

Loughborough University Institutional Repository

Fusion of LiDAR and camera sensor data for environment sensing in driverless vehicles

This item was submitted to Loughborough University's Institutional Repository by the/an author.

Citation: DE SILVA, V., ROCHE, J. and KONDOZ, A., 2018. Fusion of LiDAR and camera sensor data for environment sensing in driverless vehicles. arXiv:1710.06230v2

Additional Information:

- This is an arXiv preprint. It can be found at:
<https://arxiv.org/abs/1710.06230v2>

Metadata Record: <https://dspace.lboro.ac.uk/2134/33170>

Version: Submitted for publication

Publisher: arXiv.org

Rights: This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Full details of this licence are available at: <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Please cite the published version.

Fusion of LiDAR and Camera Sensor Data for Environment Sensing in Driverless Vehicles

Varuna De Silva, Jamie Roche, and Ahmet Konoz, *Senior Member, IEEE*

Abstract—Driverless vehicles operate by sensing and perceiving its surrounding environment to make the accurate driving decisions. A combination of several different sensors such as LiDAR, radar, ultrasound sensors and cameras are utilized to sense the surrounding environment of driverless vehicles. The heterogeneous sensors simultaneously capture various physical attributes of the environment. Such multimodality and redundancy of sensing need to be positively utilized for reliable and consistent perception of the environment through sensor data fusion. However, these multimodal sensor data streams are different from each other in many ways, such as temporal and spatial resolution, data format, and geometric alignment. For the subsequent perception algorithms to utilize the diversity offered by multimodal sensing, the data streams need to be spatially, geometrically and temporally aligned with each other. In this paper, we address the problem of fusing the outputs of a Light Detection and Ranging (LiDAR) scanner and a wide-angle monocular image sensor. The outputs of LiDAR scanner and the image sensor are of different spatial resolutions and need to be aligned with each other. A geometrical model is used to spatially align the two sensor outputs, followed by a Gaussian Process (GP) regression based resolution matching algorithm to interpolate the missing data with quantifiable uncertainty. The results indicate that the proposed sensor data fusion framework significantly aids the subsequent perception steps, as illustrated by the performance improvement of a typical free space detection algorithm.

Index Terms— Sensor data fusion, LiDAR, Gaussian Process Regression, Free space detection, autonomous vehicles, Driverless cars.

I. INTRODUCTION

From being a mainstay in science fiction, driverless vehicles, also referred to as autonomous vehicles, are steadily progressing towards becoming a commercial reality. Driverless vehicles will allow passengers or goods to be transported without the need for human intervention. The advent of driverless vehicles promise to bring with it many benefits - improved safety, better fuel efficiency, less pollution and congestion [1]. Self-driving taxis, self-parking cars, vehicle platoons and indoor personal assistant vehicles for the

disabled are few applications of autonomous vehicles on the horizon. Several major automobile manufacturers have set targets to launch commercially available fully autonomous vehicles by 2020. However, vehicles that are adequately capable to roam without the need for human intervention, is still a distant reality that requires extensive research effort for realization [2]. It is expected that the fully autonomous vehicles of the near future will be limited to a set of highly controlled settings and low speed environments [3]. The terms “Driverless vehicles” and “Autonomous vehicles” will be used interchangeably throughout this paper.

An autonomous vehicle is composed of three major technological components. The first system - sensing and perception - is responsible for sensing and understanding the surrounding environment. Information captured by the sensing and perception system is used to make basic movement decisions about direction of travel, obstacles avoidance, accelerating and deceleration. The second system - localization and mapping - enables the vehicle to know its current location at any given time. While satellite and inertial navigation based systems can be used for this purpose, these systems have their own limitations [4]. Therefore, localization is performed by mapping the surrounding environment in 3D in various settings, and comparing against the historical data. The third component is responsible for the driving policy. The driving policy refers to the decision making capability of the autonomous vehicle under various situations, such as negotiating at roundabouts, yielding to other road users and overtaking vehicles.

Furthermore, it is expected that the autonomous vehicles will be connected to other vehicles and to the surrounding infrastructure through wireless communication systems [5]. Such communication links can be utilised to exchange sensor data, and control navigation.

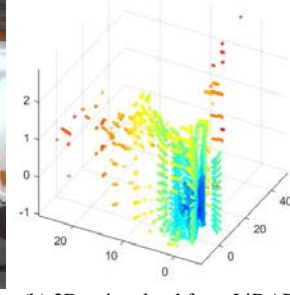
Current prototypes of driverless cars [4] utilise multiple sensors, such as Light Imaging Detection and Ranging (LiDAR), Radars, Imaging and Ultrasound Sensors – to navigate surroundings and make decisions. Radar is used for long-range sensing, while ultrasound sensors are effective at very short ranges. Imaging sensors are often used to detect traffic signals, lane marking and surrounding pedestrians and vehicles. Often, these prototypes rely on LiDAR sensors to map the surrounding environment in 3-Dimensions (3D). Data generated from each sensor need to be interpreted accurately for satisfactory operation of autonomous vehicles.

The precision of operation of an autonomous vehicle is,

This paper was received on September 2017, ...history of the paper comes here. The associate editor responsible for this submission is....

