Laser-Camera Data Discrepancies and Reliable Perception in Outdoor Robotics

Thierry Peynot and Abdallah Kassir

Abstract— **This work aims to promote integrity in autonomous perceptual systems, with a focus on outdoor unmanned ground vehicles equipped with a camera and a 2D laser range finder. A method to check for inconsistencies between the data provided by these two heterogeneous sensors is proposed and discussed. First, uncertainties in the estimated transformation between the laser and camera frames are evaluated and propagated up to the projection of the laser points onto the image. Then, for each pair of laser scan-camera image acquired, the information at corners of the laser scan is compared with the content of the image, resulting in a likelihood of correspondence. The result of this process is then used to validate segments of the laser scan that are found to be consistent with the image, while inconsistent segments are rejected. Experimental results illustrate how this technique can improve the reliability of perception in challenging environmental conditions, such as in the presence of airborne dust.**

I. INTRODUCTION

Perception is arguably one of the most critical components of an unmanned ground vehicle (UGV) system. In recent decades, great effort has been invested in the development of better and more accurate perception systems using multiple sensors, with significant success (e.g. the Darpa Urban Grand Challenge [1], [2] or the PerceptOR program [3]). However, state-of-the-art perception systems are still known to fail in a number of situations, in particular in challenging environmental conditions such as the presence of smoke or airborne dust [3], [4], [2]. The typical problem in these cases is that most perception systems tend to confuse dust clouds with actual obstacles. This confusion arises because laser data returns from dust and solid objects can be very similar. If the laser data is to be fused with data from other sensors that do not perceive this dust cloud similarly (e.g. a mm-wave radar [5] or a camera), this fusion will lead to errors in the perception system unless the *consistency* of the two sources of data is checked beforehand.

This work aims at promoting perceptual integrity for autonomous vehicles by exploiting the differences between heterogeneous sensors to detect such situations and process the sensing data appropriately prior to fusion and/or interpretation. In particular, this paper will focus on the very common case where the vehicle is equipped with a camera and one or several (2D) laser range finders (LRF). More specifically, this study aims at comparing the data from laser scans and camera images to detect inconsistencies or discrepancies and filter the data accordingly. Using the same example as before, the LRF usually detects airborne dust present in the environment, while the effect often is much less significant on a visual (or infrared) image, especially when the density of the dust cloud is low.

A laser range finder provides range and bearing data, which is intrinsically very different from the information content of a camera image (an intensity, or colour array on the camera plane). Thus, a direct comparison of the data is not possible. Nevertheless, there are some common characteristics in term of geometry in both types of data. For example, many *corners* in the range data, i.e. points corresponding to a sudden variation of range along the scan, correspond to an *edge* in the image, i.e. a sudden variation of intensity[1]. A number of applications exploit this type of commonality for the registration of visual images to laser range scans, mainly for the perception of indoor scenes (e.g. [6]). However, in these applications, *range images* are needed, which requires a 3D laser scanner, while we only use 2D lasers in this paper. Furthermore, the success of the registration is based on the hypothesis that the data contained in the laser range image and the camera image correspond. When perceiving an outdoor scene, if, for example, airborne dust or smoke appears in the environment, this hypothesis often does not hold anymore [5]. For example, smoke will not be observed by an infrared camera, and light airborne dust can significantly affect a laser scan while hardly visible in a visual image. Therefore, it is critical to check the validity of this hypothesis and to filter the data accordingly, prior to registration or fusion. Besides, in such cases where multi-modal sensors react differently in the presence of challenging environmental condition, the perception solution obtained with data from the most affected sensor can be significantly improved (see Sec. IV).

The purpose of this study is to evaluate that hypothesis by checking the redundancies in the observations made by the laser and the camera, and, for each laser point that can be projected onto the image, to estimate a likelihood that the sensing data provided by the laser do correspond to the data in the image. This likelihood will be expressed as a probability of correspondence to account for important uncertainties in the perception and comparison processes.

Sec. II describes the method used to compare laser and camera data to check for discrepancies. Sec. III focusses on the estimation of a probability of correspondence of that data. Sec. IV shows an example of application for this work

T. Peynot and A. Kassir are with the ARC Centre of Excellence for Autonomous Systems, Australian Centre for Field Robotics, The University of Sydney, NSW 2006, Australia tpeynot@acfr.usyd.edu.au

[1]However, note that the contrary is often not true. Imagine for example a planar chessboard, composed of multiple black and white squares.

and the benefit of checking for discrepancies between the two different sources of data. Finally Sec. V proposes a discussion and directions for future work.

II. A METHOD TO COMPARE LASER AND CAMERA DATA

This section describes the method used to compare laser and camera data. A critical component of this process is the ability to project laser points onto the image of the camera, whenever this projection exists. This requires a prior off-line calibration process, which is discussed in Section II-A. The different steps of the on-line comparison of laser and camera data are then described in the following Section II-B.

Note that in this paper a 2D laser range finder is used, scanning roughly horizontally with respect to the camera image. However, the method can easily be adapted to the case of a LRF set with a different orientation.

A. Uncertain Projection of Laser Points on the Image

First, a calibration process is performed to estimate the parameters of the transformation between the laser frame and the camera frame. This means calibrating the camera using, for example, the well-known camera calibration toolbox for Matlab [7] developed by Bouget et al. Then, we estimate this laser-camera transformation using a technique adapted from [8]. Once this is completed, laser points can be projected onto this image. After extending that technique, we also obtain an evaluation of the uncertainty in the transformation laser-camera. The following details how these uncertainties are estimated, based on the estimation of the projection of a laser point.

