Thesis Project

Perception and Interpretation of Dynamic Scenarios using Lidar Data and Images

Agustìn Alberto Ortega Jimènez
Institut de Robòtica i Informàtica Industrial (IRI), CSIC-UPC

Thesis Advisor: Juan Andrade Cetto
PhD Program: Automatització Avanzada y Robòtica
Institut d'Organización i Control de Sistemes Industrials (IOC)
Institut de Robòtica i Informàtica Industrial (IRI)
Universitat Politècnica de Catalunya (UPC)

June 2010

Abstract

This research is focused on the recognition and detection of dynamic objects using lidar data and image sequences with applications to mobile robotics. Analyzing scene dynamics is a challenging task because objects not only change appearance, but also become partially or completely occluded during motion. Moreover moving objects might remain still for large periods of time, making it difficult to classify them as dynamic or static without scene context. In this work we will investigate to what extent the use of dense range data, typically coming from lidar sensing devices, together with image sequences, can be used to improve segmentation, classification, recognition, and reconstruction tasks of dynamic objects and people in the scene. The fusion of lidar data and image sequences for recognition and segmentation of dynamic objects will be demonstrated through improved results for SLAM in highly dynamic scenes, with the added benefit of accurate 3D reconstruction of the dynamic objects.

Institut de Robòtica i Informàtica Industrial (IRI)
Consejo Superior de Investigaciones Científicas (CSIC)
Universitat Politècnica de Catalunya (UPC)
Llorens i Artigas 4-6, 08028, Barcelona, Spain
Tel (fax): +34 93 401 5750 (5751)
http://www.iri.upc.edu

Corresponding author:
Agustin Alberto Ortega Jimenez
tel: +34 93 401 5780
aortega@iri.upc.edu
http://www.iri.upc.es/people/aortega

Contents

1 Introduction

Scene understanding is fundamental for mobile robotics applications. We mean by scene understanding, the correct segmentation and recognition of objects classes, people or events in image and range data sequences. Scene understanding is largely motivated by context, hence segmentation, recognition and classification of objects, people or events will be largely influenced by the task at hand. In particular, for our cases of interest, mobile robot navigation and human-robot interaction, scene understanding entails the correct detection and recognition of dynamic elements in the scene, be these related to the task, or mere peripheral activity that once detected, can be safely ignored. The ability to remove this peripheral activity from sensor data will allow us to improve our simultaneous localization and mapping methods [6, 95, 40], which to a large extent have been devised to work on static data, with few exceptions [5, 32].
Lidar sensing devices provide range measurements to objects in 2d and 3d. If aggregated from multiple vantage points, these measurements can be used to build rich 3d scene representations. Lidar scanners are a common sensing alternative for the detection and recognition of objects and features in three-dimensional scenes [91, 22, 92, 1]. In Figure 1 we show a 3d point cloud, registered with our custom-built 3d range scanner. The scan has an angular resolution of 0.5 degrees, and shows static as well as dynamic objects in the scene, such as a desk, people, and mobile robots.
Lidar sensors however, are time-of-flight devices. This is, the range measurements they produce are computed by measuring the time it takes for an emitted signal to return to the device after hitting the scene. Typical lidar scanners take 25 millimeters to compute a 2d slice of range data. The time needed to compute 3d images with this type of sensors may vary from 2 to 20 seconds typically. Computing the scan shown in Figure 1 with such dense resolution, takes 18 seconds. Scene content can vary significantly in that amount of time, and cannot be segmented or interpreted robustly from one single scan. Our objective is to integrate multiple scans, as well as image data to correctly recognize and reconstruct the scene.
Camera frame rates on the other hand are much faster, typically at the rate of 25 or 30fps. Using images to detect motion might improve segmentation and recognition results that otherwise would be impossible purely from laser data. Cameras however, do not provide range information directly, but through triangulation from various vantage points. Nonetheless, the accuracy of distance computation from stereo cannot compete with that of lidar devices. Cameras also provide richer appearance information on the scene when compared to lidar sensors. This will also aid in the tasks of identification and recognition [80, 42]. The use of appearance data however, might also be a problem in dynamic scenes. Illumination changes, cast shadows or reflections may hinder segmentation and recognition results unless properly addressed. See for instance, the scene in Figure 2, which shows a typical mobile robotics scenario. Our task is to devise robust computer vision and data fusion methods that can recognize dynamic objects in these type of scenarios with unlimited dynamic content.

Thus, our research is focused in the recognition and detection of dynamic objects fusing lidar data and images. The results of our research will be demonstrated through improved SLAM in highly dynamic scenes, with the added benefit of accurate 3d reconstruction of dynamic objects.

This document is organized as follows: First, the main objective and methodology are presented. Then, the state of the art using either lidar data or images is discussed. Thirdly, the proposed methodology to carry out our research is explained, and our contributions achieved so far are detailed, together with resources available to run the real experiments. Finally, a detailed work plan is given, divided in tasks and milestones.

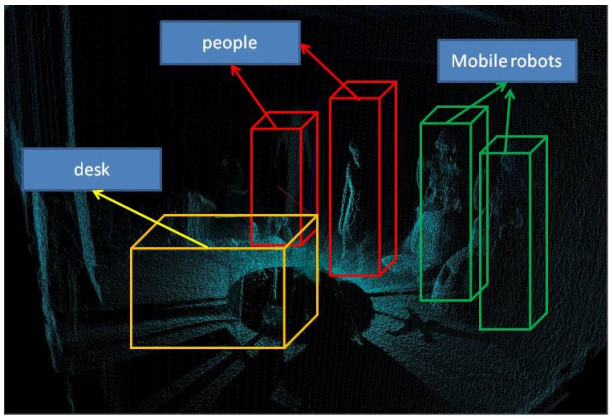


Figure 1: 3D scan of a typical mobile robotics dynamic indoor scenario.

Figure 2: Indoor dynamic image sequence of a typical mobile robotics scenario.

2 Objective

2.1 Main Objective

The main objective of our research is to contribute with novel object detection and recognition algorithms for dynamic objects using lidar data and images. Our systems shall work over scenes with unlimited dynamic content. The detection of dynamic objects will allow improvements in other tasks such as SLAM, 3d reconstruction of motion, and 3d scene reconstruction.

2.2 Methodology

To achieve our main objective we will solve the following issues:

- Define the classes of objects to be detected. Come up with the right representation (feature extraction) of dynamic objects such as cars or people, and static objects such as ground, walls, buildings, etc. for our classifiers.
- Build the right classifiers. Come up with a novel and efficient classification mechanism that merges appearance and range data, as well as spatial and temporal information for the recognition of dynamic objects in the scene.
- Compare our method with competing alternatives.

- Show that the method works both on synthetic and real data with unlimited dynamic content.
- Use our method to enhance SLAM and scene reconstruction algorithms already devised in our group of research.

3 State of the art

This section presents a review of the state of the art in methods that can use either or both lidar data and images. The reviewed topics are: sensor calibration, 3d reconstruction, feature descriptor in lidar data and images, recognition and detection of dynamic or static objects. We will focus our research in detection and recognition of dynamic objects. Table 1 presents an overview of topics that will be seen in this Section.

Topic	Application Area
Sensor Calibration	- Camera Sensor - Laser Sensor
Feature Descriptors	- Images - Lidar data
3d Segmentation	- Geometric features - Texture features
3d Recognition/Detection	- Places - People - Cars
3d Reconstruction	- Indoors - Outdoors - Objects

Table 1: Review of recent topics using lidar data and images

3.1 Sensor Calibration

We are interested in an accurate registration of laser range data with intensity images. The registration can be possible using sensor calibration that allows having a mapping between sensors and the real world. Whereas camera calibration is a mature topic and the intrinsic and extrinsic parameters can be obtained [35]. These methods need to observe a planar pattern [105, 93] to calibrate a single camera. Camera calibration can be extended to calibrate a camera network restrained to share field of view [87, 103]. Instead of using a pattern the map build from a SLAM session [95] over lidar data can be used to calibrate a non-overlapping camera network [67, 4], when is fusing the laser data and images to complete the calibration process, analysis more details in Sec. 5. Register and calibration using lidar data and images are studied in [56, 66].
Laser calibration obtains the extrinsic parameters regarding the other sensors. Homogeneous transformation to sensors such as laser-camera to register the 3d points in the image can be applied. We define the transformation between laser-camera to obtain the extrinsic parameters as follow; it is defined a pinhole camera system with projection world coordinates $P = [X, Y, Z]^T$ and image coordinates $p = [u, v]^T$. The projection of 3d point in the image is defined as