Varuna De Silva, Jamie Roche and Ahmet Konoz are with Loughborough University, Queen Elizabeth Olympic Park, London, United Kingdom. The corresponding author for this publication is Varuna De Silva and he can be reached at V.D.De-Silva@lboro.ac.uk or varunax@gmail.com.

The next few paragraphs should contain the authors' current affiliations, including current address and e-mail. For example, This work is funded by the Engineering and Physical Sciences Research Council (EPSRC) UK.



(a) Image from the wide angle camera
 (b) 3D point cloud from LiDAR
 Fig. 1. Examples of sensor data to be fused.

thus, limited by the reliability of the associated sensors. Each type of sensor has its own limitations, for example, LiDAR sensor readings are often affected by weather phenomena such as rain, fog or snow [6]. The diversity offered by multiple sensors can positively contribute to the perception of the sensed data. The effective alignment (either spatially, geometrically or temporally) of multiple heterogeneous sensor streams, and utilization of the diversity offered by multimodal sensing is referred to as sensor data fusion [7].

In this paper, we present a novel approach for fusing distance data gathered by a LiDAR sensor, with the luminance data from a wide-angle imaging sensor. The data from the LiDAR comes in the form of a 3D point cloud.

The rest of this paper is organised as follows: Section II provides an overview of the related work and associated challenges. The framework for fusion of LiDAR and Imaging sensor data are presented in Section III. Section IV describes the experimental framework and discussion of the results. Finally, we conclude the paper in Section V, with some references to possible future work.

II. SENSOR DATA FUSION FOR DRIVERLESS VEHICLES

A review of the literature that is relevant to the contributions of this paper is presented in this section, followed by the positioning of the current contribution. This section is organized in three sections: the need for data fusion and challenges it poses, relevant work in LiDAR and camera data fusion and finally on challenges addressed by the current work within the scope of driverless vehicles.

A. Challenges in Multimodal Data Fusion

Information about a system can be obtained from different types of instruments, measurement techniques and sensors. Sensing a system using heterogeneous acquisition mechanisms is referred to as multimodal sensing [8]. Multimodal sensing is necessary, because a single modality cannot usually capture complete knowledge of a rich natural phenomena. Data fusion is the process by which multimodal data streams are jointly analysed to capture knowledge of a certain system.

Lahat et al. [8], identifies several challenges that are imposed by multimodal data. These challenges can be broadly categorized in to two segments: challenges at acquisition level and challenges due to uncertainty in the data sources. Challenges due to problems of data acquisition level include: differences in physical units of measurement (non-

commensurability), differences in sampling resolutions, and differences in spatio-temporal alignment. The uncertainty in data sources also pose challenges that include: noise such as calibration errors, quantization errors or precision losses, differences in reliability of data sources, inconsistent data and missing values.

The above challenges discussed by Lahat et al. [8], were identified by considering a multitude of applications. In the next subsection, we will discuss specific challenges associated with fusing LiDAR and imaging data.

B. Fusion of LiDAR and Different types of Imaging data

LiDAR data can be fused with different types of imaging sensor data to cater for a range of applications. Terrain mapping is a popular application of LiDAR data that uses an aerial borne LiDAR scanner to identify various ground objects such as buildings or vehicles. The independent use of LiDAR scanner proves challenging in such applications due to obstructions and occlusions caused by vegetation. Therefore, while LiDAR exhibits good height measurement accuracy, it lacks in horizontal segmentation capability to delineate the building boundaries. A graph based data driven method of fusing LiDAR data and multi-spectral imagery was proposed in [9]. Authors in [9] propose, a connected component analysis and clustering of the components to come up with a more accurate segmentation algorithm.

In a substantial body of literature, LiDAR and image data fusion is considered as an extrinsic calibration process. Here fusion is regarded as the process of rigid body transformation between the two sensors' coordinate systems [10]. For the purpose of extrinsic calibration, an external object, such as a trihedral calibration rig [11], [12], a circle [13], a board pattern [14], [15] or a checkerboard pattern [16] [17] [18], is used as a target to match the correspondences between the two sensors. While such methods, yield accurate alignment, they do not address the issues related to uncertainty of sensor readings.

The problem of LiDAR and imaging data fusion can be approached as a camera pose estimation problem, where the relationship between 3D LIDAR coordinates and 2D image coordinates is characterised by camera parameters such as position, orientation, and focal length. In [19], the authors propose an information-theoretic similarity measure to automatically register 2D-Optical imagery with 3D LiDAR scans by searching for a suitable camera transformation matrix. LiDAR and optical image fusion is used in [19] for creating 3D virtual reality models of urban scenes.

The fusion of 3D-LiDAR data with stereoscopic images is addressed in [20]. The advantage of stereoscopic depth estimation, is its capability to produce dense depth maps of the surroundings by utilising stereo matching techniques. However, the dense stereo depth estimation is computationally quite complex. This is due to the requirement of matching corresponding points in the stereo images. Furthermore, dense depth estimation using stereo images suffer from the limited dynamic range of the image sensors, for instance, due to the saturation of pixel values in bright regions [21].

