



HAL
open science

Towards Improving Data Completeness in Drone-Based Waste Detection

Souheir Mehanna, Abderrahmane Maaradji, Zoubida Kedad, Mohamed Chachoua, Assia Soukane, Hakim Hacid

► **To cite this version:**

Souheir Mehanna, Abderrahmane Maaradji, Zoubida Kedad, Mohamed Chachoua, Assia Soukane, et al.. Towards Improving Data Completeness in Drone-Based Waste Detection. The 7th Conference for Information & Communication Technologies for Organization & Society (ICTO), Jul 2023, Paris, France. hal-04158804

HAL Id: hal-04158804

<https://hal.science/hal-04158804v1>

Submitted on 11 Jul 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Towards Improving Data Completeness in Drone-Based Waste Detection

Souheir Mehanna^{1,3}, Abderrahmane Maaradji², Zoubida Kedad¹, Mohamed Chachoua³, Assia Soukane², and Hakim Hacid⁴

¹ Université Paris-Saclay (UVSQ)

² ECE Engineering School

³ École des ingénieurs de la Ville de Paris

⁴ Technology Innovation Institute (TII), UAE

Abstract. This paper explores the potential applicability of a previously proposed data quality model and metrics applied to the context of drone-based waste detection. The proposed framework aims to assess the completeness of waste detection data. The paper discusses the challenges and opportunities of adapting the existing model and quality indicators to this use case and identifies potential areas for future research. The proposed framework can inform the development of a more effective data engineering solution adaptable to the context of waste detection using drones.

Keywords: Data Completeness · Drones · Waste Detection.

1 Introduction

In the era of big data and edge computing, ensuring good data quality is increasingly important. Edge computing, which involves processing data locally on devices rather than sending it to a central server, has become a popular solution for handling the massive amounts of data generated by sensors and devices. Drone-based solutions are becoming popular for various applications, including waste detection and management, due to their flexibility, cost-effectiveness, and efficiency. Drones can be equipped with various sensors and computer vision technologies to detect and classify different types of waste, such as plastic bottles, litter, and hazardous materials. However, in a cooperative drone system, multiple drones collect data in parallel under different conditions, which can lead to a deterioration in the quality of the collected data. Decisions based on poor-quality data can be misleading and result in wrong decisions. Data completeness is an important quality dimension to help better understand the quality of the data. Data completeness, among other quality dimensions, is essential to ensure reliable waste detection results in drone-based solutions. However, this can be challenging due to the mobility of drones and the need for comprehensive spatial completeness.

Different models and quality metrics for data completeness were previously proposed for different use cases and sensors. In our previous work [9], we proposed a model that identifies different facets of data completeness, such as spatial, temporal, and sensor completeness used for air quality monitoring. This paper aims to explore the applicability and relevance of the previously proposed metrics applied to the context of drone-based solutions for waste detection. As a vision paper, we discuss the applicability of adapting the existing model and quality metrics to the context of drone-based waste detection, including considerations such as the mobility of drones and the need for real-time data processing. We also identify potential areas of future

research that can further advance the development of a framework for improving data completeness in drone-based waste detection.

In the remainder of this paper, section 2 discusses existing data quality models with respect to data completeness in the context of mobile sensors. In section 3, we evaluate the effectiveness of the proposed framework in improving data completeness and the accuracy of waste detection results. Section 4 presents identified challenges, future perspectives, and conclusion.

2 Related Works

The completeness of the data refers to the degree to which all of the required data are present in the data set [2]. Data completeness can be influenced by several factors in the case of waste detection using multiple drones. Data quality issues have been addressed in many research papers. Some of them [3], [11], [6] have investigated quality dimensions and their evaluation metrics. For instance, a framework for data quality assessment was proposed in [1]. Dasu et al [5] proposed two types of data quality checks, the first monitors the data acquisition process and checks the incoming data, while the second monitors the data stream content quality and investigates data quality according to four defined types of data constraints. The approach described in [10] defines local variations and a data quality level using graph convolutional networks. Although there are several data quality evaluation frameworks, these proposals do not take into account the specific characteristics of the data in drone-based waste detection systems.

3 Proposed framework

We discuss in this section the applicability of existing data quality metrics on the image data coming from drones. Drones can face complicated external circumstances, such as weather conditions which could prevent the drone from taking video at times. Other quality difficulties, such as a low battery, can limit the number of frames the drone is able to capture, which will also deteriorate the completeness of the collected data. The work of [4] discussed the importance of data quality to ensure a decent quality of indicators. Data completeness is one of the most assessed quality dimensions because a poor completeness level of the data has a significant impact on deteriorating the quality of all the analysis that follows. Our previous works in [8], [9] studied different data completeness for time series data coming from nomadic sensors.

3.1 Drone (sensor) Completeness

In the work of [8], [9], the authors defined sensor completeness as a quality facet that captures the extent to which the measurements of a given sensor are complete over a certain time period. This completeness facet is also applicable in this context because it gives an idea about the performance of the drone compared to its expected behavior and end results. There are many factors that impact this facet of completeness. The internal battery of the drone, other internal factors of the drone such as the resolution of the collected data and the height at which they were taken, the contrast, the quality of the frames collected, etc. There are other external factors such as the weather conditions such as wind direction and speed, the presence of a sand storm, or a fog. After the data has been collected by the drone, a pre-processing

step is done to filter out all the frames that are damaged, of poor resolution, or even not taken at the correct altitude and contrast.

The authors proposed to measure the drone (sensor) completeness of a drone S_i over a period of time P as the number of acceptable available data frames from drone S_i over the required/expected number of data collected by the drone. The required number of frames is to be defined according to the requirements of the application.