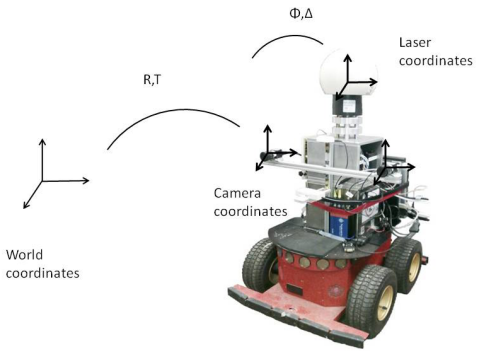


Figure 3: Relation between sensors coordinate systems

$$p \sim K(RP + t).$$

Where R and t are orientation matrix and translation known as the extrinsic parameter, and the matrix K represents the intrinsic camera parameters. Thus R, t and K can be computed with computed with a classic pattern based camera calibration method [105, 93].
Now it is defined a laser point P^f and laser scan plane as $Y = 0$ then the rigid transformation from the camera coordinate to laser coordinate system can be described by Equation 1

$$P^f = \Phi P + \Delta, \tag{1}$$

where Φ is defined such as $3x3$ the orientation matrix and translation vector Δ respect to laser coordinates system. The goal of laser calibration is to develop ways to solve and find these extrinsic camera parameters Φ and Δ which define the position and orientation of the camera with respect to the laser coordinate system. Figure 3 presents the relation between laser and camera sensors respect the world coordinate system.
Register the 3d laser range data points in the image is performed applying the transformation to obtain the same coordinate system (Eq. 1).
Laser calibration has been used for 3d reconstruction projecting 3d point to the image and assigning RGB (Red, Green, Blue) pixel value to 3d point. There are laser calibration methods that use a planar pattern and 2d lidar scan [104], this method minimize a function between image homographies H and 2d laser plane or normal N, the minimization process is non-linear and solved by

$$\min_{\Phi,\Delta} \sum_i \frac{N_i}{|N_i|}(\Phi^{-1}(P^f - \Delta) - |N_i|)^2,$$

where N_i defines the checkerboard plane in the i_th pose. A rotation Φ is parameterized by the Rodrigues formula as a 3-vector parameter, which it is in the direction of the rotation axis and has a magnitude equal to the rotation angle. They assume previous camera calibration with almost 15 planar pattern poses and the images must be unwrapped. This method has been extended to work with a set of 3d lidar built-based [94]. The method is sensible to laser outliers, and 3D planes must be selected manually doing a tedious task. Scaramuzza et al. [76] propose a method that does not need planar pattern, the method only needs to select image and lidar data points almost 4 correspondence, the re-projection error is minimized given by

$$\min_{\Phi,\Delta} \sum_i (p_i - m(\Phi, \Delta, P_i^f))^2,$$

where $m(.)$ defines projection of a 3D point P_i^f in the image; The method considers different laser-camera resolution and applies non-linear refinement, both methods need direct interaction with the user. We propose to analyze an automatic method of semi-automatic calibration using features extract from image and lidar data such lines, corners, etc; considering the methods that have been recently applied to laser calibration.

Once laser is calibrated respect to other sensors such as camera sensor, lidar data points can be registered using the extrinsic parameters. We shall have that 3d lidar data resolution used to be lower than image spatial resolution then interpolation methods for lidar data and high resolution images fusion must be considered. Interpolation generates dense data that shares the same resolution between laser data points and images. Methods such as MRF (Markov Random Fields) have been used in lidar data and images [100] to interpolate data, Diebel and Thrun [20] use a MRF to interpolate range images with low resolution and high resolution images, they obtain range images with resolution of 10x, the minimization function includes pixel information, laser depth, this method was applied for 3D indoor reconstruction besides subpixel refinement is not considered. Meanwhile Andreasson et al. [31] compare the Diebel and Thrun approach generating Voronoi diagrams in the registered 3D points in the image, they consider features such as the spatial position, color information, position, and the area, the approach is tested in outdoors improving their results. Moreover Harrison and Newman [34] propose a new method extending MRF solving the minimization function in closed-form besides includes a second curvature order improving their results and reducing the computational cost. Yang et al. [102] improve the interpolation incrementing the spatial resolution of 100x, the method is iterative and require previous depth estimation. We shall consider a robust interpolation methods that can improve the results to generate automatic 3d reconstruction.

3.2 Feature Descriptors

In the past computer vision methods have tried to understand dynamic proprieties through of images or image sequences. Computer vision has development methods such as tracking, segmentation, detection, and recognition of dynamic objects. These methods consider a set of features that contain valuable information such as motion, appearance, shape, region moments, etc. The features used to be combined with machine learning techniques to recognize dynamic and static objects. A detailed review on computer vision algorithms can be seen in Gonzalez et al. [29].

Computer vision descriptors most popular in the state of the art are: Scalable Invariant Features (SIFT's) descriptor[55] that uses a Difference of Gaussians (DoG), this descriptor stores the gradient orientation in different scales, it is rotation-scale invariant but is not invariant to affine transformations. Speed-up Robust Features (SURF's) [11] descriptor uses the determinant of the Hessian matrix, the histogram created is smaller than SIFT's and reason for this it is faster to train. Shape-Context [12] descriptor is based in gradient orientation around contour point, the values are saved in an histogram, the descriptor is sensible to the background scenes then it is needed to segment previously. Harris corner descriptor [33] is quite popular in computer vision, it has extended to be invariant to scale and widely used in for camera calibration to find the chessboard pattern corners. Ferns [69] uses template information using probabilistic methods, the method require any prior training in addition to have different orientations templates. LBP (Local Binary patterns) [2] features uses texture information, principally have been used in face recognition. HoGs (Histogram of Orientated Gradients) [18] similar than SIFT's and SURF's storing the gradient orientation in an histogram and HoF (Histogram of Flow) descriptor [19] saves the orientation of flow fields using optical flow. Curvature Scalable Space(CSS) [15] is based in image edges or contours , CSS represents object shape in the curvature space. Many of these descriptors have been extended to recognize actions, gestures, tasks, or activities using

videos or image sequences [49, 10, 77], for security, or analysis behaviors, etc.
Computer vision descriptors have been extended to work with 3d points or lidar data. In [26] an extension based in SURF's to recognize objects over 3d points is proposed, this descriptor is named Thrift that computes the determinant of a Hessian matrix in 3d, the authors test their method with a local database and propose to use with a benchmark 3D object dataset, the authors do not mention if the descriptor is scale-invariant and it is not tested considering noise levels. Rusu et al. [75] propose a method to label 3d points using a descriptor called FPFH (Fast Point Feature Histogram), this descriptor encode the local surface geometry around a 3d point, it is invariant to pose and scale, the method have been used to register 3d clouds combined with the ICP algorithm, afterwards for object recognition using CRF (Conditional Random Fields) and SVM (Support Vector Machines). Hetzel et al. [37] propose a descriptor to recognize free-form objects composed by a set of features such as pixel depth, and surface normal, the descriptor is robust to occlusion besides the histogram is smaller than the SIFT's, recognition is performed using either histogram matching or probabilistic recognition algorithms. The work proposed by Chen and Bhanu [16] create a descriptor represented by mean histogram, the keypoints are the surfaces with more variability or significant gradient changes, the descriptor is named LSP (Local Superficies Patches descriptors), in order to speed-up the authors propose to use a hash table. In [27] is proposed a 3d descriptor based in Shape Context which is computed by each point storing its normal orientation, each bin correspond to 3 angles (elevation, azimuth, and roll), the descriptor depends of the lidar data or the surface resolution, the shape context 3d is sensible to noise then pre-filtering is essential. The CSS descriptor have extended by Steine et al. [83] but reducing the dimensionality with PCA named Eigen-CSS, authors only consider range images to compute contours then used SVM classifier to recognize 3d objects, the method only works with range images but not directly with 3D points. A descriptor called integral volume [28] is defined as 3d surface point of maximum variability, the integral volume given the point with radio r in a sphere centered is computed, the descriptor is robust to noise and used to register 3d points. The spin images [41] are a free-form descriptor that uses normal surfaces, the spin images are a set of images in 3d points projected to 2d cylindrical coordinates, the coordinates are defined with respect to orientated point with radio α and elevation β, to match two spin images are performed using correlation metric, the spin image dimensionality can be reduced using PCA, the method have been tested using the ICP algorithm to register the 3d objects, the spin images have been wide spread used in pattern recognition community using 3d information, applications can be seen in [47, 52, 92]. The spin images have been used based in the SIFT's descriptor[78], they compute SIFT's in the spin image and the descriptor is named RIFT (Rotation Invariant Feature Transform), the descriptor is robust and invariant in rotation. A survey to recognize 3d objects using 3d data points is presented in [14]. Table 2 presents a survey of the 3d features descriptor listed in this Section.

