

Article

A Multimodal Fusion System for Object Identification in Point Clouds with Density and Coverage Differences

Daniel Fernando Quintero Bernal ^{*}, John Kern  and Claudio Urrea 

Electrical Engineering Department, Faculty of Engineering, University of Santiago of Chile (USACH), Av. Víctor Jara 3519, Estación Central, Santiago 9170124, Chile; john.kern@usach.cl (J.K.); claudio.urrea@usach.cl (C.U.)

* Correspondence: daniel.quintero@usach.cl

Abstract: Data fusion, which involves integrating information from multiple sources to achieve a specific objective, is an essential area of contemporary scientific research. This article presents a multimodal fusion system for object identification in point clouds in a controlled environment. Several stages were implemented, including downsampling and denoising techniques, to prepare the data before fusion. Two denoising approaches were tested and compared: one based on neighborhood technique and the other using a median filter for each “x”, “y”, and “z” coordinate of each point. The downsampling techniques included Random, Grid Average, and Nonuniform Grid Sample. To achieve precise alignment of sensor data in a common coordinate system, registration techniques such as Iterative Closest Point (ICP), Coherent Point Drift (CPD), and Normal Distribution Transform (NDT) were employed. Despite facing limitations, variations in density, and differences in coverage among the point clouds generated by the sensors, the system successfully achieved an integrated and coherent representation of objects in the controlled environment. This accomplishment establishes a robust foundation for future research in the field of point cloud data fusion.

Keywords: density differences; LiDAR; multimodal fusion; object identification; point clouds; point coverage



Citation: Quintero Bernal, D.F.; Kern, J.; Urrea, C. A Multimodal Fusion System for Object Identification in Point Clouds with Density and Coverage Differences. *Processes* **2024**, *12*, 248. <https://doi.org/10.3390/pr12020248>

Academic Editor: Wen-Jer Chang

Received: 28 December 2023

Revised: 20 January 2024

Accepted: 22 January 2024

Published: 24 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Data Fusion (DF) is a significant field in science that helps extract valuable information from raw data, reduces the dimensions and volume of the data, and improves the data flow. In recent years, there has been a drastic increase in the use of DF techniques, probably driven by the explosion of data from the Internet of Things (IoT) integrated into a wide range of real scenarios such as education, agriculture, and transportation [1,2]. Other factors that could be contributing are the availability of low-cost sensors [3] and more widespread access to specialized hardware and software. DF can be defined as a system composed of a data source, operation (technique), and purpose (regression, classification, dimensionality reduction, and clustering) [4,5]. DF is known as Multi-Modal Data Fusion (MMDF) when multiple heterogeneous data acquired from different modalities [6,7] are integrated to obtain more useful information for a specific task. The modalities can have different dimensionalities and temporalities. Dimensionality refers to 1D data (e.g., data without spatial designation), 2D data (e.g., images or videos), and 3D data (e.g., point clouds). Temporality refers to static data (e.g., data without temporal designation) or time series (e.g., video or data sampled at regular intervals) [8]. The developments presented in this document focus on the fusion of high-dimensional static data (3D) with a total of four modalities, including unorganized point clouds, where each set of coordinates represents an individual point forming $M \times 3$ matrices.

Researchers like Yang et al. offered an improved approach to fusing scattered point clouds and images in autonomous vehicles. They developed a deep learning framework called ‘pointpillar’ to frame the point clouds and calibrate the Light Detection and Ranging’s

(LiDAR) coordinate systems, as well as the camera (by projecting the LiDAR point clouds onto the camera image) [9]. Bracci et al. discussed the challenges of fusing heterogeneous point clouds, highlighting discrepancies in density and location accuracy. They propose a standard ICP algorithm and surface smoothing with a fifth-order polynomial, and suggest that semantic information can enhance the ICP results [10]. Swetnam et al. examined the fusion of data from multiple platforms for ecosystem monitoring, identifying detection biases between LiDAR and Structure from Motion (SfM). They employ three platforms (small Unmanned Aerial Systems, manned aircraft, and ground-based) for their study [11].

Bokade et al. observed that as the number of considered modalities increases, the scientific publications become scarcer. They indicate a dominant use of 1D and 2D static data [8]. P. Ghamisi et al. [12] suggested that merging datasets from diverse sources can produce more accurate results than using a single source of information, and that future research on the fusion of multiple sources involving point clouds should focus on integrating point clouds from different sources with significant differences in features, such as point density and three-dimensional accuracy. Cheng et al. [13] pointed out that inconsistencies in 3D data, due to the use of different device platforms with distinct perspectives and resolutions, generate difficulties in locating specific features of scenes in the registration process.

This study aims to examine, through quantitative and qualitative metrics, the performance of different approaches and techniques used in various stages of 3D static data fusion. This process combines information from three LiDAR devices and a ToF camera, through a box grid filter, in a system that faces certain assembly limitations, as well as differences in density and coverage of the points captured by the sensors used. The researchers conducted this study as part of a mining industry project with the aim of improving rock crushing operations. The system, which will be integrated into standard rock breakers, allows for the identification, tracking, and selective impact of rocks, optimizing mining grinding processes. This study represents an initial stage necessary to achieve this goal and harness the potential of multi-sensor technology in this important industrial sector.

2. Materials and Methods

This study explores the fusion of high-dimensional static data (3D) through unorganized point clouds, employing multiple LiDAR sensors and a ToF camera. The proposed data fusion system encompasses various stages, as outlined based on the works of Poux [14] and Xiao et al. [15]: (1) data source, (2) fusion model, (3) fusion method, and (4) application. In the following subsections, each of the stages will be described (Figure 1).

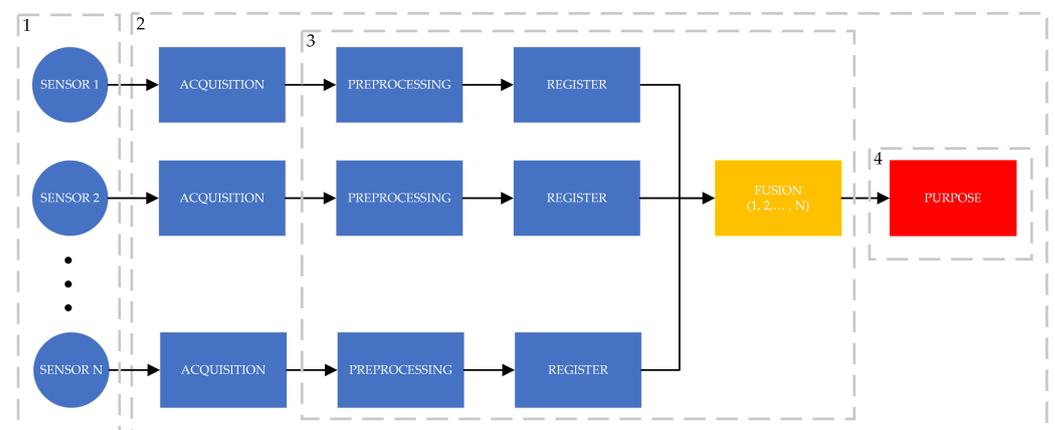


Figure 1. Stages of the proposed data fusion system.