Another, drawback of stereo based depth estimation is the

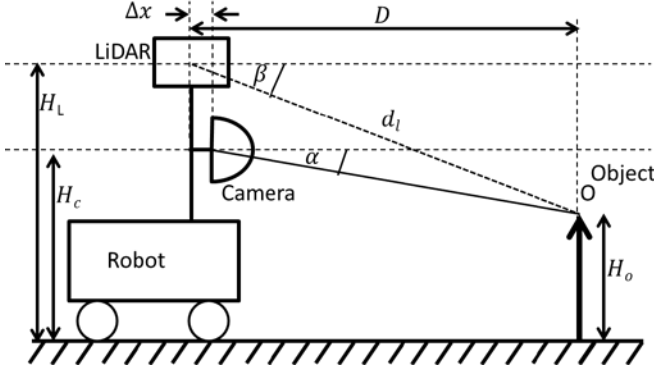


Fig. 2. Side view of the sensor setup.

limited range of depth sensing. LiDAR scanning on the other hand provides a utility to measure depth at high accuracies, albeit at lower point resolutions compared with imaging sensors. Authors in [20] propose a probabilistic framework to fuse sparse LiDAR data with stereoscopic images, which is aimed at real-time 3D perception of environments for mobile robots and autonomous vehicles. An important attribute of probabilistic methods, such as in [20] is that it represents the uncertainty of estimated depth values.

C. Challenges in data fusion addressed in this paper

In this paper, we consider LiDAR and imaging sensor data fusion in the context of autonomous vehicles. Autonomous vehicles as an application poses significant challenges for data driven decision making due to the associated safety requirements. For reliable operation, decisions in autonomous vehicles need to be made by considering all the multimodal sensor data it acquires. Furthermore, the decisions must be made in the light of the uncertainties associated with both data acquisition methods, and the utilized pre-processing algorithms.

This paper addresses two fundamental issues surrounding sensor data fusion. They are: the spatial misalignment and resolution difference in heterogeneous sensors. Apart from being different from previous contributions in the type of sensors used for data fusion, our motivations for this paper are two-fold: Firstly, we are interested in developing a more robust approach for data fusion, which accounts for uncertainty in the fusion algorithm. This will enable the subsequent perception tasks in an autonomous vehicle to operate more reliably. Secondly, we envisage situations in the future, where autonomous vehicles will be exchanging useful sensor data between each other. In such situations it would be impractical for extrinsic calibration methods to work, because there are inevitable per-unit variations that exist between sensors due to manufacturing variations. Therefore, it would be necessary for the data fusion algorithms to work with minimal calibration. Based on the above premises, we propose a robust framework for data fusion with minimal calibration.

III. THE PROPOSED ALGORITHM FOR LiDAR AND WIDE-ANGLE CAMERA FUSION

To address the challenges presented above, in this section we propose a framework for data fusion. This section

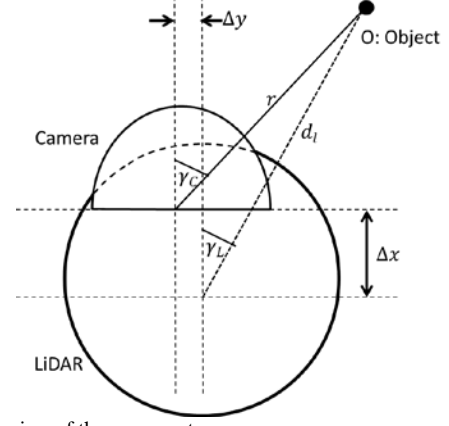


Fig. 3. Top view of the sensor setup.

describes the proposed algorithm for fusion of LiDAR data with a wide angle imaging sensor. The organization of the section is as follows: in section III.A the geometric model for alignment of the two sensor types are presented, followed by Gaussian Process based matching of resolutions of the two sensors, in section III.B.

A. Geometric alignment of LiDAR and Camera data

The first step of the data fusion algorithm is to geometrically align the data points of the LiDAR output and the 360-Degree camera. The purpose of the geometric alignment is to find the corresponding pixel in the camera output for each data point output by the LiDAR sensor. For the purpose of this derivation, consider an object O of height H_O at distance D from the robot. The sensor setup is graphically illustrated in Figure 2 and the horizontal alignment of the sensors are depicted in Figure 3.

The notation used in Figures 2 and 3 are listed as follows:

Δx = Frontal displacement of the centres of LiDAR and Camera sensor. Δy = Horizontal displacement of the centers of LiDAR and Camera sensor. d_L = distance to the object O sensed by LiDAR. β, γ_L = Latitude and longitude of object O as measured by the LiDAR, respectively. H_C = Height of the camera from the ground. H_L = Vertical height of the LiDAR, and α, γ_C = Latitude and longitude of object O as measured by the Camera, respectively.