3.3 3D Segmentation

Segments share similar features such as shape, color, etc, and are used to detect or classify image content. In computer vision different segmentation methods have been proposed. Typically, segmentation means identifying a set of pixels clustered in regions with high similarity in their RGB values, but segmentation can be performed not only over intensity values, but also using other features such as gradient, motion, etc. The most popular segmentation methods in computer vision are: the Region-Growing [23], Watershed , Mean-Shift [85], or probabilistic approaches such as MRF [100], or EM (Expectation Maximization), etc.
Segmentation using computer vision helps also to distinguish between dynamic and static objects. In [80] a method that uses Gaussian Mixtures (GMs) is proposed, they assume a constant number of GMs besides the Gaussian parameters such as the mean and variance parameters are

Feature	Descriptor
Silhouettes/Shape	- CSS
	- Shape Context
	- LSP
Free-form	- Spin images
Gradient	- Thrift
	- RIFT
	- FPFH
	- Integral Volume

Table 2: Descriptors listed in this section

updated in each time t by each image, the approach is robust to different lighting conditions. In [42], the work is extended and improved. The authors demonstrate that the background extraction problem can be divided into two densities, a dynamic model, and a model update. They test their results over outdoor traffic scenes.

In order to work with lidar data, segmentation methods have been extended to handle 3D points. The 3D segmentation in lidar data is challenging due the sparseness of data, and the large noise levels [58]. The segmentation using lidar data can help in tasks such as planning to create traversability maps, obstacle detection to avoiding collisions, navigation [57], and reconstruction of geometric features (see Sec. 3.5), as well as for segmenting dynamics objects during vehicle guidance [82].

Lidar data contains geometric information or primitives such as planes, cylinders, spheres, etc. The planes or normal segmentation is the lowest level primitive in lidar data for reconstruction or planning. There are methods related to normal computation such as the Plane-SVD (Singular value decomposition), Plane-PCA (Principal Component analysis), Vector-SVD studied [45, 46]. Normal computation using laser range data requires of robust cluster methods. It is necessary to fit a plane with 3d points, previously must be grouped in regions a set of 3d points using clustering methods.

Region growing algorithms cluster a set of planar patches taking criteria as distance, and angle between planes. Dilatation-based methods are growing adding regions of planar patches [51]; they consider 3 steps: first generating candidates or planar patches, then regions simulated the dilatation with planes, finally spurious regions or 3D points are eliminated. A method that uses the criteria of distance and curvature using growing strategies and disjoint tree sets is presented in [68], this method will be discussed more derailed in the Section 5. Planes can be computed using graph-based methods to clustered 3D planes, the graph methods used to be computational expensive and the methods are considered NP problem [61]. Applications using region growing based-methods can be seen in [88] to segment obstacles using planar surfaces and cones regions to determinate an object position. The system presented by Poppinga et al., [70] for instance contains a number of heuristics to obtain incremental plane-fitting with the assumption that nearest neighbors are taken directly from the indexes in the range image. Moreover, its secondary polygonalization step is viewpoint dependent, relying also on the neighboring associations given by the indexes of the range data. If the number of planes to detect is known a priori, EM can be used to assign points to planes in terms of normal similarity, density of points and curvature [54]. The technique is shown for indoor scenes in which planar patches are usually orthogonal to each other. For larger, sparser point distributions, such as the ones found in outdoor range data, the assumption of a priori knowledge of the number of planes is unrealistic. To this end, hierarchical EM can be used [90], incrementally reducing the number of planes with a Bayesian Information

Criterion (BIC), at the expense of higher computational cost. Contrary to region growing, one could search for region boundaries instead. A good exemplar of this technique is presented in an architectural modeling application [17], in which polyhedral models are generated from range data by clustering points according to their normal directions plotted on a Gaussian sphere. This mechanism helps overcome the sparsity of the point distribution. The assumption that the scene is made of planar regions is exploited to detect plane intersections and corners to compute plausible segmentations of building structures made of polyhedrons of low complexity.
There are cases in which not only is required to segment planes. Segmenting multiple geometric objects is presented by Ladonde et al. [48], they use GMs and EM algorithm for clustering different geometric structures; the segmentation is improved using mathematic approximation such as sphere for noise, cylinders and wires for trees, planes for roads and walls. Vandepel et al. [97] propose a method to segment lidar data finding geometric objects as wires, surfaces, and spheres, their work is extended in [96] segmenting military wires by means of cylinders computing a symmetry histogram. In addition is proposed a segmentation of multiple geometric structures that uses GMs with Random Sampling Consensus (RANSAC) to detect new objects that can appear in the scene [65]; they can distinguish between spheres, cylinders, and planes.
Segmentation of geometric features, shapes, and intensity using 3d data points can use machine learning strategies. Probabilistic method such as MRF methods applied to vision have been extended to work with 3d laser data. Munoz et al. [64] classify 3d laser points using an extension of MRF called AMRF (Associative Markov Random Field) the system works off-line using 3d points each point is labeled using an anisotropic model.
There exist methods that integrate lidar data and images for segmentation. Today is an active research in the robotic community as grasping [44], to create traversability maps [21]. Rasmussen [74] proposes a road segmentation method using color and texture clues, the segmentation is performed using as primary features a bin histogram in the RGB space, they train a Neuronal Network (NN) to generate traversability maps in outdoors with vegetation, the output is a 3D map. Barnea et al. [9] segment data using features as laser intensity, planar surfaces, and color information from images, Firstly apply image segmentation using Mean-Shift algorithm then integrate a set of geometric features to segment the 3D data. In [75] is proposed a labeling method that uses CRF similar to the MRF for segmentation. Posner et al. propose a segmentation and recognition method for people and geometric structures as walls, and ground [71].

3.4 3D Recognition and Detection

Robots in real world need to recognize, to interpret, and to classify a different class of objects for manipulation, navigation, interaction, etc. Recognition helps to interpret and understand proprieties of dynamic environment. Recognition applications using mobile robots include recognition of people, places, textures, besides of recognize dynamic and static object also for segmentation (See Sec. 3.3).
Lidar sensors have been used for object recognition lidar-based 2d. Mozos et al. [63] propose a method to classify places using 2d information, places as corridors, doors, rooms are recognized, the Adaboost (Adaptive Boosting) algorithm is used to classify features such as area, perimeter, and shape of 2d scan, this method is sensible to noise, for dynamic and natural environments. Dynamic objects recognition have dedicated some efforts to recognize people, Arras et al. [7] detect people in a 2d scan, the authors use like circle features in clustered data i.e. radio, area, etc, classification is performed with probabilistic method called Bayesian classifier, the method has false-positive samples over circular object such as chairs and baskets. To avoid the false-positive detection Mozos et al. [62] extend their work using multiple 2d lidar sensors or layers, each sensor is placed at different heights, the advantage over their previous work is robustness in occlusion, different classes such as chairs or tables but need more 2d lasers, and used to

be expensive. Static objects have been recognized with laser-based 2d, Wurm et al. [101] use scans 2d to identify glasses and no glasses for grasping application. Tracking systems have been development with a lidar-based 2d, Monteiro [60] use the 2d data to track and detect people using Viola-Jones method in images. Similarly, in [84] tracking and classification of vehicles and people are proposed for navigation. Applications such as navigation and planning can run over places with car traffic and people clouded [86] using a single lidar 2d.
When we are using laser-based 2d can lose valuable information because our world is tridimentional, Lidar built-based 3d has been proposed to get 3d information of our world. Besides applications have been chosen to interpret it, Agrawal et al. [1] propose a method to classify 3d objects, the data are micro-classified using histogram information, each histogram contains surface information such as normal orientation similar Shape-Context descriptor, the samples are labeled by an expert and K-Means algorithm is used to cluster data. Triebel et al. [91, 92] recognize places and objects using the MRF, they use spin images and reduce data dimensionality with LDA (Linear Discriminant Analysis), the data are training using a AMNS (Associative Markov Networks). In [22] are used spin images to recognize objects such as chairs, boxes, people, the learning is not supervised, they apply a clustering technique called latent Deritche allocation.
New trends have been development to recognize and identify dynamic and static objects. Wang et al. [99, 98] propose a tracking system for people and vehicles, the method only locate the dynamic object. Probabilistic approaches and unsupervised learning are used to cluster different classes of dynamic objects [32], they consider object trajectories, object speed, images templates, and points distribution of lidar data, the method is heuristics.
Incorporating lidar data and images, and other sensors that can help and improve the results of new applications to recognize objects. Mohottala et al. [59] use 3d information and images to recognize vehicles, the vehicles are segmented using surfaces and silhouettes, the results are refined with a vision method called Graph-Cut for segmentation, the image use binary features. Posner et al. [39, 72, 71] propose an unsupervised method that uses a probabilistic approach and bag of words, each features are geometric of low-level (planes and elevation maps) and contextual (pixel information). A multi-sensor approach presented by Douillard et al. [21] recognize multi-classes objects building an elevation map to detect grass and ground, the classification is performed CRF and Virtual Evidence Boosting that allows to have a expert supervision, recognize objects such as cars, people, and segment wall, grass, etc, they consider distance, angle, laser intensity, and visual features, texture, they perform their results using Viola-Jones. Indoors application to recognized static objects are presented by Marton et al. [56] for grasping, they recognize kitchen object and the training is done with a Bayesian method, The arm-based robot uses a set of sensors as laser, camera, thermal camera, and TOF camera to obtain a greater number of features to help in segmentation. Lim and Suter [52] use super-voxels or 3d regions, they applied multi-scale CRF, each super-voxels is clustered and reduced, the features to train are spin images, normal, color, etc. In [30] intensity, color, depth and surface is used to classify objects, features from the image are the gradient 3d, normal, centroid. Steder et al. [81] recognize in range images using Harris corner descriptor and planar patches, they compare their improvements versus spin images.

