Paper Reference No.: PN-202

Title: An Integrated Geospatial and Visualization Platform for Solar Radiation Mapping

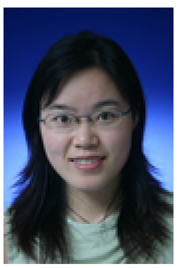
Authors: Yuxia Yao, Hock Beng Lim, Chee Keong Chan, Fook Hoong Choo

Name of the Presenter: Yao Yuxia

Mailing Address:
Intelligent Systems Centre
Nanyang Technological University
Research Techno Plaza, BorderX Block, Level 7
50 Nanyang Drive, Singapore 637553

Tel: +65-67906678
Fax: +65-63164480

Biography:



Yuxia Yao received the BSc degree in Computer Science from Huazhong University of Science and Technology, Wuhan, China in 2003, and the PhD degree from the School of Computer Engineering at Nanyang Technological University, Singapore, in 2008. She is currently a research fellow in the Intelligent Systems Center, Nanyang Technological University. Her research interests include mobile computing, data management in wireless sensor networks and spatial databases.

Hock Beng Lim is program director of the Intelligent Systems Center at Nanyang Technological University, Singapore. He received his BS in Computer Engineering, MS in Electrical Engineering, and PhD in Electrical and Computer Engineering from the University of Illinois at Urbana-Champaign, and his MS in Management Science and Engineering from Stanford University. His research interests include sensor networks and sensor grids, cyber-physical systems, ambient intelligence, parallel and distributed computing, wireless and mobile networks, computer architecture, embedded systems, performance evaluation, e-Science and high-performance computing.

Dr Chan Chee Keong received his BEng from National University of Singapore in Electrical and Electronic Engineering, MSC and DIC in Computing from Imperial College, University of London and PhD from Nanyang Technological University. Upon graduation from NUS, Dr Chan worked as an R&D Engineer in Philips Singapore for several years. He then joined NTU as a lecturer, was promoted to Senior Lecturer and then to Associate Professor. Currently, he is attached to the Information Engineering Division, lecturing in subjects related to computer systems, artificial intelligence, software engineering and cyber security.

Professor Choo graduated from the University of Leeds and UMIST, UK in 1977 and 1979 respectively. He worked as a System Design Engineer with GEC Electrical Projects in Rugby designing electrical drives for the Mining and General Industries. While at GEC, he also worked on and developed the company's first vector controlled induction machine drives. He joined Lucas Control System in Birmingham as a Project Engineer, working at the Lucas Research Centre, UK in 1983 responsible for the development of ac drives for the Machine Tool industries. He joined NTI in 1984 as a Lecturer where he is now an Associate Professor in the School of Electrical and Electronic Engineering, College of Engineering, NTU. His research interests are in the area of Power Electronics, Electrical Drives, Alternative Energy and Energy Conversion and Control covering the application of IT and embedded system to these areas and robotics. Some of the current work includes the development of elearning and web-services for the teaching of engineering and engineering projects. He co-founded and is the President of the spin-off company, ecPresence Technology Pte Ltd, which specialises in enabling web-technology in business and education. He has published numerous papers in automation, control and robotics, power electronics and drives, and in web-based online learning system.

An Integrated Geospatial and Visualization Platform for Solar Radiation Mapping

Yuxia Yao[1], Hock Beng Lim[1], Chee Keong Chan[2], Fook Hoong Choo [2]
[1] Intelligent Systems Centre, Nanyang Technological University
[2] School of Electrical and Electronics Engineering, Nanyang Technological University
[*yxyao, limhb, eckchan, efhchoo*]@ntu.edu.sg

Abstract: In this work, we design an integrated platform based on emerging geospatial and visualization technologies for handling environmental sensor data. Specifically, we focus on the application of solar radiation mapping. In recent years, geospatial map services such as Google Earth and Virtual Earth have become increasingly popular. They allow users to produce customized web-based maps with desired functionalities such as adding overlays on the map. Furthermore, the customized maps can be embedded into different websites and shared with others. This emerging technology provides a promising platform for visualizing scientific data collected from sensor networks and sharing it with interested parties.

Visualization of sensor data requires certain procedures to be carried out at the backend, such as data collection, data archiving and data processing. In this work, we have developed a sensor grid infrastructure to connect weather stations deployed in schools across Singapore, and continuously collect data from the weather stations and archive the data in a large-scale data depository.

