Static Vehicle Detection and Analysis in Aerial Imagery using Depth

Satwik Kottur[1], Dr. Gerard Medioni[2], and Dr. Jongmoo Choi[2]

[1]Indian Institute of Technology Bombay, India
[2]University of Southern California, United States

Abstract

This report proposes an approach to automatically detect static vehicles in an outdoor parking space using depth. The relevant 3D information is generated from a Digital Surface Model (DSM), which is a result of a novel and existing technique to solve camera pose estimation and dense reconstruction simultaneously. Validation using local 2D features, based on existing methods, is then done to ensure better detection rates. Further, performance of the detection system is evaluated by changing the internal parameterization of 3D model generation and the dependence is analyzed.

1 Introduction

Vehicle detection and tracking is one of the most sought-after research topics in the realm of computer vision due to the huge diversity in applications ranging from surveillance and security to traffic flow analysis and studies. Many interesting approaches have been proposed resulting in great advancements in this area. Vehicle detection in parking spaces is one such topic the current work addresses.

Monitoring parking spaces has a great importance in both civilian and surveillance applications. Using aerial imagery to accomplish this task is even more challenging due to the small angular size of vehicles based on the height of the camera. A passenger car has a typical size of around 30x10 pixels in the current setting. However, the advantages of using such images are immense and make it interesting to pursue the idea. One of the most crucial advantage being the wide area of viewing - particularly in tracking for maintaining a clear sight of the target. This report proposes, using existing works ([9], [4]), a simple detection technique to identify and detect vehicles in a parking space along with relevant analysis. A sequence of aerial images taken from an intrinsically calibrated camera is the input to the pipeline. Static vehicles in outdoor parking spaces are identified through combination of 3D and 2D features with the help of an intermediate Digital Surface Model (DSM) of the urban landscape.

The report is organized as follows. Section 2 briefly explains the related work that forms the core the present work builds around. Section 3 outlines the entire pipeline of the current work. Section 4 is mainly analysis of the dependence of system performance on internal parameterization of 3D model generation. We present conclusions and scope for improvement in the final section.

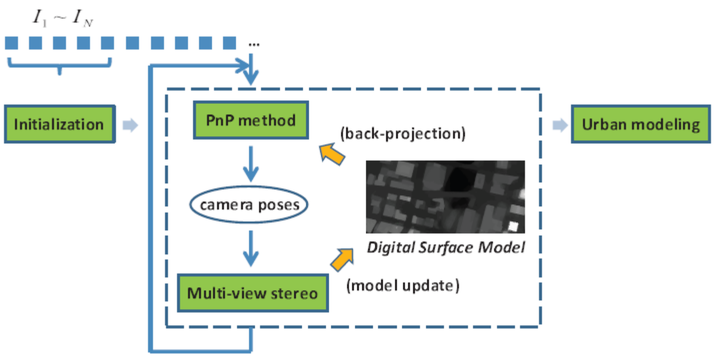

Figure 1: Flowchart for Simultaneous Camera Pose Estimation and Dense Depth Reconstruction

2 Related Work

We briefly go through the underlying framework around which the present work is based on as relevant literature can be referred to for detailed description. Primarily, the following works form the basis :

2.1 Coupled Camera Pose Estimation and Dense Reconstruction

There are mainly two problems that need to be addressed - camera pose estimation, which involves estimation of position and orientation of the camera (6D); and dense reconstruction i.e. establishing a dense depth map. Many solutions were proposed that solve these two problems sequentially : solving the Structure from Motion (SfM) problem using sparse set of 2D matches points followed by multi-view stereo algorithm for the dense depth map.

However, this work largely differs from previous works ([2], [6] ,[8] ,[3], [5]) by its attempt to solve both the problems jointly for aerial videos of urban scenes. This coupled approach seemingly performs better with higher efficiency. As seen from Fig. 1, a 3D model is updated after estimating the camera pose using a new frame as it arrives. Unlike traditional SfM solely based on matched 2D points, Perspective-n-Point (PnP) method dependent on 3D information back-projected from the model is used to estimate the camera pose. Dense point correspondences are then estimated by optical flow using stabilized frames. Such stabilization makes the correspondences robust even for large baselines. Though the camera estimation makes do with

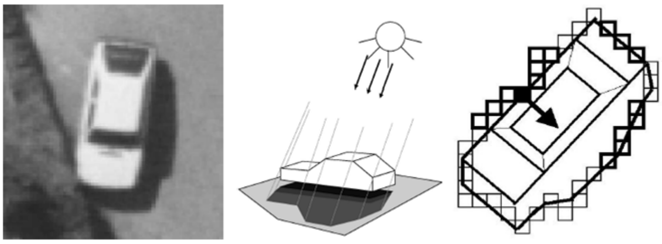

Figure 2: Wire Frame Model for Vehicle Detection

global optimization, estimated poses are as accurate as with global Bundle Adjustment making this method efficient and accurate. A novel representation for urban scenes has also been proposed and used in this work. A Digital Surface Model (DMS) composed of two elements : ground/roofs and building side-walls is maintained and updated during the process.