3.5 3D Reconstruction

Modern times in advances in compute vision and robotics demand the construction of real scenarios and objects. SLAM methods have allowed to build 3d maps using mobile robots [8, 89]. Though progress has been impressive in SLAM, there is still much to do in creating algorithms that reduce the computational cost, characterization of landmarks, non-linearity and data association. But not only SLAM methods have used for 3d reconstruction, Computer vision

methods known as SFM (Structure From Motion) have enabled the creation of 3d scenes. The scenarios can be constructed thanks to sensors that can perceives the 3d world such as, camera stereo systems, and TOF cameras, or some other device capable of capturing the depth. 3d reconstruction has applications in planning to know the free.space, navigation to be aware of the goal and the start, security where recently the trend is to create 3d environments known in 3DTV (3D Television) to cover a complete area, quality control for the reconstruction of 3d objects, and to create 3d models automatically i. e. CAD (Computer Assisted Design).
Actually lasers have become popular for 3d reconstruction since these devices are not sensitive to illumination such as stereo systems or reflections of TOF cameras. The disadvantage is that laser devices are expensive. Some methods obtain geometric features from lidar data [73], they fill gaps regions with the convex hull algorithm, finally are approximated by splines the lines of each structure. The lidar based-built 3d data have been applied to reconstruct faces [13], application for security where people need to be recognized, or application to create more realistic graphics model for games, movies, virtual environments, etc.
it is possible to reconstruct a scenario using a single camera taking and assuming knowledge of scene structure. Lee et al. [50] recognize and reconstruct 3d structures in indoors using lines segments, first finding the 3d structure over the image scene segmented by lines then many geometric hypothesis are generated, they find the best match between a set hypothesis for finally generating a 3d model projecting on images. A novel method that recover spatial layout in clutter environments is presented in [36], parametric 3d boxes are used with machine learning methods to recognize the best hypothesis, the method is designed to work in indoors. Hoeim et al. [38] propose segmentation method based in classification, the method only uses images of the scene in different orientations, the method segment the image using Felzenwal's algorithm to work with super voxels, features such as color, texture, location, shape, and geometry 3d are considered, then labeling is done with Adaboost, they authors propose application in 3d reconstruction and recognition.
3d reconstruction has been performed fusing laser and camera sensors. In the Sec. 3.3 are reviewed recent methods for extracting low-level features necessary for 3d reconstruction named geometric segmentation. Geometric segmentation extracts primitives as planes, cylinders, lines or any geometric feature meanwhile visual segmentation uses images and texture or color information to create more realistic scenarios. Stamos et al. [79] construct models using 3d range data and images matching 3d lines versus 2d lines, first are computed planes then extracting lines by means plane intersection, on the other hand segmentation lines in the image; they register is performed minimizing the lines distances, finally they project image information in the 3d data. Joung et al. [43] present a reconstruction method using a calibrated lidar-camera sensors, modified ICP uses color information for register each cloud points by robot pose is used. Stereo camera system and laser range finder for reconstruction is presented to generate more realistic indoors/outdoors 3d scenarios [53], they use SIFT's for matching each image in the sequence, the ICP algorithm is used to register the cloud points, the results are combined to improve the 3d model. Omnidirectional cameras allows to cover a complete area for reconstruction, in [25] present a 3d reconstruction system that uses omnidirectional cameras and Graph-Cut algorithm to obtain the depth in the scene.

4 Approach

An illustration of the proposed approach is shown in Fig. 4. The proposed system is divided in four modules: sensor calibration, segmentation, recognition, and applications. Sensor calibration shall allow to have the extrinsic transformation between lidar and camera, the 3d points will be registered in the image. Also, this module requires to interpolate data to have similar sampling between image and lidar data. The output of this module will be the extrinsic calibration

parameters and data interpolation. The segmentation module will help to distinguish between dynamic and static objects, and also to segment important geometric and visual features useful for the recognition module. The recognition module will classify object classes using features extracted in the segmentation module. This module will include machine learning methods and novel feature selection mechanisms. Once dynamic and static objects are detected and recognized, we can use this knowledge in applications such as SLAM in dynamics environments and scene reconstruction.

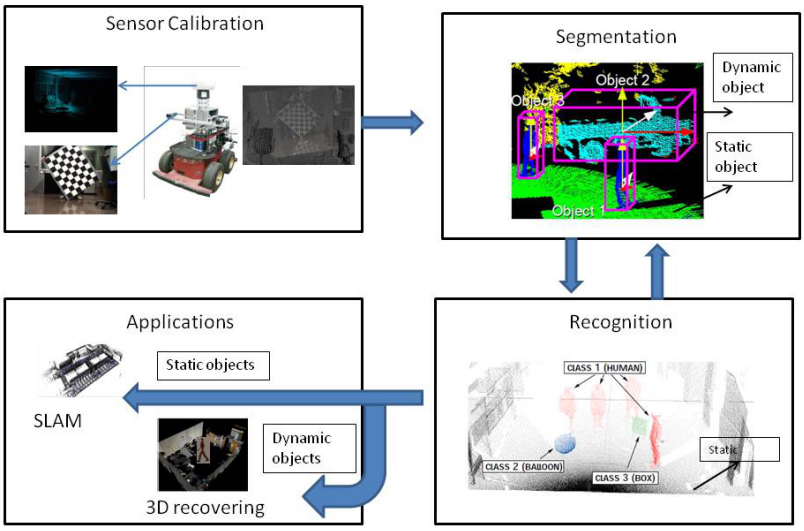


Figure 4: Our approach using lidar data and images for dynamic and non dynamic object recognition.

5 Achievements

This section presents the achievements reached in the period that covers until the thesis proposal. Contributions presented here are as follows: camera network calibration that use a laser range finder data and images,and segmentation of planar surfaces in 3d data.
The calibration procedure is explained as follows, and illustrated in Fig. 5, consists of two main steps [67, 4]. In the first step, a nominal calibration of the cameras is generated by registering the lidar data to an aerial image of the experimental site, showing both in a graphical user interface, and prompting a user to coarsely specify the camera location, orientation, height and field of view. These initial parameters allow the cropping of the entire lidar into regions of interest compatible with the field of view of each camera.
The second stage aims at refining the cameras nominal calibration by matching, in a semi-automatic manner, 3d features to the corresponding 2d features in the camera's images. The lidar data corresponding to each camera's field of view is segmented into a set of best fitting planes with large support from the point clouds and then, straight lines are computed from the intersection of perpendicular planes from the set. The extracted 3d lines are then associated with 2d image lines and this information is fed to a non-linear optimization procedure that improves both intrinsic and extrinsic camera calibration parameters. Finally, homographies of the walking areas are computed by selecting planar regions in the lidar data. The final output of the whole calibration procedure consists in a) the extrinsic camera parameters (the relative