2.1. Data Source

The alignment of the research with the objectives of the project IDeA I + D ID21I10087 has enabled the effective utilization of the LiDAR sensors and ToF cameras available in the

project. These devices were specifically chosen for their capability to generate 3D data, their high precision, and their robustness under industrial environmental conditions, which has been fundamental in achieving the research objectives. This collaboration has optimized resources and reinforced the coherence and relevance of the work, effectively integrating it within the broader framework of the project IDeA I + D ID211I0087.

Three devices with LiDAR technology and a Time of Flight (ToF) camera were employed. ToF represents the operating principle of both ToF cameras and LiDAR. ToF relies on measuring the time it takes for a signal, such as light emitted by an illumination unit, to travel to an object and return to a detection system [16,17]. Equation (1) illustrates the process for calculating the distance (d) of each point of interest on the object using the measured time ($\frac{t}{2}$) and the speed (u) of electromagnetic radiation [18]. However, a ToF camera and a LiDAR differ in their methods of light emission and detection, as well as in their spatial resolutions and typical applications.

$$d = \frac{t}{2} \times u \quad (1)$$

Table 1 provides a description of the key characteristics of the sensors, based on the information provided by the manufacturers.

Table 1. Sensor features.

Sensors	Features
OS0	A LiDAR sensor with the following characteristics: (a) $90^\circ \times 360^\circ$ field of view, (b) 35 m measurement range, (c) 0.3 m minimum measurement range, (d) 32 channels of vertical resolution, and (e) 512, 1024, or 2048 horizontal resolution. The sensor also has a typical precision of ± 0.8 cm and a maximum of ± 4 cm at 1024 at 10 Hz mode, within one standard deviation [19]
BLAZE 101	A sensor that allows image capture and 3D scanning. Here, are some features of this device: (a) $67^\circ \times 51^\circ$ field of view, (b) working range from 0.3 m to 10 m, and (c) resolution of 640×480 pixels. Additionally, the sensor exhibits an accuracy of ± 5 mm within a range of 0.5–5.5 m [20]
MRS6000	A LiDAR sensor that features: (a) 120° horizontal field of view, (b) 15° vertical field of view, (c) measurement range from 0.5 m to 200 m, and (d) angular resolution of 0.13° in the horizontal direction and 0.625° in the vertical direction [21]
MRS1000	A LiDAR sensor that offers: (a) 275° horizontal field of view, (b) 7.5° vertical field of view (with over four layers of scanning), (c) measurement limits from 0.2 m to 64 m, and (d) angular resolution of 0.25° , 0.125° , or 0.0625° [22]

The LiDAR sensors and the ToF camera used in this study exhibit notable differences in point density and spatial coverage. These variations have a significant impact on the performance and accuracy of the data fusion system. Therefore, it is crucial to conduct a thorough analysis and carefully consider these disparities during the proposed stages of the process. By appropriately addressing these differences, we ensure the attainment of more precise and reliable results in the integration of data from the various sensors.

Figure 2a illustrates a proposed controlled scenario aimed at gathering information from a rock (object of interest) located at a height of 122 cm. The sensors are positioned at an identical height of 122 cm, maintaining a horizontal distance of 100 cm from the rock. Figure 2b depicts the current physical implementation, which is based on the previously proposed scheme.

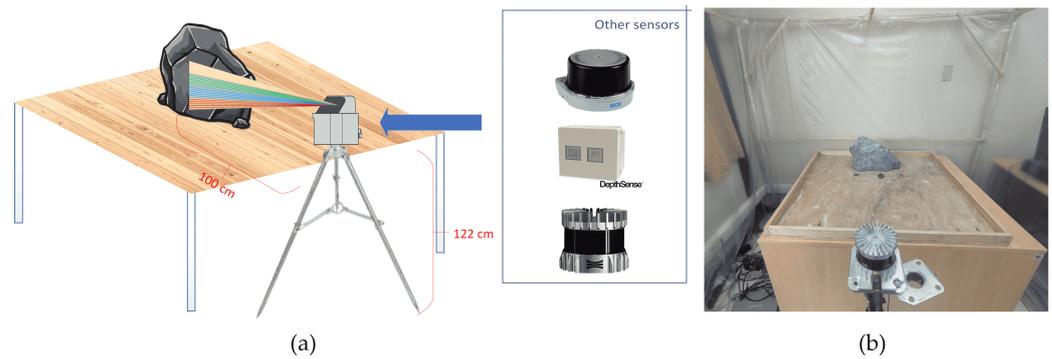


Figure 2. (a) Proposed test setup for a controlled scenario. (b) Actual assembly.

2.2. Fusion Model

The Luo and Kai architecture, depicted in Figure 3a, is utilized as the data architecture. This model operates on multiple abstraction levels (signal, feature, and decision) for multi-sensor fusion. At the signal level, raw data from sensors is directly used as input, suitable for real-time scenarios or additional preprocessing. On the feature level, relevant extracted features are fused to create enhanced features for various purposes. The decision level manages processed sensor data, producing a unified result by integrating the individual model outputs from different modalities [4,8,23].

Sensor data from $X(1)$, $X(2)$, $X(3)$, and $X(4)$ was fused as $X(1,2,3,4)$ at the signal level. Fusion occurred left to right, as shown in Figure 3a. This approach is suitable due to the data's shared structures, particularly static 3D data. Moreover, the information from each sensor is preserved, resulting in higher accuracy in the data compared to sensors operating independently [24].

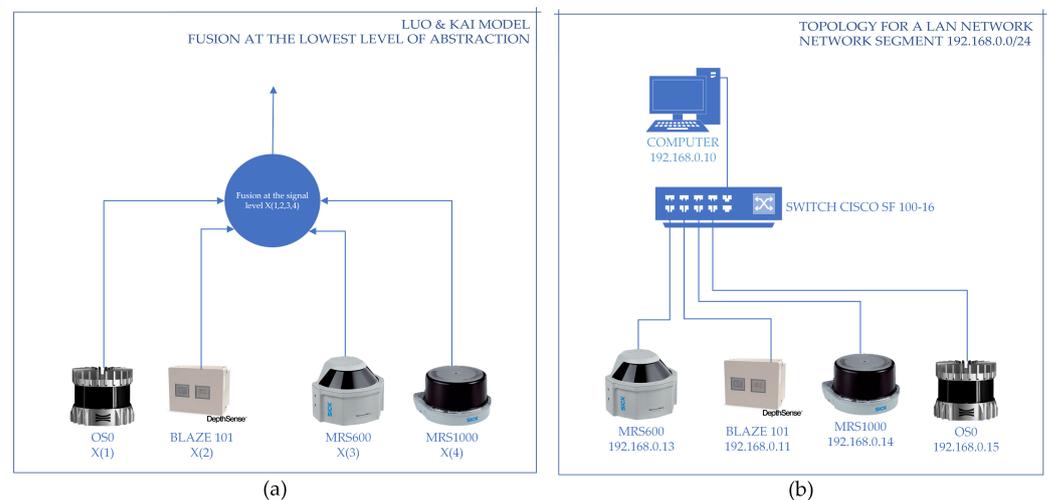


Figure 3. (a) Luo and Kai model. (b) Network topology.

