Research Article
PCA-Based Line Detection from Range Data for Mapping and Localization-Aiding of UAVs

Roberto Opromolla, Giancarmine Fasano, Michele Grassi, Al Savvaris, and Antonio Moccia

1Department of Industrial Engineering, University of Naples "Federico II", P.le Tecchio 80, 80125 Napoli, Italy
2School of Aerospace Transport and Manufacturing, Cranfield University, College Rd, Cranfield MK43 0AL, UK

Correspondence should be addressed to Roberto Opromolla; roberto.opromolla@unina.it

Received 10 November 2016; Accepted 19 February 2017; Published 4 April 2017

This paper presents an original technique for robust detection of line features from range data, which is also the core element of an algorithm conceived for mapping 2D environments. A new approach is also discussed to improve the accuracy of position and attitude estimates of the localization by feeding back angular information extracted from the detected edges in the updating map. The innovative aspects of the line detection algorithm regard the proposed hierarchical clusterization method for segmentation. Instead, line fitting is carried out by exploiting the Principal Component Analysis, unlike traditional techniques relying on least squares linear regression. Numerical simulations are purposely conceived to compare these approaches for line fitting. Results demonstrate the applicability of the proposed technique as it provides comparable performance in terms of computational load and accuracy compared to the least squares method. Also, performance of the overall line detection architecture, as well as of the solutions proposed for line-based mapping and localization-aiding, is evaluated exploiting real range data acquired in indoor environments using an UTM-30LX-EW 2D LIDAR. This paper lies in the framework of autonomous navigation of unmanned vehicles moving in complex 2D areas, for example, being unexplored, full of obstacles, GPS-challenging, or denied.

1. Introduction

Among the technical challenges which drive the research activities carried out in the field of Unmanned Aerial Vehicles (UAVs), a major issue is to improve their level of autonomy. As discussed in [1], this is equivalent to strengthen their capabilities of autonomous guidance, navigation, and control, taking into account factors like environment and mission complexity, situational awareness, and real-time implementation.

For instance, in the case of large-scale outdoor scenario, autonomous and safe navigation is ensured by the classical sensor fusion architectures integrating an Inertial Navigation System (INS) with GPS, typically indicated as GPS-INS [2, 3], which has been extensively exploited by researchers considering fixed-wing [4, 5], helicopter [6, 7], and multirotor [8] UAVs. Nevertheless, a wide range of both military and civil applications, for example, urban and indoor surveillance, infrastructure monitoring, and exploration, require a micro-UAV (MAV) to be able to navigate in more complex environments, such as unknown areas, full of static, and/or mobile obstacles in which the GPS signal may be completely absent (GPS-denied, e.g., indoor) or unreliable due to multipath, absorption, and jamming phenomena (GPS-challenging, e.g., urban or natural canyons). Hence, alternative solutions have to be found, which involve the use of exteroceptive sensors that can be active, such as RADAR [9], LIDAR [10], ultrasonic rangefinders [11], and ultra-wideband (UWB) positioning system [12], passive, for example, monocular [13] and stereovision [14] cameras operating in the visible band of the electromagnetic spectrum, or hybrid, like RGB-depth cameras [15], as they simultaneously acquire passive RGB images and depth images of the same scene in an active way [16].

Indeed, the use of these sensors has provided a boost to the capability of navigating autonomously in complex...
environments by performing Simultaneous Localization and Mapping (SLAM) [17, 18], that is, the real-time process by which a mobile marine/ground/aerial robot, moving within an unknown area, computes its own trajectory (localization) while simultaneously building a map of that environment (mapping). However, the implementation of SLAM algorithms on board MAVs still presents significant technical challenges, mainly due to the limited resources available on board for sensing and computation, as well as their complex and fast 3D dynamics [19]. On this basis, the adoption of feature-based approaches for mapping is extremely promising as it can lead to a sparse but compact representation of the observed environment, with higher speed of execution and reduced amount of data storage with respect to alternative occupancy grid techniques [20]. Among the different technological solutions suitable for SLAM (such as vision-based [21], LIDAR-based [22, 23], and RGB-depth-based [24]), LIDAR are more robust than passive vision systems to operate in indoor environment, especially in case of low- or no-light conditions. Also, they directly provide range data. However, LIDAR-based SLAM presents significant technical challenges, mainly related to the lower resolution and limited texture of the available measurements with respect to images provided by passive vision sensors. In order to deal with these issues, strong research efforts have been recently addressed to the possibility of augmenting LIDAR-based SLAM exploiting intensity information [25] or detection of robust features [26]. Instead, a different investigated solution consists in integrating passive vision and active LIDAR sensors [27, 28] in order to exploit complementary advantages and compensate their drawbacks.

