ISSN 1870-4069

A Data Fusion Approach of Physical Variables Measured through a Wireless Sensor Network

Thomas Alejandro Arias Pelayo, Guillermo Molero Castillo, Everardo Bárcenas, Rocío Aldeco Pérez

Universidad Nacional Autónoma de México,
Facultad de Ingeniería,
Mexico

thomas.arias95@gmail.com,
{gmoleroca, raldeco}@fib.unam.mx,
ebarcenas@unam.mx

Abstract. At present, current data fusion methods are a useful tool for integrating data sources, prior to data analytics, and provide a unified view of an observed phenomenon or event. This paper presents the development of the Monte Carlo method, as a data fusion mechanism, obtained from a wireless sensor network. This network of sensors was designed and installed in a closed environment of human occupation. The data collected was of physical variables, such as temperature, humidity, and dust density, which were stored in the cloud through ThingSpeak, which is an open-source platform for the Internet of Things. As a result, it was succeeded in data fused properly and the method was evaluated through the root-mean-square error. Undoubtedly, fused values can be useful, for example, for the analysis of the thermal comfort of users in closed environments, where there are minimal ventilation rates and adequate indoor air quality is needed.

Keywords: Data Fusion, Monte Carlo Method, Wireless Sensor Network, Physical Variables.

1 Introduction

Today, the automation of industrial processes, the improvement of workforce capabilities and the development of new products through Artificial Intelligence benefits society as part of the digital transformation process in this era of the fourth industrial Revolution. However, of all the technologies to consider for their potential, without a doubt, wireless sensor networks (WSN) are proving their usefulness for measuring and storing data on a specific environment, as a key technology to implement the Internet of Things (IoT), whose main characteristic is its low energy consumption and its deployment in inaccessible locations or even integrated within structures.

It is important to note that currently, wireless sensor networks are autonomous devices that work collaboratively, distributed throughout an area of interest and whose objective is to monitor physical or environmental parameters, such as temperature, sound, vibrations, pressure, movement, or pollutants. Each element of the sensor network communicates wirelessly, offering a flexible system, easy to install, and scalable in large quantities. However, as the volume of data grows, so does the need to combine data from sensors to extract useful information, allowing an understanding of the environment and making timely decisions.

Faced with this situation, data fusion systems are an efficient solution to this problem. Data fusion is the process of associating and combining data at various levels, which come from different sources, such [1]: sensors, databases, signals, decisions, and others. This field of knowledge is used in different areas, such as [2]: signal processing, statistics, inference, and artificial intelligence. It has greater advances in applications of automatic target recognition, autonomous vehicle navigation, remote sensing, virtual environments, smart homes, threat identification, among others.

The idea of combining data from various sources, such as a network of sensors, is to obtain a global and unified vision of the observed phenomenon. Therefore, data fusion provides a formalization to data integration, in such a way that the results help in the decision-making process about a certain event, object, or action [3].

This paper aims to provide a vision of the progress of the development of a method for data fusion of physical variables, obtained in a closed space of human occupation, such as a room of a house, office and other places of rest or work, through a wireless sensor network. In particular, the analysis of the three variables was carried out, such as temperature, humidity, and dust density.

The document is organized in the following way, Section 2 presents the background of data fusion, some of the most significant contributions of data fusion are mentioned, its applications are discussed, the use of wireless sensor networks and the related jobs. Section 3 describes the method established as a proposed solution. Section 4 presents the results obtained, based on an application example, and Section 5 summarizes some conclusions and future work.

2 Background

As mentioned, data fusion is about the integration of data that comes from different sources. The purpose is to have a unified and refined view of them, in order to analyze them and give a conclusion. The data fusion as a field of knowledge emerged with military and robotics applications [10].

At present, one of the areas that have given the greatest impetus to data fusion is data science, where methods are used to extract, transform and load data sources as part of the data engineering process, prior to data analytics [11, 12].

Another area, which has also benefited from data fusion, is IoT since this is a technology that depends on interconnected objects, capable of communicating for the collection of data about a certain event [13].

