

# Privacy-preserving People Detection Enabled by Solid State LiDAR

1<sup>st</sup> Andrei Günter, 2<sup>nd</sup> Stephan Böker, 3<sup>rd</sup> Matthias König, 4<sup>th</sup> Martin Hoffmann

*Campus Minden*

*Bielefeld University of Applied Sciences*

Minden, Germany

firstname.lastname@fh-bielefeld.de

**Abstract**—Detecting people in video streams involves privacy concerns since people are often unaware of being recorded. Furthermore, cameras and commodity depth sensors are prone to variances of illumination. This work proposes light detection and ranging as sensing technology for a robust and privacy preserving people detection. A people counter is implemented, which captures a coarse human shape by concatenating multiple two dimensional range scans while people pass the sensor's view. The system works under dynamic illuminated conditions and is running as an embedded node on a single core 700 MHz processing unit. An evaluation with 20 different people crossing an entrance 100 times in each direction was carried out under controlled conditions. Results show a high precision and recall of 100 percent and 99 percent respectively for counting people and determining the walking direction. Privacy is preserved since objects are represented by distance values with a low resolution of 16x16 data points.

## I. INTRODUCTION

Nowadays, many shopping centers become intelligent environments. Visitors are counted to optimize marketing strategies and resource planing. The fluctuating amount of visitors is highly related to marketing tasks like advertising, adjusting shop layouts or announcing special offers. Measuring the effectiveness of such tasks can be achieved by relating visitor counts to the regarding sales.

A common approach for counting people in a public area is based on processing image data obtained from conventional cameras. While classification algorithms provide eminent results, data acquisition with conventional cameras resides limited under certain environmental conditions. The best algorithm is not applicable when acquired data contains too much noise, e.g. data collected in a subway with insufficient illumination or outdoors affected by high solar radiation. Especially, such environments with insufficient or dynamically changing illumination require novel data acquisition techniques.

The main contribution of this work is to evaluate Leddar, which is an improved and patented subset of light detection and ranging (LiDAR), as sensing technology for object classification. LiDAR is a promising technology which allows to collect volumetric data by providing distance measurements. Furthermore, LiDAR is not prone to different illuminations, since non-visible light is emitted from the sensor into the surrounding environment.

This work distinguishes between two types of LiDAR. A mechanical LiDAR (M-LiDAR) system enables high resolu-

tion capturing by utilizing subtle coordinated rotating mechanics to direct the emitted light. Such subtle coordinated mechanics come with the cost of being complex to produce and being prone to vibrations. In contrast, solid-state LiDAR (SS-LiDAR) systems acquire data with solid state elements only. While SS-LiDAR systems capture data in lower resolution and offer a smaller field of view compared to M-LiDAR, they are more robust, simpler to produce and can be built in smaller physical size [1].

As additional contribution, this work aims to highlight the significance of applications which consider the human's privacy. Cameras violate privacy aspects since humans are recorded while being counted. A people counter for public areas is implemented to represent a use-case where privacy is a concern. A LiDAR sensor is used to acquire enough data to allow for classification of objects, but providing too little information to identify individuals. The evaluation shows that a reliable counting performance is reached while privacy is preserved.

This work is structured as follows. Section II presents state of the art of people detection with depth information. Section III describes the concept of the presented people counter. An actual implementation of the concept is presented with Section IV and corresponding results are shown in Section V. Finally, a conclusion, future work and limitations of the system are given with Section VI.

## II. RELATED WORK

Approaches using depth information to detect people or in general to classify objects can be categorized by differentiating data acquisition technologies and their resulting resolution of data. Methods analyzing high resolution images with 320x240 points of depth data and above aim to track people, recognize activities or even establish scene understanding [2], [3]. While such approaches achieve high accuracy results in detecting objects and people, they also rely on high resolution depth data hindering to preserve privacy. Prior works show that people can be identified or re-identified from depth data with high resolution [4], [5]. The aim of this work is to avoid identification and re-identification.

