

Technologies for Safe & Efficient Transportation

THE NATIONAL USDOT UNIVERSITY
TRANSPORTATION CENTER FOR SAFETY

Carnegie Mellon University

UNIVERSITY of PENNSYLVANIA

Automated Detection of Objects in Rear Camera Images

FINAL RESEARCH REPORT

Prof. Vijayakumar Bhagavatula

Contract No. DTRT-13-GUTC-26

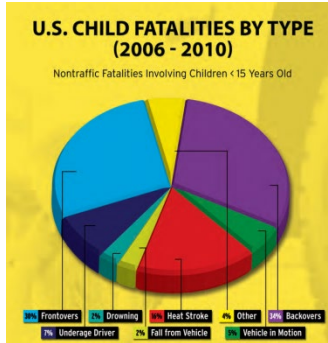
DISCLAIMER

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.

Automated Detection of Objects in Rear Camera Images

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



34% of the U.S. child fatalities were caused due to backovers in 2006-2010
www.kidsandcars.org

are becoming standard 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 the proposed research is 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 immensely useful in reducing the chances of backover accidents.

Current Research

Data Collection: The 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 captures 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 is 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. Sample images are shown in Figure 2.





Figure 2: Sample Images Captured using the Camera

Proposed Solution: 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 technique in the field of computer vision 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.

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 3.



Figure 3: 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 three different locations in the city of Pittsburgh. At each location, 11 video are collected, which gives a total of 33 test cases across the three 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 4. In each image, a black pixel denotes a non-object pixel and a white pixel denotes an object pixel.

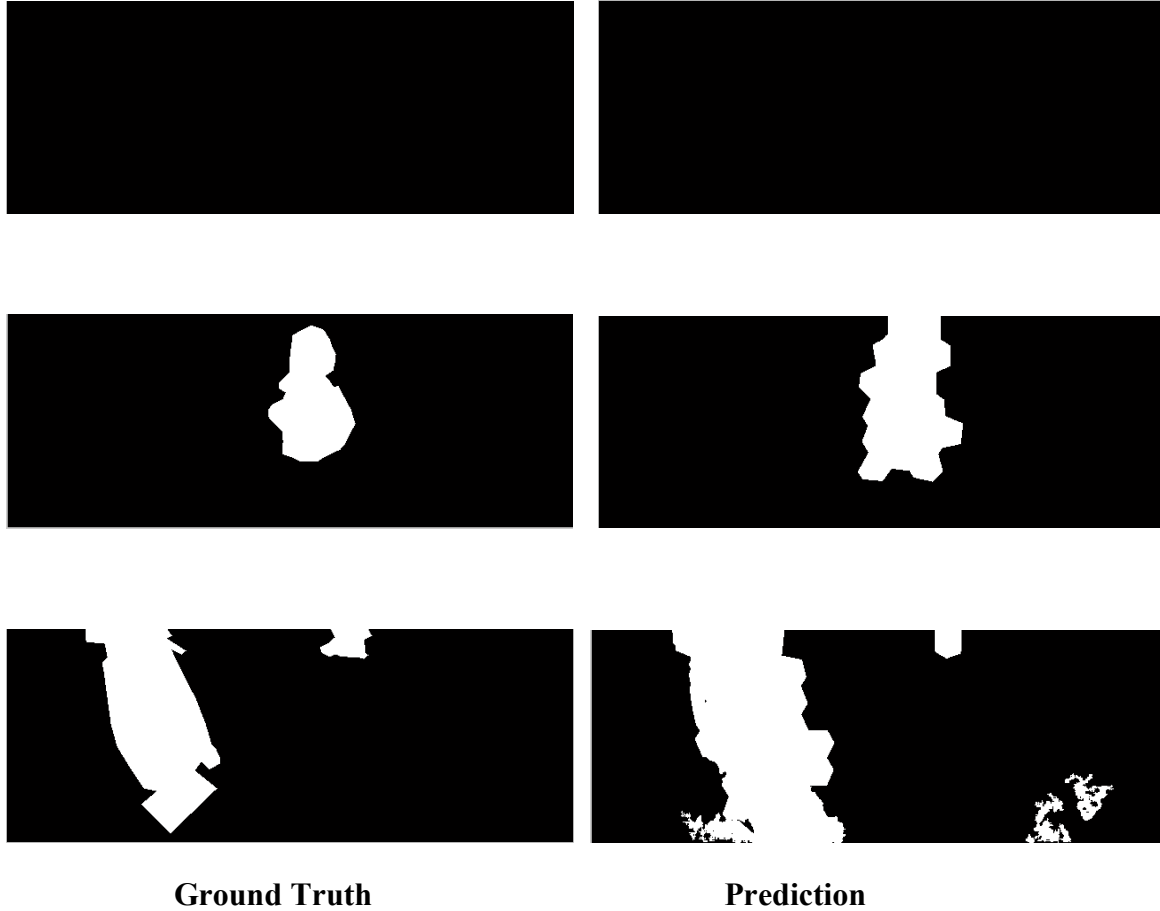


Figure 4: Sample Visual Results

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 tabulated below (all values are in percentages):

	Location 1	Location 2	Location 3
Average Pixel Accuracy	87.14	92.25	89.72
Average TPR	94.67	91.87	90.88
Average FPR	14.50	9.67	11.31

Table 1: Pixel Classification Results across Three Test Locations

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

Future Work

We plan to take up the following topics as part of our future research:

1. Expand the image database by collecting more images at different locations and weather conditions. We will make the dataset and the annotations publicly available to facilitate further research in this topic.
2. The proposed method is developed to detect objects of arbitrary shapes and sizes behind the parked car. We plan to augment our framework to detect the presence of children in particular, so that an alarm can be triggered to alert the drivers. A possible solution is to use a skin color detection algorithm to detect skin colored pixels within the segmented regions [2]. This will be taken up as part of future research.
3. Analysis of our results revealed the fact that most of the pixel-level errors are caused by illumination changes. As part of our future work, we plan to address the effect of illumination in detecting objects. We intend to exploit existing research on shadow removal for this purpose [4].
4. An interesting direction of future research is to predict the posture of the children (sitting/standing/lying) and its distance from the rear end of the car, besides predicting merely its presence. We plan to take this up as part of our future work.

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.