The values d_L, β , and γ_L are the outputs of the LiDAR sensor. The purpose of this alignment is to find the corresponding pixel in the camera output for each data point output by the LiDAR sensor. Here we assume that the main axis of the camera and the LiDAR are aligned with each other. Considering the distance to object O , we have,

$$D = d_L \cos \beta \cdot \cos \gamma_L = r \cdot \cos \alpha \cdot \cos \gamma_C + \Delta x \quad (1)$$

Considering the vertical height of the object O , we have,

$$H_O = H_L - d_L \cdot \sin \beta = H_C - r \cdot \sin \alpha, \quad (2)$$

From (1) and (2), we can calculate corresponding latitude α of the camera as following,

$$\tan \alpha = \frac{((H_C - H_L) + d_L \cdot \sin \beta) \cdot \cos \gamma_C}{d_L \cdot \cos \beta \cdot \cos \gamma_L - \Delta x} \quad (3)$$

Considering the horizontal displacement from the setup in Figure 3, we have,

$$d_L \cos \beta \cdot \sin \gamma_L = r \cdot \cos \alpha \cdot \sin \gamma_C + \Delta y \quad (4)$$

From (1) and (4), we can calculate corresponding longitude

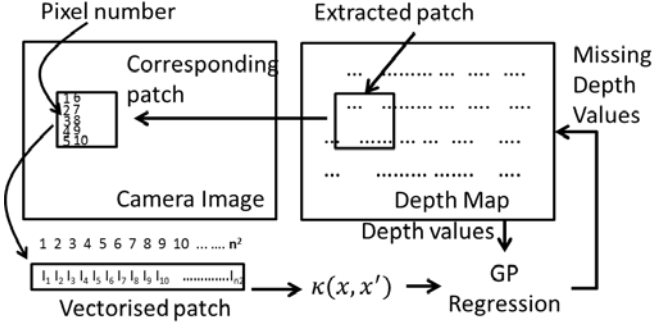


Fig. 4. The sequence of steps in Gaussian Process (GP) based resolution matching algorithm.

γ_C of the camera as following,

$$\tan \gamma_C = \frac{d_L \cos \beta \sin \gamma_L + \Delta y}{d_L \cos \beta \cos \gamma_L - \Delta x} \quad (5)$$

The Equations (3) and (5) pave way to align the data points of the LiDAR and the camera. The purpose of the calibration process is to find the parameters H_C , H_L , Δy , Δx .

Although posing minimalistic needs for calibration, the above geometric alignment process cannot be fully relied upon as a robust mechanism, because errors in calibration measurements, imperfections in sensor assembly, and per-unit variations derived from the manufacturing processes may introduce factors that deviate from the ideal sensor geometry. For example, the curvature of the 360° camera might not be uniform across its surface. Therefore, to be robust enough for such discrepancies, the geometrically aligned data ideally must undergo another level of adjustment. This is accomplished in the next stage of the framework by utilizing the spatial correlations that exist in image data.

Another problem that arises when fusing data from different sources is the difference in data resolution. For the case addressed in this paper, the resolution of LiDAR output is considerably lower than the images from the camera. Therefore, the next stage of the data fusion algorithm is designed to match the resolutions of LiDAR data and imaging data through an adaptive scaling operation.

B. Resolution matching based on Gaussian process regression

In this section we describe the proposed mechanism to match the resolutions of LiDAR data and the imaging data. In section III.B, through geometric alignment, we matched the LiDAR data points with the corresponding pixels in the image. However, the image resolution is far greater than the LiDAR output. The objective of this step is to find an appropriate distance value for the image pixels for which there is no corresponding distance value. Furthermore, another requirement of this stage is to compensate for discrepancies or errors in the geometric alignment step.

We formulate this problem as a regression based missing value prediction, where the relationship between the measured data points (available distance values) is utilized to interpolate the missing values. For this purpose we use Gaussian Process Regression (GPR) [8], which is a non-linear regression technique. GPR allows to define the covariance of the data in

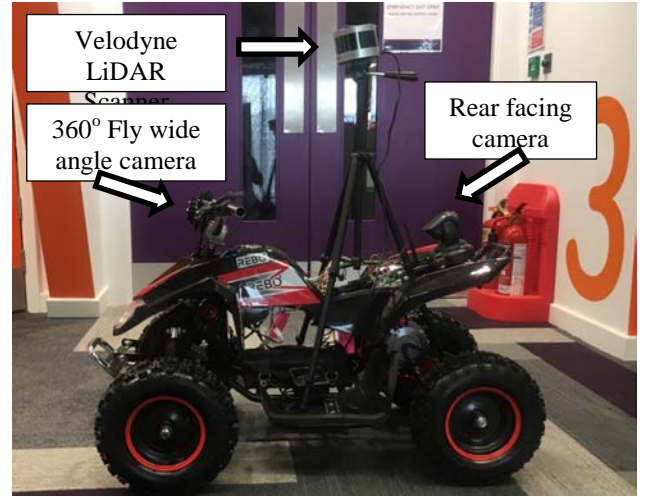


Fig. 5. A photograph of the experimental tested for data collection: an electric quad bike with a LiDAR scanner and cameras.

any suitable way. In this step, we derive the covariance from of the image data, and thereby adjusting to account for discrepancies in the geometric alignment stage. A Gaussian Process (GP) is defined as a Gaussian distribution over functions[8],

$$f(x) \sim GP(m(x), \kappa(x, x')), \quad (6)$$

where, $m(x) = \mathbb{E}[f(x)]$, and $\kappa(x, x') = \mathbb{E}[(f(x) - m(x))(f(x') - m(x'))^T]$.

The power of GPs lies in the fact that we can define any covariance function κ as relevant to the problem at hand.