### A. Privacy preserving approaches

A common approach to preserve privacy is to postprocess captured data by blurring, censoring or encrypting sensitive data [6], [7]. Such works underline that a broad range of applications remain accurate when analyzing low resolution data, but saving sensitive data for a short time remains a concern in terms of privacy. A solution was proposed by adding 3D printed optics in front of original sensor optics, which preserve privacy by preprocessing captured data [8]. There is still space for improvement since additional optics can be removed and original data might be retrieved by applying inverse functions on post- or preprocessed data. Therefore, we propose to capture data with low resolution sensors.

### B. Low resolution sensors

Preserving privacy can be accomplished by utilizing low resolution sensors, e.g. low resolution cameras [9]. Cameras are prone to different illumination conditions, therefore range sensors are preferred in this work. In robotics, some approaches aim to detect pairs of human legs or wheelchairs in two dimensional (2D) range scans [10], [11] to track people. Such approaches suffer from a high possibility of occlusions due to the adverse view of the sensor. Multiple depth sensors with a low resolution of a single pixel were mounted into the ceiling to minimize possible occlusions [12]. Using single pixel depth sensors preserves privacy, but an installation of many sensors in form of a grid with a spacing of 60 cm between each sensor was required [12]. Such an installation is hard to maintain and not feasible for many environments. We use a single sensor in entrance areas to detect people passing the sensor's view and show that our system performs comparable to state of the art techniques, while privacy is ensured with analyzing low resolution data collected by a 2D sensor.

## III. CONCEPT

The proposed system is meant to be mounted under the ceiling behind or in front of entrances. A SS-LiDAR sensor is used for two reasons. A low resolution supports in preserving privacy, and a small physical size assists in developing embedded people counting nodes cooperating with each other. Fig. 1 illustrates the installation from a top-down view.

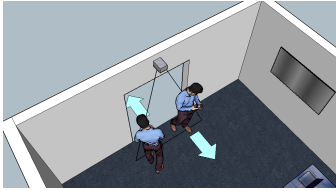


Fig. 1. Top-down view of the sensor's installation. A black triangle surrounding the people indicates the sensed area.

In order to capture human shapes, a whole entrance is covered by a sensor with an appropriate beam width. For multiple or very broad entrances more sensors are able to cooperate as a multi-sensor network to cover the entire section

of interest. The sensor is mounted tilted to distinguish walking directions of people with a single sensor.

Objects are classified by processing 2D  $(x, y)$  sensor data, whereas  $x$  indicates the horizontal position and  $y$  indicates the distance. Multiple 2D frames of objects passing the sensor's view are captured. These frame sequences are preprocessed in order to prepare them for the classification or to filter irrelevant objects. Finally, the data is classified into one of three classes. Fig. 2 represents the concept of the proposed system, whereas the classes *human:in* and *human:out* refer to the humans walking direction. More detailed information on each process is given below.

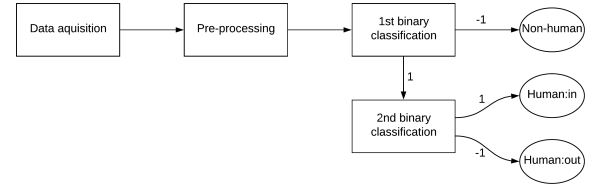


Fig. 2. Object classification of 3 classes: *non-human*, *human:in* and *human:out* by separating the classification into two binary classifications.

### A. Preprocessing

A common strategy to efficiently compute acquired data is to remove redundant information by extracting features. Our preprocessing can be described with two steps. At first, it segments multiple objects into individual samples and then objects are normalized to fit into the classification process. Multiple objects might appear in the sensor's view simultaneously, e.g. when people pass the sensor's view with a dog next to them. In such a case, the human and the dog would be segmented from the frame sequence as two individual data samples. Subsequently, each data sample is handled separately.

Data might be acquired in form of many frames, which results in large data sizes. This happens when people stop in the sensor's view or pass entrances very slow. Such long frame sequences are considered to contain a lot of redundant information. Therefore, the size for each data sample is normalized to a fixed amount. This is accomplished by sorting out redundant information.

