Automatic Adaptation of a Generic Pedestrian Detector to a Specific Traffic Scene
Meng Wang and Xiaogang Wang
Department of Electronic Engineering, The Chinese University of Hong Kong

Introduction: Pedestrian Detection in Surveillance

Examples from the INRIA dataset (left) and the MIT traffic dataset (right). There are both true positives (first row) and false positives (second row).

- A generic detector doesn’t work well on a specific scene. It may achieve a 90% detection rate on the INRIA dataset, but drop to 20% on the MIT traffic dataset, both at approximately 1 false positive per image.
- It should be a much simpler problem on a specific scene. Once the scene is fixed, the diversity of both positive and negative examples will be significantly reduced.
- Our goal: A scene-specific detector. It is trained based on a generic dataset and examples from the specific scene.

Contribution: Self-adaptation of a generic detector

- Automatically infer confident positive examples and confident negative examples from the scene examples using both context and appearance information.
- Unsupervised. No labeled example from the target dataset is required.

Context Information

1. Path models: It is more likely for positive examples to appear on pedestrian paths.
2. Sizes of pedestrians in a target scene follow a certain distribution.
3. Pedestrians usually have motions.
4. Background patches tend to repeat at the same location over time.
5. Pedestrians cannot exist on vehicle trajectories.

Path Model

- Individual paths learned unsupervisedly
- Manual Selection (Pedestrian path or not)
- Combine

- 29 individual semantic regions are learned unsupervisedly in [1]. We manually select 11 semantic regions where pedestrians usually appear. This is the only manual work in this project.
- The selected semantic regions are combined and thresholded to yield a final path model. It serves as one of the cues in selecting the pedestrian examples.
- Experiment (c) shows that the path model proves to be the most effective cue.

Experiment Results

(a)
(b)
(c)

GE: Generic detector. SP_M(N): Scene detector trained by manually labeling the first N frames out of the total 420 frames.

HOG + SVM is used as the detector. (a) Compared with the generic detector, it improves the detection rate from 21% to 62% at 10^-6 FPPW. (b) Its performance is comparable with the scene specific detector using 300 manually labeled frames. (c) Path model and motion prove to be the most effective cues in selecting confidence positive examples.