The information acquired through LiDAR sensors and the ToF camera is in the form of point clouds or 3D data. A detailed definition of what constitutes a point cloud can be found in the work of Ghamisi et al. in [12]. The term “cloud” indicates the spatial coherence and disorganized nature of the set of points (with fuzzy boundaries). A point cloud is a collection of points P_i , $i = 1, \dots, n$ in three-dimensional Cartesian space. Each point in the set P_i has three positional coordinates $(x_i, y_i, z_i)^T \in \mathbb{R}^3$. Additionally, each point may have attributes $a_{i,j}$, where $j = 1, \dots, m_i$ (where the subscript j is the number of features of point i). These features can include, for example, the color of a spectral band, a component

of the local surface normal vector, or a classification or segmentation identifier). These attributes can arise from direct measurements, as well as post-processing of the data.

Point clouds fall into two categories: organized and unorganized. These categorizations depend on how point data are stored: structured or arbitrary. Unorganized point clouds consist of a single stream of three-dimensional coordinates, with each set representing an individual point in $M \times 3$ matrices (M is the total points). In contrast, organized point clouds divide data based on spatial relationships, facilitating memory–spatial correspondence ($M \times N \times 3$ matrices, M and N for height and width, respectively; the third dimension encodes features like RGB values) [25]. Given the inherent flexibility of disorganized point clouds, characterized by their versatility, simplicity, and ability to preserve the original information, this category was chosen for the execution of this study [26].

2.2.1. Data Acquisition

A network topology, as depicted in Figure 3b, was configured with four sensors and a monitoring computer interconnected through a layer-two switch, specifically the Cisco SF-100-1618. Given that the devices were linked through a switch, the resulting configuration aligns with a star network topology. The communication protocol utilized was IPv4 (RFC 791). Regarding DF, the network is considered centralized, as each sensor in use transmits its detected information to a central fusion node [27].

The sensors and the monitoring computer communicate via IP in the 192.168.0.0/24 network segment. Static addressing was used. For data capture, the Python 3.9.15 programming language was used along with the Visual Studio Code 1.74.3 text editor. The following libraries were required to obtain data from the different sensors: `ouster-sdk` 0.7.1, `harvesters` 1.4.2, and `sick_scan_api` 2.8.15.

For all four sensors, five consecutive captures were carried out. The average number of points captured for each sensor was as follows: 202,790 for the ToF BLAZE 101 camera, 32,768 for the OS0 LiDAR, 22,128 for the MRS6000 LiDAR, and 4,404 for the MRS1000 LiDAR. Corresponding characters in Figure 4 are: (a) for the ToF BLAZE 101 camera, (b) for the OS0 LiDAR, (c) for the MRS6000 LiDAR, and (d) for the MRS1000 LiDAR.

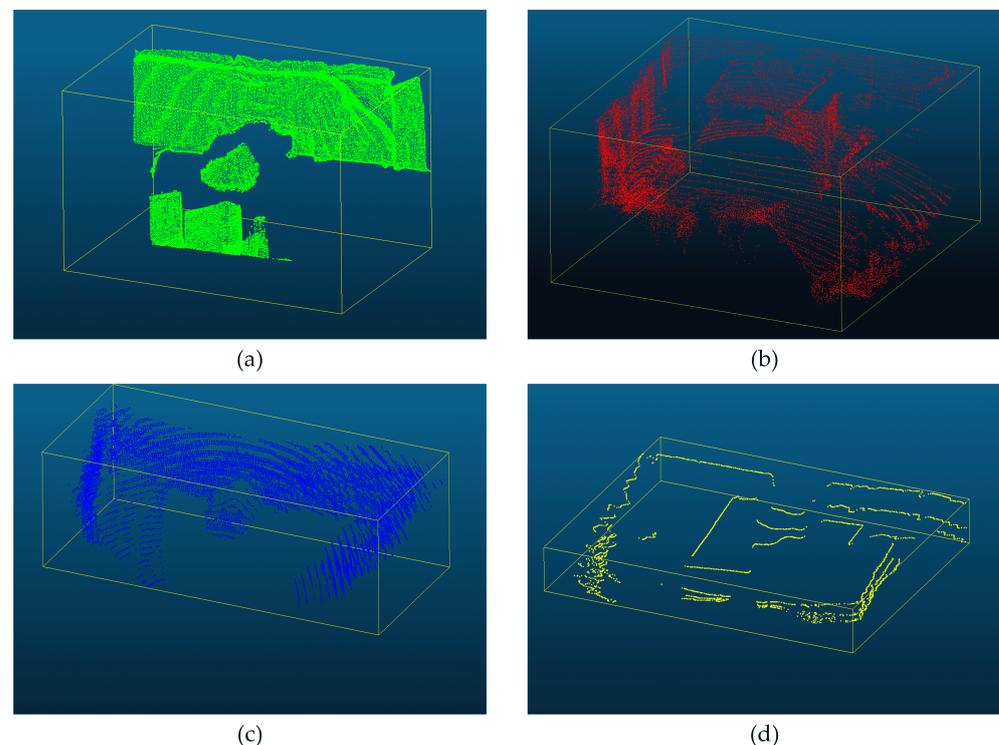


Figure 4. Capture of the 4 sensors: (a) ToF BLAZE 101 camera (BASLER), (b) OS0 LiDAR (OUSTER), (c) MRS6000LiDAR (SICK), and (d) MRS1000 LiDAR (SICK).

2.2.2. Fusion Method

This section outlines the pivotal aspects of the DF process, comprising preprocessing, registration, and applying fusion techniques. MATLAB R2023a facilitated efficient data handling and processing [4]:

- Preprocessing readies data for ensuing steps, spotlighting pertinent information. This includes processes like data cleaning, normalization, and resolution reduction.
- Data registration aligns data within a common reference frame before fusion, ensuring meaningful integration from diverse sources.
- Data fusion techniques, the core of any fusion system, involve algorithms for effective data integration, harmonizing information from various sources.