Consider a point acquired by the 2D laser, defined by a range and a bearing angle, in a frame associated to the laser. This point can also be expressed as a vector of 3D cartesian coordinates in the laser frame of reference: $\mathbf{P_l}$, and a vector of 3D cartesian coordinates in the camera frame: $\mathbf{P_c}$. The two vectors can be related by the following equation:

$$\mathbf{P_l} = \mathbf{\Phi}(\mathbf{P_c} - \mathbf{\Delta}) \tag{1}$$

where $\mathbf{\Delta} = [\delta_x, \delta_y, \delta_z]^T$ is the translation offset and $\mathbf{\Phi}$ is a rotation matrix defined by the set of three Euler angles $\mathbf{R} = [\phi_x, \phi_y, \phi_z]$:

$$\mathbf{\Phi} = \begin{bmatrix} c_y c_z & c_x s_z + s_x c_z s_y & s_x s_z - c_x s_y c_z \\ -c_y s_z & c_x c_z - s_x s_y s_z & s_x c_z + c_x s_y s_z \\ s_y & -s_x c_y & c_x c_y \end{bmatrix} \tag{2}$$

where s_i and c_i stand for $\sin(\phi_i)$ and $\cos(\phi_i)$ respectively. The extrinsic laser-camera calibration method outlined in [8] can be used to find estimates for $\mathbf{\Delta}$ and $\mathbf{R}$. For convenience, we combine the parameters of those two vectors into one vector of parameters: $\boldsymbol{\rho} = [\delta_x, \delta_y, \delta_z, \phi_x, \phi_y, \phi_z]$. The calibration process provides the best estimate of $\boldsymbol{\rho}$ based on extracted data from image-laser scan pairs. Then, we determine the variances of the elements of $\boldsymbol{\rho}$ using the Jackknife resampling method originally suggested by Quenouille [9]. The method is performed by taking Jackknife samples from the entire dataset. Let $\mathbf{O_i}$ be the observed data from the i^{th} image-scan pair, and n the total number of image-scan pairs. The i^{th} Jackknife sample $\mathbf{X_i}$ is obtained by omitting the i^{th} pair $\mathbf{O_i}$ from the dataset, i.e. $\mathbf{X_i} = [\mathbf{O_1}, \ldots, \mathbf{O_{i-1}}, \mathbf{O_{i+1}}, \ldots, \mathbf{O_n}]$, leading to a total of n different samples. For each of these samples $\mathbf{X_i}$, a different parameter vector estimate $\boldsymbol{\rho_i}$ is obtained. The parameter vector variance is then given by:

$$\boldsymbol{\sigma_\rho}^2 = \frac{n-1}{n} \sum_{i=1}^{n} (\boldsymbol{\rho}_i - \hat{\boldsymbol{\rho}})^2 \tag{3}$$

where $\hat{\boldsymbol{\rho}} = \sum_i \rho_i / n$.

Let $\mathbf{P_{li}} = [u, v]^T$ be the projection of the laser point $\mathbf{P_l} = [x, y, z]^T$ onto the image plane, where u and v are coordinates in the image frame while x, y and z are coordinates in the laser sensor frame. The projection is defined by the laser-camera transformation parameters ($\mathbf{\Phi}$, $\mathbf{\Delta}$), the intrinsic matrix of the camera $\mathbf{K}$, the set of nonlinear distortion parameters $\mathbf{d}$, and the function f, such that $\mathbf{P_{li}} = f(\mathbf{P_l}, \Phi, \Delta, K, \mathbf{d})$. The uncertainty in $\mathbf{P_{li}}$ is represented by a 2×2 covariance matrix $\mathbf{\Sigma_{li}}$. Assuming the error in $\mathbf{P_l}$ is negligible with respect to the calibration error, i.e. the laser range finder is relatively accurate, the uncertainty which propagates to $\mathbf{P_{li}}$ is mainly the uncertainty in the calibration parameters. This uncertainty is given by a 16×16 covariance matrix $\mathbf{\Sigma_p}$ representing all 16 parameters – 3 rotation parameters, 3 translation parameters, 5 camera intrinsic parameters and 5 distortion parameters. This covariance matrix can be approximated as the combination of two covariance matrices $\mathbf{\Sigma_{pt}}$ and $\mathbf{\Sigma_{pc}}$ which represent the uncertainty in the laser-camera transformation parameters ($\mathbf{\Phi}$, $\mathbf{\Delta}$) and the uncertainty in the intrinsic camera parameters ($\mathbf{K}$, $\mathbf{d}$) respectively. The resulting matrix becomes:

$$\mathbf{\Sigma_p} = \begin{bmatrix} \mathbf{\Sigma_{pt}} & 0 \\ 0 & \mathbf{\Sigma_{pc}} \end{bmatrix}. \tag{4}$$