In this framework, the aim of this paper is to present a technique for fast and robust detection of line features from range data provided by a 2D LIDAR. Indeed, these sensors, due to their limited size and weight, are particularly suited for installation on board MAVs operating in indoor areas, for example, office-like environments, but also outdoor due to their limited sensitivity to ambient light variations. Line extraction algorithms are typically characterized by two steps, namely, segmentation and line fitting. The former identifies the number of potential lines and associates the measured points to each candidate, while the latter allows extracting the line parameters from the aggregates of points. A review of the techniques for line detection from 2D LIDAR data, mostly used in the literature, is presented in [29]. Six different approaches to the segmentation task, namely, the split-and-merge algorithm [30, 31], the line-regression algorithm [32], the incremental algorithm [33] (also known as line-tracking [34]), RANSAC [35], the Hough Transform [36], and the Expectation-Maximization (EM) algorithm [37], are compared in [29] in terms of speed, complexity, correctness, and precision by means of experimental tests on real scan data collected in a large-scale office environment. In [29], each one of these techniques is further accelerated by implementing clusteringization, which allows coarsely dividing a raw laser scan into contiguous groups (clusters). With regard to line fitting, [29] states that state-of-the-art approaches rely on the least squares (LS) method.

The proposed line detection algorithm provides solutions for both segmentation and line fitting. The issue of developing and testing robust techniques for segmentation has critical importance due to the challenges of dealing with clutter in 2D environments. In this respect, an original two-level clustering algorithm is presented which includes a merging check to avoid the possibility that an edge in the scene is split in two or more parts only due to the measurements noise. This solution for segmentation can also be adapted to represent both nonlinear and piecewise linear structures. An algorithm based on the Principal Component Analysis (PCA) [38] is proposed for line fitting. Specifically, it relies on the capability of the PCA to estimate the principal directions of the 3D datasets acquired by a 2D LIDAR. PCA has already been used to detect lines from row and column edges extracted in 2D images [39]. Regarding applications with range data, PCA is used in [40] to identify feature points belonging to planar structures and in [41] to recognize echoes produced by pole-like objects and to extract the direction from road-shaped point clouds in 3D urban environments. In both [40, 41], data from 3D LIDAR systems, for example, the Multiplatform Mobile Laser Scanning [42], are processed. Such a system is not suitable for application on board MAVs. PCA-based methods have also been widely exploited in the open literature for pose estimation applications in different scenarios [43, 44]. Although PCA is not a new concept for line fitting, a detailed analysis of its performance when used with LIDAR range data is not available in the literature. On this basis, performance of the proposed PCA-based line detection algorithm is assessed by means of both numerical simulations and experimental tests. A specific analysis is carried out to compare the PCA to the LS approach for line fitting in terms of computational load and accuracy level. This is important to demonstrate the applicability of the proposed approach for implementation on board MAVs, since LS is a standard and extremely fast method to perform linear regression given a set of range 3D data.

An additional aim of this paper is to demonstrate the possibility of the PCA-based line detection algorithm to be used within a 2D SLAM process. In this respect, it is worth outlining that many UAV applications (e.g., indoor inspection and monitoring) involve the necessity of flying in 2D environments. In this context, the capability of estimating the horizontal motion and simultaneously a 2D map is relevant, also considering that the height above ground can be measured by using additional sensors suitable for the installation on board MAVs (e.g., ultrasonic rangefinders) or also by redirecting downward a narrow portion of the scan emitted by the 2D LIDAR using mirrors [19]. Hence, a line-based mapping algorithm is here presented. Specific attention is addressed to the procedures foreseen for handling the features to be stored in the updating map, as well as for improving the accuracy of the position and attitude solution provided by a localization algorithm (localization-aiding). A SLAM algorithm using a probabilistic PCA algorithm for line detection from 2D LIDAR data is presented in [45]. However, results obtained in [45] applying offline the algorithm on real data put into evidence errors in the estimated maps due to the complexity of correctly merging clusters from
subsequent point clouds. This justifies the need of further investigations to demonstrate reliability and effectiveness in performing this task. Another example of localization-aiding exploiting line detection is discussed in [46]. Specifically, the relative displacement and inclination between corresponding lines in consecutive frames are integrated within an Extended Kalman Filter. Instead, in this paper, correspondences between new lines extracted from the actual frame and lines already stored in the updating map are determined using a more robust approach based on multiple angular and position metrics, and they are adopted to directly correct the vehicle’s position and the heading angle. However, it is worth outlining that the development and performance analysis of a complete SLAM framework based on the proposed algorithms for mapping and localization-aiding is out of the scope of this paper.

This paper is organized as follows. Section 2 describes in detail the proposed line detection algorithm operating on range data, as well as the line-based mapping technique within which the algorithm is implemented, including the approach proposed for localization-aiding. Section 3 presents the numerical simulations which are carried out to compare PCA and LS methods for line fitting in terms of computational load and accuracy level. The experimental setup conceived to collect real data and the results obtained applying the PCA-based line detection algorithm as well as an example of full implementation of localization and mapping are described in Section 4. Finally, Section 5 contains the conclusion and indications concerning possible future works.