Data preprocessing: Key aspects were considered in preprocessing, notably downsampling and denoising. The approach included:

- Five captures of the scenario for each sensor (time series samples).
- Removing points at coordinates (0, 0, 0) within a one-centimeter radius and discarding invalid values (Inf or NaN coordinate values).
- Implementing denoising and downsampling in each capture.
- Subsequently, registering and combining information from the five captures of each sensor.

Table 2 offers an overview of the utilized downsampling techniques.

Table 2. Downsampling techniques.

Techniques	Description
Random	The “random” technique allows generating a randomly subsampled point cloud.
Grid Average	The “grid average” technique utilizes a box grid filter to generate a condensed point cloud. This process involves defining a bounding cube that encapsulates the complete point cloud and partitioning it into grid cells of predetermined dimensions. Points within each grid cell are aggregated by calculating the average of their coordinates, normal properties, and colors.
Nonuniform Grid Sample	The “nonuniform grid sample” technique achieves point cloud reduction by utilizing a nonuniform box grid filter. This approach involves randomly selecting a single output point from each box when a maximum number of points is set for each nonuniform grid.

Lastly, Table 3 outlines the approaches used for outlier removal during the analysis.

Table 3. Denoising techniques.

Techniques	Description
First approach	An implemented filter calculates the average distance to nearest neighboring points within a predefined neighborhood for each point. Evaluation follows to classify points as outliers. Points with an average distance to their nearest neighbors exceeding a set threshold, defined as a standard deviation from the mean of all points’ average distances, are identified as outliers.
Second approach	A median filter is separately applied to the “x”, “y”, and “z” coordinates of each point, resulting in a filtered point cloud. The filtered value of each point is determined as the median of its neighborhood’s values, with no zero-padding at the cloud edges. This approach operates solely with available points and information within each point’s neighborhood.

Data registration: The proposed system involved two sets of point cloud registrations conducted at different stages. Initially, during preprocessing, registration, and fusion of five files from each sensor were performed to accumulate a more informative scenario

perspective. Subsequently, another registration of the point clouds from the four sensors was executed to capture complementary views from these devices (this corresponds to the second registration phase).

Point cloud registration studies commonly utilize a coarse-to-fine registration strategy. Coarse registration defines initial parameters for transformation between two point clouds through feature-based methods (points, lines, surfaces, or combinations). In contrast, fine registration aims for maximal overlap using iterative approximation techniques [13,28,29].

For the first registration, coarse registration was not needed as captures from the same sensor share similar coordinate systems. In the second registration, considering system knowledge, coarse registration was performed by transforming BLAZE101, MRS6000, and MRS1000 sensor coordinates relative to the OS0 sensor. This transformation employed a pre-defined transformation matrix (Equation (2)).

$$\begin{bmatrix} R_{11} & R_{12} & R_{13} & T_x \\ R_{21} & R_{22} & R_{23} & T_y \\ R_{31} & R_{32} & R_{33} & T_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2)$$

If a point P is selected from any point cloud in the database, the transformed point, denoted as P' , can be derived using Equation (3). Here, T represents the translation vector, and R stands for the rotation matrix of international residents.

$$P' = RP + T \quad (3)$$

Table 4 outlines the techniques employed for fine registration in both proposed registrations [13].

Table 4. Register techniques.

Techniques	Description
ICP	The ICP Algorithm uses quaternion methodology, employing a 4D vector to encode rotation and angle parameters. This approach directly addresses rigid body transformation through a robust mathematical process. By selecting points from the data set and finding their corresponding points in the reference set, the algorithm minimizes the distance between these pairs to achieve transformation. The procedure iterates, recalculating the closest points until the objective function stabilizes, resulting in the registered data.
CPD	The CPD Algorithm aims to minimize the discrepancy between input point clouds by finding the optimal rigid transformation. Based on probability theory, CPD assumes coherent point distribution according to an underlying probability distribution. Using iterative methods, the algorithm estimates transformation and adjusts the probability distribution for precise alignment. This approach is valuable for handling incomplete, noisy, or differently dense data.
NDT	The NDT Algorithm transforms point cloud data into a differentiable 3D grid using a continuous probability distribution. This function models the probability of each point's position within the grid cell using a normal distribution. The Hessian matrix optimizes the normal distribution probability, enabling point cloud registration.

DF technique: The box grid filter was utilized for data fusion, allowing the fusion of point clouds within a specified 3D frame. This filter divides the three-dimensional space into uniform cells, associating each point from the clouds with its respective cell. Subsequently, an analysis is conducted within each cell to create a unified representation of its contained points, often achieved through averaging techniques [30–32]. Throughout this study, two data fusion instances were executed during the point cloud registration stages. The initial fusion

transpired after registering the five captures of each sensor in the preprocessing phase. At this point, the individual sensor point clouds were integrated to establish a unified environmental representation. Subsequently, the second fusion capitalized on the complementary perspectives of the four sensors, further refining the representation of the scenario through the combination of the previously fused point clouds (Figure 5).

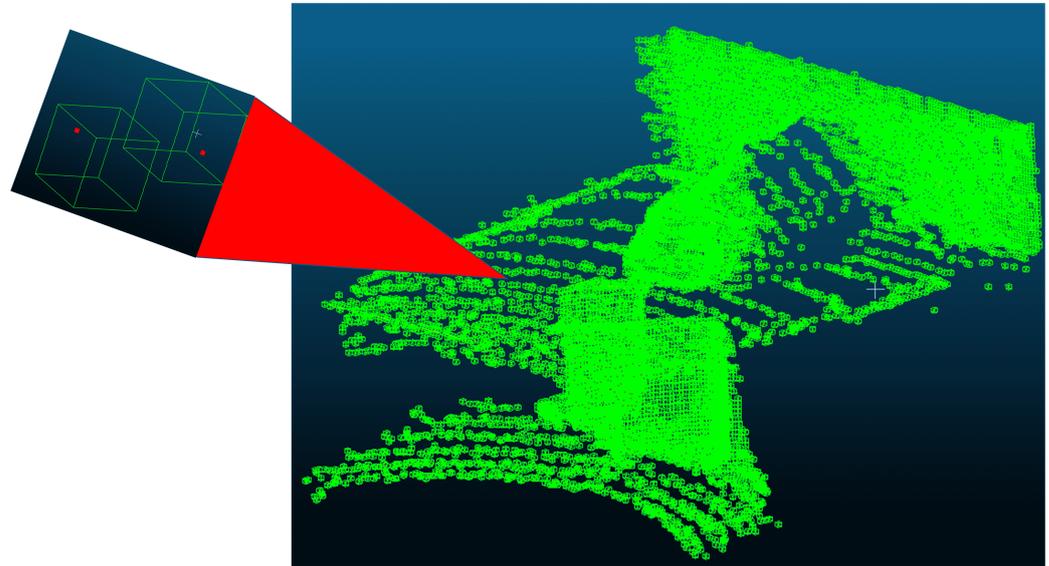


Figure 5. Data fusion performed with a box grid Filter. The image illustrates the behavior of the filter, with individual cells represented in green.

