IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 24 June 2024, accepted 23 July 2024, date of publication 31 July 2024, date of current version 15 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3435993

RESEARCH ARTICLE

Semantic Segmentation and Construction of a DataSet From a 3D Point Cloud Obtained by LiDAR Sensor

JHON SEBASTIAN APARICIO MESA[1], MARCO JAVIER SUÁREZ BARÓN[1], AND EDUARDO AVENDAÑO FERNÁNDEZ[2], (Senior Member, IEEE)

[1]School of Computer Engineering, Universidad Pedagógica y Tecnológica de Colombia, Sogamoso 152210, Colombia
[2]School of Electronics Engineering, Universidad Pedagógica y Tecnológica de Colombia, Sogamoso 152210, Colombia

Corresponding author: Eduardo Avendaño Fernández (eduardo.avendano@uptc.edu.co)

This work is supported by North Atlantic Treaty Organization (NATO) in the Science for Peace (SPS) program under grant G-5888, the funds are managed by the Universidad Pedagógica y Tecnológica de Colombia (UPTC) Clarifier project with SGI-3139.

ABSTRACT The use of LiDAR technologies offer an effective alternative for enhancing monitoring and surveillance tasks in secured areas, or hard-to-access areas. Conventional solutions allow detection and identification capabilities, which are valuable for object and pattern recognition in environments with special conditions. This paper proposes a method for structure a dataset by collecting data from LiDAR sensors and reconstructing 3D point cloud scenes into images. Furthermore, training with these data using semantic segmentation algorithms allows for the detecting of people in monitoring and surveillance contexts. The process involves data acquisition techniques, preprocessing, and the application of tools and technologies for semantic segmentation in the post-processing stage. Data is acquired from RPLidar sensor with a spatial resolution of less than 0.4 degrees per step, ensuring greater accuracy in detecting small objects. Scanning distances between 0.15 and 12 meters allow for the capture of 941 points per rotation angle. The dataset is complemented with labels and images containing binary masks of regions of interest for people detection using semantic segmentation. The main contribution of this paper is the methodological description of constructing a dataset comprising 1,254 images in various scenarios, including indoors and outdoors, with different levels of occlusion and illumination. This dataset facilitates the development of machine learning algorithms and deep learning algorithms for pattern recognition, such as Mask R-CNN (89.9% mean class accuracy) and U-NET (90.5% mean class accuracy). The results demonstrated an approximate 2% improvement in mean class accuracy compared to other state-of-the-art algorithms in semantic segmentation for point clouds.

INDEX TERMS DataSet, LiDAR, point cloud, person, semantic segmentation.

I. INTRODUCTION

The fourth Industrial Revolution has brought significant advancements in technology and communication, particularly in the development of artificial intelligence, big data, and the Internet of Things (IoT) among others [1]. These advancements have introduced several challenges in security, especially regarding information management in institutions and governments. In response, many countries have transformed their defense and security policies to better protect their nations [2]. One of the technologies implemented in security is LiDAR (Light Detection and Ranging) sensors. LiDAR operates by sending laser pulses to measure the time of flight (ToF) and phase until the pulse hits objects and returns to the source. This measurement estimates the distance and can create three-dimensional maps of the surrounding environment. In such scenarios, algorithms can be applied for the detection of objects, people, vehicles, animals, and more. In the context of monitoring and surveillance, LiDAR technology supports the control and identification of patterns, enabling alerts in high-security settings such as government buildings, military bases, or other

The associate editor coordinating the review of this manuscript and approving it for publication was Huiyan Zhang.

sensitive facilities. The goal is to provide a monitoring system with enhanced capabilities for the recognizing objects of interest and/or people, thereby improving overall security measures.

The use of LiDAR sensors offers significant advantages, such as high resolution, enabling the capture of more than 50 points per square meter on average. This is achieved through the time-of-flight concept, providing accurate information about distance and position. LiDAR's ability to return precise data results in an accurate representation of the environment or scanning area [3]. High-resolution LiDAR sensors typically capture at least 200 points per square meter, with some capturing over 100,000 points per second. This translates to a density of at least 10-20 points per square centimeter per sweep. Higher point density is especially required when objects are farther from the sensor, such as in topographic mapping applications in areas with dense vegetation. In these cases, the LiDAR must operate in low-light conditions and penetrate through of foliage to reach the ground level [4].

Given the dynamic nature of data capture processes, there are shortcomings when using image-based datasets, particularly for detecting people, necessitating unconventional technologies such as LiDAR and RADAR (Radio Detection and Ranging) sensors.

Reviewing the state of the art, initial mentions are of anomalies in environments affecting the recognition of elements of interest or those in a region of interest (ROI). This requires defining a classifier for objects in reference images, with the starting point for creating a dataset being the identification of objects to be detected and/or classified. Simon Gonzalez's work involves developing a classifier to detect abnormal patterns in diagnosing pneumonia, mass detection, and respiratory in chest X-ray images [5]. Structuring a suitable dataset allows for identifying or defining possible noise or distortions during training, resulting in a more robust system.

In [6] study, the importance and applicability of point clouds generated by LiDAR sensors in artificial intelligence, mainly in machine learning, are highlighted. Principal Component Analysis (PCA) is applied for segmentation, and a neural network is incorporated to simplify detection in various geographical environments such as vegetation, water sources, and man-made structures. This suggests using LiDAR as an advantage in detection when segmentation is applied to images.

One shortcoming of conventional cameras for image processing is their sensitivity to light intensity. LiDAR and Radar sensors are better strategies in these cases as they measure the distance between objects for proper classification, improving detection when objects are partially occluded, according to Farag [7].

Among the related datasets, Diaz's work presents CARLA, an open-source simulator for autonomous driving research. It uses a video game engine for validation in simulated environments [8], detecting objects such as people, animals, and vehicles.

Cardenas and Cevallos utilized the KITTI Road/Lane Detection dataset, which includes LiDAR sensors data for lane and road detection. This dataset allows for the classifying pedestrians, vehicles, and traffic signs, among other objects [9]. However, it does not focus exclusively on individual, limiting its application for person detection and reducing the accuracy across various metrics. Martinez's work highlights the disadvantages of this dataset, noting the lack of labels and labeling accuracy in point cloud images for some objects. This introduces a limitation in usage [10] and constitutes a major weakness in one of the most widely used datasets in the area of autonomous driving.