Let's denote a patch of size $n \times n$ extracted from the depth map D , as y_i . Pixels in an extracted patch is numbered from 1 to n^2 - increasing along the rows and columns. Some pixels of this patch has a distance value associated with it (geometric alignment stage). The objective of this regression step is to fill the rest of the pixels with an appropriate depth value.

The pixels with a depth associated to it will act as the training set $D = \{(x_i, f_i), i = 1:N\}$, where N is the number of pixels that has a depth associated with it, and x_i is the pixel number. Let $X = \{x_i, i = 1:N\}$, $f = \{f_i, i = 1:N\}$, and $X^* = \{x_j, j = 1:n^2 - N\}$, is the set of pixel numbers for which the depth map is empty. The resolution matching problem then becomes to find $f^* = \{f_j^*, j = 1:n^2 - N\}$, the depth of the pixels corresponding to X^* . By definition of the GP, the joint distribution between f and f^* has the following form,

$$\begin{pmatrix} f \\ f^* \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \mu \\ \mu^* \end{pmatrix}, \begin{pmatrix} K & K_* \\ K_*^T & K_{**} \end{pmatrix} \right) \quad (7)$$

Where, $K = \kappa(X, X)$, $K_* = \kappa(X, X^*)$, and $K_{**} = \kappa(X^*, X^*)$, are the covariance matrices defined utilising the covariance function κ , μ and μ^* is the corresponding mean vectors for f and f^* . The solution to f^* is given as the posterior predictive density as follows [22],

$$p(f^* | X^*, X, f) = \mathcal{N}(f^* | \mu^*, \Sigma^*) \quad (8) \text{ where,}$$

$$\mu^* = \mu(X^*) - K_*^T K_y^{-1} (f - \mu(X)) \quad (9), \text{ and}$$

$$\Sigma^* = K_{**} - K_*^T K^{-1} K_* \quad (10).$$

A suitable covariance function κ has to be defined to meet the objective of filling the missing values in the depth map D . So how do we define a suitable covariance function? To do so,

we make the assumption that similar pixels of the color image will have the same depth value. Similarity of the pixels is defined based on the Euclidian distance between the pixels and the grey-level of the pixel. As such we define the covariance function as,

$\kappa(x, x') = c(x, x') \cdot s(x, x')$, where the covariance between any two pixels x, x' , $\kappa(x, x')$ is the multiplication of two factors: $c(x, x')$: closeness between the two pixels in terms of spatial Euclidian distance and $s(x, x')$: similarity between the two pixels in terms of its grey-level value, defined as follows,

$$c(x, x') = \exp\left(-\frac{1}{2} \cdot \frac{\|x - x'\|^2}{K_p}\right) \quad (11),$$

$$s(x, x') = \exp\left(-\frac{1}{2} \cdot \frac{(I_x - I_{x'})^2}{K_I}\right) \quad (12),$$

where I_x denotes the gray-level value of the camera image at pixel position x . K_p and K_I controls the width of the respective kernels. The missing depth value of a pixel x , is taken to be the mean value μ^* at x given by Equation 9, and the corresponding uncertainty of the calculated pixel value is taken to be the variance at x given by Equation 10. To summarize, the GP based regression to fill the missing depth values is illustrated in Figure 4.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this section we will describe the experimental test bed from which test data were gathered, and discuss the results obtained with the data fusion algorithm described in the previous section.

A. Description of the test bed

To collect the data necessary for the experiment, a test bed was assembled as illustrated in Figure 5. The test bed is composed of a front facing wide angle camera, a rear facing camera, and a LiDAR scanner tagged to an electric quad bike. However, in this paper we are focused on fusing only the front facing wide angle camera output with the LiDAR scanner output.

The Velodyne VLP-16 LiDAR, which is used in the test bed, a compact low power light-weight optical sensor, has a maximum range of 100 metres. The sensor supports 16 channel communications taking a total of 300,000 measurements per second. Data is captured over 360° on the horizontal axis and 30° on the vertical, utilizing 16 laser/detector pairs.

The wide angle camera utilized in the setup is a 360Fly camera is enclosed in a 61mm diameter sphere with a single fish eye lens mounted on the top. The field of view is 360° on the vertical and 240° horizontal. Standard 360° video that is output from this camera is a flat equi-rectangular video displayed as a sphere.

The measurements for the geometric alignment stage is as follows: $H_C = 0.55m$, $H_L = 0.61m$, $\Delta y = 0.07m$, $\Delta x = 0.5m$.

B. Performance evaluation methodology

The purpose of data fusion is to assist subsequent data

perception tasks. Hence, the performance of a data fusion framework need to be assessed within the context of subsequent processing stages. This subsection describes the methodology employed in this work to assess the performance of data fusion frameworks. To keep within the context of driverless vehicles, we will utilize Free space detection (FSD) as a representative perception task.

FSD is the mechanism by which a driverless vehicle understands regions in the space to which it can move in to without bumping in to any obstacle. Typically data driven learning methods are utilized to train FSD algorithms. In this experiment we illustrate the effectiveness of sensor data fusion in terms of training a FSD classifier.