3.2 Spatial Completeness

Spatial completeness was defined in [8], [9] as the extent to which data sufficiently represents a specific spatial area, and it characterizes the coverage of this area considering the available measurements. In this context, spatial completeness is the coverage of the collected frames over a particular spatial area. We assume, for example, a drone is tasked with collecting video footage of the waste objects in a vast designated area A in the desert. Consider the drone is low on battery, the drone will not be able to well cover the entire area A with the desired altitudes and with the proper quality of the frames.

One of the metrics proposed to measure spatial completeness is considering a uniform distribution of the collected data over space. This means that the available data is compared to a reference considered the ideal situation. This metric applies to our drone's context as one assumption could be to consider an ideal spatial completeness to have a uniform coverage of the collected frames over the designated area A .

3.3 Temporal Completeness

Temporal completeness was defined in [8], [9] as the extent to which a considered period of time is covered by the available data. This completeness facet is applicable to our context as it is interesting to understand whether the data collected by the drone are distributed over the time it was operating or not. We take as an example poor weather conditions while a drone is operating, such conditions in a desert are quite frequent. Assume a drone is expected to be collecting measurements at regular intervals over a certain period P , and a fog or a sandstorm happens halfway through period P . The drone will not be able to collect data at the second half of the period which will degrade the temporal completeness of the data collected by the drone.

The authors of [8] proposed one metric that suggests measuring temporal completeness considering a uniform distribution of the collected data over the time period. This means that the actual available data is compared to a reference that is considered the ideal situation. This metric applies to our drone's context as one assumption could be to consider an ideal temporal completeness to have a uniform coverage of the collected frames over the time period P .

4 Conclusion and future work

In conclusion, ensuring data quality and completeness is essential for effective decision-making and planning in drone-based solutions for waste detection and management. For monitoring large areas and inaccessible locations such as landfills, and desert areas, drones offer a flexible, cost-effective, and efficient solution[7]. However, ensuring data completeness can be challenging due to the mobility of drones and the need for comprehensive spatial coverage of the monitored area. In this paper, we proposed a

framework that adapts an existing data completeness model to the context of drone-based waste detection, with the goal of improving the accuracy and reliability of waste detection results.

It is important to note the potential challenges and opportunities related to data quality and completeness in this context. One challenge is the need to ensure compatibility and integration of different sensors and technologies used by different drones to collect data in parallel, which could lead to inconsistencies in the data collected. Another challenge is related to real-time data processing and analysis in particular for the computer-vision part of such a system. However, drone-based solutions also offer opportunities to improve data quality and completeness, such as the ability to monitor previously inaccessible areas and collect data at higher frequencies and resolutions. Overall, addressing these challenges and leveraging these opportunities is critical to developing more effective waste management policies and practices, and ultimately contributing to a cleaner and healthier environment.

References

1. Aquino, G.R.C.D., de Farias, C.M., Pirmez, L.: Hygieia: data quality assessment for smart sensor network. In: Hung, C., Papadopoulos, G.A. (eds.) Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing, SAC 2019, Limassol, Cyprus, April 8-12, 2019. pp. 889–891. ACM (2019). <https://doi.org/10.1145/3297280.3297564>
2. Batini, C., Scannapieca, M.: Data-Centric Systems and Applications: Data Quality Concepts, Methodologies and Techniques. Springer (2006)
3. Batini, C., Scannapieco, M.: Data and Information Quality - Dimensions, Principles and Techniques. Data-Centric Systems and Applications, Springer (2016). <https://doi.org/10.1007/978-3-319-24106-7>
4. Brahem, M., Hafyani, H.E., Mehanna, S., Zeitouni, K., Yeh, L., Taher, Y., Kedad, Z., Ktaish, A., Chachoua, M., Ray, C.: Data perspective on environmental mobile crowd sensing. In: Intelligent Environmental Data Monitoring for Pollution Management, pp. 269–288. Elsevier (2021)
5. Dasu, T., Duan, R., Srivastava, D.: Data Quality for Temporal Streams. Tech. rep. (2016)
6. Liu, C., Nitschke, P., Williams, S., Zowghi, D.: Data quality and the internet of things. Computing (07 2019). <https://doi.org/10.1007/s00607-019-00746-z>
7. Mehanna, S., Berchiche, N., Niare, A.B., Maaradji, A., Hacid, H., Soukane, A.: Towards real-time image mining for waste detection in deserts. In: International Innovation and Technological Advances for Sustainability Conference (ITAS 2023) from March 1-3, 2023 in Doha, Qatar. SciTePress (2020)
8. Mehanna, S., Kedad, Z., Chachoua, M.: Completeness issues in mobile crowd-sensing environments. In: 16th International Conference on Web Information Systems and Technologies. SciTePress (2020)
9. Mehanna, S., Kedad, Z., Chachoua, M.: Completeness assessment and improvement in mobile crowd-sensing environments. SN Computer Science **3**(3), 216 (2022)
10. Seo, S., Mohegh, A., Ban-Weiss, G., Liu, Y.: Automatically inferring data quality for spatiotemporal forecasting. In: International Conference on Learning Representations (2018), <https://openreview.net/forum?id=ByJIWUnpW>
11. Sidi, F., Panah, P.H.S., Affendey, L.S., Jabar, M.A., Ibrahim, H., Mustapha, A.: Data quality: A survey of data quality dimensions. In: Mahmod, R., Abdullah, R., Abdullah, L.N., Sembok, T.M.T., Smeaton, A.F., Crestani, F., Doraisamy, S., Kadir, R.A., Ismail, M. (eds.) 2012 International Conference on Information Retrieval & Knowledge Management, Kuala Lumpur, Malaysia, March 13-15, 2012. pp. 300–304. IEEE (2012). <https://doi.org/10.1109/InfRKM.2012.6204995>