2.1 Wireless Sensor Network

A wireless sensor network is made up of a set of interconnected sensor nodes that cooperate to measure and control a certain physical context that surrounds them, thus allowing interaction between people and devices or machines. These sensors work within a field of action to capture and transform signals into data, which are sent through other nodes to a common output, such as the Internet or databases [14].

A sensor network uses wireless or wired communication, and can be physically organized within an area where events occur in different ways, such as [7]: a) distributed, where there is no central node, but if a node stops function, it does not affect any other; b) centralized, where all nodes are peripheral except the central node, which is the only one that emits information; c) decentralized, where there is a collective hub of nodes, if one node goes down, several nodes may stop working, but the entire network does not go down.

In recent years, along with the growth of the Internet of Things, wireless sensor networks are gaining relevance in the industry, academia, and society in general, as a global infrastructure that allows offering advanced services through the interconnection of physical and virtual objects, thanks to the interoperability of information technologies [15], which, in turn, guarantee security and privacy requirements [17]. Thus, the applications are varied, among which agriculture, smart cities, smart homes, smart environments, industry, logistics, environment, military solutions, security, health, among others, stand out.

2.2 Data Fusion

Data fusion is a field of knowledge used to combine data produced by one or more sensors or data sources so that you can have a better estimate of the amount being measured [16]. Current ideas of data fusion are an efficient solution alternative to the increasing availability of data volumes. This is to combine data from various sources to obtain a complete description of an observed phenomenon [7].

On the other side, various definitions have emerged in data fusion. One of these points out that it is a multi-level, multi-stage process for the detection, association, correlation, estimation, and combination of data from one or more sources to achieve complete estimates and evaluations [18]. Recent examples of data fusion cover a wide range of areas, such as engineering, medicine, traffic control, environment, artificial intelligence, robotics, among others.

There are, therefore, application areas that until recently have been little explored by the data fusion community, some of these are [20]: the development of systems capable of incorporating hard and soft data, systems of data fusion in dynamic environments and real-time processing, as well as the incorporation of data fusion in smart sensors.

2.3 Monte Carlo Method

The Monte Carlo method is a non-deterministic technique, used to approximate numerical estimates to a great variety of mathematical problems. This method is named in reference to the Monte Carlo Casino (Monaco), for being the capital of the game of chance, as roulette is a simple generator of random numbers [19]. This method bases its operation on simulation, which consists of repeating the characteristics and behaviors of a real system. The objective is to imitate the behavior of real variables to, as far as possible, analyze or predict simple or complex future values.

Currently, this method is a fundamental part of the Ray Tracing algorithms for the generation of 3D images [21]. The method may vary depending on the application, but in general, it follows the following steps: i) it defines the domain of the inputs; ii) it entries with random values are generated, following a probability distribution that is within the domain; iii) the inputs in the function are evaluated, and iv) the calculated values are gathered.

Being a probabilistic analysis algorithm, the possibility of error is given by the number of errors divided by the number of times the test is performed. Following the central limit theorem, Monte Carlo has an absolute error given by:

$$\frac{1}{\sqrt{N}}.$$

This indicates that if the number of trials is increased, the probability of the error tends to zero [19]. An advantage of the Monte Carlo simulation is that it can give a good approximation of the exact value. In this way, this method can be used to measure reliability in a sensor network [8]. The equation used is the following:

$$\langle b-a\rangle \frac{1}{N}\sum_{n=1}^{\infty} f(x_i),$$

where:

- $\langle b-a\rangle$ are the intervals of the function,
- N is the number of tests,
- x_i is the random value that is evaluated in the function.

2.4 Related Works

A growing problem in data fusion is dealing with data inconsistency. For this, there are various methods that can be used, such as sending the same data packet a second time. Once sent, the similarities of the two packets are compared and outliers are removed. Another option is to compare previous data measurements and then give a certain level of confidence to the sensors that were used [4]. Another method is to assign a series of weights hierarchically to the sensors [5], but none of these methods deal with the conflict, from the point of view of inconsistency, generated by using different sensors. Table 1 summarizes some identified works, as part of the literature review.

Therefore, due to the promising future of wireless sensor networks, there is a natural interest to embrace data fusion applications with this type of technology to monitor different phenomena and events efficiently.

Table 1. Related Works.