2. PCA-Based Line Detection

The PCA-based algorithm proposed for line detection is conceived to operate receiving as input only range measurements provided by a 2D LIDAR. Typically, the data collected by these sensors are given in output according to a representation in polar coordinates, that is, the scan angle (θ), which is the angular position in the LIDAR Field-of-View (FOV) of each emitted ray with respect to a reference direction (usually it is the boresight one), and the range (ρ), which is measured by exploiting the Time-of-Flight (TOF) principle. However, the measured 2D point clouds will be expressed in the following also using the representation in rectangular coordinates (x and y) which can be obtained by applying

$$\begin{bmatrix} x \\ y \end{bmatrix} = \rho \begin{bmatrix} \cos(\theta) \\ \sin(\theta) \end{bmatrix}. \tag{1}$$

Once a 2D point cloud is acquired by a 2D LIDAR, the PCA-based line detection algorithm starts with the segmentation step, which is carried out by performing a clusterization process characterized by two hierarchical levels. At the higher level, the scan is coarsely subdivided into main clusters which represent clearly separated portions of the observed scene. This is done by looking for peaks of the interpoint distance (i.e., the distance between consecutive points in the scan) as a function of θ. These peaks, indicated as break-points, represent the beginning of any individual cluster which is found moving along the scanning direction, and they are recognized as those locations where the interpoint distance is larger than a specific threshold (DTh).

The correct value of DTh can be easily selected, considering the specifications of the available sensor in terms of angular resolution and operating range, in order to avoid separating points belonging to same main cluster. For instance, the UTM-30LX-EW produced by Hokuyo is characterized by an angular resolution of 0.25° and a maximum range of 30 m. Hence, DTh is set to 15 cm considering that the maximum interpoint distance at a range of 30 m is 13.1 cm. Figure 1 shows an example of implementation of this step of the algorithm over a point cloud acquired within a 2D corridor environment, using the UTM-30LX-EW LIDAR.

Four separate main clusters are identified in Figure 1. Before moving to the next phase, it is worth mentioning that if a cluster is too small (typically less than 5 points) it is neglected from the line detection process. Once the main clusters are separated from each other, a 2nd-level (low-level) clusterization is required to deeply search for subclusters potentially corresponding to different linear structures in the observed scene. Specifically, this step looks for local peaks of the two derivatives with respect to θ of the functions $x'(\theta)$ and $y'(\theta)$ by setting, as new break-points, those locations (i) at which either (2a) or (2b) is satisfied:

$$x'(\theta_i) = \frac{dx(\theta_i)}{d\theta} + \frac{1}{N-1} \sum_{i=1}^{N} \left( \frac{dx(\theta_i)}{d\theta} - \frac{1}{N} \sum_{i=1}^{N} \frac{dx(\theta_i)}{d\theta} \right)^2, \tag{2a}$$

$$y'(\theta_i) = \frac{dy(\theta_i)}{d\theta} + \frac{1}{N-1} \sum_{i=1}^{N} \left( \frac{dy(\theta_i)}{d\theta} - \frac{1}{N} \sum_{i=1}^{N} \frac{dy(\theta_i)}{d\theta} \right)^2. \tag{2b}$$

In the equations above, where N is the number of points in the considered main cluster, the threshold assigned to identify the local peaks of $x'(\theta)$ and $y'(\theta)$ is the sum of their mean and standard deviation computed over the cluster. This original approach is particularly convenient to better describe quite complex indoor environments where, for instance, the walls are characterized by frequent corners due to the presence of doors, windows, or other obstacles. By looking at Figure 2, it is clear how the number of break-points increases after the 2nd-level clusterization is applied with respect to Figure 1.

Of course, due to the measurement noise, some of the new break-points do not correspond to the beginning of a different linear structure. This issue is solved, after line fitting, by implementing an additional step necessary to merge adjacent lines corresponding to the same linear structure. Before entering this detail, it is necessary to focus on the line fitting step, which relies on the PCA, that is, a technique able
to determine the principal directions of a multidimensional dataset by analyzing the eigenvectors and eigenvalues of its covariance matrix ($Q$). In this case, given a generic subcluster (composed of $N_{SC}$ measurements), $Q$ can be estimated by applying

$$Q = \frac{1}{N_{SC}} \left[ \sum_{i=1}^{N_{SC}} \left( \frac{x_i - \frac{1}{N_{SC}} \sum_{j=1}^{N_{SC}} x_j}{N_{SC}} \right) \left( \frac{y_i - \frac{1}{N_{SC}} \sum_{j=1}^{N_{SC}} y_j}{N_{SC}} \right) \right].$$

Hence, a line is extracted from the subcluster if the ratio, $r$, between the two eigenvalues of $Q$ (which is a measure of the elongation of the point-set) is larger than a specific threshold (ETh). Indeed, if the point-set does not have a linear structure, the value of $r$ will not be far from 1. So, the choice of ETh mainly affects the capability to declare a subcluster as line depending on its length, and it must be done considering the range measurement accuracy of the available LIDAR. Of course, the larger ETh is, the more elongated the selected point-set must be recognized as line. For instance, if ETh is set to 100, the probability to identify a 75 cm linear structure at a distance of 10 m (assuming a std of 4 cm in the direction perpendicular to the linear one) is 1.35% (evaluated over $10^5$ simulations). If ETh is increased to 380 this probability goes to zero. Specific numerical simulations (not discussed for the sake of brevity) have demonstrated that values of ETh around a few tens are adequate to robustly extract even very small linear structures (down to 10 cm) from the observed scene.