2.2 Integrating Global and Local Features for Vehicle Detection

Aerial images of urban areas have an apparent closeness of different objects that causes strong influence on each other. For example, partial occlusion of cars due to trees, shadows due to buildings, specularity on cars due to glasses or varnish, etc. A model that completely relies on radiometry might not be efficient due to such influences that mostly appear in form of local radiometric disturbances. However, a model emphasizing a structural description - as an explicit model - seems much more robust. The performance of the implicit approach depends mainly on the training data, which cannot be assured to generalize and capture the changes in illumination. viewpoint and other influences caused by neighboring objects correctly. Another advantage offered by explicit modeling is the focus on fundamental and robust features of cars facilitating the employment of hierarchy of levels of detail.

In [4], an explicit model consisting mainly of geometric features and some radiometric properties is used. A car is modeled as 3D object by a wire-frame representation. The model contains substructures like windshield, roof and hood (Fig. 2). This model is adaptive regarding the the expected saliency of the edge features which is influenced by vehicle color, vehicle orientation and view point (position in the image). Hence, this model is chosen as it offers invariability to varied number of parameters in detection of vehicles.

3 Proposed Algorithm

The detection approach starts with the Digital Surface Model(DSM) obtained from [9] as shown in Fig. 3. The entire pipeline, divided into four stages, is outlined below.

3.1 Generation of Height Map

We begin the detection pipeline with the 3D model of the urban landscape. The model implicitly has a ground plane assigned to the urban scene which is identified through Random Sampling and Consensus (RANSAC) among points with near-ground heights. Under the assumption that the ground plane mainly consists of roads and outdoor parking spaces, we restrict our search to this plane. We back-project

Figure 3: Digital Surface Model(DSM)

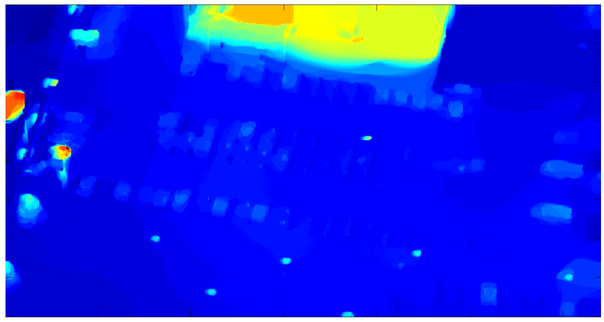

Figure 4: Height Map of the Parking Space

the heights from the Digital Surface Model(DSM) onto the image using the dense set of 2D-3D correspondences calculated intermediately in [9]. We call the height-projected image(Fig. 4) a 'height map' and use it to detect cars from depth information.

3.2 Detection from Depth

The next stage in the pipeline involves detection of possible candidates for static cars using the generated height map. We use a sliding window with a Haar-like feature to score the current location. Characterization of the window is based on the following realistic assumptions:

- **Choice of Feature**
  Modeling a car as a rectangle of raised height from the ground(or equivalently at lower depth than the immediate surroundings as seen from air), we use a bordered rectangular feature for detection. A high scoring match will be expected to have the border overlapping with the ground near the car while raised rectangle overlaps with height of the car.

- **Size of Window**
  Since we have the extrinsic parameters of the camera, the average size of a passenger car in terms of pixels can be estimated. We use a 20x10 pixels feature with a border of 5 pixels. We have observed that the exact size does not affect the detection much as long as the order is consistent.

- **Orientation of Feature**
  As the focus is on detecting static vehicles in an outdoor parking space, there are a limited number of directions of interest along with vehicles are oriented. We assume cars from the same row are parked in an orderly fashion along a single direction. However, caution must be applied to note that there can be different but a small number

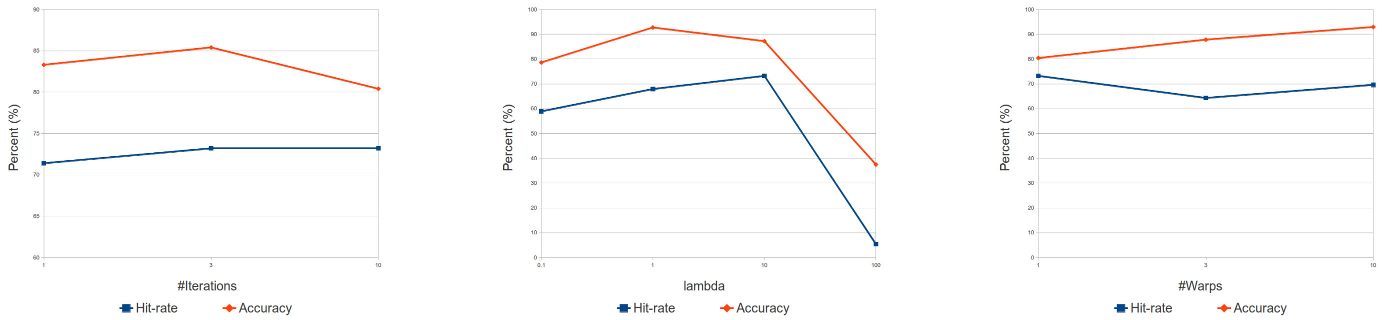

Figure 5: This is some figure side by side

Figure 6: Results of Detection from Depth

Figure 7: Wire Frame Model for Validation

of such directions. Taking advantage of this, we orient the feature in the sliding window and pick the most consistent direction.