Author	Description	Method used	Limitations
Motro *et al.* (2006) [6]	It is a data source integration system, called Fusionplex.	The process consists of finding inconsistencies in data tuples, that is, inconsistencies are detected in the instances and eliminated according to the criteria.	High computational cost to resolve conflicts in commercial applications.
Kumar *et al*. (2009) [4]	Detects data inconsistency in sensors before it is sent.	Uses a neural network to detect the inconsistency of data through its variance.	It does not analyze the inconsistency of the data.
Frikha *et al*. (2015) [5]	Assigns a hierarchy to sensors to eliminate subjectivity and reduce inconsistency. The higher the hierarchy, the greater the weight.	The sensors were compared in pairs using matrices. The weights of the sensors are added.	It is required to define the criteria to assign the initial hierarchies.
Bakr *et al*. (2017) [7]	Three data fusion approaches are presented to deal with inconsistency: a) model-based, b) redundant, and c) fusion-based.	The treatment of inconsistency of data from various sources is described.	Lacks a process of fusing data from multiple sources.

Table 2. Characteristics of the NodeMcu board.

	Characteristic	Value
1	Power supply	5 V
2	Output voltage on pins	3.3 V
3	Processing speed	80 - 160 MHz
4	Size	29.1 x 63.3 mm

3 Method

The working method defined for data fusion of physical variables, measured in a closed environment of human occupation, was divided into four stages: a) installation of the sensor network, b) data collection, c) data fusion, and d) evaluation of the fusion method.

a) Installation of the sensor network

In this stage, two sensor nodes were installed at the ends of a room. The NodeMcu board (ESP8266) was programmed to send data to the ThingSpeak platform, which allows storing, visualizing, and analyzing data in real-time, using cloud services for IoT applications. Using the cloud has advantages, since it allows access to the data collected from anywhere in the world, using any device with an Internet connection. In addition, it provides useful tools for users. Table 2 summarizes the main characteristics of the NodeMcu board, the required values of voltage, speed, and size. This board is used for the development of mobile IoT applications.

Table 3. Characteristics of the DHT11 sensor.

	Characteristic	Value
1	Power supply	3.3 – 5 V
2	Output voltage on pins	3.3 V
3	Humidity range	20 – 90% with 5% precision
4	Temperature range	0 – 50ºC with ± 2ºC of precision
5	Size	14 x 18 x 5.5 mm

Table 4. Characteristics of the dust sensor.

	Characteristic	Value
1	Power supply	5 V
2	Current consumption	Icc: MAX 20 mA
2	Pulse cycle	10 ± 1*ms*
3	Size	46.0 x 30.0 x 17.6 mm

Table 5. Extract of data collected through the implemented sensor network.

Date	Temperature (ºC)	Humidity (%)	Dust Density (mg/m^3)
2020-05-30 19:54:38	28.9	67	0.02
2020-05-30 19:59:38	29.1	69	0.01
2020-05-30 20:04:39	29.0	71	0.02
…	…	…	…
2020-06-23 09:37:12	32.1	64	0.01
2020-06-23 09:42:13	32.2	65	0.01
2020-06-23 00:47:13	32.2	66	0.02

The DHT11 sensor includes a humidity and temperature measurement component that connects to an 8-bit microcontroller to ensure quality and fast response using low voltage. Table 3 summarizes the main characteristics of the DHT11 sensor, it indicates the power values required to make it work, its size, and the range of humidity and temperature values that it can measure.

On the other side, the GP2Y1010AU0F sensor was used to detect reflected light from dust and smoke particles. This type of sensor can distinguish between dust and smoke particles by means of a pulse pattern in the output voltage. Table 4 presents some of its most important characteristics, such as the pulse cycle it uses to detect dust particles, size and feed.

b) Data collection

To send data to the ThingSpeak channels and maintain the integrity of the data, the platform has a function called “thingSpeakRead” that imports the non-null values. The data that comes from the two nodes were grouped. To carry out this task, two channels were created, where the data was stored and public access was granted for it to be read by any user [22, 23]. The data was collected in a period of five weeks, from

May to June 2020. Table 5 shows, as an example, an extract of the captured data.

c) Data Fusion

As part of the data fusion process, the Monte Carlo method was used to verify the variation in the data that was obtained through the sensor nodes. This method was chosen for its ability to numerically operate complex systems, maintaining the input and output relationship, with a certain degree of uncertainty. Subsequently, the data were fused for each of the collected variables, that is, temperature, humidity, and dust density.