### B. Object classification

Nowadays, machine learning is widely used for classification tasks. The initial problem of classifying objects into one of three classes is tackled by solving two binary classification problems. A first classification decides whether the object is *human* or *non-human*. Then, a second classification determines whether the human is walking *in* or *out* of an area of interest. Both binary classification problems are tackled by using a support vector machine (SVM). A SVM is suitable for binary classification with a limited feature space while providing efficient processing and sound classification results [13].

#### IV. IMPLEMENTATION

The system works as an embedded node with a power consumption of 4.7 W. A 700 MHz single core processor computes distances delivered by a LiDAR sensor. In this case a Leddar M16 is used. The Leddar M16 can be categorized as solid-state flash LiDAR sensor, which is a subset of LiDAR. Multiple range measurements are captured simultaneously by calculating the distance from the time of flight of a single LED flash. A single sensor provides 16 range measurements, and operates with a frequency of 50 Hz within a beam of  $\beta = 45^\circ$ . The beam is equally split into 16 segments, resulting in a beam width of  $\beta_x = 2.8125^\circ \mid x \in \{1..16\}$  per segment. A distance  $y_x$  is measured for each segment separately. Fig. 3 depicts a front-view of a single flash on a planar surface. The sensor is placed at a height of 2.6 meters, this ensures an accurate counting for an entrance width of 1.6 meters. Broader entrances could be covered by multiple sensors. Frame sequences of 16 frames per object are used to classify objects passing through an entrance, leading to a fixed size of 16x16 data points per object. Experiments showed that a size of 16 frames is sufficient for a reliable detection.

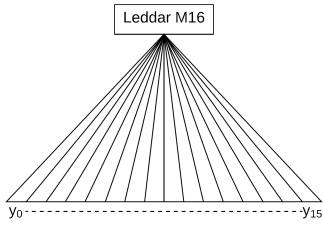


Fig. 3. Front-view of a Leddar M16 data acquisition of a planar surface, showing 16 segments with one range measurement  $y_x$  per segment.

##### A. Segmentation

Fig. 4 illustrates the raw data acquisition with a tilted top-down view. It shows the distance measures of two people walking close one after another. Both objects are segmented into single frame sequences as shown for one object in Fig. 4b. The object after segmentation is still represented by  $k = 41$  frames. In this case,  $k$  is reduced to 16 key frames in order to fit into the object classification task via SVM.

##### B. Key frame extraction

Key frames are extracted from each object to enable an efficient processing. Therefore, a change detection is applied to examine each data sample for redundant information. In a sequence of frames  $f_K \mid K \in \mathbb{N}$ , a frame  $f_{k+1}$  is considered to be redundant, when its prior frame  $f_k$  already holds similar information. Thus, a threshold  $\delta_k$  represents the difference between two frames  $f_k$  and  $f_{k+1}$  by comparing the according range measurements of each segment, this leads to a difference of  $\delta_k = \sum_{x=0}^{15} (y_{x,k+1} - y_{x,k})^2$ . If  $\delta_k$  reaches a certain value, then the frame  $f_{k+1}$  is declared as redundant frame, otherwise it is declared to be a key frame. Finally, 16 of the most significant key frames are used to represent the scanned object.

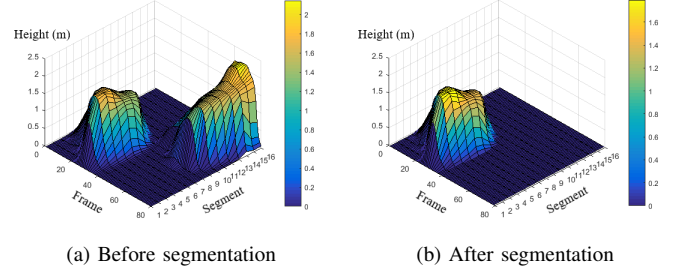


Fig. 4. Segmentation of two close people walking one after another: (a) shows the raw data before segmentation and (b) illustrates one of the two objects after segmentation.

TABLE I  
SVM CONFIGURATION FOR BOTH CLASSIFICATION PROBLEMS

| Classification       | $n$ | $C$   |
|----------------------|-----|-------|
| human — non-human    | 5   | 0.025 |
| human:in — human:out | 2   | 0.1   |

##### C. Object classification

The object classification was accomplished by collecting a dataset and training both SVMs. Then, a k-fold cross validation was used to determine the configurations.

