Technologies for Safe & Efficient Transportation

THE NATIONAL USDOT UNIVERSITY TRANSPORTATION CENTER FOR SAFETY

Carnegie Mellon University | **UNIVERSITY of PENNSYLVANIA**

Enhanced pedestrian and vehicle detection using surround-view camera system

FINAL RESEARCH REPORT

Prof. Vijayakumar Bhagavatula

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation's University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

Enhanced pedestrian and vehicle detection using surround-view camera system
Prof. Vijayakumar Bhagavatula

Introduction and Motivation

An alarming number of fatalities occur every year due to backover i.e. drivers not noticing objects behind their vehicles while reversing. Due to the increase in the size of the vehicles and the associated blind spots, the number of backover fatalities has been on the rise. According to www.kidsandcars.org, at least 50 children are backed over by cars every week in the US. Further, according to the Bureau of Labor Statistics, over 70 workers died from backover incidents in 2011. In an attempt to reduce the number of such incidents, rearview cameras are becoming standard in many vehicles (in US, 44 percent of 2012 models came with rear cameras standard, according to the automotive research firm Edmunds) to help the drivers see the objects in the rear blind zone. However, according to www.kidsandcars.org, it was found in an Oregon research study that only one in five drivers used a rearview camera when it was available. The objective of this research was to develop automated object detections algorithms for videos captured using rearview cameras. Accurate detection of undesired objects behind a parked vehicle can be used to trigger alarms, which can be useful in reducing the chances of backover accidents. In this report, we summarize the major achievements of this research project.

Data Collection

Data was collected using a high definition GoPro Hero video camera mounted on the rear side of a 2011 Toyota Camry (above the license plate) as shown in Figure 1. The camera captured images of resolution 1280 by 720 at a frame rate of 30 fps. Videos were collected at different locations in Pittsburgh at different times of the day (between 10am and 6pm).

Figure 1: Data Collection Set-up

The objective of this research was to detect the presence of objects, particularly children, behind a parked vehicle. To simulate the situation, child mannequins were used as subjects for data collection. Two mannequins were used and data was collected with different orientations (facing the car, facing away from the car), distances from the rear side of the car (1-3ft) and postures of the mannequins (sitting, standing, lying) under natural conditions, to mimic a real world situation.

Figure 2: Sample Images Captured

Technical Approach

We are given a collection of images depicting a clear background (captured at the time of parking). Our goal is to detect objects of arbitrary shapes and sizes close to the rear side of the vehicle. We use two different algorithms (background subtraction and ground surface classification) to address this problem and combine the predictions from both the frameworks to compute the final prediction. These are detailed below.

1. *Object Detection using Background Subtraction*

Background subtraction is a computer vision technique to detect objects from static cameras. The fundamental idea behind this approach is that of detecting objects from the difference between the current frame and a reference frame called the "background image" or the "background frame". We use the Mixture of Gaussians (MoG) approach for background subtraction in our work. In this strategy, each pixel in the image is modeled using a mixture of Gaussians and an online approximation is used to update the model [3]. The distributions are then analyzed to determine which are most likely to result from a background process. A given pixel is classified as background if the Gaussian distribution that represents it most effectively is a part of the background model and vice versa.

2. *Object Detection using Ground Surface Classification*

The idea behind this approach is to train a classification model to recognize ground surface pixels. Given a test image, each pixel is classified as ground/non-ground to detect the presence of any objects. The SLIC algorithm [1] is used to extract features from superpixels. Superpixels provide much better semantic information and greatly reduce complexity than individual pixel features. Histogram of oriented gradients (HoG) features are extracted from patches centered at the centroid of the superpixels [5]. The patch size is kept larger than the superpixel size, to capture more contextual information. A HoG spatial pyramid feature is constructed by concatenating the HoG features from co-located patches with different sizes. Support Vector Machine (SVM) is used as the underlying classification model. A sample super-pixelized image is depicted in Figure 3.
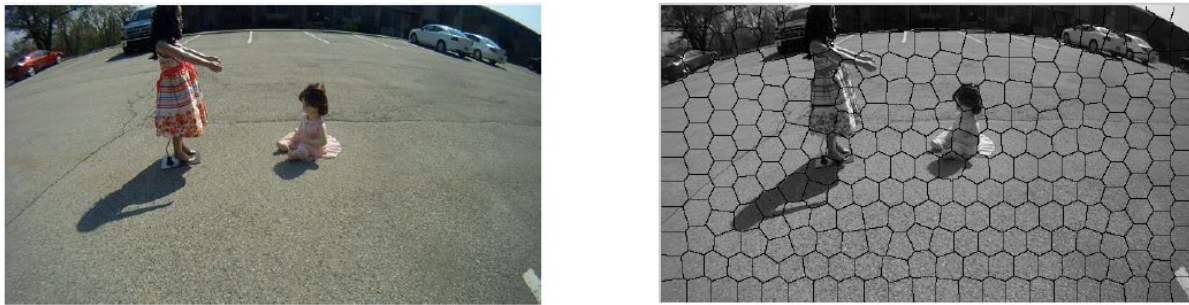
Figure 3: An Image and the Corresponding Super-pixelized Image

3. *Information Fusion*

Given the predictions from the background subtraction and ground surface classification modules, a fusion scheme needs to be adopted to produce the final result. Our empirical results demonstrated that the OR fusion scheme (predicting a pixel to be an object pixel if either of the modalities

classify it as an object pixel) produces the best results. We therefore present the results of this scheme below.

Results:
In this section, we present qualitative and quantitative results of the proposed framework.

Ground Truth Annotation: We use the MIT LabelMe Tool for ground truth annotation of the objects in the images. Sample annotations are shown in Figure 4.