The Monte Carlo method was chosen since it allows us to identify what is going to happen and its probability of occurrence. In addition, it is important to point out that through this method it was possible to identify possible conflicts generated in the data, since if this situation is ignored, the resulting data could represent, in an erroneous way, the real data obtained in the experimentation.

d) Evaluation of the fusion method

The Root Mean Square Error (RMSE) was used to evaluate the fusion method. Through this measurement, the difference between the estimated (predicted) values by the Monte Carlo method and the values captured by the sensors is calculated. The formula used was the following:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{n}\left(\frac{d_i - f_i}{\sigma_i}\right)^2},$$

4 Results

The results of the data fusion for each of the variables analyzed are shown below.

a) Temperature

The probability distributions of the values obtained for temperature were plotted, as well as the data fusion. Fig. 1 shows the measurements of sensors 1 (blue) and 2 (green), which are close to the 50th percentile of each. While the values of the data fusion (red) are between the measurements of the two sensors, with a slight similarity to the values obtained from sensor 2.

b) Humidity

The probability distributions of the humidity percentages were plotted, as well as the data fusion (Fig. 2). The most probable values for sensor 1 (blue) and sensor 2 (green) were found to be within the values of the data fusion (red). Sensor 1 (blue) measurements have average behavior relative to sensor 2 and data fusion.

c) Dust Density

Fig. 3 shows the probability distribution of the dust density percentages measured by the two sensors, as well as the values of the data fusion. Sensor 2 (green) values were found to be unusual, as they are skewed to the right.

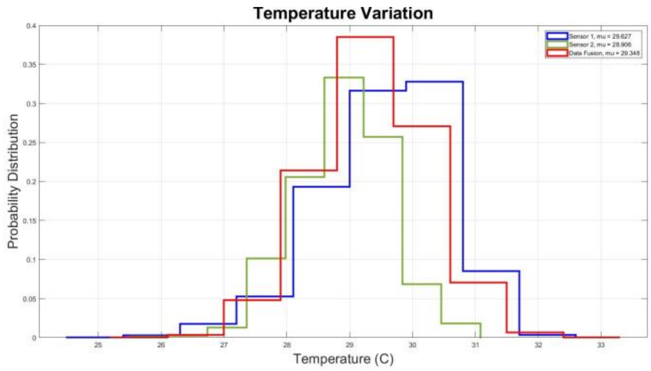

Fig. 1. Probability distribution of temperature data fusion.

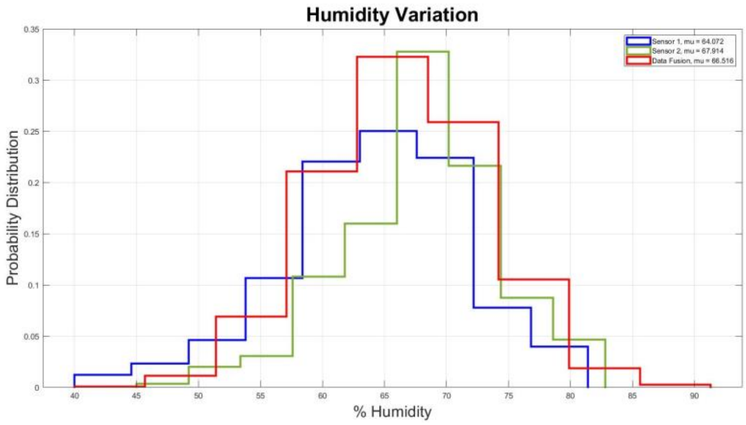

Fig. 2. Probability distribution of humidity data fusion.