1) *Dataset*: A dataset of 2000 objects was collected, which is split into 1000 samples representing humans and 1000 samples representing non-human objects, such as: animals, carts and randomly generated measurements. The 1000 samples representing humans are also labeled with *human:in* and *human:out* in order to cover the second classification problem. The collected data of adult humans shows a broad variety in terms of differences in physical condition, e.g. the dataset comprises differences in heights from 165 cm to 190 cm. Fig. 5 shows some examples of the captured objects.

2) *Training and Testing*: Both SVMs have been trained and tested by using a k-fold cross validation with  $k = 10$  in order to achieve a sound performance for the SVM. The resulting parameters are summarized for both classification problems in Table I, whereas  $n$  represents the dimension of the polynomial kernel and  $C$  represents the soft margin constant.

#### V. EVALUATION

Before we collected the final dataset used throughout this evaluation, we recognized that people cross entrances with certain behaviours by analyzing initial datasets used for training. A majority of people passing the entrance showed similar and reappearing patterns of walking directions depicted in Fig. 6. Our final dataset was adapted to contain the same amount of each walking pattern, which resulted in an improvement of correct detections. This ensures that people passing the entrance in a straight line are recognized with the same probability as people passing the entrance diagonal or curved.

A test in a controlled environment was carried out to evaluate the people counter. 20 people with many differences in physical condition were instructed to pass the entrance. They were not told to behave in a certain manner nor to stick

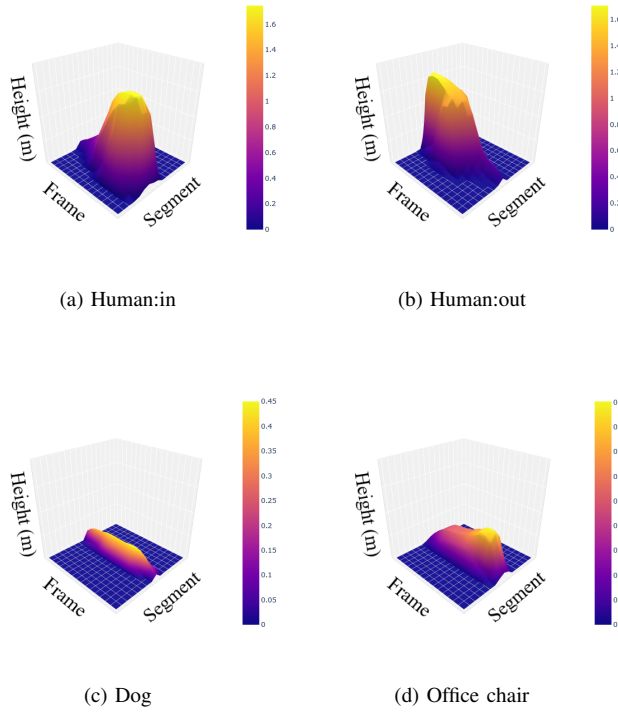


Fig. 5. Examples of captured objects in the dataset: (a) and (b) show 3D plots of humans with respect to their walking direction, (c) shows a dog and (d) shows an office chair.

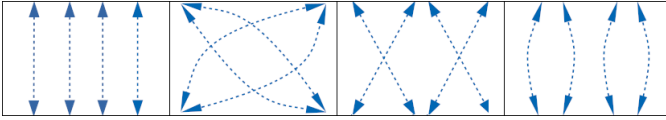


Fig. 6. Patterns of walking directions, which a majority of people chose while passing the sensor.

to a certain walking direction. 13 of 20 people do not appear in the dataset used for training. Each person passed the entrance 5 times in each direction resulting in 10 entrance crossings per person. Table II summarizes the classification results.

