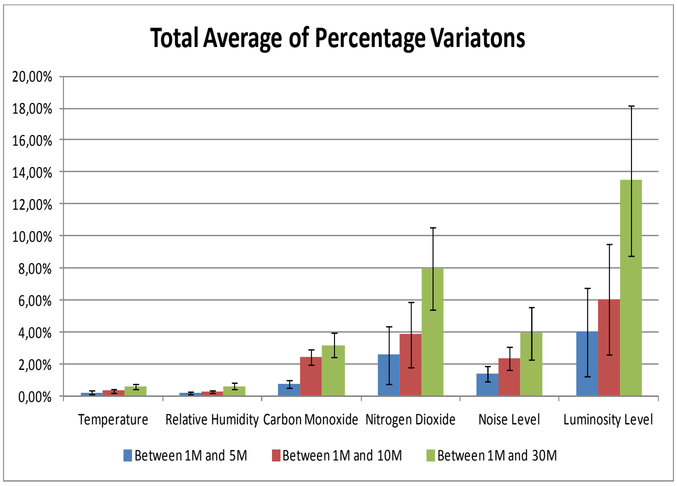


Figure 2: Percentage change average total of all interest metrics with confidence intervals (CI).

Furthermore, a most significant difference in the percent variation was noticed in the average $\overline{10M}$ and $\overline{30M}$ of luminosity level with 6.05% and 13.47% variation respectively and also in the average $\overline{30M}$ of dioxide nitrogen, with a total average variation of 7.97%.

Analyzing the individual samples of two quantities, we observed that there were significant variations in occasional moments that pushed the total averages. We intend to have an in-depth investigation of these phenomena in future experiments involving the same quantities. We aim at understanding if the observed trend of samples with atypical variation represents a normal and expected behavior of quantity, or its reflect failure sensor readings or circumstantial events in the sampling location during the observation interval.

In this sense, we intend to expand the study here considering the basis of environmental samples collected by sensors in a collaborative project [12], which will allow evaluating the dynamic behavior of almost 350 measurements points distributed worldwide.

V. CONCLUSION

This paper presents an approach for deploying a platform for urban sensing based on the use of traditional sensors through the use of mobility agents. Applicable in different contexts and from the perspective of different types of measurement data, our proposal has as main objective address some of the scalability aspects of the collection and distribution of urban data in order to facilitate the offering of open data repositories for the cities.

We performed some experiments to investigate the feasibility of data collection using a system of infrequent sampling, which is an important aspect of our approach. Initial results indicate that the use of more sporadic sampling of environmental measures do not imply significant losses in accuracy and may represent a promising alternative for the collection of urban data.

For the next steps of our research, we are, at the time, adapting meteorological micro stations to incorporate GPS modules and allow sampling of quantities including georeferenced info. The goal of our future experiments is to realize the sampling in a mobile form, obeying predetermined routes. Thus, the frequency of the collection at a same point (or region) will be determined by the frequency in which a sensor can pass over it, which is a function of path length and the number of sensors used simultaneously. The grouping of samples for calculating the average of each point will be done using the concept of spatial grouping ("spatial clustering") [13] in order to accommodate any differences from the point of reading caused by speed variation of vehicles in each round. We also intend to make comparisons by sampling between mobile readings and readings taken with stationary sensors to assess if there is a significant loss in accuracy.

We believe that pragmatic strategies as Urban Data Collectors may allow the initial offering of services in an extensive range of application domains based on the availability of data collection of relatively low-cost sensing platforms. Despite the numerous technical challenges involved, our feeling is that the ecosystem of science and technology in many cities already fulfills the potential to build and operate such platforms, opening up the possibility of offering numerous large-scale digital services, context-sensitive, for citizens as much as for municipal managers.

REFERENCES

[1] Jong Hoon Ahnn; Uichin Lee; Hyun Jin Moon, "GeoServ: A Distributed Urban Sensing Platform," Cluster, Cloud and Grid Computing (CCGrid), 2011 11th IEEE/ACM International Symposium May 2011.

[2] Roche, S., Rajabifard, A. "Sensing places' life to make city smarter," In*Proceedings of the ACM SIGKDD International Workshop on Urban Computing* (UrbComp '12). ACM, New York, NY, USA, 41-46. 2012.

[3] Zanella, A; Bui, N.; Castellani, A; Vangelista, L.; Zorzi, M., "Internet of Things for Smart Cities," Internet of Things Journal, IEEE , Feb. 2014

[4] Bellavista, P.; Cardone, G.; et al., "Convergence of MANET and WSN in IoT Urban Scenarios," Sensors Journal, IEEE , Oct. 2013.

[5] Cranshaw, J. "Whose 'City of Tomorrow' is It? On Urban Computing, Utopianism, and Ethics," UrbComp 2013. 2013.

[6] Barone, R.E.; Giuffrè, T.; Siniscalchi, S.M.; Morgano, M.A; Tesoriere, G., "Architecture for parking management in smart cities," Intelligent Transport Systems, August 2014

[7] Cardone, G.; Bellavista, P.; Corradi, A; Foschini, L., "Effective collaborative monitoring in smart cities: Converging MANET and WSN for fast data collection," Kaleidoscope 2011: (K-2011), Proceedings of ITU , vol., no., pp.1,8, 12-14 Dec. 2011.

[8] Akyildiz, IF.; Weilian Su; et al.., "A survey on sensor networks," Communications Magazine, IEEE , Aug 2002

[9] Pareit, D.; Lannoo, B.; Moerman, I.; Demeester, P., "The History of WiMAX: A Complete Survey of the Evolution in Certification and Standardization for IEEE 802.16 and WiMAX," Communications Surveys & Tutorials, IEEE , Fourth Quarter 2012

[10] http://www.3gpp.org/technologies/keywords-acronyms/99-hspa

[11] Acrobotic Industries LLC. "Smart Citizen Kit," www.acrobotic.com. September 2014.

[12] Fab Lab Barcelona."Smart Citizen Project," www.smartcitizen.me. September 2014.

[13] Cao, Z., Wang, S., Forestier, G., et al. "Analyzing the composition of cities using spatial clustering," 2nd ACM SIGKDD International Workshop on Urban Computing (UrbComp '13). ACM, 2013.