Due to the harsh operating environment of sensor networks, raw sensor data are usually noisy. Hence, the data from sensor networks must be further processed before it can be visualized. In this paper, we discuss various data pre-processing issues such as data cleaning that removes abnormal data, data aggregation that aggregates data according to user defined data resolution, and data interpolation that estimates data at locations where physical sensors are not available. We have developed techniques to address these issues under the sensor grid infrastructure.

We have designed an integrated geospatial and visualization platform for generating solar radiation maps. This system provides support for generating contour maps of solar radiation distribution over Singapore. A web portal has been developed to allow users to interactively visualize the solar radiation data. The visualization results present an insight on the solar pattern in Singapore and facilitate various applications in solar energy harvesting.

1. Introduction

In the era of Web 2.0, online geospatial maps and satellite photos add depth to the Web by connecting it with real-world topography. Such maps enable one to chart a trip and get around in real time. Their aerial views of grasslands, mountaintops, and city grids also provide a new perspective of the planet. An important characteristic of such online maps is that they provide freedom for users to interact with the map and to do mashups of their own creations. For example, in Google Maps, a set of APIs is provided to the users to integrate Google Maps into their websites with their own data sets. The information provided from data sets can be very comprehensive. With online web-based maps, users can easily share their real-time data with the rest of the world.

In this work, we focus on the visualization of real-time environmental data collected from sensor networks. Environmental sensor networks usually comprised of a set of sensors deployed in the field and wirelessly connected. They are increasingly being used for many important applications requiring the interaction between users and the physical world. They allow the physical environment to be measured at high resolutions, and greatly increase the quantity and quality of real-world data and information for applications. Important applications of sensor networks include environmental and habitat monitoring, weather monitoring and forecasting, health-care for patients, military and homeland security surveillance, tracking of goods and manufacturing processes, safety monitoring of physical structures and construction sites, smart homes and offices, and many others.

The "Combined-Cycle Solar Energy Self-Sustaining Membrane Distillation (MD) and Membrane Distillation Bioreactor (MDBR) Water Production and Recycling System" project is a large-scale research effort funded by the National Research Foundation (NRF) of Singapore under the Competitive Research Programme (CRP) scheme. Global economic growth has caused excessive carbon emissions from burning fossil fuels, which has been blamed for global warming and climate change. This together with urban migration has caused the shortage of fresh water supplies in major cities around the world. Water is a basic necessity for society. However in many places water demand exceeds supply from conventional sources such as surface and ground waters. Increasingly water is being provided from sources that require significant purification, such as sea water and waste water. This Solar CRP project aims to develop and demonstrate self-sustaining water production and recycling technology based on solar energy. The technology would be suitable for Singapore and many other water scarce regions of the world. Solar energy can be used as a source of clean energy to drive the operation of MD systems for water production and to drive the operation of MDBR systems for water purification. In the design of such systems, solar radiation is one of the most important parameters. But unfortunately, solar radiation pattern in Singapore is not well studied and understood. Comprehensive solar radiation data analysis, data mining and data modeling are required to optimize the operation of solar panels to harness maximum solar energy.

The National Weather Study Project (NWSP) [1] is a community-based environmental initiative in Singapore that aims to promote the awareness about weather patterns, climate change, global warming and the environment. In this project, hundreds of mini weather stations have been deployed in the schools to monitor the weather parameters such as temperature, humidity, wind speed, wind direction, solar radiation, rain, etc. The data are streamed into a large-scale data archive for further processing. The collected data, fine-grained both spatially and temporally, can be used for weather pattern analysis and decision support systems.

The Solar CRP project requires the visualization of solar radiation pattern in Singapore. As part of the project, we design an integrated platform based on emerging geospatial and visualization technologies for handling the solar radiation data. In particular, our focus will be the solar radiation data which are collected from the weather stations deployed under the NWSP.

The rest of this paper is organized as follows. We describe the system architecture in Section 2. Section 3 discusses the challenges related to sensor data management and preprocessing in such large-scale systems. The implementation details of our integrated geospatial and visualization platform are presented in Section 4. Finally, Section 5 concludes the paper.

2. System Architecture

With a large number of sensor devices deployed over a wide area to collect micro-level data, each sensor usually produces massive amounts of complex and diverse data for various applications. A major challenge is to collect and share real-time data from these heterogeneous sensor sources and to process these data, manage them and visualize them.