2.2.3. Application

The IDeA I + D ID21I10087 project focuses on the mining industry with the aim of improving rock crushing operations. The proposed system, intended for integration in standard breaker hammers, is designed to facilitate the identification, tracking, and precise selection of rock impacts. This innovation is expected to optimize processes at mining plants, contributing to increased efficiency and reduced costs. Moreover, the implementation of a standardized kit will allow easy integration into various types of machinery, capitalizing on multi-sensor technology to enhance productivity and quality in this vital sector. The results presented in this document, in line with the specific objectives of IDeA I + D ID21I10087, represent an essential initial phase for future analyses. These will focus on selecting suitable 3D dimensionality modalities for environments with dust, light variations, and fog. Later, 2D and 1D dimensionality modalities will be integrated to enrich the MMDF system. Upon the complete development of the MMDF system, various methods to determine the most effective impact points based on the specific shapes of rocks will be explored.

Parallel research has delved into similar domains. Lampinen et al. introduced an autonomous rock-breaking system validated in a large-scale mining environment. This setup involves a commercial rock breaker equipped with high-precision joint angle encoders and a 3D visual perception system based on YOLOv3. This autonomous approach minimizes human intervention, advancing automation in mining. However, their research indicated that the sensorization system might not be suitable for the challenging environmental conditions expected in the target application [33]. Another pertinent investigation by Correa et al. proposed a strategy to address issues associated with impact hammer operations in underground mining. These challenges include ore passage blockages, leading to fragmentation delays and hindering the mining production chain. Traditional hammer operation also faces operator perception limitations, as visual camera-based sensors offer only a 2D representation of the environment. Notably, their experiments were conducted

under laboratory conditions with two LiDARs, limiting the assessment of adverse condition impact on result accuracy [34].

3. Results

This section presents the results obtained in the various stages of the MMDF system, which include downsampling, denoising, registration, and fusion. Throughout the study, various techniques and approaches were tested and evaluated for each of these stages, with the goal of determining the most effective ones for the proposed system. For data analysis, a combination of the open-source software CloudCompare version 2.13.alpha and the Matlab platform was used, tools that facilitated the generation of the metrics presented in this section.

It is important to highlight that the results presented in this document represent a key initial phase of our research, which is aligned with the objectives of the project IDeA I + D ID21I10087. The current focus is on selecting modalities for sensors that provide 3D information, such as LiDAR and ToF cameras, in environments where visibility may be affected by dust, fog, or extreme light variations. The fusion of point clouds within this approach will allow for precise data analysis in complex mining environments. Additionally, as part of later stages, it is planned to integrate 1D and 2D dimensionality modalities to develop an advanced data fusion architecture. This will facilitate the precise identification of impact points in rocks for mining comminution tasks, thereby contributing to improved operations and efficiency in the mining sector.

Before employing downsampling and denoising, initial filtering was conducted on the point clouds to remove invalid data points and those located at coordinates (0, 0, 0). The point cloud from LiDAR MRS1000 witnessed a reduction of 25.84%, while LiDAR OS0 experienced a decrease of 16.66%. In contrast, LiDAR MRS6000 saw a reduction of 1.65%, and the ToF camera remained unaltered at 0%.

3.1. Downsampling

Experimental individual tests were conducted on several captures to ascertain the parameters of each utilized downsampling technique (Table 5):

- Random technique: A parameter was defined to govern the proportion of input data to be retained by the function. Visual inspection guided the determination of the reduction value for the point cloud that balanced the data's representativeness with the object of interest's clarity, specific to the used sensors.
- Grid Average technique: A parameter was set to dictate the size of each cell within the three-dimensional grid. An empirical value of 1 cm (0.01) was adopted based on the dimensions of the point clouds derived from the four sensors.
- Nonuniform Grid Sample technique: A parameter of 6 was chosen, reflecting the maximum number of points permissible within each grid cell during the resolution reduction process. This value was selected as it meets the minimum requirement for applying this technique.

Table 5. Downsampling technique parameters and grid average results for four sensors.

Sensors	Random	Grid Average (m)	Nonuniform Grid Sample	Average Points	Points Removed
OS0	0.8	0.01	6	27.309	3.493
BLAZE 101	0.4	0.01	6	202.785	176.033
MRS6000	0.8	0.01	6	21.763	9.162
MRS1000	0.8	0.01	6	3.266	1.165

The choice of the downsampling technique was based on two essential criteria: the data distribution histogram and the correlation between visual inspection and the reduction in the number of points. For the histogram, the number of neighbors that a point has within

a radius of 1 cm (0.01) was considered. Figure 6 presents four histograms of a capture from the ToF camera, to which the three previously mentioned preprocessing techniques were applied. With the Grid Average method, a favorable balance was achieved between the representation of points on the object of interest and their reduction.