In the review of related work, there are few LiDAR sensor datasets available for human detection. Most are not labeled or do not include segmentation mask files. Thus, it is proposed to build a dataset using image scenes obtained with LiDAR sensors focused on people detection. This dataset will be include labeling file for each images and allow visualization of the segmentation mask related to these objects, simplifying training processes based on semantic segmentation. This work contributes to the methodological development of new research, creating structured datasets to improve security and surveillance in vulnerable scenarios.

The use of technological elements such as LiDAR offers significant advantages over traditional surveillance tools, like cameras, LiDAR technology can operate effectively under varying light conditions, including changes in contrast changes, interference, occluded objects and changes in perspective [11]. LiDAR technology generates point clouds by measuring the distance between an object and the sensor using laser light pulses [12]. A point clouds is a collection of points representing coordinates in three-dimensional space [13], making it an ideal tool for data acquisition.

Constructing a dataset is a specialized process where the objectives and characteristics of the images to be captured are defined. For this project, the focus is on detecting people. The images are captured using specified technological elements, and additional relevant information, such as descriptions and characteristics, is added for each image. Preprocessing tasks, including size adjustment, color normalization, noise removal, and data enhancement, are then performed. For this specific case, labeling tasks are required for semantic segmentation.

Semantic segmentation is a deep learning technique that divides an image into meaningful sections or regions, each corresponding to an object that can be classified into various classes or categories. A dataset has been constructed based on images obtained from 3D point clouds captured by a LiDAR sensor. This dataset contains information related to scenes where people appear in various poses and angles. Additionally, each scene includes a segmentation mask file, allowing for the identification of objects within the scene,

resulting in a more comprehensive dataset compared to existing ones.

This work evaluates the performance of Mask-RCNN and UNET for the "Person" class using the newly generated dataset, demonstrating an improvement in person recognition in indoor environments.

II. MATERIALS AND METHODS

This research focuses on identifying the necessary elements to structure a dataset for the semantic segmentation task aimed at detecting the "Person" class. The primary component for the dataset construction is a point cloud obtained from a LiDAR sensor. The methodology for this work is shown in Figure 1.

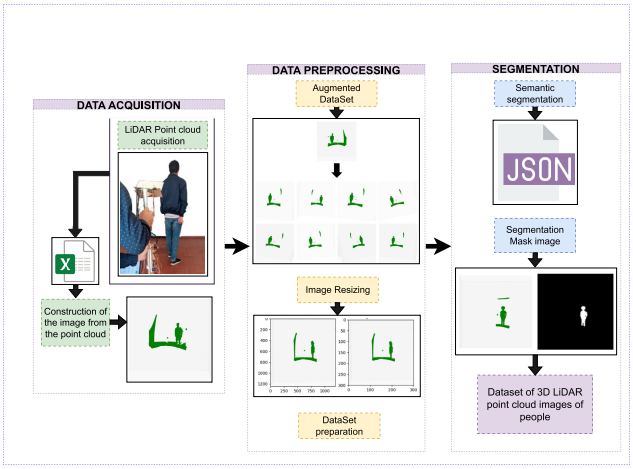

FIGURE 1. Methodology applied in the research.

The methodology consists of 3 phases: (1) Initial phase: Acquisition of the point cloud in XLSX format and processing it to obtain an image of the projected point cloud, (2) Preprocessing phase: Preprocessing the images and structuring the initial dataset, including data augmentation, (3) Segmentation phase: Labeling the dataset for segmentation and creating binary mask images for semantic segmentation.

A. DATA ACQUISITION

1) LIDAR POINT CLOUD ACQUISITION

To understand the data collection process, it is imperative to emphasize the advantages of LiDAR technology. Table 1 presents general data and specifications of the LiDAR sensor used in data acquisition.

TABLE 1. RPLIDAR S1 specifications.

Specification	Value
Detection distance	0.15-12 meters
Detection angle	360 degrees
Scanning frequency	9200 Hz
Angular resolution	0.15 degrees
Detection accuracy	< 2% @ 1 meter
Communication speed	115200 bps (UART), 1 Mbps (CAN)
Power consumption	< 2.5W

TABLE 2. Point cloud variable characteristics.

Name	Data Type	Variable Type	Description
Angle	Text	Nominal	Current LiDAR position angle concerning the initial angle (0, no movement).
Distance	Text	Nominal	The distance at which the light beam of the sensor is reflected by some surface is given in millimeters.
Tilt	Text	Nominal	Degree of tilt of the LiDAR within the Y-axis.

The 2D LiDAR sensor used for image capture is the RPLIDAR S1. This sensor provides distance and angle measurements (with respect to the sensor) of the scenario and, along with an Inertial Measurement Unit (IMU) that provides the tilt angle measurements, generates a three-dimensional point cloud by superimposing captures. The sensor's characteristics are listed in Table 2.

The data acquisition uses cylindrical coordinates (three-dimensional coordinate system) to describe the position of a point in space. The parameters considered are the angle θ (rotation angle of the LiDAR), the radial distance "r" (distance from the light beam reflection point to the origin), and the perpendicular height (degree of inclination relative to the Y-axis). To acquire data from scenes with people, various positions (standing, walking, crouching) and angles are used to obtain a 360-degree reference. Accessories such as hats, different arm positions, and poses holding a cell phone are also included. Figure 2 illustrates the data acquisition process, using a mobile base that places the LiDAR 1 meter above the ground, allowing vertical angular displacement for a complete sweep of the captured scenes.

It's important to emphasize that during the capture process, the use is limited to 90 degrees so that the point cloud obtained using the LiDAR focuses on the center of interest, thus providing a greater emphasis on the point cloud. Considering that the capture was conducted in environments with minimal movement in the scene, in order to have greater control over the obtained point clouds.

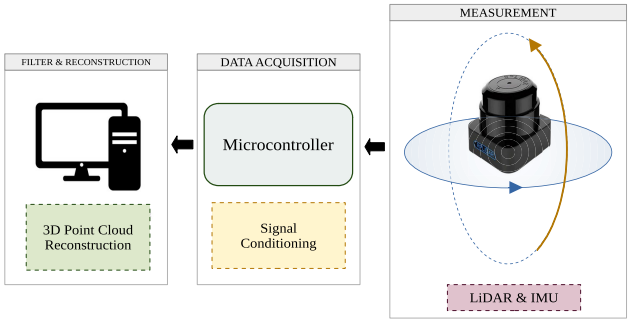

FIGURE 2. Scheme for data acquisition and point cloud reconstruction.