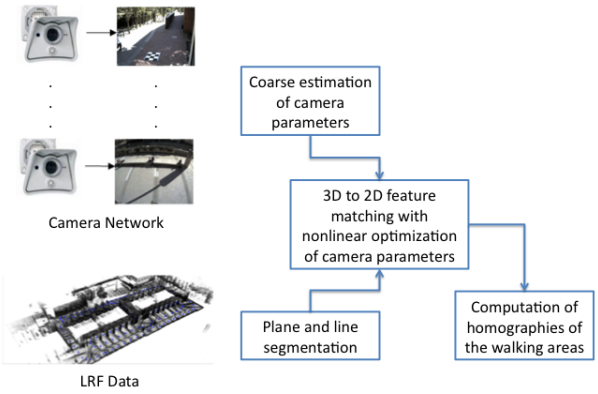
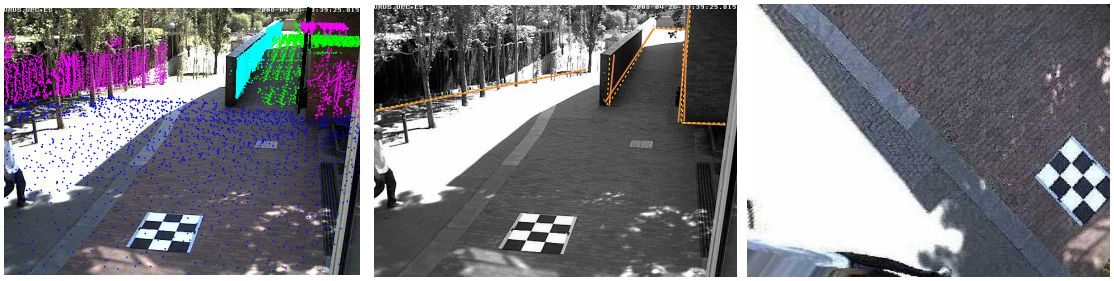


Figure 5: Distributed camera network calibration methodology.

position and orientation in the world frame), b) the intrinsic parameters (focal distance, image center, and aspect ratio) and c) the homographies of the walking areas.
Given the nominal calibration, the 3d straight lines extracted from the lidar data can now be projected in the image and guide the user to select the corresponding 2d image lines. This 3d-2d association allows improving the nominal calibration by minimizing a cost function containing the camera projection matrix P.
Let $p_i = [u_i \ v_i]^T$ denote points that belong to an image line and $P_i = [X_i \ Y_i \ Z_i \ 1]^T \ i = 1, .., n$ denote the corresponding 3d points on the matching line in the lidar data
The cost function is defined as:

$$\hat{\vartheta}_j = \arg\min_{\vartheta_j} \sum_i \|m_i - h\left(P_{proj}(\vartheta_j) \cdot P_i\right)\|^2 \tag{2}$$

where h is a de-homogenization function, $P_{proj}(\vartheta_j) = K[R|t]$ is the projection matrix of the j-th camera and ϑ_j are the calibration parameters, namely focal length and principal point, plus the extrinsic parameters for position and orientation. Calibration results of the camera network calibration are shown in the Fig. 6.



(a) Segmented range data projected to a camera image.

(b) Plane boundaries used for camera calibration.

(c) Calibration results are used to recover an orthographic view of the scene.

Figure 6: Application of the segmentation method for the calibration of an outdoor camera network. Plane boundaries and plane intersections are projected to the image of one of the cameras in the network. A nonlinear optimization of the projection error is used to refine the calibration parameters.

Segmentation is an important task in recognition techniques this will allow to work with different primitives as planar surfaces. Tractability on the other hand is possible by extracting higher order primitives from the extremely large point sets these mapping algorithms produce and, relying on these primitives, to pursue higher level tasks.

To fit normals with 3d data points is defined the error between a fitted planar patch and the lidar data values for the kNNs (k Nearest Neighbors) defined such as

$$\epsilon = \sum_{i \in K} (P_i^\top n - d)^2. \tag{3}$$

Where P_i is a set of clustered points, $n = (n_x, n_y, n_z)^\top$ is the local surface normal at p, K is the set of kNNs to p, and d the distance to the origin. This error can be re-expressed in the following form

$$\epsilon = n^\top \underbrace{\left(\sum_{i \in K} p\, p^\top \right)}_{Q} n - 2d \underbrace{\left(\sum_{i \in K} p^\top \right)}_{q} n + |K|^2 d^2. \tag{4}$$

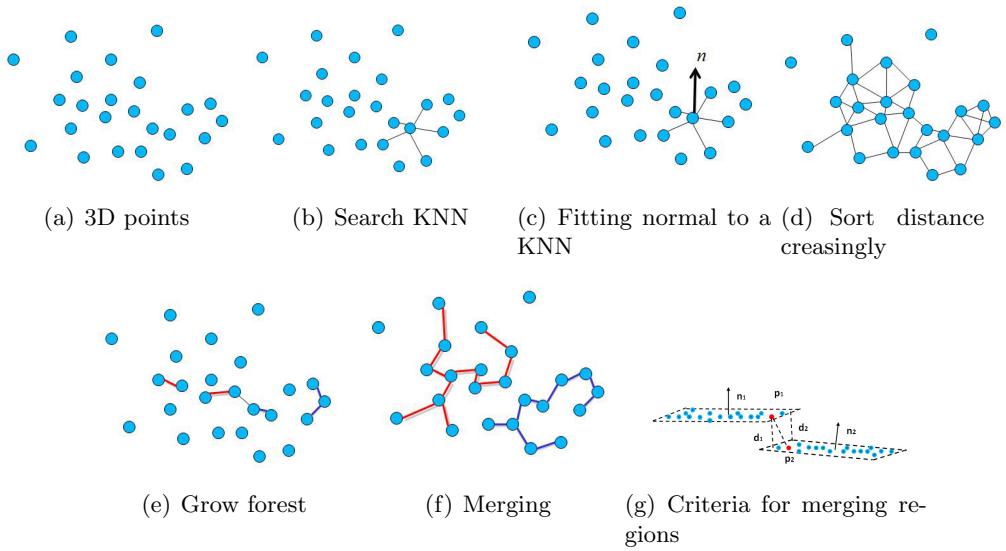
Using the follows Langrian

$$l\left(n^\top, d, \lambda\right) = \epsilon + \lambda(1 - n^\top n),$$

the local surface normal that best fits the patch K is the one that minimizes the above expression [3]. Deriving l with respect to n and d, and setting the derivatives to zero, it turns out that the solution is the eigenvector associated to the smallest eigenvalue of

$$\left(Q - \frac{q\, q^T}{|K|^2} \right) n = \lambda n. \tag{5}$$

The algorithm proceeds as follows. First, the entire data set is preprocessed to compute local normal orientation of fitted planar patches for each point with respect to its kNNs. Then, distances between nearest neighbors are computed. These distances are then sorted in increasing order and the resulting list is processed to create a forest of trees by merging neighboring points according to point distances and to the angle between their normals. The complete process of our algorithms is visualized in the Fig. 7.



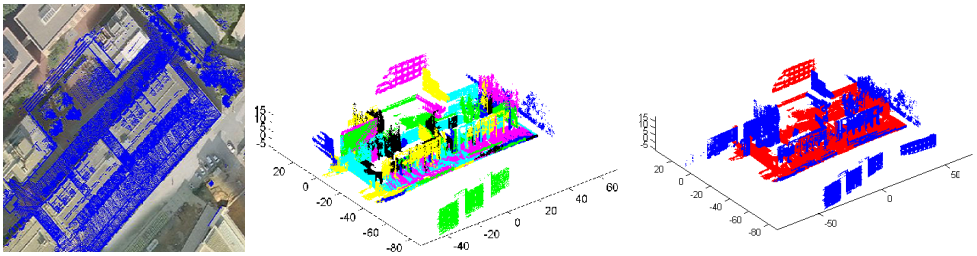
(a) 3D points (b) Search KNN (c) Fitting normal to a KNN (d) Sort distance increasingly

(e) Grow forest (f) Merging (g) Criteria for merging regions

Figure 7: Complete process of our algorithm

The segmented planes can then be used to produce traversability maps (see Fig. 8), to aid in the calibration of a camera network, or to generate VR models of the scene. The proposed algorithm

is very efficient since its computational complexity is $O(n \log n)$ on the number of points in the map. Our method builds upon Felzenszwalb's algorithm for 2d image segmentation [24], and extends it to deal with non-uniformly sampled 3d range data.



(a) Barcelona Robot Lab (b) Plane segmentation using our method (c) An application to create traversability maps

Figure 8: Segmentation of planar patches .