Figure 4: Sample Ground Truth Annotations

Evaluation Metrics: We use three evaluation metrics to analyze the performance of our framework: **(1) Pixel Accuracy** – the percentage of pixels that get predicted correctly; **(2) True Positive Rate (TPR)** – the percentage of object pixels, that get classified as object pixels (higher the better) and **(3) False Positive Rate (FPR)** – the percentage of non-object pixels that get classified as object pixels (lower the better).

Experimental Setup: We present results on the data collected at eleven different locations in the city of Pittsburgh. At each location, 11 video are collected, which gives a total of 121 test cases across the eleven locations. For a particular test case, only the background image from that location is used for training the system (to mimic a real-world situation); images from other locations are also included in the training set.

Qualitative Results: Sample results from our framework are depicted in Figure 5. In each image, a black pixel denotes a non-object pixel and a white pixel denotes an object pixel.

Ground Truth   Prediction
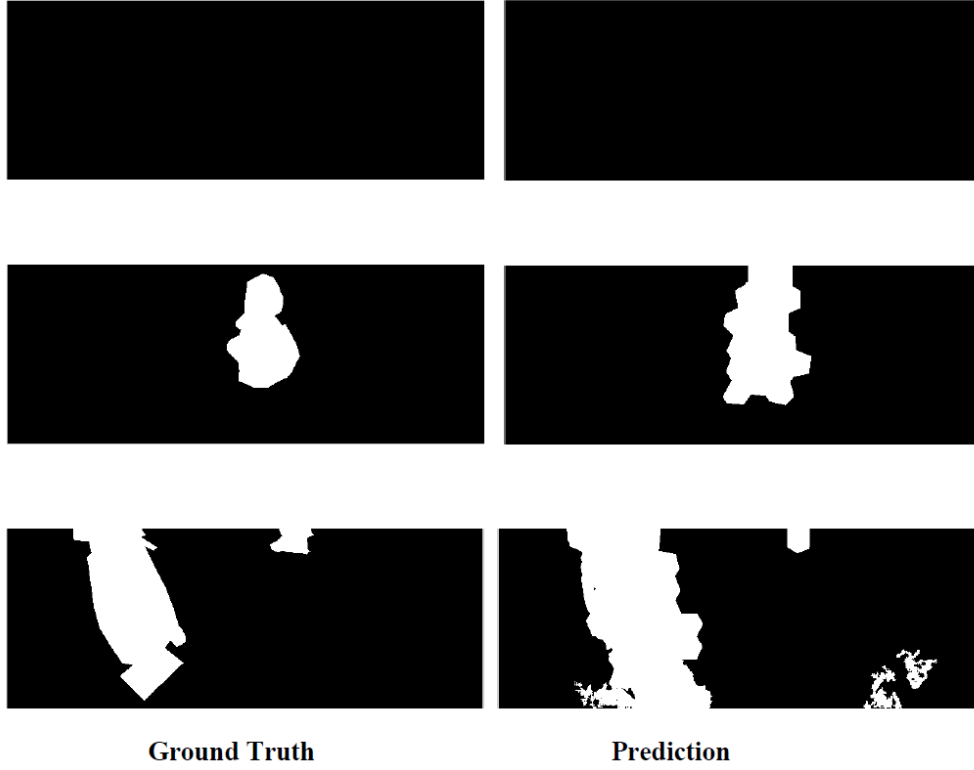
Figure 5: Sample Visual Results

Quantitative Results: We analyze the results separately for each location. At each location, we present the values averaged across all test cases in that location. The results are shown in Table 1 (all values are in pixels and percentages).

Locations	Accuracy	TPR	FPR
Banksville Plaza	87.72	93.78	13.20
China Buffet King	86.85	91.67	13.41
CMU East Campus Garage 1	93.75	90.86	8.52
CMU East Campus Garage 2	89.91	95.18	12.05
Crane Village Aptmnts 1	83.49	94.45	18.51
Days Inn Banksville	80.03	94.08	20.10
Dollar Tree (Banksville)	85.21	91.88	14.36
Eating Park (Banksville)	80.57	94.96	20.51
Mintt Restaurant	89.41	94.29	12.05
Crane Village Aptmnts 2	86.68	93.61	15.09
Tamarind Restaurant	89.11	92.73	12.64
Mean ± sd	86.61 ± 4.10	93.41 ± 1.42	14.59 ± 3.71

Table 1: Pixel Classification Results across Three Test Locations

The proposed framework therefore shows tremendous promise in accurately identifying object and ground pixels.

References
[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua and S. Susstrunk, "*SLIC Superpixels Compared to State-of-the-Art Superpixel Methods,*" IEEE Transactions on Pattern Analysis and Machine Intelligence (**PAMI**), 2012.
[2] M. Jones and J. Rehg, "*Statistical Color Models with Application to Skin Detection,*" IEEE Conference on Computer Vision and Pattern Recognition (**CVPR**), 1999.
[3] C. Stauffer and W. Grimson, "*Adaptive background Mixture Models for Real-time Tracking,*" IEEE Conference on Computer Vision and Pattern Recognition (**CVPR**), 1999.
[4] Q. Wu and B.V.K. Vijaya Kumar, "*Strong Shadow Removal Via Patch-based Shadow Edge Detection,*" IEEE International Conference on Robotics and Automation (**ICRA**), 2012.
[5] Z. Yu, W. Zhang and B.V.K. Vijaya Kumar, "*Robust Rear-View Ground Surface Detection with Hidden State Conditional Random Field and Confidence Propagation,*" IEEE International Conference on Image Processing (**ICIP**), 2014.