As illustrated in Figure 6(b), the data fusion stage produces a depth map indicating the distance from the LiDAR to each pixel in the color image. For the purpose of this study, free space is defined as any point in space, which is at the same level as the wheels of the test bed. We assume the floor on which the vehicle moves is flat. Therefore, any pixel representing a point in space at the same level as of the bottom of the wheels, is considered a “free space” point.

For performance evaluations that proceed, we utilize three image segments that are to be fused. Utilizing the fused output, which is a distance map as shown in Figure 6(b), the free space points are identified. The logical mask that represents free space points is referred to as the “free space mask”. The free space mask is then compared to the ground truth mask. The ground truth for the three image segments are manually marked. Similarly, to measure the performance of different techniques discussed in the proceeding sections, a free space detection mask is obtained and compared against the ground truth.

The number of pixels that do not match the ground truth is obtained by a simple ‘xor’ operation of the masks. The proportion of pixels that do not match the ground truth is taken as the ultimate measure of performance.

C. Visual illustration of fusion results

The Figure 6 provides a visual illustration of the inputs and outputs of the data fusion process. Figure 6(a) is the data samples to be fused. The markers on the image indicate the geometrically aligned distance data points. Figure 6(b) is the result of GP based resolution matching experiment, and Figure 6(c) is the corresponding uncertainty associated with the depth value interpolations. Note that, to the middle of the images, there is a laser reading that is missing. Although GP based regression fills those regions with reasonable values, the uncertainty is also high. Furthermore, some areas at which there are sharp colour discontinuities, there is a high level of uncertainty, for example on the carpet.

Also note that data from one of the lidar scanners are missing in the middle of Figure 6(a). In such situations, data fusion algorithm still finds a depth value, but it is represented with high level of uncertainty, as seen in Figure 6(c).



Fig. 6. Visual illustration of the outputs of the data fusion framework.

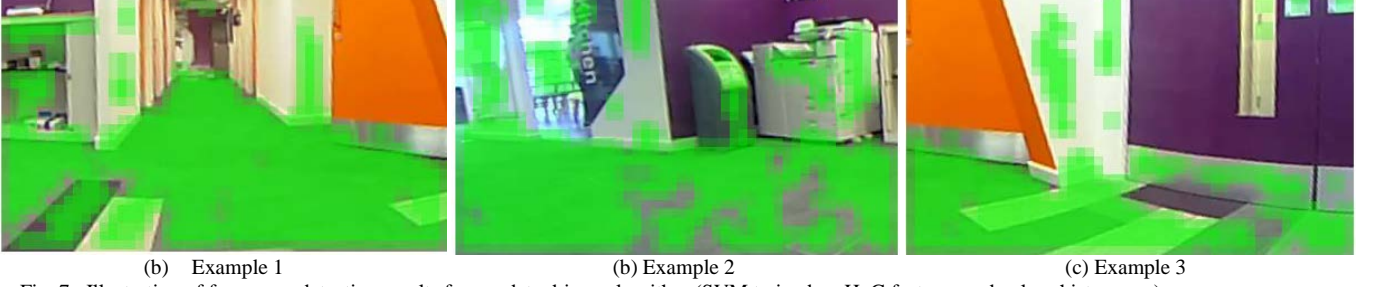


Fig. 7. Illustration of free space detection results from a data driven algorithm (SVM trained on HoG features and colour histograms).



Fig. 8. Illustration of free space detection results from the proposed fusion framework.

D. Performance evaluation

In this section we evaluate the performance of the proposed data fusion methodology by considering the approach described in section IV.B. The FSD based on the proposed data fusion framework is referred to as the “LiDAR+Camera Fusion based FSD” in Table I.

1) Comparison of multi-modal data fusion with single sensor based approach for image understanding.

Firstly, we consider the performance improvements gained by sensor data fusion as compared to single sensor. The single sensor based approach is simulated as a FSD algorithm that learns from examples of free space from color images.

For this purpose, we collect a training set of image patches from the camera image and assign appropriate labels as free space or not. 140 image patches of size 16×16 are collected from the three example images to train a Support Vector Machine (SVM) classifier. Histogram of Oriented Gradients (HoG) features and histogram of the HSV (Hue, Saturation, Value) space, are extracted from these training image patches, and the SVM is trained from those feature vectors. This SVM based FSD classifier is referred to as the “HoG and HSV features based SVM” in Table I, which is a single-sensor (imaging) approach for FSD. Furthermore,

we also compare against an SVM classifier trained from only HoG features, referred to as “HoG features based SVM”.

The corresponding FSD results are illustrated in Figures 7 and 8, where the detected free space is marked in green. Note that, in the examples in Figure 7, which corresponds to the “HoG and HSV features based SVM”, there are many flat regions (not much texture details) that are falsely classified as free space. The data fusion based FSD algorithm has a low false detection level.

To summarize the performance, the comparison of the percentage of pixels that are incorrectly classified for each of the examples is given Table I. Considering the results in Table I, it is quite evident that the data fusion based FSD algorithm outperforms the SVM classifier by at least 60%. As such, this experiment demonstrates the advantages of multimodal sensor data fusion as compared to single sensor based approaches. Furthermore, it is also evident from the results in Figures 7, 8 and Table I, that the proposed framework presented in this paper performs very well to align, and match the two heterogeneous data streams.

2) Comparison of different algorithms for resolution matching.