6 Work Plan and Calendar

This section presents the work plan and timetable of our research work. Detailed diagram that explain each task that will be completed in the period from 2007 to 2012 is seen in the Table 3. The period is divided in 11 tasks that are explained as follows:

- T1 Courses: This time interval were cursed 5 signatures suggested by the doctoral committee, this period consist of 12 months including the summer period of 2008.
  Term: 12 months.
- T2 state of the art review in computer vision methods: this task includes a review in computer vision method to recognize objects.
  Term: 4 months.
- T3 Short stay and programing method to calibrate camera network: Contribution explained for network camera calibration in the Section 5.
  Term: 3 months.
- T4 Programming method to segment planar patches over 3d lidar data: Segmentation of planar surfaces using 3d lidar data (See Section 5).
  Term : 3 months.
- T5 Summer school on computer vision, machine learning method, and problem definition.
  Term: 3 months.
- T6 Review in the state of the art to recognize dynamic objects using lidar data and images.
  Term: 9 months.
- T7 Experiments definitions and sensors calibration.
  Term: 3 months.
- T8 Analysis of proposed and recent methods that work using lidar data and images.
  Term: 6 months.
- T9 Experiments using the proposed method with simulated dataset.
  Term: 3 months.

- T10 Real experiments using Segway RPM-400 mobile platform.
  Term: 6 months.
- T11 Writing, review and thesis defense to obtain the PhD degree.
  Term: 6 months.

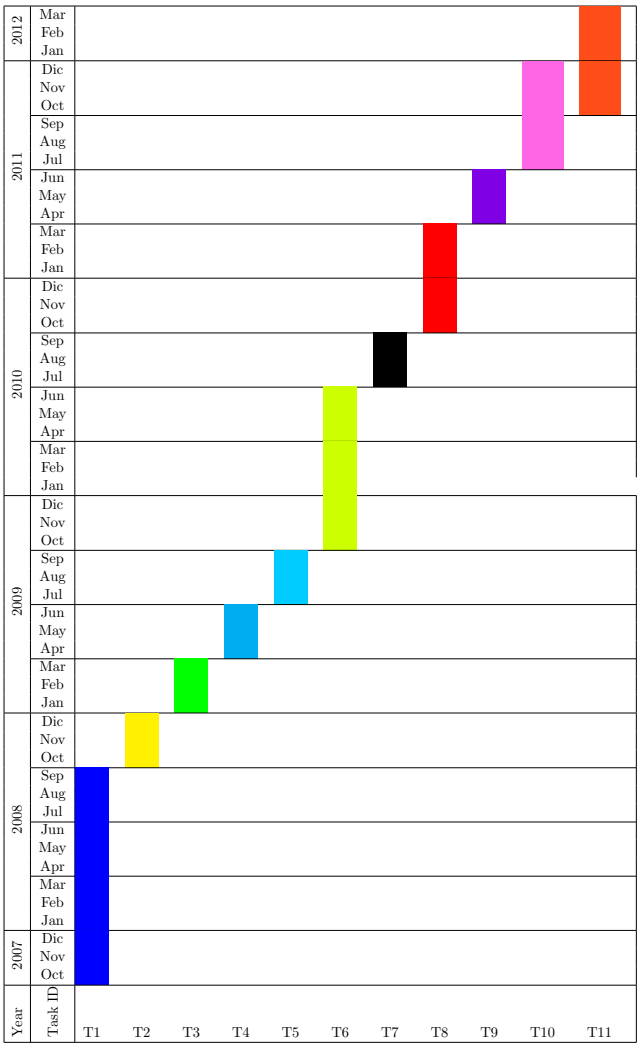


Table 3: Work plan diagram

References

[1] A. Agrawal, A. Nakazawa, and H. Takemura. MMM-classification of 3d range data. In *Proc. IEEE Int. Conf. Robot. Automat.*, pages 2269–2274, Kobe, May 2009.

[2] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE Trans. Pattern Anal. Machine Intell.*, 28(12):2037, 2006.

[3] J. Andrade-Cetto and A. C. Kak. Object recognition. In J. G. Webster, editor, *Wiley Encyclopedia of Electrical and Electronics Engineering*, supplement 1, pages 449–470. John Wiley & Sons, New York, 2000.

[4] J. Andrade-Cetto, A. Ortega, E. Teniente, E. Trulls, R. Valencia, and A. Sanfeliu. Combination of distributed camera network and laser-based 3d mapping for urban service robotics. In *Proc. IEEE/RSJ IROS Workshop Network Robot Syst.*, pages 69–80, Saint Louis, Oct. 2009.

[5] J. Andrade-Cetto and A. Sanfeliu. Concurrent map building and localization on indoor dynamic environments. *Int. J. Pattern Recogn. Artif. Intell.*, 16(3):361–374, May 2002.

[6] J. Andrade-Cetto and A. Sanfeliu. The effects of partial observability when building fully correlated maps. *IEEE Trans. Robot.*, 21(4):771–777, Aug. 2005.

[7] K.O. Arras, O.M. Mozos, and W. Burgard. Using boosted features for the detection of people in 2d range data. In *Proc. IEEE Int. Conf. Robot. Automat.*, pages 3402–3407, Rome, Apr. 2007.

[8] T. Bailey and H. Durrant-Whyte. Simultaneous localisation and mapping (SLAM): Part II state of the art. *IEEE Robot. Automat. Mag.*, 13(3):108–117, Sep. 2006.

[9] S. Barnea and S. Filin. Segmentation of terrestrial laser scanning data by integrating range and image content. In *Proc. Int. Society for Photogrammetry and Remote Sensing*, 2008.

[10] D. Batra, Tsuhan Chen, and R. Sukthankar. Space-time shapelets for action recognition. In *Proc.IEEE Workshop on Motion and video Computing*, pages 1–6, jan 2008.

[11] H. Bay, T. Tuytelaars, and L. Van Gool. SURF: Speeded up robust features. In *Proc. 9th European Conf. Comput. Vision*, volume 3951 of *Lect. Notes Comput. Sci.*, pages 404–417, Graz, 2006.

[12] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Machine Intell.*, 24(4):509–522, apr 2002.

[13] V. Blanz, K. Scherbaum, and H. P. Seidel. Fitting a morphable model to 3d scans of faces. In *Proc. IEEE Int. Conf. Comput. Vision*, pages 1–8, Rio de Janeiro, Oct. 2007.

[14] B. Bustos, D. Keim, D. Saupe, T. Schreck, and D.V. Vranić. Feature-based similarity search in 3d object databases. *ACM Comput. Surv.*, 37(4):345–387, 2005.

[15] C. C. Chang, C. Y. Liu, and W. K. Tai. Feature alignment approach for hand posture recognition based on curvature scale space. *Neurocomputing*, 2008.

[16] H. Chen and B. Bhanu. 3d free-form object recognition in range images using local surface patches. *Pattern Recogn. Lett.*, 28(10):1252–1262, 2007.

[17] J. Chen and B. Chen. Architectural modeling from sparsely scanned range data. *Int. J. Comput. Vision*, 78(2-3), 2008.

[18] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proc. 19th IEEE Conf. Comput. Vision Pattern Recog.*, pages 886–893, San Diego, Jun. 2005.

[19] N. Dalal, B. Triggs, and C. Schmid. Human detection using oriented histograms of flow and appearance. In *Proc. 9th European Conf. Comput. Vision*, volume 3951 of *Lect. Notes Comput. Sci.*, pages 428–441, Graz, 2006.

[20] J. Diebel and S. Thrun. An application of markov random fields to range sensing. In *Proc. of Conf. on Neural Information Processing Systems*, Cambridge, dec 2005.

[21] B. Douillard, A. Brooks, and F.T. Ramos. A 3d laser and vision based classifier. In *Proc. of the 5th Int. Conf. on Intell. Sensors, Sensor Networks and Information Processing*, 2009.

[22] F. Endres, C. Plagemann, C. Stachniss, and W. Burgard. Unsupervised discovery of object classes from range data using latent dirichlet allocation. In *Robotics: Science and Systems V*, Seattle, USA, Jun. 2009.

[23] P. F. Felzenszwalb and D. P. Huttenlocher. Image segmentation using local variation. In *Proc. 12th IEEE Conf. Comput. Vision Pattern Recog.*, pages 98–104, Santa Barbara, Jun. 1998.

[24] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. *Int. J. Comput. Vision*, 59(2):167–181, Sep. 2004.

[25] S. Fleck, F. Busch, P. Biber, and W. Strasser. Graph cut based panoramic 3d modeling and ground truth comparison with a mobile platform - the wagele. *Image and Vision Computing*, 27(1-2):141–152, 2009.

[26] A. Flint, A. Dick, and A. Hengel. Thrift: Local 3d structure recognition. In *Proc. Conf. of the Australian Patt. Recog. Soc. on Digital Image Comp. Tech. and App.*, pages 182–188, dec 2007.