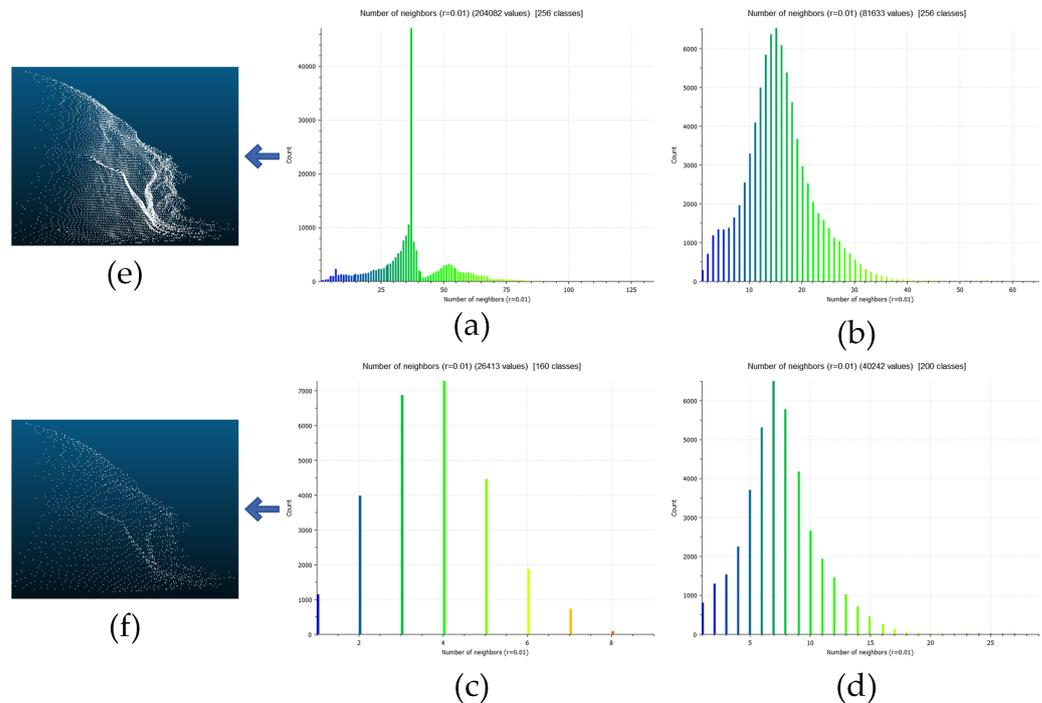


Figure 6. Block diagram of the system components. The figure shows the histogram and the amount of data obtained from a capture of the ToF camera (a), to which three downsampling techniques were applied: (b) Random, (c) Grid Average, and (d) Nonuniform GridSample. In (e), the original point cloud is visualized, and in (f), the subsampled point cloud using the Grid Average method is displayed.

Table 5 shows the results obtained from the sensor captures when applying the Grid Average technique during downsampling. A reduction of 86.81% was achieved for the ToF camera, 42.10% for the LiDAR MRS6000, 35.67% for the LiDAR MRS1000, and 12.79% for the LiDAR OS0.

3.2. Denoising

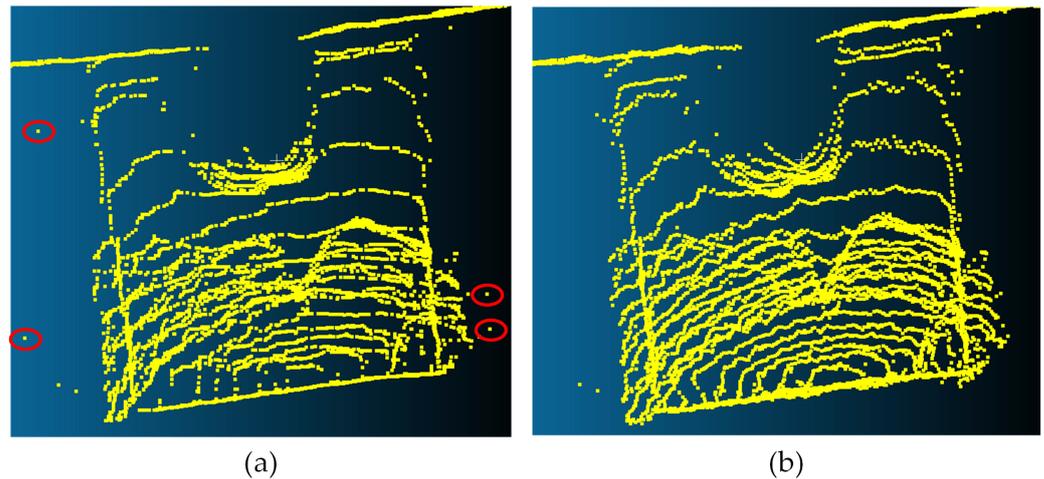
Concerning the two adopted denoising methods, the criteria underlying the parameter selection for each of them are outlined below (Table 6):

- First approach: A neighborhood parameter was assigned to compute the average distance to the nearest neighbors of all points. To avoid excessive data reduction—an objective incongruent with this preprocessing phase, where the focus is solely on outlier elimination—a neighborhood value of 1 was chosen.
- Second approach: Given the point cloud's unorganized nature, a radial neighborhood method was employed, forming a sphere around each point. For the radius value, a percentage of the smallest dimension dispersion of the point cloud was selected to maintain accuracy while reducing noise.

The first approach was favored due to its visual alignment with the goal of retaining object-of-interest points while eliminating outliers (Figure 7). Table 6 displays the outcomes of employing the first denoising approach on sensor captures. Reduction percentages were as follows: 4.21% for LiDAR MRS6000, 1.43% for LiDAR MRS1000, 0.99% for LiDAR OS0, and 0.43% for the ToF camera.

Table 6. Denoising approach parameters and average filtered points for four sensors.

Sensors	First, Approach		Second, Approach (m)	Average Points	Average Filtered Points
	First, Parameter	Second, Parameter			
OS0	1	1	0.0481	23.817	235
BLAZE 101	1	1	0.0433	26.752	114
MRS6000	1	1	0.0520	12.600	530
MRS1000	1	1	0.0512	2.101	30

**Figure 7.** Point cloud from the OS0 sensor, filtered through two denoising approaches. (b) First approach and (a) second approach.

3.3. Registration

As previously established, it was not necessary to perform a coarse registration for the first type of registration due to the similarity of the coordinates obtained from the same sensor. However, about the second type of registration, a transformation matrix based on prior empirical knowledge of the assembly was used. The transformation matrices for the BLAZE101, MRS6000, and MRS1000 sensors about the OS0 sensor can be consulted in Equation (4), Equation (5), and Equation (6), respectively. It is important to note that due to the specific capture characteristics of the LiDAR OS0 sensor, it was necessary to raise this sensor approximately 20 cm above compared to the other sensors. Before performing the empirical registration, a translation was performed on the “z” axis for the point clouds of the OS0 sensor, which can be observed in Equation (7).

$$\begin{bmatrix} 0.9972 & -0.0663 & -0.0348 & 0.0140 \\ 0.0662 & 0.9978 & -0.0023 & 0 \\ 0.0349 & 0 & 0.9994 & 0.0260 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (4)$$

$$\begin{bmatrix} 0.9972 & -0.0663 & -0.0348 & 0.0140 \\ 0.0662 & 0.9978 & -0.0023 & 0 \\ 0.0349 & 0.9994 & 0 & 0.0260 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (5)$$

$$\begin{bmatrix} 0.9972 & -0.0663 & -0.0348 & 0.0140 \\ 0.0662 & 0.9978 & -0.0023 & 0 \\ 0.0349 & 0 & 0.9994 & 0.0260 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (6)$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0.1850 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (7)$$

The results of the first fine registration, which involved the registration of five captures from the same sensor, are shown in Table 7. The chosen parameters for these tests are outlined below:

- ICP: All parameters were retained at their default values.
- CPD: All parameters were maintained at their default settings. However, the registration aimed to achieve a rigid transformation to prevent any deformation in the analyzed objects.
- NDT: All parameters were kept at their default values. For this technique, it was necessary to specify the dimension of the three-dimensional cube that voxelized the reference point cloud (captures from the OS0 sensor) as a positive scalar number. The following values were assigned: 0.5 for the OS0 sensor, and 0.1 for the BLAZE 101, MRS6000, and MRS1000 sensors.