The purpose of this experiment is to experimentally justify the utilization of Gaussian process (GP) regression for resolution matching framework in section III.B. As described

TABLE I
PERFORMANCE OF DATA FUSION FRAMEWORK: PROPORTION OF PIXELS
FALSELY CLASSIFIED AS FREE SPACE

Algorithm	Example 1	Example 2	Example 3
HoG & HSV feature based SVM	0.16	0.26	0.17
HoG features based SVM	0.25	0.32	0.27
Proposed LiDAR + Camera Fusion based FSD	0.06	0.10	0.05

in section III.B, we approached the resolution difference issue in sensor data fusion as a missing value interpolation problem. We compare the performance of GP regression method with two other methods: tensor factorization for incomplete data [23] and robust smoothing based on discrete cosine transform [24]. The results are compared based on the performance of FSD algorithm, and are summarized in Table II. As illustrated the proposed GP regression based approach perform consistently well.

Although recommended as a framework in [8], tensor factorization based approaches such as [23] are not very suitable for high-dimensional imaging applications. The proposed tensor based methods in literature are mainly targeted at low dimensional signals. Although slightly lower in performance compared to the proposed method in two test cases, the DCT based smoothing approach for grid data [24] performs well for the resolution matching scenario. This means the algorithm in [24] too is able to capture the high dimensional non-linear spatial variations. However, the downside of this approach is that it does not provide the uncertainty of the estimations.

E. Limitations of the current method

We identify several limitations of the proposed method. Firstly, the proposed method works only on a single image frame and does not utilize the diversity offered by temporal redundancies. Another, limitation is that the proposed data fusion method does not adequately address the issue of seeing through glass. Both types of the optical sensors used in this study, would not recognize see through glass as an object but identifies objects beyond the see through glasses. This would become problematic for applications such as FSD. The reflectivity data that are generated in the LiDAR, which is not utilized in this method, might prove beneficial towards solving this issue. We will consider the above problems and address the limitations in our future work.

V. CONCLUSIONS AND FUTURE WORK

This paper addresses the problem of fusing the outputs of a LiDAR scanner and a wide-angle monocular image sensor. The first part of the proposed framework spatially aligns the two sensor data streams with a geometric model. The resolutions of the two sensors are quite different, with the image sensor having a much denser spatial resolution. The two resolutions matched, in the second stage of the proposed

TABLE II
COMPARISON OF DIFFERENT RESOLUTION MATCHING ALGORITHMS
AGAINST THE PROPOSED GAUSSIAN PROCESS FRAMEWORK

Algorithm	Example 1	Example 2	Example 3
Proposed Gaussian Process Framework	0.06	0.09	0.05
Tensors factorization for missing values [23]	0.21	0.17	0.22
Robust DCT smoothing for grid data [24]	0.05	0.12	0.06

framework, by utilizing a Gaussian Process regression algorithm that derives the spatial covariance from the image sensor data. The output of the GP regression, not only provides an estimation of the corresponding distance value of all the pixels in the image, but also indicates the uncertainty of the estimation by way of standard deviation. The advantages of the proposed data fusion framework is illustrated with a free space detection algorithm, which improved its performance by more than 60% compared to single sensor based detection capability. The future work planned, includes extension of the sensor fusion framework to include multiple cameras, radar scanners and ultra sound scanners. Furthermore, we will research methods for robust free space detection based on the data fusion framework.

REFERENCES

- [1] T. Litman, "Autonomous vehicle implementation predictions: Implications for Transport Planning," *Victoria Transport Policy Institute*, vol. 28, 2017.
- [2] L. Jones, "Driverless when and cars: where?," *Engineering & Technology*, vol. 12, no. 2, pp. 36–40, Mar. 2017.
- [3] T. Simonite, "Prepare to be underwhelmed by 2021's autonomous cars," *MIT Technology Review*. [Online]. Available: <https://www.technologyreview.com/s/602210/prepare-to-be-underwhelmed-by-2021s-autonomous-cars/>. [Accessed: 19-Jul-2017].
- [4] A. Broggi *et al.*, "PROUD #x2014;Public Road Urban Driverless-Car Test," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 6, pp. 3508–3519, Dec. 2015.
- [5] G. Pau, "Quickly Home Please: How Connected Vehicles Are Revolutionizing Road Transportation," *IEEE Internet Computing*, vol. 17, no. 1, pp. 80–83, Jan. 2013.
- [6] R. H. Rasshofer, M. Spies, and H. Spies, "Influences of weather phenomena on automotive laser radar systems," *Advances in Radio Science*, vol. 9, pp. 49–60, Jul. 2011.
- [7] R. C. Luo, C.-C. Yih, and K. L. Su, "Multisensor fusion and integration: approaches, applications, and future research directions," *IEEE Sensors Journal*, vol. 2, no. 2, pp. 107–119, Apr. 2002.
- [8] D. Lahat, T. Adali, and C. Jutten, "Multimodal data fusion: an overview of methods, challenges, and prospects," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1449–1477, 2015.
- [9] S. A. N. Gilani, M. Awrangjeb, and G. Lu, "Fusion of LiDAR data and multispectral imagery for effective building detection based on graph and connected component analysis," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XL-3/W2, pp. 65–72, Mar. 2015.
- [10] J. Li, "Fusion of LiDAR 3D points cloud with 2D digital camera image," Oakland University, 2015.
- [11] X. Gong, Y. Lin, and J. Liu, "Extrinsic calibration of a 3D LIDAR and a camera using a trihedron," *Optics and Lasers in Engineering*, vol. 51, no. 4, pp. 394–401, 2013.
- [12] X. Gong, Y. Lin, and J. Liu, "3D LIDAR-Camera Extrinsic Calibration Using an Arbitrary Trihedron," *IEEE Sensors*.
- [13] H. Alismail, D. L. Baker, and B. Browning, "Automatic calibration of a range sensor and camera system," in *3DIMPVT 2012*, 2012.