2) CONSTRUCTION OF THE IMAGE

As previously mentioned, the data are represented in cylindrical coordinates, which are converted to Cartesian

coordinates to facilitate the 3D representation of the point cloud.This conversion is done using equations that translate cylindrical coordinates to Cartesian coordinates. Equation 1 presents the formula for calculating "theta," the angle related to the LiDAR position, with a coverage range between 0° and 360°:

$$\theta = \frac{angle \cdot \pi}{180°} \tag{1}$$

Next, the tilt angle is converted into a numerical value using Equation 2, in which ϕ is calculated by converting the difference between the LiDAR tilt angle and a reference angle during the scan. This value is calculated as follows:

$$\phi = \frac{tilt \cdot \pi}{180°} \tag{2}$$

Finally, using the reflection distance between the origin and the surface on which the light beam is reflected, the coordinate are calculated in an orthogonal system (x, y, z). Each axis a different equation, combining sine and cosine functions, resulting in Equations 3, 4, and 5, for the 'x, y, z' axes, respectively:

$$x = r \cdot \sin\phi \cdot \cos\theta \tag{3}$$
$$y = r \cdot \sin\phi \cdot \sin\theta \tag{4}$$
$$z = r \cdot \cos\phi \tag{5}$$

The resulting values are stored in an array and, using Python graphics libraries, 3D images are created from the point cloud. The images are preprocessed by applying noise reduction filters, resulting in an initial dataset of 249 images. Priority was given to scenarios involving people. The information related to the dataset is presented in Table 3.

TABLE 3. Characteristics by category initial dataset.

Class	Number of Images	Number of objects per class	
Person	233	77.92%	1
		15.66%	2
Empty / Other	16	6.42%	0

Figure 3 shows examples of the data: one person (A), two persons (B), and empty scenes (C). To have a visual understanding of the plots obtained by the LiDAR point cloud.

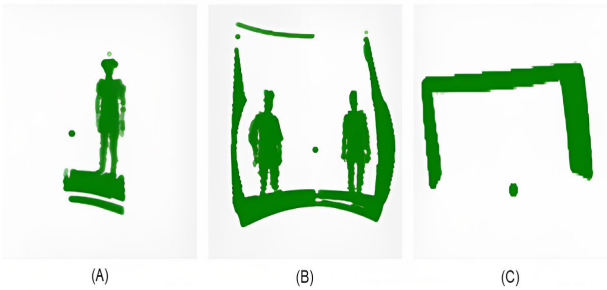

FIGURE 3. Visualization of scenario types.

3) KALMAN FILTER FOR MEASUREMENT NOISE

For an a posteriori coordinate denoted by x_{k-1}, y_{k-1} and x_k, y_k the a priori coordinate for a point k, the displacement of the IMU is ΔD, and the tilt angle information calculated by the inertial navigation system is θ [16]. The displacement dynamics between each pair of points can be expressed as follows:

$$x_k = x_{k-1} + \Delta D \cdot \cos\phi \tag{6}$$
$$y_k = y_{k-1} + \Delta D \cdot \sin\phi \tag{7}$$

where the state variables correspond to:

$$\vec{x} = [x, y, \phi] \tag{8}$$

The Kalman filter provides a robust response for noise rejection [17]. With M_k and P_k as the covariance matrices, Φ_k as the fundamental matrix, Q_k as the covariance matrix associated with the system noise, V_k as the uncertainty matrix of the measurements, and Z_k as the matrix of measurements provided by the sensors, the Kalman filter process is shown in Figure 4.

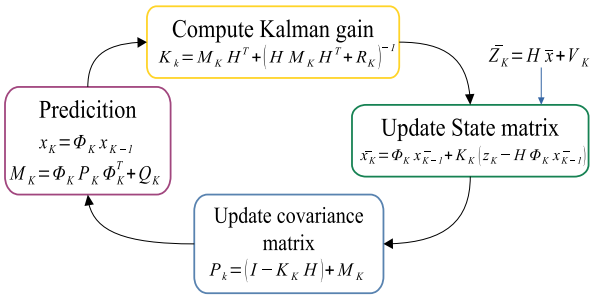

FIGURE 4. Scheme for EKF. Taken from [17].

B. DATA PREPROCESSING

1) AUGMENTED DATASET

In machine learning, a large dataset is crucial for training algorithms to achieve better predictions and detections. Therefore, the dataset obtained in the previous section was augmented using a script that employs the SMOTE (Synthetic Minority Over-sampling Technique) statistical technique. This technique extends the dataset from pre-existing instances without altering the number of majority values [15].

This script uses the open-source Keras library, which allows the creation of neural networks through an intuitive application programming interface (API) and is compatible with frameworks such as TensorFlow, which works through modular blocks that simplify its use [18].

The use of Keras is supported by the **'ImageDataGenerator()'** method, which generates images through random transformations while ensuring that the center of interest in the image remains intact. Parameters considered include::

- **Rotation range:** Image rotation cannot exceed 15 degrees.
- **Rescale:** The image resizing should less than 3px.

- **Crop Range:** Image crop range cannot exceed 15%.
- **Zoom range:** Image zoom range cannot exceed 15%.
- **Horizontal Rotate:** Allows movement of the image along the 'X' axis.
- **Fill mode:** Fill the image if necessary using the nearest pixel method, where the image is filled from the nearest pixels.

These random configurations enable the generation of at least 8 times the original number of images. The script defines these configurations to be randomized and employs a loop to load images from the initial dataset, generating diverse samples and storing them in a predetermined location with specific names. The created samples are validated, allowing for the selection of the most appropriate ones to augment the dataset and ensure its quality. Figure 5 illustrates these transformations for a specific case.