If the condition on $r$ is satisfied, the parameters suitable to univocally identify the line can be computed, for example, on the basis of the classical Cartesian representation shown in

$$y = m_L x + n_L,$$

where $x$ and $y$ are the coordinates of a generic point of the line, $m_L$ is the angular coefficient, and $n_L$ is the known term. As the direction of the line ($\alpha_L$) is given by the eigenvector (whose components are $\lambda_x$ and $\lambda_y$) corresponding to the maximum eigenvalue of $Q$, $m_L$ is derived using

$$m_L = \tan(\alpha_L) = \frac{\lambda_y}{\lambda_x},$$

while $n_L$ is obtained by assuming that the centroid of the subcluster (whose coordinates are $x_C$ and $y_C$) belongs to the line, as shown in

$$n_L = y_C - m_L x_C.$$
observed linear structure). This is done by projecting the first and last points of the subcluster in the line direction. Finally, the previously mentioned merging step allows joining two adjacent subclusters if the corresponding lines satisfy the conditions highlighted in

\[
\begin{align*}
(\alpha_{t1} - \alpha_{t2}) < \alpha_T \cap &\left[\left(\alpha_{t1} - \alpha_{CC}\right) < \alpha_T\right] \cap \left[\left(\alpha_{t2} - \alpha_{CC}\right) < \alpha_T\right] \\
\cup &\left[\left(\alpha_{t1} - (180 - \alpha_{CC})\right) < \alpha_T\right] \cap \left[\left(\alpha_{t2} - (180 - \alpha_{CC})\right) < \alpha_T\right],
\end{align*}
\]

where \(\alpha_{CC}\) is the orientation of the direction of the segment which links their centroids, \(\alpha_T\) is an angular threshold, and the subscripts “1” and “2” are used to distinguish between the two lines. Also, the symbols \(\cap\) and \(\cup\) are used to indicate the intersection and union operators from set theory, respectively. Small values of \(\alpha_T\) (e.g., \(0.05^\circ\)) allow the method to recognize even slight variation in the orientation of the observed linear structure. As a result, Figure 3 is obtained by applying this line fitting approach (setting \(ETH\) to 50 and including the merging step) to the clustered dataset shown in Figure 2.

2.1. Line-Based Mapping and Localization-Aiding. The algorithm proposed for line-based mapping is conceived to build a 2D map of the observed environment in an incremental way, that is, by progressively adding and updating lines as they are detected through subsequent acquisitions of range data. Of course, in accordance with the concept of SLAM, an adequate solution in terms of position and attitude must be provided by a localization algorithm. Indeed, the first step consists in applying this pose solution to convert the acquired point cloud from the sensor reference frame (SRF) to the inertial frame (e.g., East-North-Up or North-East-Down with fixed origin) in which the trajectory of the moving platform and the map of the observed scenario are determined. Afterwards, the PCA-based line detection algorithm is implemented and the extracted line segments are used to update the map. Specifically, each of this segment is a potential candidate to contribute to the map generation process, but a decision-making step is required to establish whether or not it corresponds to a linear structure already contained (at least partially) in the map. To this aim, each of the new detected lines is compared to all the existing ones in the map using two metrics, namely, \(\Delta \alpha_T\) and \(\Delta \rho_L\). The former represents the difference between the two compared lines in terms of orientation (\(\alpha_L\)) which can be easily obtained from (5). The latter, instead, is a distance metric derived by summing two contributions as shown in

\[
\Delta \rho_L = \Delta d_L + \Delta C_L, \tag{8}
\]

\(\Delta d_L\) is the difference between the two compared lines in terms of the distance from the map origin \(d_L\) which is computed using

\[
d_L = \frac{-y_C + \tan(\alpha_L) x_C}{\sqrt{1 + \tan^2(\alpha_L)}}, \tag{9}
\]

Instead, \(\Delta C_L\) is the distance between the centroids of the two compared lines, whose coordinates are \((x_{C1}, y_{C1})\) and \((x_{C2}, y_{C2})\), respectively. \(\Delta C_L\) can be derived using

\[
\Delta C_L = \sqrt{(x_{C1} - x_{C2})^2 + (y_{C1} - y_{C2})^2}. \tag{10}
\]

Given a line detected in the current dataset, if it is not possible to find an already stored segment for which both metrics are kept below a minimum threshold assigned taking localization errors into account (e.g., 15˚ and 1 m, resp.), it is classified as a new element of the map. Conversely, the detected line is matched to the line in the map for which \(\Delta \rho_L\) is the minimum. Indeed, among the two metrics, \(\Delta \rho_L\) is certainly the most important to ensure correct line matches, as in many indoor scenarios (e.g., corridors); it is likely to find more than one linear structure with the same orientation \(\Delta \alpha_T\) is similar) but located at different positions. Once all the possible matches are found, the corresponding elements of the map must be updated. This is done by projecting the two ends of the new line on the direction of the old one, so that the two updated ends are the ones for which the length of the line is the maximum.