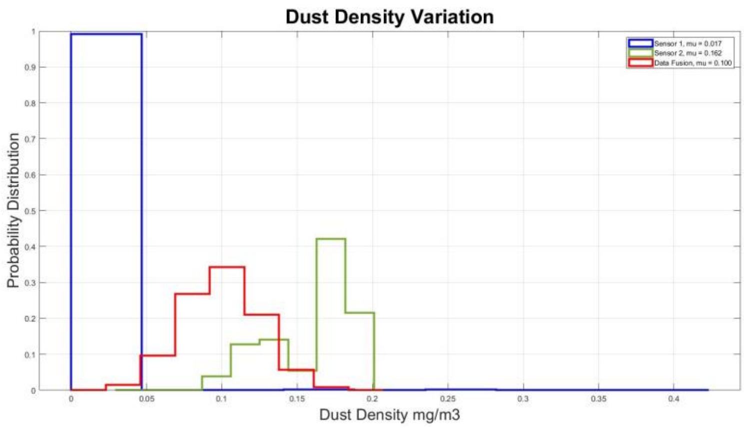

Fig. 3. Probability distribution of dust density data fusion.

Table 6. Mean square error estimation.

Variable	Sensor	RMSE
Temperature	Sensor 1	1.3892
	Sensor 2	1.2586
Humidity	Sensor 1	10.1487
	Sensor 2	9.2745
Dust density	Sensor 1	0.0895
	Sensor 2	0.0735

This may be due to sensor 2 being located above a bookcase and that the amount of dust that was raised was greater due to the cleaning carried out in this part of the room. In addition, it was observed that sensor 1 (blue) presented less variation, with a probability of density less than 0.05, that is, almost 99%. Although most of the data obtained in sensor 1 varied from 0.00 to 0.05, there were also some cases of variability between 0.07 to 0.4. This difference, with respect to the other sensor, is due to the fact that in that area of the room there is almost no air current and no cleaning was done.

On the other side, Table 6 shows the results obtained from the mean square error measurements with respect to the data fusion. Low levels of error were reached in temperature (1.38 and 1.25%) and dust density (0.08 and 0.07%); while, for humidity higher percentages of error were observed, that is, 9.27 and 10.14%. This may be because some days it rains and others don't. When it rains, the humidity increases, that is, there is high atmospheric saturation, which can give unusual values. Consequently, the precision level for temperature was 98.7%, humidity 90.3%, and dust density 99%. Given that 99% of the density levels of the powder were between 0.00 and 0.05, the RMSE value for the two sensors that measured this variable was low, that is, almost 1% error.

It is important to mention that the accuracy of the Monte Carlo method increases as more simulations are performed. Therefore, the computational cost also increases.

5 Conclusion

A network of sensors was implemented in a rest and work area, given the current context of the COVID-19 disease, such as a room in a home, in order to obtain data on temperature, humidity and dust density. For this, a NodeMcu development card (ESP 8266) with Wi-Fi connection capabilities was used in order to connect to the ThingSpeak platform to collect, visualize and analyze data from the sensors.

DHT11 sensors were used to measure temperature and humidity, and GP2Y1010AUOF sensors were used to measure dust density. The data obtained by the sensors can be consulted in real time through the two channels created in ThingSpeak. Sensor data can also be downloaded through the ``Extract Data" function on the same channel

The values obtained for temperature, humidity, and density of the powder were correctly fused using the Monte Carlo method. The values of the sensors and the values obtained through the method used were compared. In addition, low error levels were reached, which were evaluated by means of the root-mean-square error, which allowed to verify the accuracy of the obtained values. As future work, it will be sought to analyze based on the fused values of temperature, humidity, and dust density, if the observation environment, in this case, a room, is within the thermal comfort zone defined according to the ANSI/ASHRAE 55 standard. This type of analysis is useful to identify minimum ventilation rates and other measures to provide adequate indoor air quality in human-occupied spaces.

It will also consider expanding the implementation with different techniques, such as Artificial Neural Networks, Kalman Filter, among others, with the purpose of comparing the results and obtaining better and better-fused data.

Acknowledgment. This work was supported by UNAM-PAPIIT IA105320.

