Dynamic object detection fusing LIDAR data and images

Institut de Robòtica i Informàtica Industrial

Agustin Ortega and Juan Andrade-Cetto
Institut de Robòtica i Informàtica Industrial, CSIC-UPC,Barcelona,Spain

CSIC — Universitat Politècnica de Catalunya BarcelonaTech

We present a method to segment dynamic objects on point clouds using images and 3D laser data. Per-pixel background classes are adapted online as Gaussian Mixtures independently for each sensor. The learned classes are fused labeling pixels/voxels that belong to either the background, or the dynamic objects We pay special attention in the calibration and synchronization modules to reach accuracy in registration and data association. We show results of people segmentation in indoor scenes using a Velodyne sensor at a high frame-rate .

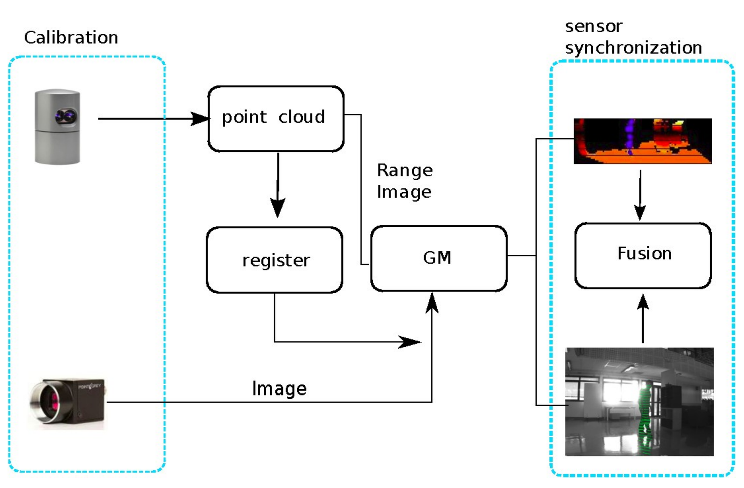
Problem definition

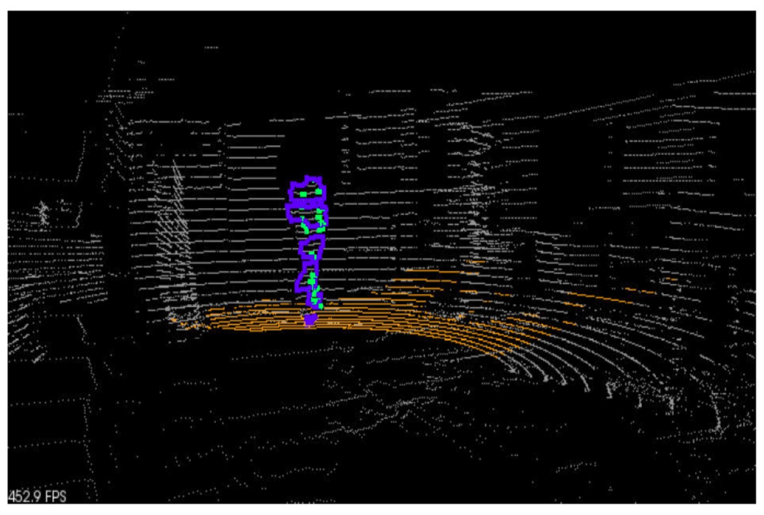
We address the problem of eliminating moving bodies from high acquisition rate 3D range scans using image data. This work is an extension of the work presented in [1].

Range points are classified according to temporally consistent corresponding images. Per pixel class distributions are learned on-line as Gaussian Mixture [2].

We use a fusion approach for range and intensity images inspired by the adaptive Mixture of Local Experts (MLE) architecture[4].