The matches between new detected lines and old elements of the map can be used for localization-aiding, that is, to improve the accuracy of the position and attitude estimates provided by a localization algorithm. Here, a strategy for localization-aiding is proposed which exploits angular information extracted from the previously mentioned line matches. Specifically, if \(N_{ML}\) is the total number of line matches at a generic time \(t_k\), the heading solution at that time can be corrected as shown in

\[
\gamma_{NEW} (t_k) = \gamma_{OLD} (t_k) + \Delta \gamma_{MAP}, \tag{11}
\]
where the subscripts “OLD” and “NEW” indicate the original (provided by the localization algorithm) and updated estimates of the heading angle (\( \gamma \)), respectively, while the correcting factor (\( \Delta \gamma \)) is the weighted average of the angular offsets between the matched lines, as shown in

\[
\Delta \gamma_{\text{MAP}} (t_k) = \frac{\sum_{i=1}^{N_{\text{ML}}} w_{\text{ML}}^i \Delta \gamma_{\text{ML}}^i}{\sum_{i=1}^{N_{\text{ML}}} w_{\text{ML}}^i}.
\]  

(12)

The values of the weights (\( w_{\text{ML}} \)) in (12) are determined using the distance metric \( \Delta \rho \) as shown in (13), thus giving more relevance to those matches which are closer in terms of location occupied in the map

\[
w_{\text{ML}} = \left( \frac{1}{\Delta \rho_k} \right)^2.
\]  

(13)

At this point, also the position can be updated using

\[
\mathbf{T}_{\text{NEW}} (t_k) = \mathbf{T} (t_{k-1}) + \begin{bmatrix} \cos (\Delta \gamma_{\text{MAP}}) & \sin (\Delta \gamma_{\text{MAP}}) \\ -\sin (\Delta \gamma_{\text{MAP}}) & \cos (\Delta \gamma_{\text{MAP}}) \end{bmatrix} \left( \mathbf{T}_{\text{OLD}} (t_k) - \mathbf{T} (t_{k-1}) \right),
\]  

(14)

where \( t_{k-1} \) is the previous time instant and the subscripts “OLD” and “NEW” are again used to indicate the original (provided by the localization algorithm) and updated estimates of the vehicle’s position (\( \mathbf{T} \)).

3. Numerical Simulation Results

This section presents description and results of the numerical simulations carried out to assess absolute performance of the PCA when used for line fitting and to get a comparison with the traditional LS method [47], for which a fast version is implemented [48]. The main aspects which are deemed relevant to this analysis are the attained accuracy level and the computational load.

Since the goal of these simulations is to assess line fitting performance of the PCA method over sets of points distributed according to a linear pattern, a realistic reproduction of the operation of a range sensor (e.g., 2D LIDAR) is not carried out. In this respect, the simulations simply rely on the generation of a set of \( n \) points, randomly distributed along a linear direction identified by the angular coefficient (\( m_L \)), so that the point coordinates (\( x' \) and \( y' \)) can be determined using

\[
x' = x_S i,
\]

\[
y' = m_L x' + \nu_i,
\]  

(15)

where \( x_S \) is the point separation and \( \nu_i \) is the \( i \)th extraction from a normal distribution with zero mean and standard deviation equal to \( \sigma \). An example of simulated dataset is shown in Figure 4.

The effect on the results of considering different settings for \( x_S \) and \( \sigma \) is analyzed. Indeed, \( x_S \) is a measure of the angular resolution of the range data (although its variation can also reproduce the effect of modifying the distance of the sensor from the scene given a fixed resolution), while \( \sigma \) simulates the sensor noise in the measured distance. On the other hand, \( m_L \) is set to 0 as line fitting performance is independent of the orientation of the linear dataset. As a consequence of this choice, the length of the simulated linear dataset (\( L \)), which also has an impact on line fitting performance, is determined by the selected values of \( n \).

Once the setting parameters of the simulated range data are chosen, both PCA and LS methods are applied obtaining the corresponding lines, each one identified by an angular coefficient (\( m_{\text{PCA}} \) and \( m_{\text{LS}} \), resp.) and a constant term (\( n_{\text{PCA}} \) and \( n_{\text{LS}} \), resp.). Finally, the line fitting accuracy is evaluated as the standard deviation of the distance of the assigned points from the estimated line (\( \text{Err}_{\text{PCA}} \) and \( \text{Err}_{\text{LS}} \), resp.), according to

\[
\text{Err}_{\text{PCA}} = \frac{1}{n} \sum_{i=1}^{n} (y^i - m_{\text{PCA}} x^i - n_{\text{PCA}})^2,
\]  

(16a)

\[
\text{Err}_{\text{LS}} = \frac{1}{n} \sum_{i=1}^{n} (y^i - m_{\text{LS}} x^i - n_{\text{LS}})^2.
\]  

(16b)

It is worth mentioning that, for any selected configuration of the setting parameters, the results are averaged on 10000 simulations thus having the possibility of carrying out meaningful statistical analyses. Firstly, different values of \( x_S \) are considered, ranging from 5 mm to 10 cm, thus being consistent with typical 2D LIDAR specification in terms of angular resolution (e.g., 0.1°–0.25°) taking also their typical operational range (up to a few tens of meters) into account. Instead, \( \sigma \) is kept fixed to 4 cm, which is also in line with typical range noise of a 2D LIDAR. The analyzed configurations and the corresponding simulation results are summarized in Table 1 where the computational time (\( t_{\text{PCA}} \) and \( t_{\text{LS}} \), resp.) is expressed in milliseconds (ms).