References

1. Kale, D., Aparadh, S.: a study of a detection and elimination of data inconsistency in data integration. International Journal of Scientific Research in Science, Engineering and Technology, 2(1), pp. 532−535 (2016)
2. Shahbazian, E., Bergeron, P., Duquet, J., Jouan, A., Valin, P.: Data fusion applications for military and civilian purposes developed DND/LM Canada decision support testbed. In: Conference Record of the Thirty-Third Asilomar Conference on Signals, Systems, and Computers, 1, pp. 420−424 (1999)
3. Durrant-Whyte, H., Henderson, T.: Multisensor data fusion. In: Siciliano B., Khatib O. (eds) Springer Handbook of Robotics. Springer, Berlin, Heidelberg (2008)
4. Kumar, M., Garg, D.: Multi-sensor data fusion in presence of uncertainty and inconsistency in data. In: Milisavljevic (Ed.), Sensor and data fusion, pp. 225−244, Intech (2009)
5. Frikha, A., Moalla, H.: Analytic hierarchy process for multi-sensor data fusion based on belief function theory. European Journal of Operational Research, 241(1), pp. 133−147 (2015)
6. Motro, A., Anokhin, P.: Fusionplex: resolution of data inconsistencies in the integration of heterogeneous information sources. Information Fusion, 7(2), pp. 176−196 (2006)
7. Bakr, M., Lee, S.: Distributed multisensor data fusion under unknown correlation and data inconsistency. Sensors, 17(11), pp. 1−31 (2017)
8. Yang, Q., Chen, Y.: Monte carlo methods for reliability evaluation of linear sensor systems. IEEE Transactions on Reliability, 60(1), pp. 305−314 (2011)
9. Schiavon, S., Hoyt, T., Piccioli, A.: Web application for thermal comfort visualization and calculating according to ASHRAE Standard 55. Building Simulation, 7(4), pp. 321−334 (2014)
10. Muñoz, J.: Estimación Bayesiana para el tratamiento de la inconsistencia en la fusión de flujos de datos de fuentes físicas. Doctoral Thesis, Universidad Veracruzana (2020)
11. Jagadish, H., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J., Ramakrishnan, R., Shahabi, C.: Big data and its technical challenges. Communications of the ACM, 57(7), pp. 86−94 (2014)
12. Zhu, X., Wu, G., Ding, W.: Data mining with big data. IEEE Transactions on Knowledge and Data Engineering, 26(1), pp. 97−107 (2014)
13. Cecchinel, C., Jimenez, M., Mosser, S., Riveill, M.: An architecture to support the collection of big data in the internet of things. In: IEEE World Congress on Services, pp. 442−449 (2014)

14. Akyildiz, I., Su, W., Sankarasubramaniam, Y., Cayirci, E.: Wireless sensor networks: a survey. Computer Networks, 38(4), 393−422 (2002)
15. Ramos, A., Varum, T., Matos, J.: Compact Multilayer Yagi-Uda Based Antenna for IoT/5G Sensors. Sensors, 18(9), 2914 (2018)
16. Wah-Ng, G., Tan, C., Poh-Ng, T., Siow, S.: Assessment of Data Fusion Systems. In: 9th International Conference on Information Fusion, 1−8 (2006)
17. Contreras-Masse, R., Ochoa-Zezzatti, A., García, V., Elizondo, M.: Selection of IoT Platform with Multi-Criteria Analysis: Defining Criteria and Experts to Interview. Research in Computer Science, 148(11), 9−19 (2019)
18. Blasch, E., Steinberg, A., Das, S., Llinas, J., Chong, C., Kessler, O., Waltz, E., White, F.: Revisiting the JDL model for information exploitation. In: Proceedings of the 16th International Conference on Information Fusion, IEEE, 129−136 (2013)
19. Illana, J.: Metodos Monte Carlo. Departamento de Física Teórica y del Cosmos, Universidad de Granada, Spain, 1−47 (2013)
20. Barnaghi, P., Tonjes, R., Holler, J., Hauswirth, M., Sheth, A., Anantharam, P.: Real Time IoT Stream Processing and Large-scale Data Analytics for Smart City Applications. In: European Conference on Networks and Communications, 1−36 (2014).
21. Kajiya, J.: The Rendering Equation. ACM SIGGRAPH Computer Graphics, 20(4), 143−150 (1986)
22. Channel for sensor 1, https://thingspeak.com/channels/1010164/private_show. Last accessed in August 10, 2020
23. Channel for sensor 2, https://thingspeak.com/channels/1010156/private_show. Last accessed in August 10, 2020