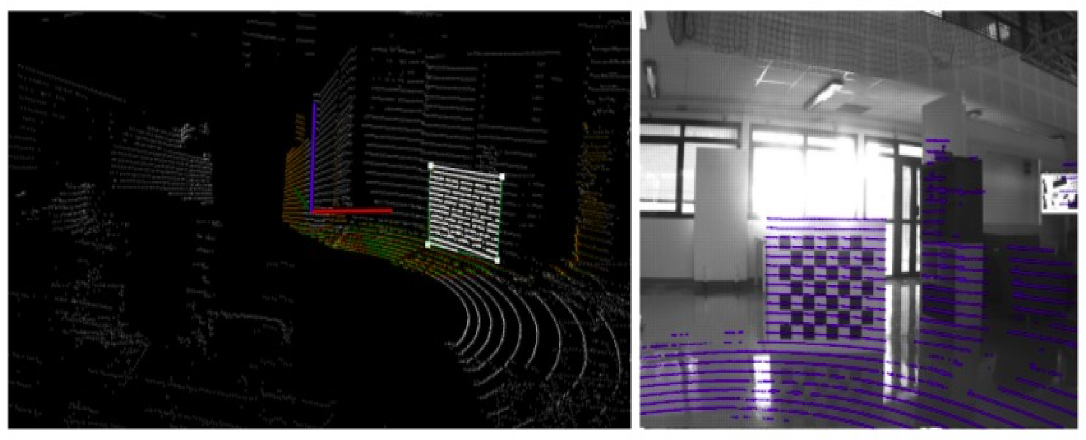
Approach





Sensor calibration

We perform laser-camera extrinsic calibration selecting image and 3D points on the point cloud and computing the camera pose estimation using Lu's method [3] from these correspondences.



Sensor synchronization

Synchronization method named Approximate Time Policy algorithm in ROS is used. The method compares the timestamps between consecutive point clouds and that of the images to align the two data streams. Timestamp between consecutive point clouds is t_{cloudi}, and for the images t_{framej} then compare and set to lie within a reasonable threshold $Ts=1/10$ that is in milliseconds then we have the following:

$$|t_{\mathrm{cloud}_i} - t_{\mathrm{frame}_j}| \leq T_s$$

Segmentation of moving objects

The probability of each image pixel to be of the static object class is modeled as a weighted mixture of K Gaussian distributions[2].

$$P(x|\mathcal{S}) = \sum_{k=0}^{K} \omega_k \eta\left(\mu_k, \Sigma_k\right)$$

This mixture is iteratively refined from image data

$$\begin{aligned} \omega_k(t+1) &= (1-\alpha)\omega_k(t) + \alpha \\ \mu_k(t+1) &= (1-\rho)\mu_k(t) + \rho x \\ \Sigma_k(t+1) &= (1-\rho)\Sigma_k(t) + \rho(x-\mu(t))^T(x-\mu(t)) \end{aligned} \qquad \rho = \alpha\eta\left(\mu_k, \Sigma_k\right)$$

Static class

We choose as the static object class, the first $S<K$ ordered distributions which add up to a factored weight ω_s, where

$$S = \mathrm{argmin}_S\left(\sum_{i=1}^{s} \omega_i \geq \omega_s\right)$$

Point classification

For each point in the image, its corresponding pixel values for the subset of matching synchronized images is measured and compared to the full set of pixel values for the whole sequence by means of the learned mixture distribution.

A range point is classified as static if all matching pixels in the synchronized image subset are within the learned static class.

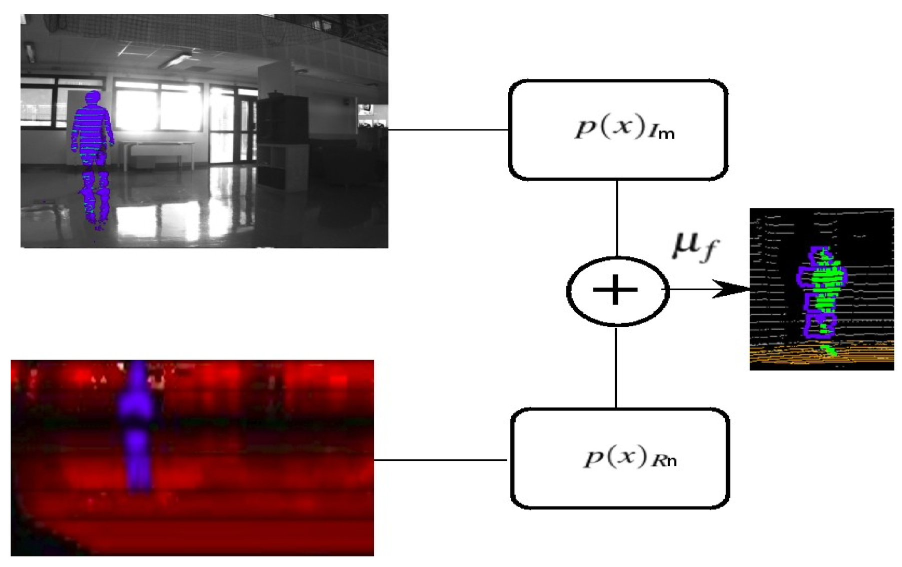
Fusion approach

Our proposed fusion scheme is inspired by the adaptive (MLE) architecture [4].

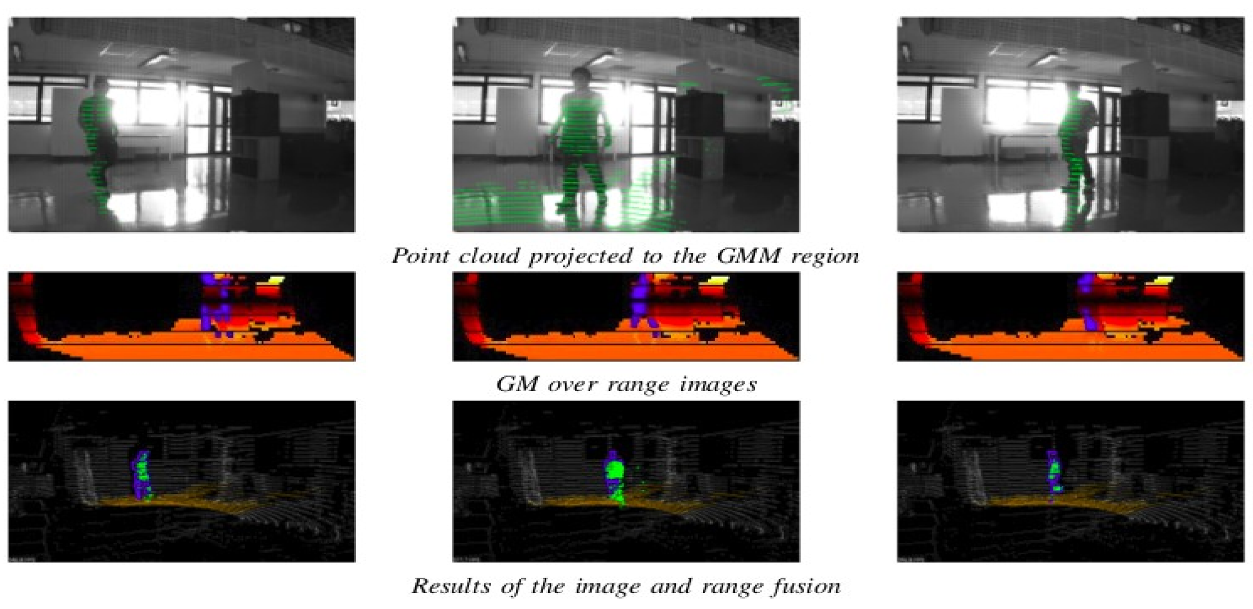
$$\mu_f = \sum_{j=0}^{i} g_j(x) \dot{p}_j(x).$$

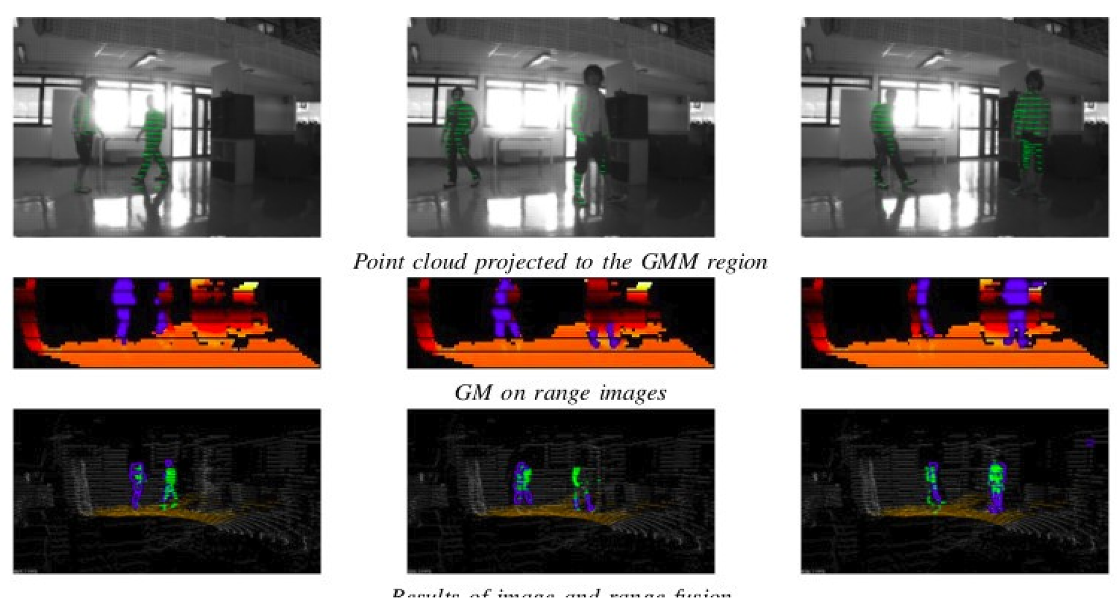
where p is the probability of detection, j ={Rn,Im} for Rn range image, and intensity image Im probabilities, and g is our learned weighting function. In probabilistic form as the following:

$$p(y|x,\alpha) = \sum_{i=0}^{i} g_j(x) p(y|x,\alpha_i)$$



Experiments





Conclusions

We present a method to segment dynamic objects from 3D point clouds using 3D scans and a camera. We use data fusion based in GMM and the scheme proposed here is inspired by the adaptive Mixture of Local Experts.

The method has been tested in a static indoor scenario that contents moderate dynamic object in the scene. We can apply this approach in SLAM or 3D reconstruction to improve the results eliminating spurious points as dynamic.

Acknowledgments

This work acknowledges support from a PhD scholarship from the Mexican Council of Science and Technology (CONACYT) and a Research Stay Grant from the Agencia de Gestio d'Ajuts Universitaris i de Recerca (AGAUR) of The Generalitat of Catalonia (2012 CTP00013) for A. Ortega as well as the Spanish Ministry of Economy and Competitiveness Project PAU+ (DPI-2011-27510).

The authors thanks S. Lacroix and A. Maligo for their help with the facilities and experiments, and to the Laboratoire d'Analyse et d'Architecture des Systemes(LAAS) - CNRS, Toulouse, France for the research stay opportunity.

References

[1] A. Ortega, J. Andrade-Cetto,Segmentation of dynamic objects from laser data, 5th ECMR, pp. 115-121,, Orebro, Sweden, Sept. 2011
[2]C. Stauffer and W. Grimson. Learning patterns of activity using real-time tracking. IEEETrans. Pattern Anal. Mach. Intell., 22(8):747–757, 2000
[3]C.P. Lu, G.D. Hager, and E. Mjolsness. Fast and globally convergent pose estimation from video images. IEEE Trans. Pattern Anal. Mach. Intell., 22:610–622, 2000.
[4]L. Spinello and K. O. Arras. Leveraging RGB-D data: Adaptive fusion and domain adaptation for object detection. Proc. IEEE Int. Conf. Robotics Autom., Saint Paul, May 2012.