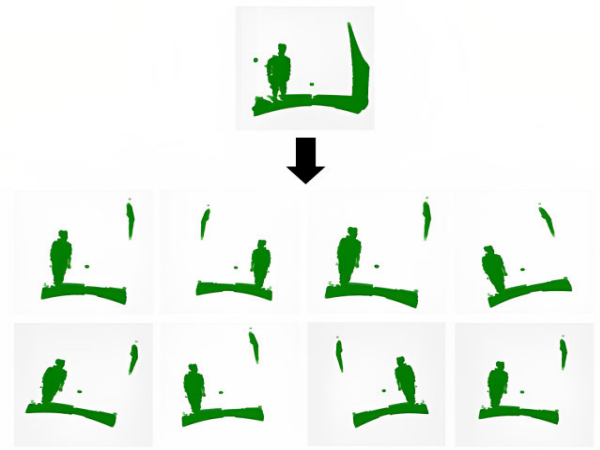

FIGURE 5. Image input and output representation of the data augmentation algorithm.

2) IMAGE RESIZING

As part of the data preparation, images must be rescaled to ensure uniformity in size for training, which helps maintain model performance and reduce computational costs. A script resizes the images in a directory to 300 × 300 pixels and overwrites them. Figure 6 compares an original image with a resized image.

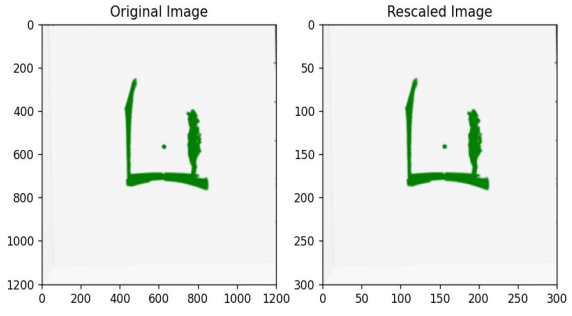

FIGURE 6. Image resizing.

TABLE 4. Characteristics of the variables belonging to the DataSet.

Name	Data Type	Variable Type	Description
Point cloud	Text	Nominal	Name of the image representing the point cloud.
Format	Text	Nominal	Format of the set (png, jpg).
Area	Text	Nominal	Size in length and width of the obtained image.
Date	Text		Date the file was obtained.
Source	Text	Nominal	Source from which the file was obtained (LiDAR).
Object type	Text	Categorical	Type of object involved in the image (person, animal, etc).
# of objects	Text	Nominal	Number of objects that are inside the image, depending on the previously mentioned variable.
Place	Text	Categorical	Type of location where the file was obtained (indoor, outdoor).
Ilumination	Text	Categorical	Type of illumination at the time the point cloud was taken (natural, artificial).
Similar objects	Text	Nominal	Objects with similar characteristics to each other (e.g., uniformed persons).
# Of similar objects	Text	Nominal	In case of similar objects, the number of additional similar objects.
Occlusion	Text	Nominal	Level of occlusion of the object of interest within the image (Scale 0-4).
X-axies movement	Text	Nominal	Presence of motion within the X-axis.
Y-axies movement	Text	Nominal	Presence of motion within the Y-axis.
Z-axies movement	Text	Nominal	Presence of motion within the Z-axis.
Degrees	Text	Nominal	Value in degrees in which it was increased to make the shots of the image.
# Of layers	Text	Nominal	Number of layers used to capture the image.
Aditional objects	Text	Nominal	Presence of additional objects within the center of interest.
Environment	Text	Categorical	Type of environment (Controlled, semi-controlled, uncontrolled).

3) DATASET PREPARATION

After capturing data from the point cloud, a dataset is constructed based on the characteristics and variables mentioned in Table 4. Images are generated from the point cloud, and an initial dataset is created featuring different scenarios, including people in everyday situations. Scenarios are categorized by the level of movement around the point of interest during sampling: uncontrolled (abrupt movements), semi-controlled (slight movements), and controlled (no movement). Factors such as the occlusion level of the point of interest and other aspects are also considered.

C. SEGMENTATION

Image segmentation is crucial in computer vision for assigning a label to each pixel in an image, indicating which class or category each pixel belongs to. Segmentation groups pixels with the same label into a segment, facilitating image analysis and extraction of relevant information [19]. The concept of semantic segmentation gained prominence in 2014 with the work of Long, Shelhamer, and Darrell, who introduced the Fully Convolutional Network (FCN) architecture [20].

There are several methods of image segmentation, such as binary, semi-automatic, and semantic segmentation. Semantic segmentation is considered the most appropriate technique in many cases, as it takes into account the specific characteristics of the data set to achieve a more accurate and detailed segmentation.

Binary segmentation is a simple segmentation method that assigns a label of "1" or "0" to each pixel, based on whether it belongs to an object of interest in the image. Semi-automatic segmentation, combines manual and automatic segmentation to obtain accurate segmentation. Semantic segmentation, an advanced method that uses deep learning to analyze object characteristics and assign semantic labels to each pixel, producing accurate and detailed segmentations..

1) SEMANTIC SEGMENTATION

Semantic segmentation assigns a label or category to each pixel in an image, offering advantages over other object detection models by allowing precise delimitation of objects by pixel category. Labels are used to define the input data's class or category, while segmentation masks hide or highlight certain parts of the input data during segmentation [21].

Tags are used to define the input data with a certain class or category. Under the concept of semantic segmentation, labels are used to label each pixel in the image with the class to which it belongs, e.g., background, object, border, and others. These labels are used to train and evaluate image segmentation models, such as convolutional neural networks.

On the other hand, segmentation masks are used to hide or highlight certain parts of the input data during the segmentation process. In image segmentation, masks are used to hide pixels that are not part of the object to be segmented and highlight the pixels that are. Segmentation masks are used in conjunction with labels, as they help to focus the model's attention on the important parts of the image.

So, a **'JSON'** file of labels is made in which the masks are contained by way of pixel location on the polygons, this labeling is done manually since trying to do it by deep learning techniques presents many errors in terms of accuracy, therefore, the labeling is done in the Visual Geometry Group (VGG) Image Annotator tool, it is a free and open-source tool for image annotation. It allows users to label and annotate objects in images, which is useful for machine learning and computer vision tasks [22].

In this tool the class **'Person'** is defined, and labels are created using the Visual Geometry Group (VGG) Image Annotator tool, which is free and open-source for image annotation.

