Real-time Pedestrian Detection in Consecutive Frames

Andrey Zaytsev, Jerry Guo, Shubhendra Chauhan, Yiming Wang
Department of Computer Science, College of Engineering, University of Illinois at Urbana-Champaign

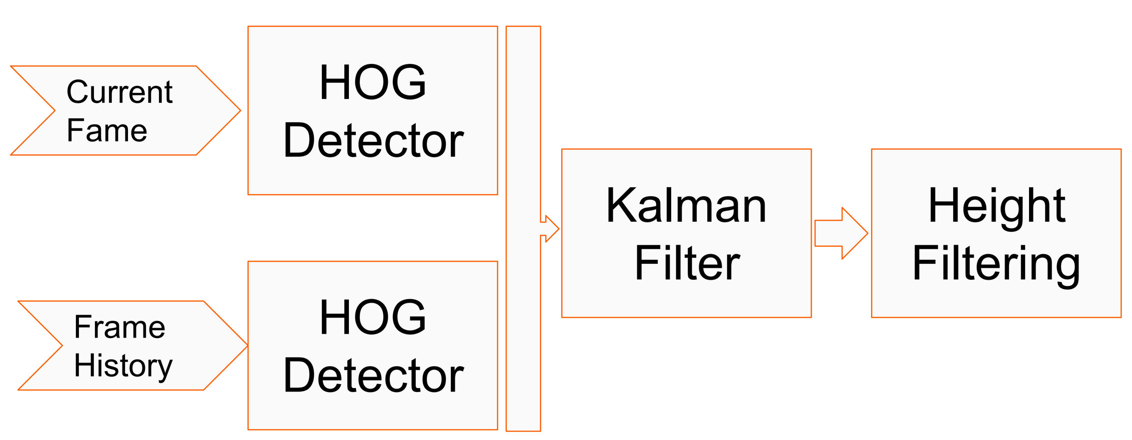
INTRODUCTION

Example output of a pedestrian detector. The detected pedestrians are enclosed in yellow boxes with a provided score by the HOG detector.

The problem of pedestrian detection is a vital part of autonomous driving. Cars need to know where the pedestrians are not only for the purposes of safety, but also for tasks such as picking passengers up, getting background information about the environment, as well as getting more accurate location information.
For instance, if there are a lot of people standing on the curb, one may infer that the car is driving through an airport. In addition, seeing a lot of pedestrians cross the road in front of the car allows us to conclude that there must be a red light even if it is obscured by some other objects.

Challenges
In contrast to a static camera, where people can be easily detected using any trivial background subtraction algorithms, pedestrian detection is a more complex problem. Since a camera mounted on a car is constantly moving, methods of detecting moving objects perform poorly. One needs to combine the information contained in individual frames with the video sequence to detect people.

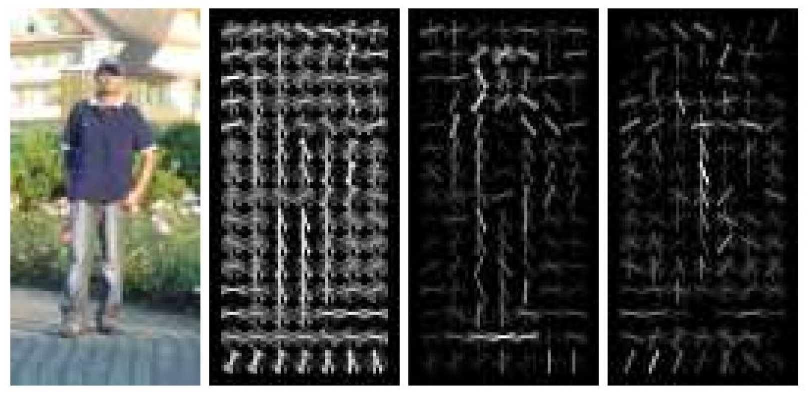
APPROACH



In order to improve the accuracy of a static person detector, we make use of the continuity of the frames in a video. We use Histogram of Oriented Gradients (HOG) person detector as our base.
In addition to that, we use object tracking to avoid losing objects that were previously detected. This is implemented through Kalman filters.
In order to filter out false positives, we use automatic height estimation, which is based on Vanishing Points (VPs) and a derived horizon line.

METHOD

Histogram of Oriented Gradients (HOG) Detector



We used HOG pedestrian detector as the baseline. The detector uses the distribution of intensity gradients or edge directions to describe local object appearance and shape. It then counts the occurrence of gradient orientations in localized portions of an image. The output feature vectors are fed to an SVM classifier. HOG detector is particularly suitable for human detection in images since it operates on local cells and is invariant to geometric and photometric transformations.