Table 7. First registration. RMSE error of the three techniques used for fine registration performed for the four sensors.

Sensors	ICP (m)	CPD (m)	NDT (m)
OS0	0.0187	0.0174	0.0178
BLAZE 101	0.0029	0.0033	0.0037
MRS6000	0.0092	0.0090	0.0093
MRS1000	4.4457×10^{-16}	3.2150×10^{-16}	7.8324×10^{-5}

According to the RMSE values obtained, the CPD technique was selected for the registration process of the point clouds of the three LiDARs. For the ToF camera, the ICP technique was chosen for registration.

In the second point cloud registration phase, a fine registration test labeled as “A” was conducted, retaining the conditions outlined earlier. However, for the NDT technique, the default cube size of 0.5 was employed (Table 8). Despite achieving minimal RMSE error through the ICP registration method to align Blaze-101 sensor data with respect to the OS0 sensor, the outcomes did not align with anticipated expectations. Consequently, an additional test, labeled “B”, was undertaken. In test “B”, the transformation matrix from coarse registration was employed as an initial matrix for fine registration. This approach was exclusive to ICP and NDT techniques due to their compatibility with this calculation approach. The goal was to enhance registration accuracy and achieve more satisfactory results.

Table 8. Second registration. RMSE error of the three techniques used for fine registration performed on the sensors with respect to the LiDAR OS0.

Sensors	ICP (m)		CPD (m)	NDT (m)	
	A	B	A	A	B
OS0—BLAZE 101	0.1894	0.1237	0.2802	0.2367	0.2713
OS0—MRS6000	0.1583	0.1027	0.2867	0.2382	0.1401
OS0—MRS1000	0.1637	0.0600	0.2382	0.1858	0.0665

Despite test “B” achieving a significant reduction in RMSE and improving data alignment during fine registration, it was determined that the results did not fully meet the expectations set for this stage of the process. Consequently, it is acknowledged that test

“B” represents a promising methodology for the automation of registration, but further research is needed to tailor it to the specific requirements of this study.

3.4. Fusion

For the first fusion, a three-dimensional frame dimension was chosen in the filter that did not increase the average number of points of the captures of each sensor. As for the second fusion, a default value of 1 cm was used, considered appropriate based on the spatial distribution of the point clouds used (Table 9).

Table 9. Three-dimensional frame dimension adjusted for the four sensors during the two point cloud fusions performed.

Sensors	First Fusion (m)	Second Fusion (m)
OS0	0.026	0.010
BLAZE 101	0.010	0.010
MRS6000	0.015	0.010
MRS1000	0.010	0.010

Figure 8a presents a detailed isometric view, highlighting the data fusion process from the four sensors. This view emphasizes the comprehensive integration of sensor data. Figure 8b offers a frontal view of the object of interest within the controlled environment, serving as a basis for comparison. Finally, Figure 8c shows the frontal view of the result achieved after the fusion of data from the four sensors.

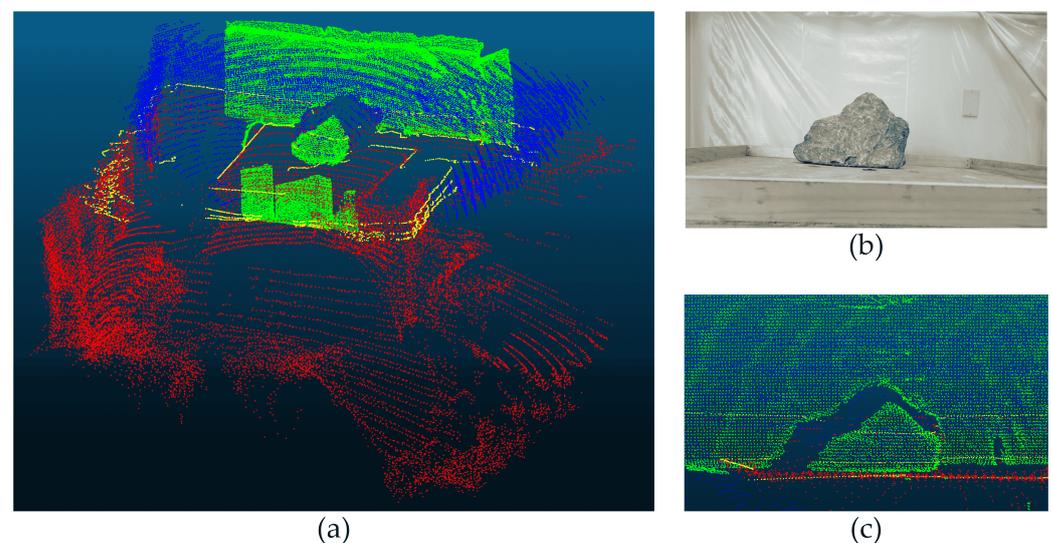


Figure 8. (a) Isometric view of the fusion of the four sensors used. (b) Front view of the object of interest in the controlled scenario. (c) Front view of the fusion of the four sensors used.

4. Discussions

In this section, we analyze the outcomes of the point cloud fusion process in the multimodal system. This analysis encompasses quantitative and qualitative results, along with discussions about limitations and avenues for future enhancements:

- Unorganized point clouds have attributes such as generality, simplicity, and the capacity to preserve original data. However, they possess limitations like irregular and dispersed distribution, hampering efficient processing and analysis due to the lack of order. This can undermine computational efficiency and hinder the utilization of inherent geometric properties, limiting advanced analysis and technique application based on data structure.