It can be clearly stated that the computational time provided by the PCA (largely less than 1 ms) is comparable to the result of the LS approach. Also, there are no significant differences in the accuracy level. Indeed, the difference between the root mean squares (rms) of the estimation error provided by the two methods is always negligible if compared to the absolute performance attained by the PCA. This result

![Figure 4: Example of simulated point datasets. Setting parameters: \( m_L = 0; n = 1000; x_S = 4 \) cm; \( \sigma = 50 \) cm.](image-url)
Table 1: Comparison between PCA and LS approaches for line fitting. The same noise level ($\sigma = 4$ cm) is considered for all the analyzed configurations of the setting parameters. Results are averaged over 10000 simulations.

<table>
<thead>
<tr>
<th>$x_S$ (m)</th>
<th>$n$</th>
<th>$L$ (m)</th>
<th>$t_{PCA}$ mean (ms)</th>
<th>$t_{LS}$ mean (ms)</th>
<th>$Err_{PCA}$ rms (m)</th>
<th>$(Err_{PCA} - Err_{LS})/Err_{PCA}$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.005</td>
<td>200</td>
<td>1</td>
<td>0.10</td>
<td>0.02</td>
<td>0.040</td>
<td>2.41 $\cdot 10^{-1}$</td>
</tr>
<tr>
<td>0.005</td>
<td>1000</td>
<td>5</td>
<td>0.12</td>
<td>0.04</td>
<td>0.040</td>
<td>9.60 $\cdot 10^{-3}$</td>
</tr>
<tr>
<td>0.005</td>
<td>2000</td>
<td>10</td>
<td>0.59</td>
<td>0.34</td>
<td>0.040</td>
<td>2.00 $\cdot 10^{-5}$</td>
</tr>
<tr>
<td>0.01</td>
<td>100</td>
<td>1</td>
<td>0.09</td>
<td>0.02</td>
<td>0.040</td>
<td>2.41 $\cdot 10^{-1}$</td>
</tr>
<tr>
<td>0.01</td>
<td>500</td>
<td>5</td>
<td>0.11</td>
<td>0.03</td>
<td>0.040</td>
<td>9.60 $\cdot 10^{-3}$</td>
</tr>
<tr>
<td>0.01</td>
<td>1000</td>
<td>10</td>
<td>0.12</td>
<td>0.04</td>
<td>0.040</td>
<td>2.41 $\cdot 10^{-3}$</td>
</tr>
<tr>
<td>0.025</td>
<td>40</td>
<td>1</td>
<td>0.09</td>
<td>0.02</td>
<td>0.039</td>
<td>2.41 $\cdot 10^{-1}$</td>
</tr>
<tr>
<td>0.025</td>
<td>200</td>
<td>5</td>
<td>0.10</td>
<td>0.02</td>
<td>0.040</td>
<td>9.60 $\cdot 10^{-3}$</td>
</tr>
<tr>
<td>0.025</td>
<td>400</td>
<td>10</td>
<td>0.10</td>
<td>0.03</td>
<td>0.040</td>
<td>2.40 $\cdot 10^{-3}$</td>
</tr>
<tr>
<td>0.05</td>
<td>20</td>
<td>1</td>
<td>0.09</td>
<td>0.02</td>
<td>0.038</td>
<td>2.42 $\cdot 10^{-1}$</td>
</tr>
<tr>
<td>0.05</td>
<td>100</td>
<td>5</td>
<td>0.09</td>
<td>0.02</td>
<td>0.040</td>
<td>9.60 $\cdot 10^{-3}$</td>
</tr>
<tr>
<td>0.05</td>
<td>200</td>
<td>10</td>
<td>0.10</td>
<td>0.03</td>
<td>0.040</td>
<td>2.40 $\cdot 10^{-3}$</td>
</tr>
<tr>
<td>0.1</td>
<td>10</td>
<td>1</td>
<td>0.09</td>
<td>0.02</td>
<td>0.036</td>
<td>2.41 $\cdot 10^{-1}$</td>
</tr>
<tr>
<td>0.1</td>
<td>50</td>
<td>5</td>
<td>0.09</td>
<td>0.02</td>
<td>0.039</td>
<td>9.60 $\cdot 10^{-3}$</td>
</tr>
<tr>
<td>0.1</td>
<td>100</td>
<td>10</td>
<td>0.08</td>
<td>0.02</td>
<td>0.036</td>
<td>2.41 $\cdot 10^{-1}$</td>
</tr>
</tbody>
</table>

Table 2: Comparison between PCA and LS approaches for line fitting. Effect of the noise level ($\sigma = 4$ cm; $x_S = 5$ cm; $L = 1$ m). Results are averaged over 10000 simulations.