We have developed a sensor grid framework, called the Scaleable Proxy-based aRchItecture for seNsor Grid (SPRING) [2]. It is a flexible framework that is not constrained by the characteristics and requirements of specific target applications. By using proxies as interfaces between the sensor networks and the grid fabric, the SPRING architecture can support a wide range of sensor devices, including the less computationally powerful ones. Furthermore, the SPRING architecture is scalable and capable of integrating multiple heterogeneous sensor networks with the grid. This situation is likely to occur for an environmental monitoring network over a wide area.

A National Weather Sensor Grid (NWSG) [3] is designed as the key infrastructure for environmental monitoring sensor networks. The NWSG is built upon the SPRING framework and has several important features. First, it connects mini weather stations geographically distributed across Singapore to automatically collect and aggregate weather data in real time. Second, the weather data are stored in a large-scale data archive that uses distributed data storage resources. Third, the NWSG integrates computational resources for the compute-intensive processing of weather data. Fourth, the weather data can be conveniently accessible and shared via the web through mash-ups, blogs, and other user applications. Techniques and tools are being developed to efficiently publish, query, process, visualize, archive and search the vast amount of weather data. Finally, the NWSG provides a scalable platform to handle hundreds of weather stations and it can also integrate different types of sensors besides weather stations.

The NWSG has been implemented with user level, middleware and core services following a grid-based architecture. A proxy-based architecture is used to integrate the weather sensors with the grid infrastructure. In order to facilitate a seamless integration of heterogeneous sensor resources and grid environments, the NWSG uses an ontology-based design with the resource brokerage and meta scheduler services implemented at the middleware level to help in resource discovery and job planning. Since the NWSG is targeted at a wide range of users ranging from the scientific community to the public, it implements user authentication, job management and load balancing services to minimize job execution response time and maximize throughput.

Figure 1 shows various components of the proposed sensor grid infrastructure. A Virtual Organization (VO) is a resource-sharing participant of the sensor grid. Typical examples of VOs are the networks for metropolitan traffic monitoring, environmental monitoring, a smart home, an elderly care place and a corporate office. A VO may own one or more of the resources such as sensor resources, compute/data resources, grid-enabled devices, etc. The key idea is to enable each VO to access and share the resources of other VOs through distributed resource brokerage and user authentication. The sensors in each VO collect real time data which is archived at the proxies for user queries. The data collection policy is decided and implemented by the proxy node based upon the request coming from a user or an application. The computational and data resources in different VOs are shared to enable users and applications from anywhere to access archived data and to run sophisticated jobs that require archived or real time data from any amount of sensor and computational resources.

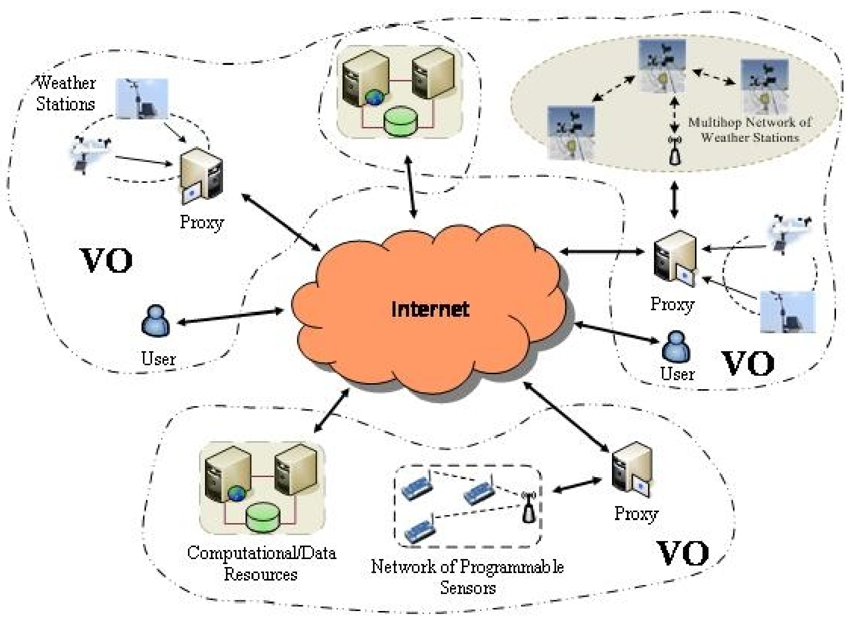


Figure 1: Sensor Grid Design Components

This architecture is different from conventional grid infrastructures as it provides access to real-time sensor data from heterogeneous and geographically distributed sensors or sensor networks. In addition, the compute and data resources utilized to execute a user job may also be part of different VOs. Since access to heterogeneous resources requires discovery, authentication, scheduling and execution of user requests, the sensor grid maintains VO-level ontology and implements the middleware needed to integrate sensor networks with grid infrastructures.