[27] A. Frome, D. Huber, R. Kolluri, T. Bulow, and J. Malik. Recognizing objects in range data using regional point descriptors. In *Proc. 8th European Conf. Comput. Vision*, Prague, 2004.

[28] N. Gelfand, N. J. Mitra, L.J. Guibas, and H. Pottmann. Robust global registration. In *Proc. of the 3th Eurographics symposium on Geometry processing*, page 197, Vienna, 2005.

[29] R.C. Gonzalez, R. E. Woods, and S. L. Eddins. *Digital Image Processing Using MATLAB*. Upper Saddle River, 2003.

[30] S. Gould, P. Baumstarck, M. Quigley, A.Y. Ng, and D. Koller. Integrating visual and range data for robotic object detection. In *Proc. 10th European Conf. Comput. Vision*, volume 5302 of *Lect. Notes Comput. Sci.*, Marseille, 2008.

[31] Henrik H. Andreasson and R. Triebel nd A. Lilienthal. Vision based interpolation of 3d laser scans. In *Proc. of the Int. Conf. on Autonomous Robots and Agents*, pages 455–460, New Zeland, 2006.

[32] D. Hahnel, R. Triebel, W. Burgard, and S. Thrun. Map building with mobile robots in dynamic environments. In *Proc. IEEE Int. Conf. Robot. Automat.*, pages 1557–1563, Taipei, Sep. 2003.

[33] C. G. Harris and M. Stephens. A combined corner edge detector. In *Proc. Alvey Vision Conf.*, pages 189–192, Manchester, Aug. 1988.

[34] A. Harrison and P. Newman. Image and sparse laser fusion for dense scene reconstruction. In *Proc. of the Int. Conf. on Field and Service Robotics*, Cambridge, jul 2009.

[35] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, 2nd edition, 2004.

[36] V. Hedau, D. Hoiem, and D. Forsyth. Recovering the spatial layout of cluttered rooms. In *Proc. IEEE Int. Conf. Comput. Vision*, Kyoto, Oct. 2009.

[37] G. Hetzel, B. Leibe, P. Levi, and B. Schiele. 3d object recognition from range images using local feature histograms. In *Proc. 15th IEEE Conf. Comput. Vision Pattern Recog.*, Kauai, Dec. 2001.

[38] Derek Hoiem, Alexei A. Efros, and Martial Hebert. Geometric context from a single image. In *Proc. IEEE Int. Conf. Comput. Vision*, Beijing, Oct. 2005.

[39] P. Newman I. Posner, M. Cummins. Fast probabilistic labeling of city maps. In *Robotics: Science and Systems IV*, Zurich, Jun. 2008.

[40] V. Ila, J. M. Porta, and J. Andrade-Cetto. Information-based compact Pose SLAM. *IEEE Trans. Robot.*, 26(1):78–93, Feb. 2010.

[41] A. Johnson and M.Hebert. Using spin images for efficient object recognition in cluttered 3d scenes. *IEEE Trans. Pattern Anal. Machine Intell.*, 21(1):433–449, May 1999.

[42] Dar-Shyang Lee Jonathan, Jonathan J. Hull, and Berna Erol. A bayesian framework for gaussian mixture background modeling. In *Proc. Int. Conf. on Image Processing*, 2003.

[43] J. Joung, K. An, J. Kang, M. Chung, and W. Yu. 3d environment reconstruction using modified color icp algorithm by fusion of a camera and a 3d laser range finder. In *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, pages 3082–3088, Saint Louis, Oct. 2009.

[44] J.A.P. Kjellander and M. Rahayem. Planar segmentation of data from a laser profile scanner mounted on an industrial robot. *The Int. Journal of Adv. Manuf. Tech.*, 45(1):181–190, 2009.

[45] K. Klasing, D. Althoff, D. Wollherr, and M. Buss. Comparison of surface normal estimation methods for range sensing applications. In *Proc. IEEE Int. Conf. Robot. Automat.*, pages 1977–1982, Kobe, May 2009.

[46] K. Klasing, D. Wollherr, and M. Buss. Realtime segmentation of range data using continuous nearest neighbors. In *Proc. IEEE Int. Conf. Robot. Automat.*, pages 2011–2016, Kobe, May 2009.

[47] K. Lai and D. Fox. 3d laser scan classification using web data and domain adaptation. In *Robotics: Science and Systems V*, Seattle, USA, Jun. 2009.

[48] J.F. Lalonde, N. Vandapel, D. Huber, and M. Hebert. Natural terrain classification using three-dimensional ladar data for ground robot mobility. *J. of Field Rob.*, 23(1):839–861, nov 2006.

[49] I. Laptev and T. Lindeberg. Space-time interest points. In *Proc. IEEE Int. Conf. Comput. Vision*, pages 432–439, Nice, Oct. 2003.

[50] D. C. Lee, M. Hebert, and T. Kanade. Geometric reasoning for single image structure recovery. In *Proc. 23nd IEEE Conf. Comput. Vision Pattern Recog.*, Florida, Jun. 2009.

[51] M. Li, P. Fu, and S. Sun. 3d laser scanning data segmentation based on region dilation strategy. In *Proc. of the 2009 Ninth International Conference on Hybrid Intelligent Systems*, pages 313–316, Washington, 2009.

[52] E.H. Lim and D. Suter. 3d terrestrial lidar classifications with super-voxels and multi-scale conditional random fields. *Computer Aided Design*, 41(10):701–710, 2009.

[53] L. Liu, G. Yu, G. Wolberg, and S. Zokai. Multiview geometry for texture mapping 2d images onto 3d range data. In *Proc. 20th IEEE Conf. Comput. Vision Pattern Recog.*, pages 2293–2300, New York, Jun. 2006.

[54] Y. Liu, R. Emery, D. Chakrabarti, W. Burgard, and S. Thrun. Using EM to learn 3D models of indoor environments with mobile robots. In *Proc. 18th Int. Conf. Machine Learning*, pages 329–336, Williamstown, Jul. 2001.

[55] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, Nov. 2004.

[56] Z. Marton, R. Rusu, D. Jain, U. Klank, and M. Beetz. Probabilistic categorization of kitchen objects in table settings with a composite sensor. In *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, pages 4777–4784, Saint Louis, Oct. 2009.

[57] F. Maurelli, D. Droeschel, T. Wisspeintner, S. May, and H. Surmann. A 3D Laser Scanner System for Autonomous Vehicle Navigation. In *Proc. of the 14th Int. Conf. on Advanced Robotics*, 2009.

[58] A.S. Mian, M. Bennamoun, and R. Owens. Three-dimensional model-based object recognition and segmentation in cluttered scenes. *IEEE Trans. Pattern Anal. Machine Intell.*, 28(10):1584–1601, Oct. 2006.

[59] S. Mohottala, S. Ono, M. Kagesawa, and K. Ikeuchi. Fusion of a camera and a laser range sensor for vehicle recognition. In *Proc. IEEE CVPR Workshops*, pages 16–23, Florida, Jun. 2009.

[60] G. Monteiro, C. Premebida, P. Peixoto, and U. Nunes. Tracking and classification of dynamic obstacles using laser range finder and vision. In *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Beijing, Oct. 2006.

[61] F. Moosmann, O. Pink, and C. Stiller. Segmentation of 3d lidar data in non-flat urban environments using a local convexity criterion. In *IEEE Intelligent Vehicles Symposium*, 2009.

[62] O.M. Mozos, R. Kurazume, and T. Hasegawa. Multi-part people detection using 2d range data. *Int. Journal of Social Robotics*, pages 1–10, 2007.

[63] O.M. Mozos, C. Stachniss, and W. Burgard. Supervised learning of places from range data using adaboost. In *Proc. IEEE Int. Conf. Robot. Automat.*, pages 1730–1735, Barcelona, Apr. 2005.

[64] D. Munoz, N. Vandapel, and M. Hebert. Directional associative markov network for 3-d point cloud classification. In *4th Int. Symposium on 3D Data Processing, Visualization and Transmission*, June 2008.

[65] P. Nunez, P. Drews, R. Rocha, M. Campos, and J. Dias. Novelty detection and 3d shape retrieval based on gaussian mixture models for autonomous surveillance robotics. In *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, pages 4724–4730, Saint Louis, Oct. 2009.

[66] P. Nunez, P. Drews Jr, R. Rocha, and J. Dias. Data fusion calibration for a 3d laser range finder and a camera using inertial data. In *Proc. European Conf. Mobile Robotics*, Dubrovnik, Sep. 2009.