<table>
<thead>
<tr>
<th>$\sigma$ (m)</th>
<th>$Err_{PCA}$ rms (m)</th>
<th>$(Err_{PCA} - Err_{LS})/Err_{PCA}$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.005</td>
<td>0.005</td>
<td>4.00 $\cdot 10^{-3}$</td>
</tr>
<tr>
<td>0.01</td>
<td>0.009</td>
<td>1.50 $\cdot 10^{-2}$</td>
</tr>
<tr>
<td>0.05</td>
<td>0.048</td>
<td>3.79 $\cdot 10^{-1}$</td>
</tr>
<tr>
<td>0.1</td>
<td>0.096</td>
<td>1.56</td>
</tr>
<tr>
<td>0.2</td>
<td>0.204</td>
<td>6.85</td>
</tr>
</tbody>
</table>

is not affected by variation of density of the simulated dataset, which is represented by $x_S$. As expected, the longer the dataset is (for a given value of $x_S$), the higher the associated value of $r$ becomes, and consequently the closer PCA and LS accuracy levels get. Indeed, the difference reduces of two orders of magnitude, that is, from $10^{-1}$% to $10^{-3}$%, as $L$ goes from 1 m to 10 m. Additional numerical simulations are carried out to show how the results presented above can be affected by a variation in the level of measurement noise. Specifically, $x_S$ is set to 5 cm and $n$ is set to 20, so that $L$ is 1 m, thus reproducing a quite sparse dataset characterized by a linear spatial distribution. Instead, five different values of $\sigma$ are considered ranging from 5 mm (best-case) to 20 cm (worst-case). Simulation results are collected in Table 2.

Before discussing the absolute accuracy level attained by the PCA and the comparison with the LS performance, it is worth outlining that three of the selected values of $\sigma$, that is, 5 mm, 1 cm, and 5 cm, are consistent with the nominal levels of performance that scanning LIDAR, exploiting the TOF principle for distance measurement, are able to ensure. Indeed, high-performance systems may be characterized by range accuracies from around 1 cm down to even submillimeter level, while, for the majority of scanning LIDAR, $\sigma$ never exceeds the interval going from 3 cm to 5 cm. Conversely, the remaining values to which $\sigma$ is set in the simulations (i.e., 10 cm and 20 cm) are representative of worst-case conditions which may only occur if the sensor receives too much energy due to, for instance, direct sun illumination in outdoor environment. On this basis, results in Table 2 show that the PCA is able to carry out line fitting with the same accuracy level as the LS method, even operating on sparse data characterized by large nonnominal values of the range measurement noise.

Finally, it is interesting to carry out further simulations to evaluate the impact on line fitting accuracy caused by the presence of outliers in the considered dataset. Since we are just dealing with the problem of line fitting, outliers must be modeled as measured points which have been erroneously attributed to a given cluster. Specifically, outliers are generated by randomly extracting (according to a uniform probability distribution) a percentage of points ($P_O$) in the dataset for which the displacement from the reference direction is equal to $3\sigma$. Again, it is important to stress that even if these simulated data are not representative of real laser scan data, they can be considered adequate for the analysis of line fitting performance as the segmentation necessary to separate clusters belonging to different linear structures is not addressed. For this analysis, $x_S$ is set to 2.5 cm and $n$ is set to
200, so that $L$ is 5 m, while $\sigma$ is 4 cm. Simulation results are collected in Table 3, where four different values, namely, 1%, 5%, 10%, and 20%, are considered for $P_O$. Results in Table 3 show that the PCA-based line fitting technique has comparable performance to LS and ensures cm-level accuracy even setting $P_O$ to 20%, which in the analyzed case corresponds to 40 outliers over the total number of 200 points composing the dataset.

### 4. Tests on Experimental Data

The scope of this section is to present the results obtained applying the proposed algorithms and techniques on real range data. To this aim, it is necessary to describe the experimental setup conceived for data collection as well as to clarify in which scenarios the experimental tests are carried out. As regards the setup, the selected 2D LIDAR is the UTM-30LX-EW produced by Hokuyo, for which the main specifications and a full characterization can be found in [49] and [50], respectively. This sensor is installed on a portable platform together with one autopilot, that is, the Pixhawk produced by 3D robotics, which is used to obtain reference information regarding the attitude of the platform, and one embedded board, that is, the Nitrogen6X produced by Boundary-Devices. This latter component is used to register data from both the Pixhawk (using USB connection) and the LIDAR (using Ethernet connection) by exploiting the corresponding nodes of the Robot Operating System (ROS) [51], that is, the mavros and the urg_node, respectively. This ensures simultaneous recording of range data from the LIDAR and inertial data from the Pixhawk, together with their timestamps, within the same bag-file. The data acquisition is commanded by a laptop exploiting a Wi-Fi connection. A conceptual representation of the experimental setup is shown in Figure 5.

This setup has been first used to perform static acquisition of range data within indoor cluttered environments considering both limited-size areas, like corridors or offices, and wider scenarios such as halls. Figure 6 shows a portion of the scaled-planimetry of the ground floor of Building 83 (which is part of the School of Engineering) inside the Cranfield University Campus. In this planimetry, red polygons are used to highlight the areas where LIDAR data have been collected.

Each horizontal sequence of images in Figure 7 shows partial (clusterization) and final (line fitting) outputs obtained implementing the PCA-based line detection algorithm over point clouds collected within the previously defined indoor scenarios. These results demonstrate the robustness of the proposed line detection approach. Indeed, it has been able to extract a large number of linear structures characterized by largely variable lengths, that is, ranging from a few meters down to a few tens of centimeters. The standard deviation of the distance between the points belonging to each cluster and the corresponding detected line is of cm-level. Also, the algorithm appears to be able to deal with the presence of clutter in the observed scene. This is particularly clear by looking at Figures 7(b), 7(c), and 7(e).