This backend architecture facilitates the developing of integrated geospatial and visualization platform to visualize and publish the environmental sensor data.

3. Data Management and Preprocessing

Data streamed into the database are usually raw data which are noisy and incomplete in nature. The data with such anomalies is meaningless for scientific applications and is unable to provide the desired information that the scientists expect. For example, in the Solar CRP project, the scientists require a visual solar radiation map in Singapore that can provide richer details of solar radiation's intensity and variation over a large spatio-temporal domain. Thus, the raw data should be first cleaned, aligned, interpolated and aggregated before it can be used to generate a visual solar radiation map. The following sections detail the issues and techniques related to each of these data processing steps

3.1 Data Cleaning

Sensor data are sometimes noisy due to environmental disturbances or due to faults in the sensor hardware. There may even be incomplete, missing or inaccurate data [4]. Thus, data cleaning is an important step for improving the quality of data. We consider both offline cleaning of the historical data as well as online cleaning of the real-time streaming data. Based on signal processing and statistical techniques [5], we have developed data cleaning algorithms to clean the data obtained from a single weather station. Besides the time series data from a single weather station, the data from a group of weather stations in close geographic vicinity can be used to improve the accuracy of data cleaning. In this way, both the temporal and spatial attributes of the data are utilized for data cleaning. Some of the research challenges include:

- Design and implementation of data cleaning algorithms based on advanced signal processing and statistical techniques.
- Design and implementation of techniques to remove the outliers from both the archived and real time data. Adaptive statistical models are needed in this case to cater for changes in the data distribution.
- Design and implementation of group-level data cleaning algorithms. We will study the optimization of weather station grouping that should be grouped such that the variances in the data from stations in the same group are minimized.
- Design and implementation of model-based data cleaning algorithms. Due to different characteristics of the weather data (e.g., temperature and rain have different patterns over time), different models may be needed to characterize their patterns. Further study on weather data pattern is needed to support data cleaning.
- To address the issue of incomplete or missing data, algorithms to derive correct data readings should be developed. One possible approach is to combine adaptive data sampling techniques with data cleaning.

In this work, we have implemented a k-distances algorithm to remove the outliers/spikes from the solar readings in a single weather station. For this purpose, a sliding window based approach is used to obtain a set of readings and identify the outliers in the set. For each reading in the window, its Euclidian distance from the rest of the readings in the window is calculated. The average distance is then compared with a threshold which is determined based on the correlation between the data. If the average distance is larger than the threshold, the reading is classified as an outlier and is removed from the data set. A new reading averaged from the remaining samples in the sliding window is used to replace the outlier. The efficiency of this algorithm is affected by the size of the sliding window and the threshold. The limited space in this paper does not allow us to provide extensive details of the algorithm.

3.2 Data Alignment and Interpolation

Data alignment issue arises due to the different sampling rate configured for various sensors. Thus, given a specific point in time, the availability of the data from each sensor cannot be guaranteed. To address this issue, it is required to align the data over time dimension. We have implemented data aggregation algorithms that aggregate data from each sensor in a certain time window, e.g., within 30 minutes.

In sensor networks deployed over a wide area, sensor nodes (weather stations) are deployed at discrete sites. In this case, it is not possible to obtain sensor data from each geographical point in the area. This requires interpolating sensor data over the entire geographical space. The quality of the interpolation results highly depends on the data that are involved in the interpolation. In addition to employing existing interpolation methods like inverse distance interpolation and kriging algorithms, our systems offers other characteristics that can be exploited to improve the accuracy of interpolated data. The rationale behind employing system characteristics is to find the spatial dependence between the locations void of a real sensor and the locations that are involved in interpolation process with real sensor data.

In the current implementation, the data are aligned by aggregating every 30 minutes' data. For interpolation, the inverse distance algorithm is employed.

4. Integrated Geospatial and Visualization Platform

In this section, we present the integrated geospatial and visualization platform. There are three components in this platform: a back-end query engine, a visualization and geospatial engine, and a web interface, as shown in Figure 2.