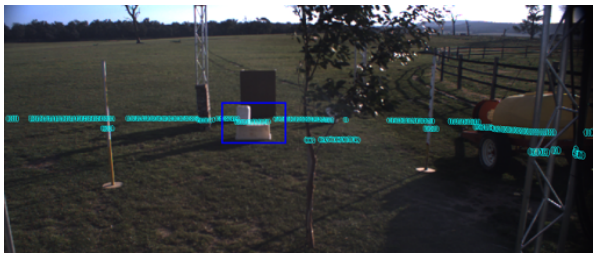
Using the Jackknife method, $\mathbf{\Sigma_{pt}}$ is approximated by a diagonal matrix with the individual elements of $\boldsymbol{\sigma_\rho}^2$ placed along the diagonal. $\mathbf{\Sigma_{pc}}$ can be calculated using the method outlined in [10, §5.2]. Finally, $\mathbf{\Sigma_{li}}$ can be expressed using the following equation [11]:

$$\mathbf{\Sigma_{li}} = \mathbf{J_f} \mathbf{\Sigma_p} \mathbf{J_f}^T \tag{5}$$

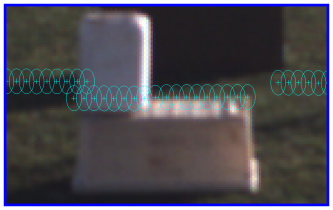
where $\mathbf{J_f}$ is the Jacobian matrix of f with respect to the 16 parameters listed above.

Therefore, given a laser point $\mathbf{P_l}$, we can compute its estimated projection $\mathbf{P_{li}}$ onto the image, with the square root of the diagonal terms of $\mathbf{\Sigma_{li}}$ giving the standard deviations σ_u and σ_v of the coordinates of $\mathbf{P_{li}}$ in pixels along the u and v axes of the image. We can then define a neighbourhood of uncertainty on the projection as an ellipse centered on $\mathbf{P_{li}}$ whose major and minor radii are $r_u = 3\sigma_u$ and $r_v = 3\sigma_v$, thus representing 95% of probability that the projection lies in this ellipse, considering a Gaussian distribution. Fig. 1 shows an example of laser scan projection obtained. However, for convenience, a rectangular neighbourhood $M_{u,v}$ of size $n_u \times n_v$ will be used in practice, where $n_u = (2\lceil r_u \rceil + 1)$,

$n_v = (2\lceil r_v \rceil + 1)$, and $\lceil r_u \rceil$ (resp. $\lceil r_v \rceil$) is the nearest higher integer to r_u (resp. r_v).



(a)



(b) Zoom on the blue area in (a)

Fig. 1. Projection of laser points (cyan +) on camera image, with associated uncertainty represented as cyan ellipses.

B. Comparing Laser Projected Points with Image Content

Once the parameters of the projection of any laser point on the camera image are estimated, the on-line process of comparing the data can be realised. This process has been designed to be simple enough, in terms of computation, to be applicable on-line without limiting the rate of acquisition of the data. The following describes the sequence of steps involved.

1) Data acquisition: A synchronised pair of image and laser scan is acquired. In practice, both sources of data are continuously acquired, and accurately time-stamped. Since the frame rate of the camera is lower than the scanning rate of the LRF, for each image the closest laser scan (in time) is used for the comparison. Only a gray-scale image is needed, therefore, a conversion of the image may be required if a colour camera is used.

2) Edge detection: Edges in the image are detected using a Sobel filter[2]. Without loss of generality, the configuration of the laser w.r.t. the camera is considered such that the laser is scanning horizontally. Consequently, the Sobel filter is set to detect vertical edges in the image. The filtered image can be obtained by convolution of the original gray-scale image with the following mask:

$$\begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}. \tag{6}$$

Note that the image can be filtered to remove noise, prior to the edge detection, if needed. If the laser device is set with a significantly different configuration w.r.t. the camera, edges in the corresponding perpendicular direction will have to be detected instead.

[2]Another edge detection filter, such as Canny, could be used instead.

3) Range filtering: A smoothing filter can be applied to the range values of the laser scan if needed, as long as it does not affect the discontinuities, see [12, §24.2].

4) Range Gradient: A gradient of range is computed on the laser scan data. E.g., it can be obtained by convolution of the laser scan data with the following mask [12, §24.2]:

$$\begin{bmatrix} -1/2 & 0 & 1/2 \end{bmatrix} \tag{7}$$

5) Corner extraction: Points corresponding to *corners* in the laser scan are identified. In this paper, the term *corner* refers to a point where the norm of the range gradient is superior to a pre-defined threshold: *gradThreshold*, experimentally set to be clearly above the noise floor. Note that in many cases, two successive points of the laser scan correspond to corners, i.e. one laser point on either side of the laser scan discontinuity, forming a pair of successive corners.

6) Scan segmentation: The laser scan is segmented in regions of consecutive points that have similar values of range, separated by corners. Note that most segments are composed of all the points that are between two successive laser corners. However, some segments may contain only one corner.

7) Candidate corner selection: The laser corner points that should be kept for the matching process are selected. Hereafter, they will be referred to as *candidate* corners.

Since the laser is scanning in a plane which is approximately perpendicular to the image plane, among the laser corners found above, only the ones that are "in front" (i.e. at a shorter range) are likely to have a corresponding edge in the image. Imagine a vertical pole in the middle of the scene perceived by both sensors. At both sides of the pole, there will be sudden variations of range in the laser scan. The projections of laser points taken just before and after the range discontinuity can have a significantly different locations on the image, due to the difference of perspective of the two sensors[3] (e.g. see the right edge of the bottom plastic box in Fig. 1(b)). This indicates that the laser corners considered for matching with the visual data should be the ones actually belonging to the object, i.e. the points "in front" (shorter range) rather than the ones further in the back.