## VI. CONCLUSION

This work presented a people counter based on a LiDAR sensor. LiDAR performs robust in environments with dynamic illumination and allows to develop privacy preserving applications since only range data is collected. It was shown that a classification of humans with their proper walking direction can be accomplished. This is achieved by capturing a set of

2D range scans while objects pass the area of interest and analyzing the resulting data. Therefore, a LiDAR sensor was installed in a top down view and data was collected to train two SVMs. The first SVM classifies whether the object is human or not, while the second SVM is used to determine the walking direction. An evaluation presented reliable counting metrics for a test run with 20 different people each passing the entrance 10 times.

The presented dataset consists of only 2000 objects containing 1000 humans. This leads to limitations of the system. It is expected that people wearing hats, carrying objects or sitting in wheelchairs are not going to be detected properly, since the current dataset is not representing such data samples. Future work on preprocessing data to separate close objects properly from each other as well as improving the dataset is required to tackle these limitations. Further, a comprehensive evaluation in a real world environment such as a shopping center, is required to verify the accuracy. The presented results are promising to refine this system into a robust, accurate and privacy preserving application for counting numerous people in frequently visited places.

## REFERENCES

- [1] C. Niclass, A. Rochas, P. . Besse, and E. Charbon, "Design and characterization of a cmos 3-d image sensor based on single photon avalanche diodes," *Journal of Solid-State Circuits*, vol. 40, no. 9, pp. 1847–1854, 2005.
- [2] C. R. Qi, W. Liu, C. Wu, H. Su, and L. J. Guibas, "Frustum pointnets for 3d object detection from rgb-d data," in *Conference on Computer Vision and Pattern Recognition*, 2018, pp. 918–927.
- [3] M. Herzog and K. Dietmayer, "Training a fast object detector for lidar range images using labeled data from sensors with higher resolution," in *Intelligent Transportation Systems Conference*, 2019, pp. 2707–2713.
- [4] A. Dubois and J. Bresciani, "Person identification from gait analysis with a depth camera at home," in *Engineering in Medicine and Biology Society*, 2015, pp. 4999–5002.
- [5] A. Wu, W. Zheng, and J. Lai, "Robust depth-based person re-identification," *Transactions on Image Processing*, vol. 26, no. 6, pp. 2588–2603, 2017.
- [6] C. Thorpe, F. Li, Z. Li, Z. Yu, D. Saunders, and J. Yu, "A coprime blur scheme for data security in video surveillance," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 12, pp. 3066–3072, 2013.
- [7] K. Akkaya, V. Baboolal, N. Saputro, S. Uluagac, and H. Menouar, "Privacy-preserving control of video transmissions for drone-based intelligent transportation systems," in *Conference on Communications and Network Security*, 2019, pp. 1–7.
- [8] F. Pittaluga and S. J. Koppal, "Pre-capture privacy for small vision sensors," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 11, pp. 2215–2226, 2017.
- [9] M. Eldib, F. Deboeverie, W. Philips, and H. Aghajan, "Discovering activity patterns in office environment using a network of low-resolution visual sensors," *Journal of Ambient Intelligence and Humanized Computing*, vol. 9, no. 2, pp. 381–411, 2018.
- [10] A. Leigh, J. Pineau, N. Olmedo, and H. Zhang, "Person tracking and following with 2d laser scanners," in *International Conference on Robotics and Automation*, 2015, pp. 726–733.
- [11] L. Beyer, A. Hermans, T. Linder, K. O. Arras, and B. Leibe, "Deep person detection in two-dimensional range data," *Robotics and Automation Letters*, vol. 3, no. 3, pp. 2726–2733, 2018.
- [12] I. Bhattacharya and R. J. Radke, "Arrays of single pixel time-of-flight sensors for privacy preserving tracking and coarse pose estimation," in *Winter Conference on Applications of Computer Vision*, 2016, pp. 1–9.
- [13] Chih-Wei Hsu and Chih-Jen Lin, "A comparison of methods for multiclass support vector machines," *Transactions on Neural Networks*, vol. 13, no. 2, pp. 415–425, 2002.

TABLE II  
CLASSIFICATION RESULTS FOR A CONTROLLED TEST ENVIRONMENT.

| Class     | Precision | Recall |
|-----------|-----------|--------|
| Human     | 1.0       | 0.995  |
| Human:in  | 1.0       | 1.0    |
| Human:out | 1.0       | 0.99   |