2) SEGMENTATION MASK

Constructing an image segmentation mask involves identifying and delimiting different regions within an image related to objects or labels. These masks visually separate image parts and assign labels to each class of interest, aiding object recognition.

To prepare the dataset with the corresponding labels, the **'utils'** dataset library is used, applying transformation techniques to create lists of images with information about their corresponding masks. Each mask stores regions of interest as polygons, transforming these regions by setting values of interest to 1 and the rest of the image to 0, creating binary masks for each region of interest.

A Gaussian smoothing filter is then applied to the images using the 'skimage.filters' library, with a "sigma" parameter of 1.2 to produce smoother segmentation mask images. The processed images are stored using the Python library 'Scikit-image'. Figure 7 presents an original image (A) and its corresponding segmentation mask (B).

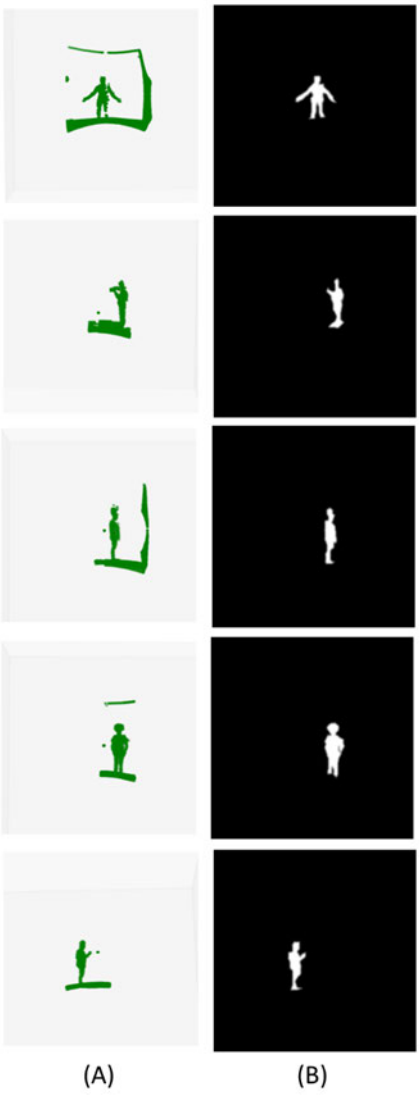

FIGURE 7. Results segmentation mask.

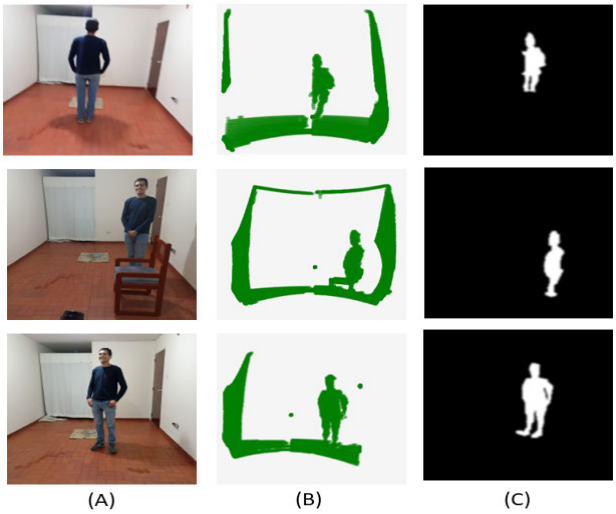

FIGURE 8. Results of the dataset.

III. RESULTS AND DISCUSSION

During data acquisition, various scenarios were considered, encompassing both indoor and outdoor environments, across controlled, semi-controlled, and uncontrolled settings. This approach ensures that the dataset includes diverse image types, making the results less sensitive to potential variations in detection across different environments.

Figure 8 illustrates examples of results obtained during dataset acquisition for the class **'Person'**. Section (A) shows a photograph of the scene, section (B) displays the image obtained from the 3D LiDAR point cloud, and section (C) presents the corresponding segmentation mask image.

The dataset comprises 1245 images, with 1165 images featuring people and 80 images depicting other objects or environments where people are absent. Table 5 details the total number of classified images based on predefined characteristics for each image.

TABLE 5. Classification and total images of the dataset.

Object Type	Place	Environment	Number of images
Persons	Indoor	controlled	840
		Semi-Controlled	225
		Uncontrolled	0
	Outdoor	controlled	40
		Semi-Controlled	60
		Uncontrolled	0
Other objects	Indoor	controlled	40
		Semi-Controlled	40
		Uncontrolled	0
	Outdoor	controlled	0
		Semi-Controlled	0
		Uncontrolled	0
		Total Persons	1165
		Total Other objects	80
		Total	1245

The dataset is characterized by important detection parameters such as object type, scene location, environment, source, number of objects, and occlusion. Image normalization ensures uniform dimensions of 300 × 300 pixels, alongside preprocessing techniques that significantly increase dataset volume. This augmentation enhances dataset diversity, enabling more effective learning and reduced execution times in Machine Learning applications.

The dataset, with predefined variables and preprocessing steps, is publicly available on the **'Zenodo'** platform. Zenodo facilitates sharing of image datasets among researchers and data scientists, promoting exploration and development of various projects related to data science and machine learning.

This publication includes acquired images, segmentation mask images organized in different folders, a file containing semantic segmentation labels, and a CSV file providing additional information for each image. The dataset is assigned a Digital Object Identifier (DOI) (https://doi.org/10.5281/zenodo.10547234), ensuring accessibility for further research and development of machine learning models. The augmented dataset results from applying the SMOTE oversampling technique, creating new records for different defined classes within the dataset. This augmentation improves overall dataset balance, leading to enhanced performance of machine learning models trained on this dataset. The produced dataset focuses extensively on individuals, demonstrating superior reliability compared to existing datasets. Notably, this dataset includes binary segmentation masks specifically related to individuals, providing an additional advantageous feature. The use of cutting-edge technologies such as LiDAR in image acquisition offers significant advantages over traditional methods. The resultant dataset encompasses 2D point cloud data, contributing to increased reliability in object detection training. This approach represents a substantial advancement in terms of accuracy and reliability compared to conventional datasets, thereby reinforcing its applicability for object detection training purposes.