#### 4.1. Localization and Mapping Experiments

Dynamic acquisitions have also been realized with the experimental setup transported by a human operator, thus partially reproducing the 3D dynamics which characterizes the flight of MAVs. The 2D maze shown in Figure 8 is the test area prepared for the dynamic experiment.

The major goal of these tests is the preliminary assessment of the algorithms presented in Section 2.1 for mapping and localization-aiding. This is done by implementing...
the entire localization and mapping process, in which the LIDAR/Inertial Odometry algorithm presented in [52] is in charge of the localization step. The structure of this localization and mapping process is summarized in Figure 9.

Once LIDAR and inertial data are collected exploiting the experimental setup described in the previous subsection, the localization and mapping algorithm are run offline in MATLAB environment. An example of application of this localization and mapping approach is shown in Figure 10.

By looking at Figure 10, it is clear that the line-based mapping technique allows generating a sparse but accurate representation of the observed scene. Indeed, all the edges of the real map (black lines) are adequately represented by the detected lines, in terms of length, location, and orientation. In this respect, the localization-aiding step has a fundamental role since it allows bounding potential drifts of the error in the estimated trajectory which could cause the edges of the estimated map to be characterized by position and/or angular offsets with respect to the real corresponding linear structure.

In order to better highlight the attained advantages, Figure 11(a) shows the correction provided by the localization-aiding step, according to (12), to the time behavior of the heading angle with respect to the solution given by the Pixhawk ($\gamma_{FCU}$). As a consequence, (14) can be applied thus making the estimated trajectory not to drift eastward as it occurs if the localization-aiding step is not implemented (as in [52]). This is highlighted in Figure 11(b).

5. Conclusions

This paper presented a new technique for line detection from range data (2D point clouds) provided by a 2D LIDAR, which is of interest to Unmanned Aerial Vehicles which need to carry out autonomous navigation applications, such as localization and mapping, in 2D cluttered environments.

As regards the two steps composing the algorithm, segmentation consisted of a hierarchical-level clusterization approach suitable for quickly subdividing the measured point cloud into distinct aggregates of points representing potential candidates to be extracted as lines. Line fitting relied on the Principal Component Analysis (PCA) instead of exploiting the classical least squares (LS) linear regression as most of the state-of-the-art approaches.

Numerical simulations were realized to assess line fitting performance of the PCA in comparison with a fast LS solution. The comparison was done in terms of accuracy and computational load, and the simulations considered the variability of dataset density and length, and robustness against noise and outliers. Although the algorithms were implemented offline in MATLAB, results allowed stating that the PCA is compatible with real-time implementation as it kept the run-time below 1ms and close to the LS performance, even considering large clusters of 2000 points. Also, PCA ensured the same accuracy level of LS in spite of the variability of the characteristics of the simulated dataset.

Performance of the entire line detection process was evaluated over real LIDAR data acquired by using a purposely realized experimental setup. Runs over static acquisitions performed in different indoor environments, such as corridors, offices, and halls, demonstrated the effectiveness of the proposed method in extracting most of the linear structures in the scene with cm-level accuracy and its robustness when dealing with the presence of high level of clutter.

In order to demonstrate the applicability of the proposed line detection algorithm for autonomous navigation in 2D environments, it was integrated within an innovative line-based mapping technique including also a new method for localization-aiding based on the extraction of angular information from the updating map. Dynamic acquisitions within a purposely prepared 2D test area were carried out to obtain a first performance assessment of the line-based mapping algorithm and to show the advantages provided by the introduction of the localization-aiding method. Indeed, the results showed the capability of this approach to generate a sparse but cm-level accurate representation of the observed scene and to correct the heading and position solution provided, respectively, by an inertial unit and by a localization algorithm. Future works will be aimed at testing the proposed methods installing the experimental setup on board of a MAV, in order to fully assess performance in terms of localization and mapping capabilities.
LIDAR measurements
Clusterization, 1st level

LIDAR measurements
Break-points

Clusterization, 2nd level

Line fitting

LIDAR measurements
Detected lines

Figure 7: Continued.
Figure 7: From (a) to (e), step-by-step results from the PCA-based line detection algorithm over LIDAR point clouds collected within indoor cluttered areas. (a)-(b) Corridors; (c)-(d)-(e) halls.

Figure 8: Test area for localization and mapping experiments. The "start" and "end" points of the travelled path are highlighted in black.

Figure 9: Flow diagram of the localization and mapping process implemented on the data from the dynamic test.
Figure 10: LIDAR-based localization and mapping process. The mapping algorithm includes the implementation of the localization-aiding approach.

Figure 11: Effect of the localization-aiding step in terms of heading angle (a) and estimated trajectory (b).
Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was carried out in the frame of Program STAR-Linea 2, financially supported by UniNa and Compagnia di San Paolo.

References


Submit your manuscripts at https://www.hindawi.com