In practice, the *candidate* corner points are selected using the following process. Consider a laser corner point c, belonging to the laser scan segment labeled s_i. c is considered as a candidate corner if the range at c is smaller than the range of the previous segment[4] (s_{i-1}) *or* the next one (s_{i+1}). More specifically, c is considered "in front" of segment s_{i-1} (resp. s_{i+1}) if the range at the last (resp. first) corner belonging to s_{i-1} (resp. s_{i+1}) is higher than the range at c.

8) Projection: The laser scan points are projected on the image and the associated neighbourhoods of uncertainty are computed (see Sec. II-A).

9) Matching test: For each selected laser point projected on the image that corresponds to a candidate corner, we look for a matching edge in the image within the neighbourhood of this projected point.

[3]The sensors are physically mounted at different locations on the vehicle.

[4]It is assumed that the laser points are sorted in increasing order of angle.

Consider a specific candidate corner. The best estimate of the projection of this point on the Sobel-filtered image is a pixel position (u, v) in this image, with an associated neighbourhood $M_{u,v}$, obtained as described before. This "matching" process looks for a visual edge in this neighbourhood, i.e. points of that neighbourhood with an intensity superior to a pre-defined threshold *imedgeThreshold* in the Sobel-filtered image. In practice, in this embodiment we are interested in vertical edges, so we consider there is an edge in the neighbourhood if there are at least m pixels with an intensity higher than *imedgeThreshold* that belong to different consecutive lines and are *8-adjacent* [13, §2.5]. Choosing $m > 1$ allows to exclude a number of edges due to noise or artefacts and focus more on edges due to the presence of objects. We used $m = 2$ to generate the results presented in Sec. IV.

10) Correspondence likelihood: Knowing the result of the matching process, for each segment of the laser scan that has a projection on the image and contains candidate corners, i.e. for each *candidate segment*, a probability of correspondence of the laser data with the camera data can be evaluated. This is detailed in Sec. III below.

III. LIKELIHOOD OF CORRESPONDENCE

For each laser point of candidate segments, the objective is to estimate the probability that the laser information corresponds to the information in the image acquired at the same time by the camera. This probability mainly depends on:

- The laser and camera measurement uncertainties. They can be represented by a sensor model and estimated thanks to a process of calibration of the intrinsic parameters. For the camera, they were estimated in Sec. II-A. However, for the laser, in our application they are typically negligible compared to the uncertainties below.
- The uncertainty in the laser-camera transformation, i.e. in the projection of the laser points onto the image, represented in this work by a pixel neighbourhood associated to each projected point (see Sec. II-A).
- For candidate laser corners, the likelihood that they correspond to an edge in the image.
- For other laser points of candidate segments, the likelihood to be in the "same category" (i.e. consistent or inconsistent with the camera data) as the candidate corners of this segment.

A. Probability Evaluation

This section details the calculation of the probability of correspondence for a laser point corresponding to a candidate corner. Let us consider the estimated projection (u, v) of such laser point in the image and the corresponding neighbourhood $M_{u,v}$, as defined in Sec. II-A. We define the following events:

- A: The laser and camera information correspond. A is a boolean.
- B: A visual edge is found in the neighbourhood of projection. B is a boolean.
- C: The projection of the laser point is as computed using the transformation between the laser frame and the camera frame initially estimated by the calibration process (Sec. II-A).

For the current laser corner, we can express the probability that the local laser and camera information correspond (A), knowing the location of the projection of the laser data on the image (C), and knowing whether a visual edge was found in $M_{u,v}$ (B). In particular, if such a visual edge was found (i.e. $B = 1$), this can represent the probability that this visual edge actually matches with the laser corner associated to the current neighbourhood: $P(A = 1|B = 1, C)$. Using the Bayes theorem, this probability can be expressed as:

$$P(A|B, C) = \frac{P(C|A, B)P(A|B)}{P(C|B)}$$

Developing $P(A|B)$ using the Bayes formula again gives:

$$P(A|B, C) = \frac{P(C|A, B)}{P(C|B)} \frac{P(B|A)P(A)}{P(B)}$$

where:

- $P(B|A)$ is the probability that a visual edge is found in the neighbourhood, knowing the laser and camera information do correspond. It describes the likelihood of the assumption that if the laser and camera information do correspond, any laser candidate corner should correspond to a visual edge (see Sec. I). This assumption is thought to be generally true, with a few exceptions in quite extreme cases, which means this probability is close to 1, i.e. knowing the data do correspond ($A = 1$), a visual edge should exist in the neighbourhood of the corner. Therefore, a visual edge should be found in the neighbourhood unless it does not encapsulate the actual position of the point.
- $P(A)$ is the *a priori* probability that the laser data and camera data correspond. This will be set to a fixed uncertain value at this point, as we do not have a priori knowledge on this event. It might be possible to consider estimating that value based on statistics over data taken in a "typical" environment, but this would only be valid for this type of environment and some ground truth about the correspondence of the data would be needed.
- $P(B) = P(B|A)P(A) + P(B|\bar{A})P(\bar{A})$, with $P(\bar{A}) = 1-P(A)$. $P(B|\bar{A})$ is the *a priori* probability of finding a visual edge in a neighbourhood $M_{u,v}$. This can be considered as the probability to find a visual edge anywhere in the image (i.e. choosing a location randomly), using the very same process described before. Let $nbPos$ be the minimum number of pixel positions in the image needed to have the neighbourhood (or mask) M of size $n_u \times n_v$ cover the whole image, and $nbEdges \leq nbPos$ the number of positions where a visual edge was found within M. We can then express the *a priori* probability $P(B|\bar{A})$ as:

$$P(B|\bar{A}) = nbEdges/nbPos$$

Intuitively, this term tells us how *informative* it is to find a visual edge in such a neighbourhood. Indeed, if there is a lot of structure (i.e. many edges) distributed over the image this will mean B is not very informative, while if there are only a few edges that can be found overall the image ($P(B|\bar{A})$ low), this means that when we have a laser corner and $B = 1$, the likelihood that the visual edge found matches the corner is much higher.