- **Location Variation Constraint** Based on the rectangular box model for the car, we impose a local variation constraint. The standard deviation of the depth values in the feature is evaluated and is assigned a negative impact to the overall score. The idea is to suppress objects like lamp posts, etc. which have high score from the feature due to larger height. This constraint enforces more or less constant height within the feature.

Under the above assumptions, we evaluate score at each pixel and select the locations of possible positives(Refer to Fig. 6). The pipeline then proceeds to the next stage for validation.

3.3 Validation using 2D features

The 3D wire frame model (Refer Fig. 7) is employed for validation on the possible locations obtained till the present stage. We project the wire frame on to the image based on the camera position and viewing angle. However, to keep the model simple, we do not consider the shadow cast by the car and restrict ourselves to the geometric model. Calculation of score is done as illustrated in [4] and low scoring positions are discarded.

3.4 Enumeration and Localization

This is the final stage of the entire pipeline. After the validation, we finally count and locate the static vehicles as predicted by the proposed approach.

4 Results

To evaluate the system, we use Brown Dataset [1] to build the Digital Surface Model (DSM) and perform detection and counting. Though aerial images of any outdoor parking space can be fed as input to the system, performance metrics presented in this paper are for Site 3 from the above data. Identifying that optical flow calculation is the decisive step in DSM generation, we analyze the influence of internal parameterization of flow computation on the detection system in terms of functioning and performance. Primarily, we focus on three parameters - number of iterations, smoothness co-efficient and number of warps.

- **Number of Iterations**
  Number of iterations, as in case of any optimization problem, plays a crucial role in obtaining the solution. However, its influence on detection looks subtle. Hit rate slightly improves on increasing the number of iterations. However, accuracy goes down for large number of iterations.

- **Smoothness Constraint**
  Regularization is perhaps the most important consideration in optical flow evaluation. The behavior of the system is heavily dependent on the smoothness co-efficient as expected. Low values of smoothness term lead to noisy depth maps and lower accuracy

[1] `http://vision.lems.brown.edu/project_desc/Object-Recognition-in-Probabilistic-3D-Scenes`

and hit rates. On the other side, high values tend to smoothen the height considerably. This results in lose of information about small height variations as in case of vehicles. Hence, both hit rate and accuracy go down drastically for higher values of regularization co-efficient.

- **Number of Warps**
  Coarse-to-fine warping techniques are generally employed in optical flow calculations based on experimentation. However, [7] attempts a mathematical support for the idea. In our experiment, the dependence of hit rate and accuracy on warping is minimal as they show slight improvement with increasing number of warps.

5 Conclusion

In this report, we put forward a simple detection algorithm to identify static vehicles in an outdoor parking lot. Importantly, identification is based mainly on 3D information in the form of height maps obtained from Digital Surface Model (DSM) of the urban landscape. We also make use of available 2D information for validation and present the final detection results. Subsequently, relationship between the pipeline and internal parameterization of landscape model generation is analyzed and presented. Identifying of empty parking places would be an interesting problem to pursue and solve. Incorporating better features for detection and validation can be explored for better performance of the overall system.

References

[1] R. Hartley B. Triggs, P. McLauchlan and A. Fitzgibbon. Bundle adjustment - a modern synthesis. *Vision algorithms:theory and practice,2000.*

[2] Y. Furukawa and J. Ponce. Accurate camera calibration from multi-view stereo and bundle adjustment. *IJCV,2009.*

[3] Y. Lin H. Liao and G. Medioni. Aerial 3D reconstruction with line-constrained dynamic programming. In *ICCV,2011.*

[4] Stefan Hinz. Integrating local and global features for vehicle detection in high resolution aerial imagery. *ISPRS Archives,2003*, XXXIV(3/W8).

[5] J. Frahm A. Akbarzadeh P. Moredohai B. Clipp C. Engels D. Gallup S. Kim P. Merrell et al M. Pollefeys, D. Nist r. Detailed real-time urban 3D reconstruction from video. *IJCV,2008.*

[6] S. Lovegrove R. Newcombe and A. Davison. Dtam:dense tracking and mapping in real-time. In *ICCV,2011.*

[7] Nils Papenberg Thomas Brox, Andrs Bruhn and Joachim Weickerte. High accuracy optical flow estimation based on a theory for warping. *In Proc. 8th European Conference on Computer Vision, 2004.*

[8] S. M. Seitz Y. Furukawa, B. Curless and R. Szeliski. Manhattan-world stereo. In *CVPR,2009.*

[9] G. Medioni Z. Kang. Online coupled camera pose estimation and dense reconstruction for urban modeling. In *Unpublished.*