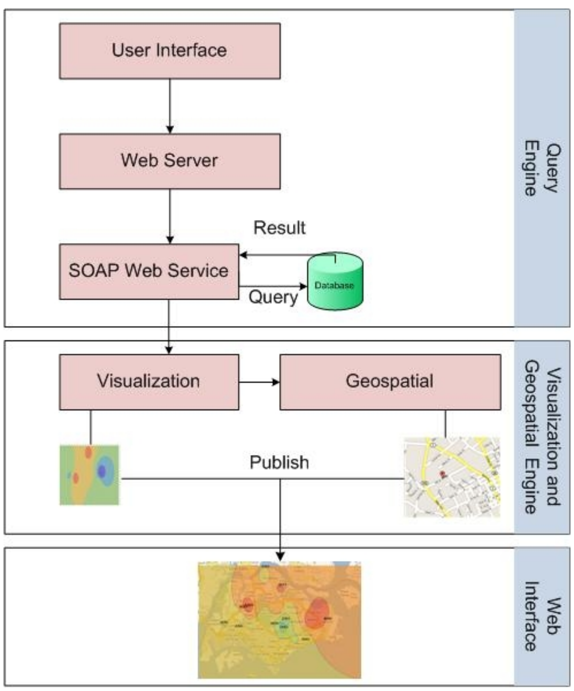


Figure 2: System Architecture

4.1 Query Engine

Users issue queries via a web interface by specifying the time duration of data that must be used to generate the solar radiation contour map. This user request is passed over to the server side query engine by the web server. First, raw solar radiation data from all the weather stations are obtained in the XML format by using the data download web service provided by the sensor grid. After parsing the XML data, the data cleaning procedure is carried out that removes the abnormal spikes from the data using k-distance technique. Missing values are then filled up by doing a moving average of adjacent readings along time dimension. To interpolate the data points for locations void of real sensors, we have divided Singapore's geographical map into a set of grid cells with a size of 0.01° by 0.01°. The solar reading at each grid location was estimated by using inverse distance interpolation method.

4.2 Visualization and Geospatial Engine

In this work, we use Wolfram Mathematica 7.0 [6] and Google Maps [7] to develop an integrated visualization and geospatial platform.

The visualization engine is implemented using Mathematica, which is a tool many scientists often use for scientific computing purposes and is powerful in optimized computations and visualization. Mathematica is split into two parts, the kernel and the "front end". The kernel interprets expressions and returns result expressions. The Mathematica Front End provides a GUI which allows the creation and editing of Notebook documents which can contain program code with pretty printing, formatted text together with results including typeset mathematics, graphics, GUI components, tables and sounds. In this work, we have only used the kernel component. Mathematica provides numerous plotting functions such as line plots, contour plots, 2-D and 3-D plots. It satisfies the requirements from most of the scientific applications.

The geospatial engine is implemented by using Google Maps, a powerful open source geospatial visualization service. It provides various APIs for the users to add overlays over geographical map of the region. In our application, weather stations at different schools are displayed on the map. This is implemented by overlaying a "kml" document provided by Google Maps. The real-time data generated from the weather stations are displayed as an html and refreshed regularly. Google Maps also provides services to overlay images, such as the images generated by Mathematica plotting services. By specifying the boundaries of the images (in the format of longitude and latitude), Google Maps can overlay the images onto exact locations on the map. In addition, AJAX/JQuery data request calls are used to obtain most updated data from the server side.

4.3 Web Interface

We have developed a web portal to showcase the platform. It enables the users to input their requirements and display the visualization results. Moreover, other related services such as data download are also provided on the portal.

4.4 Implementation of Solar Radiation Mapping Services

The solar radiation mapping service is developed based on the integrated geospatial and visualization platform. The interpolated data points at each grid cells passed from the query engine are sent to Mathematica. We use one of the Mathematica plotting functions called ListcontourPlot. The average solar radiation value of the selected time period is shown at different contour lines. After the map is generated, it is embedded within Google Maps and published on the portal.

As an example, Figure 3 shows the solar radiation map for Singapore on March 2nd, 2009. The solar radiation map indicates that on that day, the solar intensity is higher at three locations in Singapore; namely, around Changi Airport, Choa Chu Kang and Simpang area. Such a solar radiation map will help the scientists (in the Solar CRP project) and also the general public to understand more about the solar radiation pattern in Singapore. Currently, preliminary results on the daily solar radiation patterns in Singapore have been obtained. Further research will be conducted to establish more extensive results, including the solar radiation statistics and maps on weekly, monthly, and annual basis in addition to the daily patterns.