- $P(C|A, B)$ is the likelihood of the laser point projection on the image, knowing the laser and camera data correspond, and whether a visual edge was found in the neighbourhood of the laser point projection. This is related to the uncertainty in the projection of the laser point on the image, i.e. to the neighbourhood (or mask) $M_{u,v}$, interpreted as an array of size $n_u \times n_v$ containing the values of probability of this projection.
  - If a visual edge was found ($B = 1$), and the location in the neighbourhood of the closest pixel belonging to this visual edge is (i, j), then $P(C|A, B = 1)$ can be seen as the value of the Gaussian in the mask at the same location:

$$P(C|A = 1, B = 1) = M_{u,v}(i, j)$$

  - Knowing the data correspond ($A = 1$), under the assumption mentioned above, there should be a corresponding visual edge. Thus, if no visual edge was found ($B = 0$), this means the location where to look for this edge, i.e. the actual projection of the laser point, is not in the estimated neighbourhood. Therefore, $P(C|A = 1, B = 0)$ can be interpreted as the probability that the projection of the laser point is *outside* of this neighbourhood, i.e.:

$$P(C|A = 1, B = 0) = 1 - \sum_{i=1}^{n_b} \sum_{j=1}^{m_b} M_{u,v}(i, j)$$

- $P(C|B) = P(C|B, A)P(A) + P(C|B, \bar{A})P(\bar{A})$, where $P(C|B, \bar{A})$ corresponds to the likelihood of the projection of the laser point onto the image, knowing that the laser and camera data do not correspond ($A = 0$) and that an edge was found or not (B). If the data do not correspond, then B does not provide more information for C, thus we can say that $P(C|B, \bar{A}) = P(C|\bar{A})$. This term corresponds to the confidence in the calibration, taken as the best chance for the projection, i.e. the probability at the centre of the neighbourhood $M_{u,v}$.

B. Decision

Once a likelihood of correspondence has been computed for each candidate corner, a decision on the validity of the segments containing such corners, i.e. candidate segments, should be taken before integration and possible fusion of the sensing data in the perception system:

- If the segment is bounded by two pairs of matching edges/corners, the points of this segment are likely to correspond to the information in the image, therefore this segment is *validated*. This means the laser data can be combined (or associated) with the image data.
- If the segment is bounded by two non-matching edges/corners, the laser segment is unlikely to correspond to the information in the image, so it is *rejected*. This means that in this area the laser and the camera are probably not seeing the same elements in the environment. For example, the laser may be detecting the dust cloud while the camera images are not particularly affected by it (see Sec. IV). Therefore, their data are considered *inconsistent* and should not be combined or fused.
- If the segment is bounded by a matching edge/corner and a non-matching edge/corner, or contains only one candidate corner, an *unknown* status is attributed to this segment, as no decision can be taken at this point.

Fig. 2. The Argo UGV sensing the *static* trial area

IV. EXPERIMENTAL RESULTS

This section aims at showing how this method of comparing laser-camera data can be used to improve the quality and integrity of perception for an unmanned ground vehicle in challenging environmental conditions. In particular, the case of the presence of airborne dust clouds, which affect the sensing of the laser and the camera in different ways, is considered.

A. Experiment Setup

The datasets introduced in [5] and presented in more details in [14] are used. They were acquired by a UGV equipped with multiple sensors (Figs. 2 & 3), including a colour camera, an infrared camera, and 2D LRFs (SICK LMS 291). The visual camera is a Prosilica Mono-CCD gigabit ethernet camera, acquiring 3×8 bit images of resolution 1360×1024 at 10 frames per second. In this section, the LRF used is the one labeled as *LaserHorizontal* in Fig. 3. In some of the *static*[5] datasets, a scene (see Fig. 2) is observed by all sensors onboard the motionless UGV, including lasers and cameras, first in clear conditions, and then in the presence of various environmental conditions such as airborne dust, smoke or rain. The laser scans obtained in clear conditions can be used as a *reference scan*, to make comparisons with the scans obtained in other environmental conditions. In those datasets, all objects are static except for the small tree

[5]using the notation in [14]

in the middle of the image (see Fig. 5), which is slightly but randomly changing shape over time due to wind. Thus, to be able to use the reference scan as a 'ground truth', the laser points corresponding to the tree were systematically removed from all scans used to evaluate the results in this section.

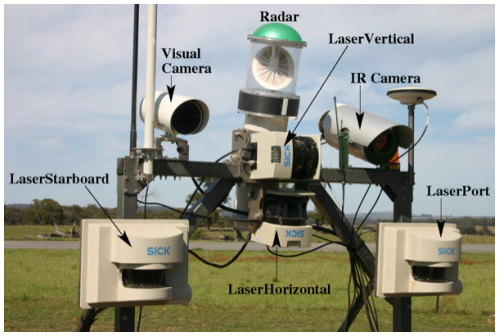


Fig. 3. Sensor Frame on the Argo. The data used for the experimental results presented in this paper are the *Visual Camera* (top left) and the *Laser Horizontal* (bottom).