- Preprocessing, involving the removal of invalid values and origin-based points, is crucial. This step showcased point cloud reduction without distorting the object's general shape in the scene.
- In the downsampling phase of the MMDF system, various techniques were evaluated to identify the most suitable one. Three main methods were considered: Random, Grid Average, and Nonuniform Grid Sample. The Random Technique proved efficient in reducing the density of the point cloud. Meanwhile, the Grid Average Method was particularly effective in preserving the shape of the point cloud. The Nonuniform Grid Sample Technique, in turn, maintained the accuracy of normals through the original data, crucial for applications dependent on this information. The selection of the downsampling technique was based on an analysis of the data distribution histogram and the correlation between visual inspection and point reduction. Among these, the Grid Average method was found to be the most effective for the system, achieving an ideal balance between maintaining the clarity of the object of interest and the representativeness of the data.
- With regard to the adopted denoising methods, two main approaches were used. The first approach employed a neighborhood parameter to calculate the average distance to the nearest neighbors of each point, choosing a neighborhood value of 1 to avoid excessive data reduction and focusing solely on outlier elimination. This method was preferred due to its visual alignment with the goal of retaining object-of-interest points while eliminating outliers. The reduction percentages with this approach were 4.21% for LiDAR MRS6000, 1.43% for LiDAR MRS1000, 0.99% for LiDAR OS0, and 0.43% for the ToF camera. The second approach, particularly effective when point density is high, retained the point count but smoothed the data to remove outliers, using a radial neighborhood method that formed a sphere around each point. Despite its effectiveness, this method struggled with dispersed or distant outliers.
- In the first registration, no algorithms were needed for coarse registration due to similar coordinate systems among same-sensor samples. While the CPD technique demonstrated a slightly lower RMSE error in the first fine registration, no significant distinction was apparent when compared to the ICP and NDT techniques. Furthermore, it is important to highlight that the CPD technique resulted in higher computational costs. For the second coarse registration, manual adjustment of BLAZE-101, MRS6000, and MRS1000 sensor point clouds relative to the LiDAR was performed based on system knowledge. The study employed the LiDAR OS0 as a reference due to its suitable point density and coverage. In the second fine registration tests (A and B), expectations were not fully met, but test B showed promise in addressing discrepancies in the first (test A) and achieving better alignment of Blaze-101, MRS100, and MRS6000 sensors with OS0. Additionally, MRS1000 and OS0 alignment was observed, mainly due to similar data capture in certain areas, aiding registration.
- This study is part of the IDeA I + D ID21I10087 project, which focuses on improving rock crushing operations in the mining industry. Our work significantly contributes to this objective by implementing multiple sensor technology for precise rock identification and selection, even under challenging visibility conditions such as high and low light, fog, and dust. The results obtained are a crucial initial phase, laying the foundation for future analyses focused on selecting appropriate 3D dimensionality modalities for these challenging environments. Additionally, the integration of 2D and 1D dimensionality modalities will be explored, with the aim of fully developing the MMDF system and determining the most effective impact points based on the specific shapes of the rocks.

5. Conclusions

The article has successfully presented a multimodal point cloud fusion system that efficiently integrates LiDAR sensors and a ToF camera. This system has demonstrated its ability to effectively identify objects in a controlled environment, even with significant

differences in point density and coverage. The techniques used for resolution reduction and noise removal have been key to improving data quality before fusion, which in turn has increased the system's accuracy. Furthermore, the registration techniques have ensured precise alignment of sensor data in a common coordinate system, which is essential for a coherent and detailed representation of objects.

The system's ability to effectively merge data opens new possibilities for comprehensive views and detailed analysis. This includes identifying suitable impact areas in the crushing of mining rocks and the possibility of conducting analysis for the selection of modalities in adverse conditions, as well as evaluating the integration of 1D and 2D dimensionality modalities.

This system holds great potential for industrial applications, especially in improving milling processes, aiming for higher efficiency and cost reduction. Looking forward, there is an expectation to expand the system's use to more diverse and complex environments, significantly broadening its applicability. Future research will focus on exploring more sensors and developing advanced techniques to further improve precision and potential in object identification.

Author Contributions: Conceptualization, D.F.Q.B. and J.K.; methodology, D.F.Q.B. and J.K.; software, D.F.Q.B. and J.K.; validation, D.F.Q.B. and J.K.; formal analysis, D.F.Q.B. and J.K.; investigation, D.F.Q.B. and J.K.; resources, D.F.Q.B. and J.K.; data curation, D.F.Q.B. and J.K.; writing—original draft preparation, D.F.Q.B. and J.K.; writing—review and editing, D.F.Q.B. and J.K.; visualization, D.F.Q.B. and J.K.; supervision, J.K. and C.U.; project administration, J.K. and C.U.; funding acquisition, J.K. and C.U. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been supported by Agencia Nacional de Investigación y Desarrollo (ANID), Chile, through the IDeA I + D ID21I10087 project and the Vicerrectoría de Investigación, Innovación y Creación of the University of Santiago of Chile (USACH), Chile.

Data Availability Statement: The datasets presented in this article are not readily available because the data are part of an ongoing study.

Acknowledgments: We acknowledge and appreciate the support received from ANID-Subdirección de Capital Humano/Doctorado Nacional/2022-21220739 for this research.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Ding, W.; Jing, X.; Yan, Z.; Yang, L.T. A survey on data fusion in internet of things: Towards secure and privacy-preserving fusion. *Inf. Fusion* **2019**, *51*, 129–144. [\[CrossRef\]](#)
- Ullah, I.; Youn, H.Y. Intelligent Data Fusion for Smart IoT Environment: A Survey. *Wirel. Pers. Commun.* **2020**, *114*, 409–430. [\[CrossRef\]](#)
- Gramsch, E.; Oyola, P.; Reyes, F.; Vásquez, Y.; Rubio, M.A.; Soto, C.; Pérez, P.; Moreno, F.; Gutiérrez, N. Influence of particle composition and size on the accuracy of low cost PM sensors: Findings from field campaigns. *Front. Environ. Sci.* **2021**, *9*, 751267. [\[CrossRef\]](#)
- Meng, T.; Jing, X.; Yan, Z.; Pedrycz, W. A survey on machine learning for data fusion. *Inf. Fusion* **2020**, *57*, 115–129. [\[CrossRef\]](#)
- Canalle, G.K.; Salgado, A.C.; Loscio, B.F. A survey on data fusion: What for? In what form? What is next? *J. Intell. Inf. Syst.* **2021**, *57*, 25–50. [\[CrossRef\]](#)
- Zhang, Y.; Sidibé, D.; Morel, O.; Mériaudeau, F. Deep multimodal fusion for semantic image segmentation: A survey. *Image Vis. Comput.* **2021**, *105*, 104042. [\[CrossRef\]](#)
- Chen, P.Y.; Lin, H.Y.; Pai, N.S.; Huang, J.B. Construction of Edge Computing Platform Using 3D LiDAR and Camera Heterogeneous Sensing Fusion for Front Obstacle Recognition and Distance Measurement System. *Processes* **2022**, *10*, 1876. [\[CrossRef\]](#)
- Bokade, R.; Navato, A.; Ouyang, R.; Jin, X.; Chou, C.A.; Ostadabbas, S.; Mueller, A.V. A cross-disciplinary comparison of multimodal data fusion approaches and applications: Accelerating learning through trans-disciplinary information sharing. *Expert Syst. Appl.* **2021**, *165*, 113885. [\[CrossRef\]](#)
- Yang, Q.; Liu, F.; Qu, J.; Jing, H.; Kuang, B.; Chai, W. Multi-sensor fusion of sparse point clouds based on neuralnet works. In *Journal of Physics: Conference Series*; IOP Publishing: Bristol, UK, 2022; Volume 2216, p. 012028.
- Bracci, F.; Drauschke, M.; Kühne, S.; Márton, Z.C. Challenges in fusion of heterogeneous point clouds. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *42*, 155–162. [\[CrossRef\]](#)