[67] A. Ortega, B. Dias, E. Teniente, A. Bernardino, J. Gaspar, and Juan Andrade-Cetto. Calibrating an outdoor distributed camera network using laser range finder data. In *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, pages 303–308, Saint Louis, Oct. 2009.

[68] A. Ortega, I. Haddad, and J. Andrade-Cetto. Graph-based segmentation of range data with applications to 3d urban mapping. In *Proc. European Conf. Mobile Robotics*, pages 193–198, Dubrovnik, Sep. 2009.

[69] M. Ozuysal, M. Calonder, V. Lepetit, and P. Fua. Fast keypoint recognition using random ferns. *IEEE Trans. Pattern Anal. Machine Intell.*, 32(3):448 –461, mar 2010.

[70] J. Poppinga, N. Vaskevicius, A. Birk, and K. Pathak. Fast plane detection and polygonalization in noisy 3D range images. In *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, pages 3378–3383, Nice, Sep. 2008.

[71] I. Posner, D. Schroeter, and P. Newman. Describing composite urban workspaces. In *Proc. IEEE Int. Conf. Robot. Automat.*, pages 4962–4968, Rome, Apr. 2007.

[72] I. Posner, D. Schroeter, and P. Newman. Online generation of scene descriptions in urban environments. *Robot. Auton. Syst.*, 56(11):901–914, 2008.

[73] S. Pu and G. Vosselman. Knowledge based reconstruction of building models from terrestrial laser scanning data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(6):575 – 584, 2009.

[74] C. Rasmusse. combining laser range, color, and texture cues for autonomous road following. In *Proc. IEEE Int. Conf. Robot. Automat.*, Washington, May 2002.

[75] R.B. Rusu, A. Holzbach, N. Blodow, and M. Beetz. Fast geometric point labeling using conditional random fields. In *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, pages 7–12, Saint Louis, Oct. 2009.

[76] D. Scaramuzza, Ahad Harati, and R. Siegwart. Extrinsic self calibration of a camera and a 3d laser range finder from natural scenes. In *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, San Diego, Nov. 2007.

[77] P. Scovanner, S. Ali, and M. Shah. A 3-dimensional sift descriptor and its application to action recognition. pages 357–360, 2007.

[78] L. J. Skelly and S. Sclaroff. Improved feature descriptors for 3d surface matching. In *Proceedings of the SPIE*, volume 6762, 2007.

[79] I. Stamos and P.K. Allen. 3-d model construction using range and image data. In *Proc. 14th IEEE Conf. Comput. Vision Pattern Recog.*, volume 1, page 1531, Hilton Head, SC, Jun. 2000.

[80] C. Stauffer and W. Grimson. Learning patterns of activity using real-time tracking. *IEEE Trans. Pattern Anal. Machine Intell.*, 22(8):747–757, 2000.

[81] B. Steder, G. Grisetti, M. Van Loock, and W. Burgard. Robust on-line model-based object detection from range images. In *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Saint Louis, Oct. 2009.

[82] D. Steinhauser, O. Ruepp, and D. Burschka. Motion segmentation and scene classification from 3d lidar data. In *Proc. IEEE Intell. Vehicles Symposium*, pages 398–403, June 2008.

[83] S. Stiene, K. Lingemann, A. Nuchter, and J. Hertzberg. Contour-based object detection in range images. In *Proc. of the Third Int. Symposium on 3D Data Processing, Visualization, and Transmission*, pages 168–175, 2006.

[84] D. Streller, K. Furstenberg, and K. Dietmayer. Vehicle and object models for robust tracking in traffic scenes using laser range images. In *Intelligent Transportation Systems*, pages 118–123, 2002.

[85] Q. Sumin and H. Xianwu. Hand tracking and gesture gecogniton by anisotropic kernel mean shift. In *Proc. Int. Conf. on Neural Networks and Signal Processing*, pages 581–585, 2008.

[86] Hartmut Surmann, Andreas Nüchter, and Joachim Hertzberg. An autonomous mobile robot with a 3d laser range finder for 3d exploration and digitalization of indoor environments. *Robot. Auton. Syst.*, 45(3-4):181–198, 2003.

[87] T. Svoboda, D. Martinec, and T. Pajdla. A convenient multicamera self-calibration for virtual environments. *Presence:Teleop. Virtual Env.*, 14(4):407–422, 2005.

[88] A. Talukder, R. Manduchi, A. Rankin, and L. Matthies. Fast and reliable obstacle detection and segmentation for cross-country navigation. 2:610–618, June 2002.

[89] S. Thrun. Robotic mapping: A survey. In G. Lakemeyer and B. Nebel, editors, *Exploring Artificial Intelligence in the New Millenium*, World Scientific Series in Robotics and Intelligent Systems. Morgan Kaufmann, 2002.

[90] R. Triebel, W. Burgard, and F. Dellaert. Using hierarchical EM to extract planes from 3D range scans. In *Proc. IEEE Int. Conf. Robot. Automat.*, pages 4437–4442, Barcelona, Apr. 2005.

[91] R. Triebel, K. Kersting, and W. Burgard. Robust 3d scan point classification using associative markov networks. In *Proc. IEEE Int. Conf. Robot. Automat.*, pages 2603–2608, Orlando, May 2006.

[92] R. Triebel, O.M. Mozos, and W. Burgard. Collective classification for labeling of places and objects in 2d and 3d range data. In *Data Analysis, Machine Learning and Applications: Proc. of the 31st Annual Conference of the Gesellschaft Fur Klassifikation EV, Albert-ludwigs-universitst Freiburg, March 7-9, 2007*, page 293, 2008.

[93] R. Tsai. A versatile camera calibration technique for high accuracy 3d machine vision metrology using off-the-shelf tv cameras. *IEEE J. Robot. Automat.*, 3(4):323–344, Aug. 1987.

[94] R. Unnikrishnan and M. Hebert. Fast extrinsic calibration of a laser rangefinder to a camera. Technical Report CMU-RI-TR-05-09, Robotics Institute, Pittsburgh, July 2005.

[95] R. Valencia, E.H. Teniente, E. Trulls, and J. Andrade-Cetto. 3D mapping for urban service robots. In *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, pages 3076–3081, Saint Louis, Oct. 2009.

[96] N. Vandapel and M. Hebert. Finding organized structures in 3-d ladar data. *Selected Topics in Electronics And Systems*, 42:161, 2006.

[97] N. Vandapel, D. Huber, A. Kapuria, and M. Hebert. Natural terrain classification using 3-d ladar data. In *Proc. IEEE Int. Conf. Robot. Automat.*, pages 5117–5122, New Orleans, Apr. 2004.

[98] C-C. Wang, C. Thorpe, and S. Thrun. Online simultaneous localization and mapping with detection and tracking of moving objects: theory and results from a ground vehicle in crowded urban areas. In *Proc. IEEE Int. Conf. Robot. Automat.*, volume 1, pages 842–849, Taipei, Sep. 2003.

[99] Z. Wang, S. Huang, and G. Dissanayake. D-SLAM: A decoupled solution to simultaneous localization and mapping. *Int. J. Robot. Res.*, 26(2):187–204, 2007.

[100] G. Winkler. *Image Analisys Random Fields and Markov Chain Monte Carlo Methods A Mathematical Introduction*. 2003.

[101] K.M. Wurm, R. Kümmerle, C. Stachniss, and W. Burgard. Improving robot navigation in structured outdoor environments by identifying vegetation from laser data. In *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Saint Louis, Oct. 2009.

[102] Q. Yang, R. Yang, J. D., and D. Nister. Spatial-depth super resolution for range images. In *Proc. 21st IEEE Conf. Comput. Vision Pattern Recog.*, Minneapolis, Jun. 2007.

[103] T. Yokoya, T. Hasegawa, and R. Kurazume. Calibration of distributed vision network in unified coordinate system by mobile robots. In *Proc. IEEE Int. Conf. Robot. Automat.*, pages 1412–1417, Pasadena, May 2008.

[104] Q. Zhang and R. Pless. Extrinsic calibration of a camera and laser range finder (improves camera calibration). In *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, pages 725–728, Saint Louis, Oct. 2009.

[105] Z. Zhang. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Machine Intell.*, 22(11):1330–1334, 2000.

Acknowledgements

This research is supported by the Mexican Council of Science and Technology (CONACyT) and conducted at the Institut de Robótica i Informática Industrial (IRI), a Joint University Institute of the Universitat Politécnica de Catalunya (UPC) and the Consejo Superior de Investigaciones Científicas (CSIC). Our work is embarked within the national projects PAU (DPI-2008-0622), and MIPRCV Consolider-Ingenio 2010.