Note that with the calibration made for this experimental test, the ceiling of the maximal standard deviation for the projection on the image was 3 pixels, thus the largest size of neighbourhood used was 19×19 pixels. The threshold on the gradient (*gradThreshold*) was 0.5 and the threshold on the edge intensity in the Sobel image (*imedgeThreshold*) was 40.

B. Laser scan filtering in challenging conditions

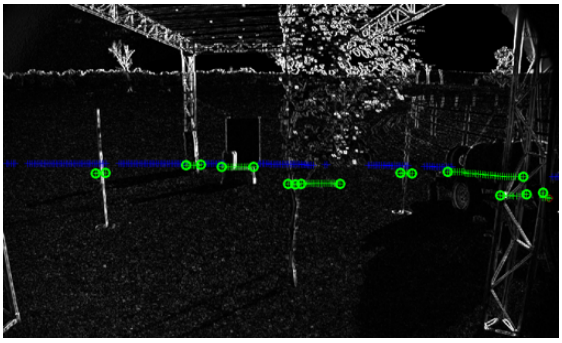
Fig. 4 shows the laser scan points projected on the image in *clear*[6] environmental conditions. The *validated* segments are shown in green, while the *rejected* segments are shown in red. For the other segments, shown in blue, no decision could be taken, as they are not candidate segments. Note that the laser scan segments corresponding to both laser poles (on the left and on the right of the image), the brick pile and the boxes were *validated*.

Fig. 5 shows the same type of illustration but in the presence of some light airborne dust. It can be seen that most laser points hitting the small dust cloud on the left of the image are considered as *inconsistent* with the image data. This is because the laser hits the dust cloud, whereas the same scene can still quite clearly be identified in the image; more specifically, the structural information in the Sobel-filtered image hardly changed. In particular, the laser scan does not hit the vertical pole on the left of the image, while the corresponding laser scan segment had been validated in the previous illustration, when no airborne dust was present (Fig. 4).

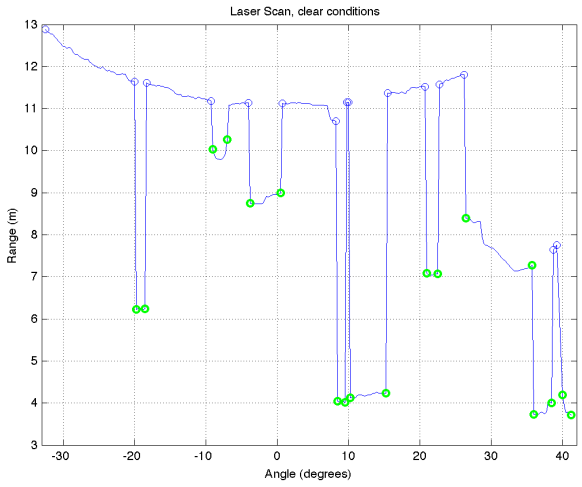
C. Error Analysis

This section proposes an analysis of the results obtained after the comparison of laser and camera data, and the validation or rejection of some segments of the laser scan. As it was mentioned before, the process only concerns the

[6]in the absence of airborne dust, smoke or rain



(a) Sobel image with projected laser points.



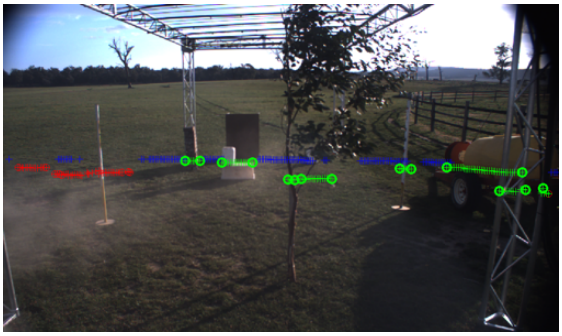
(b) Corresponding laser scan. The circles show all *corner* points, with the *candidate* corners in green.

Fig. 4. Example of laser points projected on Sobel-filtered image of the scene (a) and the corresponding laser scan (b) in *clear* conditions. In the image: green circles illustrate *matching* edges/corners while red circles show corners found in the laser scan for which no matching edge was found in the neighbourhood of projection on the image Green/red crosses correspond to laser points that belong to a *validated/inconsistent* scan segment, respectively. Dark blue crosses correspond to points in segments that do not contain candidate corners (i.e. no evaluation of these points could be made).

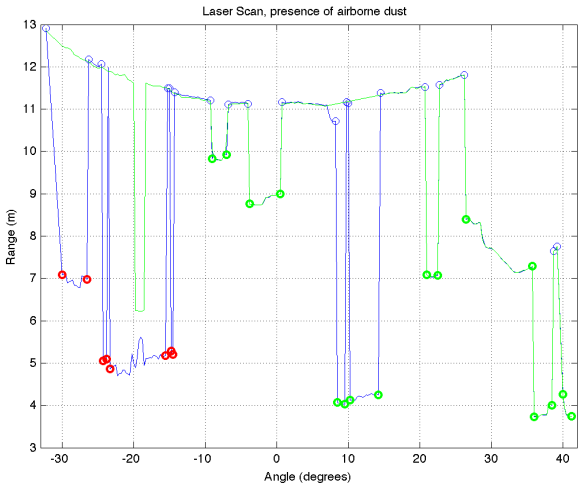
points within the segments of the laser scan that contain candidate corners (i.e. candidate points). Therefore, the statistics generated hereafter are calculated using only these candidate points, as they are the only points for which a decision can be taken, i.e. the only points that can be affected by this process.