Object Tracking

Kalman filter is used to track the detected objects. It has two phases: 'predict' and 'update'. In the predict phase, we make an estimate of state using process model. In the update phase, we combine current prediction and current observed information to refine the state estimate. The update step is done with the help of measurement model.

Using object tracking in our method allows us to leverage the results at the previous frames to improve the accuracy of the HOG detector by detecting people who may otherwise be "lost."

Process Model

$$\mathbf{x}_k = A\mathbf{x}_{k-1} + \omega_k \qquad A = \begin{bmatrix} 1 & dt & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & dt \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad E[\omega_k \omega_k^T] = Q$$

$$\mathbf{x}_k = \begin{bmatrix} x & \dot{x} & y & \dot{y} \end{bmatrix}_k$$

Measurement Model

$$\mathbf{z}_k = H\mathbf{x}_k + \nu_k \qquad H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \qquad E[\nu_k \nu_k^T] = R$$

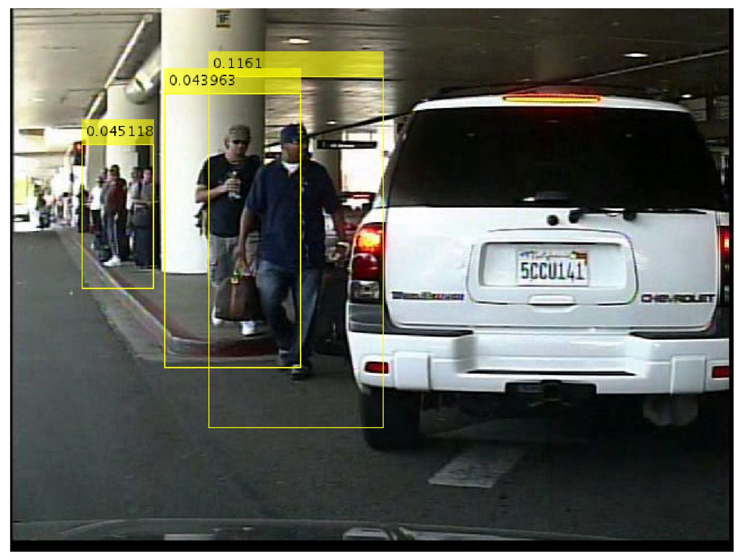
Automatic Height Estimation

Automatic vanishing point detection uses the algorithm provided in the homework 3, which utilizes line segmentation and RANSAC to estimate the vanishing points and based on them infer the horizon line. It then uses a fixed object of a known height to estimate the height of any given object in the scene using the image cross ratio.

RESULTS

Our method of combining HOG pedestrian detector with Kalman filters and height estimation achieves an improvement over the baseline method of HOG detection. We evaluated it visually using the videos from the Caltech Pedestrian Detection Dataset. We will now illustrate the improvements over the baseline HOG detector.

Often pedestrians get "lost" by the HOG detector. This happens because it does not have any information about the results of the previous frame.
In comparison, our method uses object tracking, which allows us to detect people on consecutive frames even though we get no detection on a single frame alone. This is illustrated with the following two frames:



First frame with three people getting detected correctly by HOG

Second frame where two of the people are "lost" by the HOG detector

Second frame with our approach, which still detects the "lost" people

RESULTS (CONTINUED)

In addition to keypoint tracking, we can see the effect of height estimation, which filters out false predictions that turn out to be too small to be a person.

A bag is getting falsely detected as the second tracked person

The bag is filtered out by our height detection algorithm due to being too small

CONCLUSIONS & FUTURE WORK

By running our experiments and observing the improvement over traditional person detection methods, we conclude that it is vital to make use of the continuous nature of frames in videos. We specifically used a trivial HOG detector as opposed to state-of-the-art methods as our baseline in order to clearly see the improvement gained by object tracking and height estimation.

In the future, it is necessary to evaluate our method of pedestrian detection not only on the entire Caltech dataset, but also compare it with the current state-of-the-art methods of pedestrian detection, which use methods such as deep convolutional neural networks and cascades. It may turn out to be beneficial to combine those methods with ours.

ACKNOWLEDGEMENTS

We used the Matlab libraries provided by the Computer Vision and Mapping Toolboxes.
In addition, the frames are taken from the Caltech Pedestrian Detection Dataset videos.
The HOG detector was trained on the INRIA dataset.

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN