Urban Traffic Monitoring from Aerial LIDAR Data with a Two-Level Marked Point Process Model

Attila Börcs and Csaba Benedek
Distributed Events Analysis Research Laboratory, Computer and Automation Research Institute
H-1111, Budapest, Kende utca 13-17, Hungary, \{borcs,bcsaba\}@sztaki.hu

Abstract

In this paper we present a new model for joint extraction of vehicles and coherent vehicle groups in airborne LIDAR point clouds collected from crowded urban areas. Firstly, the 3D point set is segmented into terrain, vehicle, roof, vegetation and clutter classes. Then the points with the corresponding class labels and intensity values are projected to the ground plane, where the optimal vehicle and traffic segment configuration is described by a Two-Level Marked Point Process (L²MPP) model of 2D rectangles. Finally, a stochastic algorithm is utilized to find the optimal configuration.

1. Introduction

Automatic traffic monitoring is a central goal of urban traffic control, environmental protection and aerial surveillance applications. Complex traffic analysis needs a hierarchical modeling approach: at low level individual vehicles should be detected and separated, meanwhile at a higher level we need to extract coherent traffic segments, by identifying groups of corresponding vehicles, such as cars in a parking lot, or a vehicle queue waiting in front of a traffic light. In this paper, we introduce a joint probabilistic model for vehicle detection and traffic segmentation in airborne LIDAR data, which contains point position, intensity and echo information.

LIDAR based vehicle detection methods in the literature follow generally either a grid-cell- or a 3D point-cloud-analysis-based approach [6]. We propose a hybrid model here, where the initial point cloud is classified via 3D features, but the optimal object configuration is extracted in a 2D lattice, after ground plane projection. We model a traffic scene by a Marked Point Process (MPP) [3], which is an efficient Bayesian tool to characterize object populations, through jointly describing individual objects by various data terms, and using information from entity interactions by prior geometric constraints. However, conventional MPP models offer limited options for hierarchical scene modeling, since they usually exploit pairwise object interactions, which are defined on fixed symmetric object neighborhoods. In a traffic situation we often find several groups of regularly aligned vehicles, but we must also deal with junctions or skewed parking places next to the roads (Fig. 3), where many differently oriented cars appear close to each other. In addition, the coherent car groups may have thin, elongated shapes, therefore concentric neighborhoods are less efficient. For this reason, we propose here a Two-Level MPP (L²MPP) model, which partitionates the complete vehicle population into vehicle groups, called traffic segments, and extracts the vehicles and the optimal segments simultaneously by a joint energy minimization process. Object interactions are differently defined within the same segment and between two different segments, implementing adaptive object neighborhoods. This model extends our single level MPP method [2] proposed for vehicle detection. In addition, we present here an improved point cloud segmentation algorithm, and provide a detailed quantitative evaluation on four datasets of 471 vehicles, considering two reference methods [4, 5].

2. Point Cloud Preprocessing

We have developed a Markov Random Field (MRF) model for point cloud segmentation, which utilizes various 3D descriptors. For featuring the terrain class, we estimate the dominant plane of the input cloud using the RANSAC algorithm, and calculate the distance of each point from this plain. Regarding the roof class, we assume that roof points form large connected regions of the cloud, which are composed of segments with uniform surface normals. Local point cloud density is also calculated to recognize sparse clutter regions (like most...
3. L²-Marked Point Process Model

The inputs of this step are the label and intensity maps over the pixel lattice $S$, which were extracted in the previous section. The detection is mainly based on the label map, but additional evidences are extracted from the intensity image, where several cars appear as salient bright blobs due to their shiny surfaces. We assume that each vehicle $u$ can be approximated from top view by a rectangle, which is described by five parameters: $c_u$ and $e_\theta$ center coordinates in the lattice $S$, $e_L$, $e_t$ side lengths and $\theta$ orientation (Fig. 1(c)). Let be $R_u \subset S$ the set of pixels corresponding to $u$. Note that with replacing the rectangle shapes for paralleleograms, the “shearing effect” of moving vehicles may also be modeled [6], but in the considered test data this phenomenon could not be reliably observed. Let $H$ be the space of $u$ objects. We define a neighborhood relation $\sim$ in $H$: $u \sim v$ iff the distance of the object centers is smaller than a threshold. We describe the scene by a Two-level Marked Point Process (L²-MPP) model: a global configuration $\omega$ is a the set of $k$ traffic segments, $\omega = \{\psi_1, \ldots, \psi_k\}$, where each traffic segment $\psi_i (i = 1 \ldots k)$ is a configuration of $n_i$ vehicles, $\psi_i = \{u_1^i, \ldots, u_{n_i}^i\} \in \mathcal{H}^{n_i}$. Here we prescribe that $\psi_i \cap \psi_j = \emptyset$ for $i \neq j$, while the $k$ set number and $n_1, \ldots, n_k$ set cardinality values may be arbitrary (and initially unknown) integers. We mark with $u \sim \omega$ if $u$ belongs to any $\psi_i \in \omega$, i.e. $\exists \psi_i \in \omega : u \in \psi_i$. $\Omega$ denotes the space of all the possible $\omega$ global configurations. Taking an inverse approach, an energy function $\Phi(\omega)$ is defined, which can evaluate each $\omega \in \Omega$ configuration based on the observed data and prior knowledge. Therefore, the energy can be decomposed into a data term and a prior term: $\Phi(\omega) = \Phi_d(\omega) + \Phi_p(\omega)$, and the optimal $\omega$ is obtained by minimizing $\Phi(\omega)$.

3.1 Data-dependent energy terms

Data terms evaluate the proposed $u$ vehicle candidates based on the input label- or intensity maps, but independently of other objects of the population. The data modeling process consists of two steps. First, we define different $f(u) : H \rightarrow \mathbb{R}$ features which evaluate a vehicle hypothesis for $u$ in the image, so that ‘high’ $f(u)$ values correspond to efficient vehicle candidates. The following features are utilized by our model:

- $f^{vo}(u)$ vehicle evidence feature: the ratio of the number of vehicle classified pixels within the proposed rectangle $R_u$ of object $u$ (see Fig. 1(d)-(e))
- $f^{eb}(u)$ external background feature: the ratio of background classified pixels around the proposed rectangle, within the $T^{(i)}$ external regions of Fig. 1(d)
- $f^{i}(u)$ intensity feature: ratio of the vehicle colored pixels within $R_u$ (Fig. 1(f))

In the second step, we construct $\varphi_d(u)$ data driven energy subterms for each feature $f$, by attempting to satisfy $\varphi_d(u) < 0$ for real objects and $\varphi_d(u) > 0$ for false candidates. For this purpose, we project the feature domain to $[-1,1]$ with a monotonously decreasing nonlinear $Q(f, d^d_f)$ function [3, 2], whose zero value is equal to parameter $d^d_f$. With other words, $d^d_f$ is the object acceptance threshold for feature $f$, which can be set based on manually annotated training data in a straightforward way. Once we obtained the $\varphi_d(u)$ subterms, the joint data energy of object $u$ is derived as $\varphi_d(u) = \max(\min(\varphi_d^{vo}(u), \varphi_d^{eb}(u), \varphi_d^{i}(u)))$. 
Alternate the within the \( \omega \) where

\[ \hat{A} \]

large

\[ \psi \]

secondly

\[ u \]

\[ d \]

\[ \sim \]

\[ R \]

\[ \hat{d} \]

\[ \delta \]

\[ \varphi \]

\[ \beta \]

\[ \hat{d} \]

\[ \omega \]

\[ \delta \]

\[ \text{false positive objects.} \]

\[ \text{considering the} \]

\[ \text{Angle constraints. We have} \]

\[ \{ \}

\[ \} \]

\[ \}

\[ \}

\[ \] \]

\[ \} \]

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}

\[ \}
Table 1. Obj. and pix. level F-rates (in %) by the DP [4], hX [5] and the proposed \(L^2\)MPP (\(2M\)) methods, and the Group Classification Rate (GR) of the \(L^2\)MPP model.

<table>
<thead>
<tr>
<th>Set</th>
<th>NV*</th>
<th>Object level %</th>
<th>Pixel level %</th>
<th>GR %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>DP</td>
<td>hX</td>
<td>(2M)</td>
</tr>
<tr>
<td>#1</td>
<td>78</td>
<td>78</td>
<td>68</td>
<td>96</td>
</tr>
<tr>
<td>#2</td>
<td>91</td>
<td>90</td>
<td>93</td>
<td>98</td>
</tr>
<tr>
<td>#3</td>
<td>132</td>
<td>70</td>
<td>74</td>
<td>83</td>
</tr>
<tr>
<td>#4</td>
<td>170</td>
<td>85</td>
<td>87</td>
<td>89</td>
</tr>
<tr>
<td>All</td>
<td>471</td>
<td>83</td>
<td>82</td>
<td>91</td>
</tr>
</tbody>
</table>

*NV = Number of real Vehicles in the test set

Figure 3. Detection result with four clusters. Vehicles of different segments are displayed with different colors, background is interpolated for visualization.

For comparison, we have selected a grid-cell-based algorithm from [4], called DEM-PCA (DP); and a recent state of the art method [5], which uses h-maxima (hX) transform followed by watershed segmentation. Some qualitative results are shown in Fig. 3 and 4 (best viewed in color), and the quantitative evaluation is provided in Table 1. Since the reference methods do not deal with vehicle grouping, only the car detection rates are compared: the proposed \(L^2\)MPP model surpasses the references both at object and at pixel levels.

6. Conclusions and future work

This paper has proposed a novel Two-Level MPP model for joint extraction of vehicles and traffic segments in aerial point cloud data. The efficiency of the approach has been tested with real-world LIDAR measurements, and its advantages versus two reference methods have been demonstrated. Note that in the proposed model, the vehicles are grouped based on similar orientation, but we have experienced that the method can deal with car groups on slightly curved roads as well. As future work, we plan to extend the prior terms of our method to handle more complex vehicle arrangement patterns such as strongly curved exit ramps or roundabouts.

References