In this section, *priorError* is defined as the normalised sum of errors for all candidate laser points, i.e. the sum of all those errors divided by the number of terms in the sum. *postError* is defined as the normalised sum of errors for all validated laser points, i.e. the sum of errors when keeping only the validated points, divided by the number of validated points. Indeed, both these errors need to be normalised so that they can be compared.

Fig. 6 shows the evolution of *priorError* and *postError* for a portion of a dataset with presence of airborne dust. It can be seen that *priorError* becomes significant as soon as airborne dust appears, even in the form of a small light cloud. In these conditions, *postError* hardly increases, indicating that



(a) Image with projected laser points.



(b) Corresponding laser scan (blue line) and the *reference* scan (green line). The red circles are *rejected* corners, while the green ones were *validated*.

Fig. 5. Example of laser points projected on the original image of the scene in the presence of light dust (a). Both corresponding current scan (blue) and reference scan (green) are shown in (b). The laser scans clearly show that the laser is significantly affected by the dust cloud on the left hand side, which results in the rejection of a number of points (red crosses in (a)), as this dust is hardly visible in the image.

the validation of laser points is particularly efficient. Indeed, while some parts of the laser scan are quite affected by the presence of dust, most of the time this phenonemon is hardly visible in the image, creating clear discrepancies between the data sources. As a result of this validation/rejection process, the filtered laser scan is much more appropriate for fusion with the colour information of the image, and also for fusion with other sensors that are not affected by dust.

Fig. 7 shows the validation rate (*rateValidated*) and the rejection rate (*rateRejected*) obtained for the same dataset as before, with the presence of airborne dust. The validation rate is the number of points validated by our system that do correspond to points with negligible error divided by the number of validated points (i.e. the *true positive* rate). The rejection rate is the number of rejected points that do have a non-negligible error divided by the number of rejected points (i.e. the *true negative* rate).

This evaluation was also made on a larger dataset of 900 images, i.e. 90s at 10 frames per second, with the presence of dust in various amounts and density (from localised and light to spread over most of the image and

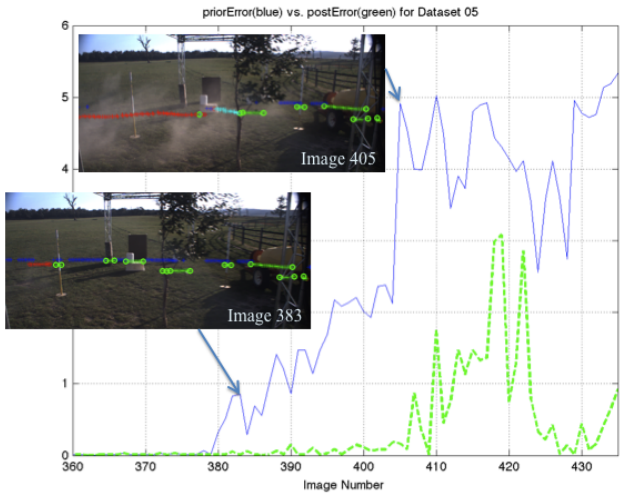


Fig. 6. Error of raw data (in blue) vs. after validation of the laser scan segments (in dashed green). The error is the averaged sum of the differences with the reference scan. Initially, the environment is clear, until airborne dust appears at image 377. At first, the variable dust clouds remain small and light: e.g., at Image 383 a small light dust cloud can hardly be seen on the left hand side of the image, although the laser scan is already affected and points are successfuly rejected. From Image 405 dust become denser and larger.

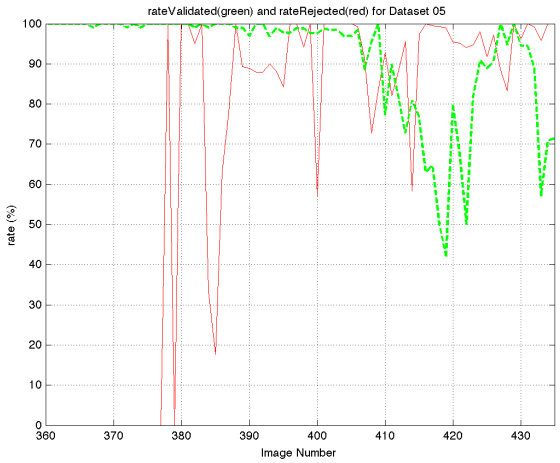


Fig. 7. Rate of validation (%, in dashed green) and of rejection (%, in red) of laser points for the same dataset as in Fig. 6. Note that up to image 377 *rateRejected* is equal to zero, as no points are rejected. The peaks where the rate of rejection is dropping correspond to situation where very few points are actually validated and/or rejected.

dense) during about 78s. Over this sequence of 780 images, the average reduction of error obtained, i.e. the average of $(priorError - postError)/priorError$, was 65%, with a validation rate $rateValidated = 84\%$ and a rejection rate $rateRejected = 73\%$.

V. CONCLUSION

In this paper, a method to check the consistency between 2D laser and camera data has been proposed. It provides an estimated likelihood of correspondence at some points of the laser scan (the candidate corners) and a mechanism of validation and rejection of laser scan segments. It was shown that rejecting segments of the laser scan that show discrepancies with the camera data while validating other

segments with data believed to be consistent can significantly mitigate the effects in the scan that typically cause interpretation errors in regular perceptions systems. Furthermore, using this data filter can also contribute in mitigating failures in perception systems by enforcing the consistency of data from heterogeneous sensors before combination or fusion.

As an example, in Sec. IV this technique was used to mitigate the effects of challenging environmental conditions such as airborne dust on a laser scan. The data comparison performed is particularly interesting when the camera image is not affected by dust as much as the laser scan. In this case, if a trusted laser-camera calibration is available, this technique can also be used to filter the laser data which will be used for obstacle detection or mapping.

Although the current study was focussed on the case of a laser and camera sensors mounted on a UGV, any perception system including this pair of sensors, with sufficient overlap between their fields of view, could benefit from this technique.

A. Discussion

Due to the poor actual intrinsic correlation between the laser data and the camera data, an actual likelihood of correspondence could only be evaluated for the candidate corners. Once a decision has been taken for those specific points, a simple mechanism was proposed to extend this decision to the other points of candidate segments, assuming there is high correlation between points in a segment. This method proved to be quite efficient, as illustrated in the presence of airborne dust in Sec. IV. However, the extent of the validity of this propagation remains to be determined.

With the technique proposed in this paper, it was not possible to evaluate the segments of the laser scan that do not contain candidate corners (typically the points from the background of the scene). Therefore, two main 'strategic' choices are possible. The perception system can be very conservative by rejecting all those points (at the risk of a significant loss of data), or accept all those segments for interpretation and fusion. Future work will study the possibility to extend that estimation of a likelihood of correspondence to all laser points that do not correspond to candidate corners. In the absence of clear common characteristics of the information on those points with the visual information, this will probably require a mechanism of inference.

In addition to detecting inconsistencies between laser and camera data, being able to identify their causes could be extremely valuable. It is believed that the main reasons for the presence of such inconsistencies are:

- The presence of an element in the scene that provokes a different effect on the laser and the camera data (e.g. airborne dust or smoke). This element can cause local inconsistencies or a global one over the whole laser scan (or image) if the element is well spread. These inconsistencies are typically variable over time.
- A significant misalignment, i.e. a situation where the available estimated transformation between the laser and the camera is not correct anymore. This will typically affect the whole set of data consistently over time, which might allow to distinguish this cause from the previous one mentioned.
- The difference of perspective of the two sensors, which also cause inconsistencies that are systematic in time but whose amplitude depend on the location of the perceived points. Therefore, a more local analysis may contribute to distinguishing this cause from the previous one.

B. Future Work

Future work will aim at determining if the evaluation of the likelihood of correspondence is possible on the points of the laser scan that are not candidate corners. This technique will also be demonstrated using an infrared camera. One of the main applications in that case is the detection of segments of laser scans affected by the presence of smoke, as it usually has no visible effect on infrared images.

ACKNOWLEDGMENTS

This work was supported in part by the Centre for Intelligent Mobile Systems (CIMS), funded by BAE Systems as part of an ongoing partnership with the University of Sydney, and the ARC Centre of Excellence programme, funded by the Australian Research Council (ARC) and the New South Wales State Government.

REFERENCES

[1] M. Montemerlo et al., "Junior: The stanford entry in the urban challenge," *Journal of Field Robotics*, vol. 25, no. 9, pp. 569–597, 2008.

[2] C. Urmson et al., "Autonomous driving in urban environments: Boss and the urban challenge," *Journal of Field Robotics*, vol. 25, no. 8, pp. 425–466, 2008.

[3] A. Kelly et al., "Toward reliable off road autonomous vehicles operating in challenging environments," *International Journal of Robotics Research*, vol. 25, pp. 449–483, May/June 2006.

[4] J. Leonard et al., "A perception driven autonomous urban vehicle," *Journal of Field Robotics*, vol. 25, no. 10, pp. 727–774, Sept. 2008.

[5] T. Peynot, J. Underwood, and S. Scheding, "Towards reliable perception for unmanned ground vehicles in challenging conditions," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2009.

[6] C. Hantak and A. Lastra, "Metrics and optimization techniques for registration of color to laser range scans," in *Proceedings of the Third Internation Symposium on 3D Data Processing, Visualization and Transmission*, 2006.

[7] J.-Y. Bouguet, *Camera Calibration Toolbox for Matlab*, http://www.vision.caltech.edu/bouguetj/calib_doc/.

[8] Z. Qilong and R. Pless, "Extrinsic calibration of a camera and laser range finder (improves camera calibration)," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2004.

[9] M. Quenouille, "Approximate tests of correlation in time-series," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 11, no. 1, pp. 68–84, 1949.

[10] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003.

[11] J. Underwood, "Reliable and safe autonomy for ground vehicles in unstructured environments," Ph.D. dissertation, The University of Sydney, 2009.

[12] D. A. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*. Prentice Hall, 2002.

[13] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Pearson Prentice Hall, 2008.

[14] T. Peynot, S. Terho, and S. Scheding, "Sensor data integrity: Multisensor perception for unmanned ground vehicles," Australian Centre for Field Robotics (ACFR), The University of Sydney, Tech. Rep. ACFR-TR-2009-002, 2009.