- [14] Y. Park, "Calibration between color camera and 3D LIDAR instruments with a polygonal planar board," *IEEE Sensors*, vol. 14, no. 3, pp. 5333–5353, 2014.
- [15] G. Moreno and A. Ivan, "LIDAR and panoramic camera extrinsic calibration approach using a pattern plane," *Pattern Recognition*, pp. 104–113, 2013.
- [16] Z. Lipu, "A new minimal solution for the extrinsic calibration of a 2D LIDAR and a camera using three plane-line correspondences," *IEEE Sensors Journal*, vol. 14, no. 2, pp. 442–454, 2014.
- [17] Z. Lipu and Z. Deng, "Extrinsic calibration of a camera and a lidar based on decoupling the rotation from the translation," in *Intelligent Vehicles Symposium (IV)*, 2012.
- [18] Z. Lipu and Z. Deng, "A new algorithm for the extrinsic calibration of a 2D LIDAR and a camera," *Measurement Science and Technology*, vol. 25, no. 6, 2014.
- [19] A. Mastin, J. Kepner, and J. Fisher, "Automatic registration of LIDAR and optical images of urban scenes," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 2639–2646.
- [20] W. Maddern and P. Newman, "Real-time probabilistic fusion of sparse 3D LIDAR and dense stereo," in *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*, 2016, pp. 2181–2188.
- [21] S. M. Abbas and A. Muhammad, "Outdoor RGB-D SLAM performance in slow mine detection," in *Proceedings of ROBOTIK 2012*, 2012.
- [22] K. P. Murphy, *Machine Learning: a Probabilistic Perspective*. MIT Press, 2012.
- [23] E. Acar, D. M. Dunlavy, T. G. Kolda, and M. Mørup, "Scalable tensor factorizations for incomplete data," *Chemometrics and Intelligent Laboratory Systems*, vol. 106, no. 1, pp. 41–56, Mar. 2011.
- [24] D. Garcia, "Robust smoothing of gridded data in one and higher dimensions with missing values," *Computational Statistics & Data Analysis*, vol. 54, no. 4, pp. 1167–1178, Apr. 2010.

AUTHOR BIOGRAPHIES



Varuna De Silva received his Ph.D. from University of Surrey in United Kingdom in 2011. He was a post-doctoral researcher in the same institute between 2011-2013, working on video and image processing for Multiview video broadcasting applications. Since 2013 November, he worked as a senior

algorithms developer for image signal processors at Apical Ltd, UK (Now part of ARM Plc). In April 2016, he joined Loughborough University, as a lecturer in digital engineering. He is the recipient of IEEE Chester Sall award from the Consumer electronics society in 2011, and Vice-Chancellors award for the best Post graduate research student of the year 2011. His current research interests include sensor data processing and artificial intelligence for driverless vehicles and applied data science.



Jamie Roche Aubrey James Roche is currently reading for a PhD in Intelligent Mobility after having undertaken an M.Sc. in Mobile Internet at Loughborough University London. Prior to coming to London Aubrey worked as a Consulting Forensic Engineer for seven years with Denis Wood & Associates in Dublin Ireland. In 2008 Aubrey graduated from Trinity Collage Dublin with a M.Sc. in Bioengineering. Prior to undertaking

his Bioengineering degree Aubrey worked as a project engineer with Irish Rail, a Water Sanitation and Hygiene (WASH) Engineer with Help International (NGO) and a graduate Engineer at Schering Plough Pharmaceuticals. In 2002 Aubrey graduated from Dublin City University with a B.Eng. in Mechatronics. His current research interests are in cognitive approaches to multimodal sensor data perception.



Ahmet Kondo was awarded his PhD in Signal Processing and Communication from the University of Surrey in 1987 where he was employed as Lecturer and Reader in 1988 and 1995 respectively. He was promoted to Professor in Multimedia Communication Systems in 1996. Ahmet took part in setting up of the world renowned Centres CCSR and

the I-Lab at the University of Surrey before joining Loughborough University London as the Director for the Institute for Digital Technologies. Ahmet has been involved in the coordination of large national and International research projects funded by the European Commission. He has been collaborating with major EU Industries and Universities. He coordinated FP6 VISNET, FP7 DIOMEDES and ROMEO projects. He contributed as a core partner to SUIT, NEWCOME and MUSCADE. Currently he is the coordinator of CLOUDSCREENS. Ahmet has successfully supervised more than 75 PhD students and is currently supervising 6. He has published more than 400 papers and has 7 patents.