In Figure 9, section (A) highlights the distribution of images related to individuals versus images depicting other objects or scenes where people are not the focus. It is evident that a significant percentage (94%) of images are related to people, while a smaller percentage (6%) pertains to scenes without people, which is useful for validation and testing purposes. This balance ensures robustness in the "Person" category detection, while also acknowledging scenarios without people (0.4%), crucial for comprehensive validation. Regarding the distinction between interior and exterior scenarios, section (B) of Figure 9 shows that 94% of scenes are indoors, and 6% are outdoors. This distribution underscores the predominance of indoor scenes but assures no compromise in model performance due to meticulous management and control during dataset creation using LiDAR technology.

The management of environments within the scenarios was categorized into Controlled, Semi-Controlled, and Uncontrolled environments, with respective percentages of 94.3%, 5.3%, and 0.4%, as depicted in section (C) of Figure 9. For training a model, it is ideal to predominantly include

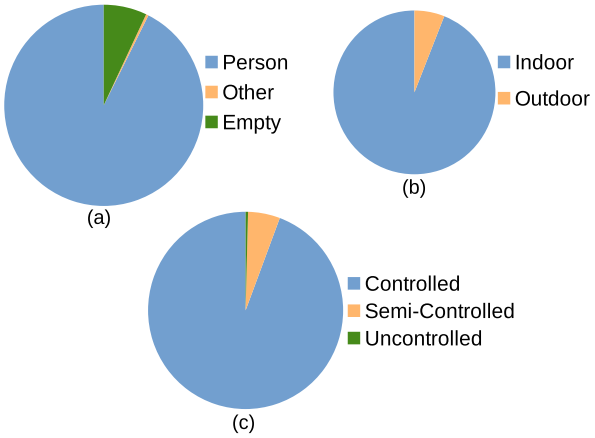

FIGURE 9. Identification of main elements of the DataSet (a) Object Type, (b) Scene, (c) Enviroment.

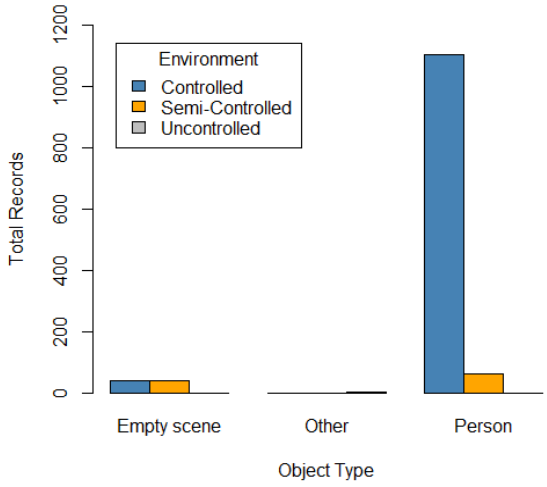

FIGURE 10. Types of environments related to object type.

scenes from Controlled and Semi-Controlled environments. These settings provide better focus on the subject of interest in each image. However, consideration for Uncontrolled environments in future training datasets remains relevant.

Figure 10 illustrates the distribution of images across different environments and object types. Predominantly, the dataset includes Controlled environments and scenarios featuring people, while also encompassing other scenarios with various objects and semi or uncontrolled scenes. This diversity improves the accuracy of neural network training using this dataset.

Figure 11 examines the relationship between occlusion levels and the object type's center of interest. It reveals that 93.58% and 6.42% of images exhibit occlusion levels 0 and 1, respectively, demonstrating a predominant proportion within the dataset. This robust distribution enhances detection accuracy, aligning with the project's objectives.

The dataset was meticulously constructed using data augmentation, resizing, and semantic segmentation, ensure the inclusion of a high-quality dataset focused on the ''Person'' category. This dataset offers advantages over

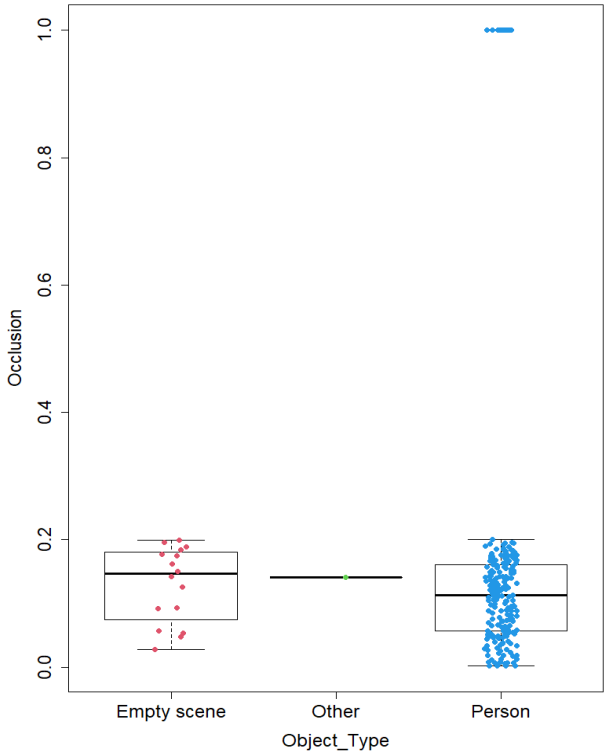

FIGURE 11. Relationship between occlusion and object type.

existing datasets by providing well-labeled, segmented data with additional information beneficial for various machine learning tasks.

Applications in security, particularly in detection tasks using deep learning techniques, can benefit significantly from this dataset. It supports the development of advanced surveillance and security systems leveraging current software and hardware technologies, thereby enhancing task management within these applications, which is a primary goal of artificial intelligence.

Regarding 3D point cloud datasets, particularly those focusing on the ''Person'' category, significant progress has been achieved. While existing datasets like UWA, PartNet, S3DIS, and ScanNet predominantly cover buildings, streets, and squares, data for complex outdoor object scenes remain limited. Therefore, the created dataset emphasizes capturing outdoor scenes focusing on the ''Person'' category [23], [24], [25].

Furthermore, the incorporation of semantic segmentation in this dataset offers advantages over other mentioned datasets, providing precise object delineation. This precision significantly enhances the performance and accuracy of neural networks in tasks such as classification and detection.

Comparison of mean class accuracy among several state-of-the-art semantic segmentation algorithms indicates superior performance of the U-Net and Mask R-CNN algorithms trained with this dataset approach. Particularly, the Mask R-CNN algorithm trained with the created dataset achieved the highest mean class accuracy compared to other state-of-the-art resultss.