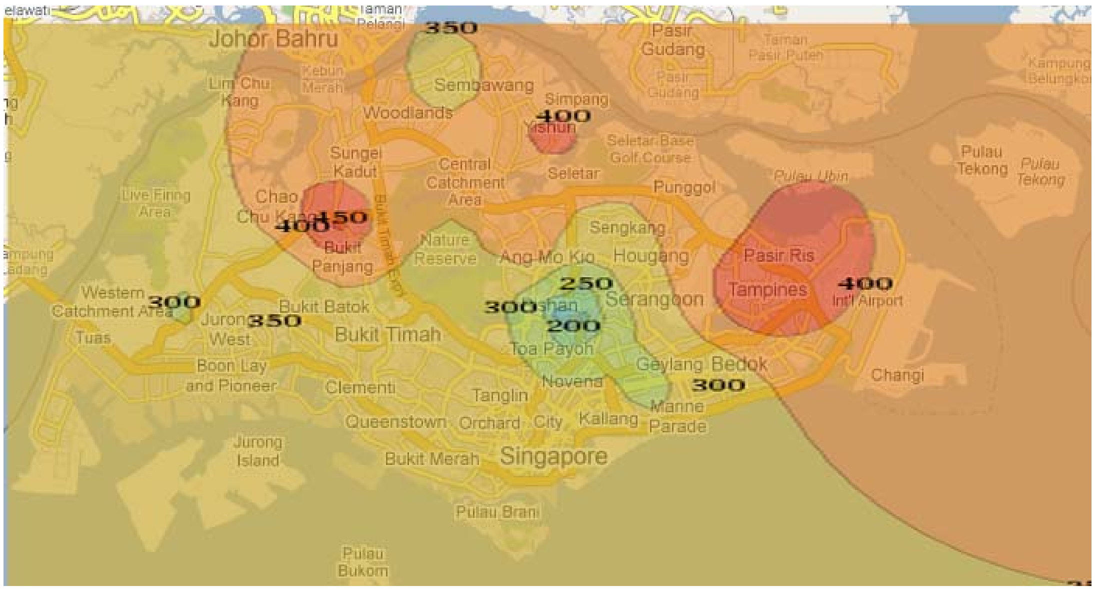


Figure 3: Solar Radiation Map for Singapore

5. Conclusion

In this paper, we have developed an integrated geospatial and visualization platform for solar radiation mapping. A sensor grid based infrastructure is used to collect solar radiation data from weather stations deployed in schools across Singapore. The data are archived and processed at the server side. Various sensor data management and preprocessing issues have been addressed. We make use of Mathematica and Google Maps to develop the integrated geospatial and visualization platform. The visualization results are published via a web portal for convenient access by scientists and the general public.

Acknowledgements

This work was supported by the Singapore National Research Foundation (NRF) under grant number NRF-G-CRP-2007-02.

References

[1] "National Weather Study Project", http://nwsp.ntu.edu.sg

[2] H. B. Lim, Y. M. Teo, P. Mukherjee, V. T. Lam, W. F. Wong, and S. See, "Sensor grid: Integration of wireless sensor networks and the grid," *Proc. of the 30th IEEE Conference on Local Computer Networks (LCN 2005)*, pp. 91-98, Nov 2005.

[3] H. B. Lim, M. Iqbal, W. Wang, and Y. Yao, "The National Weather Sensor Grid: A large-scale cyber-sensor infrastructure for environmental monitoring," To appear in the *International Journal of Sensor Networks (IJSNet), Special Issue on Environmental Sensor Networks*, Inderscience, 2009.

[4] K. Ni, et. al., "Sensor network data fault types", ACM Trans. Sensor Networks, Vol 5, Issue 3, pp. 1-29, May 2009.

[5] F. Koushanfar, M. Potkonjak, A. Sangiovanni-Vincentelli, "Error models for light sensors by statistical analysis of raw sensor measurements", Sensors, 2004. Proceedings of IEEE, pp. 1472-1475, Oct 2004.

[6] Mathematica 7.0, http://www.wolfram.com/

[7] Google Maps, http://maps.google.com/