TABLE 6. Accuracy comparison between State of Art Point Cloud semantic segmentation algorithms.

Method	Mean Class Accuracy
PointNet	86.0
VRN(Single view)++	88.98
PointCNN	88.1
Ours(Mask R-CNN)	89.9
Ours(U-Net)	90.5

The following table compares mean class accuracy among several state-of-the-art semantic segmentation algorithms applied to a specific dataset [26], including PointNet, VRN (Single view)++, PointCNN, Mask R-CNN, and U-Net. Notably, implementations of Mask R-CNN and U-Net demonstrate notable accuracy improvements over existing methods (see Table 6).

While the dataset focuses on the ''Person'' category and provides accurate segmentations, this focus may limit the model's ability to generalize to scenarios not represented within the dataset. Additionally, the advanced algorithms employed, such as Mask R-CNN and U-Net, require substantial computational resources. Therefore, optimizing these algorithms for environments with limited computational power is crucial for their broader applicability.

Although the ablation study demonstrated significant improvements in accuracy, further evaluation is needed to assess the robustness and generalization of the models in real-world scenarios. Testing in diverse and challenging environments will offer a more comprehensive understanding of model performance. Furthermore, while data augmentation provides benefits, it may not cover all possible variations encountered in real-world data. This limitation could affect the model's performance when dealing with highly variable or unexpected data.

IV. CONCLUSION

The final Dataset has 1254 images featuring people captured from the LiDAR point cloud. Enhancements in image quality could be achieved by employing higher resolution LiDAR or increasing the number of scans per scene to augment the data volume for improved object recognition or detection.

Generating a dataset that combines LiDAR images with a CSV file provides detailed information about each image in the dataset. It includes segmentation masks, essential for certain model readings requiring binary masks (1's and 0's) or image formats for effective training.

Future work involves applying this dataset to machine learning models to validate its efficacy in intrusion detection. Additionally, expanding the dataset to include various objects such as vehicles and weapons will be crucial for enhancing intrusion detection capabilities in different institutions and buildings.

ACKNOWLEDGMENT

This research activity is part of the CLARIFIER project, which operates within the international alliance among the University of Ljubljana in Slovenia, the Telecommunications, Computer Engineering, and Photonics Institute (TeCIP) of Italy, and the UPTC aims to simplify the development of a Photonic Integrated circuit as demonstrator to perform monitoring and surveillance tasks with the support of the North Atlantic Treaty Organization (NATO), and with end-users, such as Siot Ingeniería, CODALTEC, which develops technology for the Colombian Ministry of Defense, and Colombian Air Force.

REFERENCES

[1] R. Martínez and A. Palma. *Revolución Tecnológica E Inclusión Social: Reflexiones Sobre Desafíos Y Oportunidades Para La Política Social En América Latina*. Accessed: Apr. 11, 2023. [Online]. Available: https://hdl.handle.net/11362/45901

[2] I. M. G. Urbini, ''Estrategia de ciberseguridad distribuida, aplicando el concepto de operación de inteligencia,'' Ph.D. dissertation, Dept. Fac. Inform., Universidad Nacional de La Plata, Argentina, 2022, doi: 10.35537/10915/147421.

[3] D. R. Santana. (2022). *Uso Extensivo De La Tecnología LiDAR Pública En Análisis De Inundabilidad. El Caso Del Guadalete*. Accessed: Apr. 11, 2023. [Online]. Available: http://hdl.handle.net/10366/149621

[4] M. P. A. Rodríguez. (2022). *Análisis Comparativo de Modelos Digitales de Terreno Obtenidos Por Tecnologia LiDAR Con Aeronave no Tripulada y Por Fotogrametria Con UAV en Zona de Montaña*. Universidad Antonio Nariño, Bogotá, Colombia. Accessed: Apr. 11, 2023. [Online]. Available: http://repositorio.uan.edu.co/bitstream/123456789/5967/1/2022_MiryamPaolaAvellaRodricadguez.pdf

[5] L. S. González. (2021). *Redes Basadas En Aprendizaje Profundo Para La Detección De Anomalías En Radiografías De Tórax*. Universidad Nacional de Mar del Plata. Accessed: Apr. 11, 2023. [Online]. Available: http://rinfi.fi.mdp.edu.ar/handle/123456789/530

[6] F. Patricia Medina and R. Paffenroth, ''Machine learning in LiDAR 3D point clouds,'' 2021, *arXiv:2101.09318*.

[7] W. Farag, ''Kalman-filter-based sensor fusion applied to road-objects detection and tracking for autonomous vehicles,'' *Proc. Inst. Mech. Eng., I, J. Syst. Control Eng.*, vol. 235, no. 7, pp. 1125–1138, Aug. 2021, doi: 10.1177/0959651820975523.

[8] A. Díaz Prieto. (2022). *Detección De Objetos 3D Mediante LiDAR En Sistemas Embebidos*. Universidad del País Vasco, España. Accessed: Apr. 11, 2023. [Online]. Available: http://hdl.handle.net/10810/57135

[9] A. V. Cárdenas García and A. E. Cevallos Villacís. (2022). *Análisis Comparativo De Algoritmos De Detección De Objetos Con Sensor Láser LiDAR 3D*. Universidad Politécnica Salesiana, Guayaquil, Ecuador. Accessed: Apr. 11, 2023. [Online]. Available: http://dspace.ups.edu.ec/handle/123456789/22839

[10] Á. Martínez Ballester. (2022). *Implementación Y Estudio De Funcionamiento De Un Detector De Objetos En Nubes De Puntos Tridimensionales*. Universidad Miguel Hernández de Elche, España. Accessed: Apr. 11, 2023. [Online]. Available: https://hdl.handle.net/11000/27926

[11] L. López-Avila and N. Acosta-Mendoza, ''Detección de anomalías basada en aprendizaje profundo: Revisión,'' *Revista Cubana de Ciencias Informáticas*, vol. 13, pp. 107–123, Sep. 2019. [Online]. Available: https://www.redalyc.org/articulo.oa?id=378365913008

[12] A. R. Rey. (2022). *Detección De Líneas Eléctricas En Nubes De Puntos LiDAR a Través De Un Filtro Basado En Histogramas Locales*. Universidade da Coruña, España. Accessed: Apr. 11, 2023. [Online]. Available: http://hdl.handle.net/2183/31909

[13] E. L. Cuevas. (2022). *Reconstruir Lo Construido: El Auge De La Nube De Puntos*. UPM, España. Accessed: Apr. 11, 2023. [Online]. Available: https://oa.upm.es/70664/

[14] L. P. Enríquez, R. D. Hernández, and L. A. Robles, ''Implementación de CNN basada en una arquitectura VGG16 para detección y clasificación de árboles mediante la segmentación semántica en imágenes aéreas,'' *Res. Comput. Sci.*, vol. 7, no. 151, pp. 157–170, 2022, Accessed: Apr. 11, 2023. [Online]. Available: https://www.rcs.cic.ipn.mx/rcs/2022_151_7/

[15] M. Torres-Vásquez, J. Hernández-Torruco, B. Hernández-Ocaña, and O. Chávez-Bosquez, "Balanceo de datos del Síndrome de guillain-Barré utilizando SMOTE para la clasificación de subtipos," *Res. Comput. Sci.*, vol. 148, no. 7, pp. 113–125, Dec. 2019, doi: 10.13053/rcs-148-7-9.
[16] Z. Li, Z. Su, and T. Yang, "Design of intelligent mobile robot positioning algorithm based on IMU/odometer/LiDAR," in *Proc. Int. Conf. Sens., Diag., Prognostics, Control (SDPC)*, Beijing, China, Aug. 2019, pp. 627–631, doi: 10.1109/SDPC.2019.00118.
[17] O. J. Montañez, M. J. Suarez, and E. A. Fernandez, "Application of data sensor fusion using extended Kalman filter algorithm for identification and tracking of moving targets from LiDAR–radar data," *Remote Sens.*, vol. 15, no. 13, p. 3396, Jul. 2023, doi: 10.3390/rs15133396.
[18] DG Ionos. *Keras: Biblioteca De Código Abierto Para Crear Redes Neuronales*. Accessed: May 23, 2022. [Online]. Available: https://www.ionos.es/digitalguide/online-marketing/marketing-para-motores-de-busqueda/que-es-keras/
[19] J. J. Padilla-Arballo, S. Martínez-Díaz, M. A. Castro-Liera, and J. E. Luna-Taylor, "Detección de cambio en superficie costera mediante la segmentación de imágenes aéreas utilizando redes neuronales convolucionales," *Pädi Boletín Científico de Ciencias Básicas e Ingenierías del ICBI*, vol. 10, pp. 136–144, Oct. 2022, doi: 10.29057/icbi.v10iespecial4.9290.
[20] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," 2014, *arXiv:1411.4038*.
[21] PI Orellana Rueda. (2020). *Segmentación Semántica Y Reconocimiento De Lugares Usando Características CNN Pre-entrenadas*. Universidad de Chile. Accessed: Apr. 11, 2023. [Online]. Available: https://repositorio.uchile.cl/handle/2250/173733
[22] S. K. Baliarsingh, P. P. Dev, A. Bandyopadhyay, A. K. Dash, and R. Pradhan, "A smartphone-based deep learning framework for early detection of oral cancer signs," in *Proc. Int. Conf. Emerg. Syst. Intell. Comput. (ESIC)*, Bhubaneswar, India, Feb. 2024, pp. 181–186, doi: 10.1109/esic60604.2024.10481662.
[23] J. Zhang, X. Zhao, Z. Chen, and Z. Lu, "A review of deep learning-based semantic segmentation for point cloud," *IEEE Access*, vol. 7, pp. 179118–179133, 2019, doi: 10.1109/ACCESS.2019.2958671.
[24] Z. Song, Z. He, X. Li, Q. Ma, R. Ming, Z. Mao, H. Pei, L. Peng, J. Hu, D. Yao, and Y. Zhang, "Synthetic datasets for autonomous driving: A survey," *IEEE Trans. Intell. Vehicles*, vol. 9, no. 1, pp. 1847–1864, Jan. 2024, doi: 10.1109/TIV.2023.3331024.
[25] G. Xu, W. Liu, Z. Ning, Q. Zhao, S. Cheng, and J. Nie, "3D semantic scene completion and occupancy prediction for autonomous driving: A survey," in *Proc. 4th Int. Conf. Comput. Artif. Intell. Technol. (CAIT)*, Dec. 2023, pp. 181–188, doi: 10.1109/cait59945.2023.10469138.
[26] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic graph CNN for learning on point clouds," 2018, *arXiv:1801.07829*.

JHON SEBASTIAN APARICIO MESA received the degree in computer systems engineering from the School of Computer Engineering, Universidad Pedagógica y Tecnológica de Colombia (UPTC), in 2023. He was a young Researcher for the SPS G5888 Clarifier research project in cooperation with the University of Ljubljana and the TeCIP of Italia for NATO. He is currently with quality assurance and works as a QA Analyst at a core company. He is a Researcher in the field of artificial intelligence, deep learning, and neural networks.

MARCO JAVIER SUÁREZ BARÓN received the M.Sc. degree in information management from Colombian School of Engineering Julio Garavito, in 2012, and the Ph.D. degree in strategic planning and technology management from UPAEP-Mexico, in 2016. He was with ECI University-Colombia and UPAEP Mexico. He is currently an Associate Professor with UPTC Colombia. He is also a Research Associate with the Ministry of Science, Technology and Innovation, Colombia. His research interests include machine learning, AI, and deep learning. He serves on the editorial board for numerous technology journals.

EDUARDO AVENDAÑO FERNÁNDEZ (Senior Member, IEEE) received the M.Sc. degree in tele-informatics from Universidad Distrital Francisco José de Caldas, in 2005, and the Ph.D. degree in electronics engineering from Universidad de Antioquia, Colombia, in 2019. He is currently a Full Professor with the Engineering Faculty, UPTC Colombia, Sogamoso, Boyacá, Colombia. He is also a Junior Researcher with the Ministry of Science, Technology, and Innovation, Colombia (MinTIC). His research interests include machine learning and deep learning for applications in the IoT for agriculture and underground mining, and optical, wireless, and digital telecommunications.