Evolving Cascades of Voting Feature Detectors for Vehicle Detection in Satellite Imagery

Krzysztof Krawiec   Bartosz Kukawka   Tomasz Maciejewski

Institute of Computing Science
Poznan University of Technology, Poznań, Poland

CEC 2010

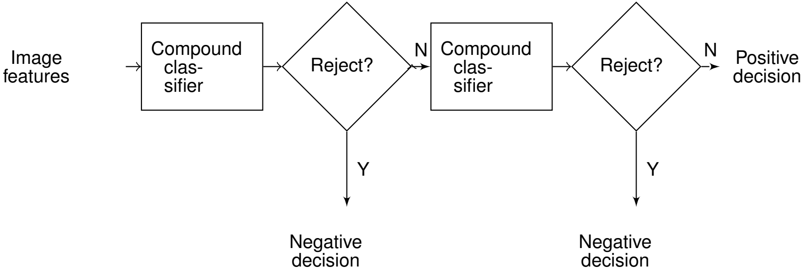
The objectives

- Use a compound GP classifier to detect vehicles in satellite imagery (visible band)
- Characteristics of the task:
  - extreme disproportion of the positive and negative classes,
  - heterogeneity of the positive class,
  - low spatial resolution,
  - uncontrolled lighting,
  - partial occlusions,
  - strong sunlight reflexes from wind shields, man-made objects that closely resemble cars (air conditioning equipment on rooftops, cargo containers, etc.)

Major problem

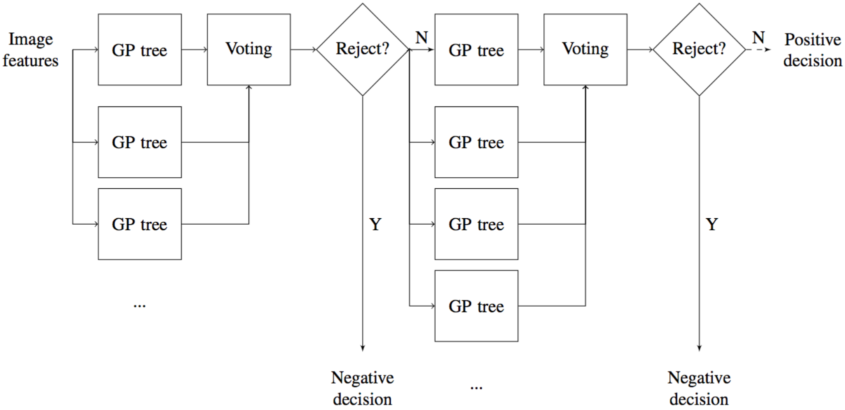
- Binary classification task with [extremely] low share of positive examples
  - cars occupy around 1.5% of image area,
  - a priori probability of the positive class: 0.0001
- The idea: use a *cascade* of classifiers.
- Each classifier:
  - processes only the examples classified as positive by all its predecessors,
  - is trained to retain (accept) all (almost all) positive examples, while rejecting as many negative examples as possible,

Cascade of detectors



- Only examples that pass through *all* cascade nodes are classified as positive
- Famous representative: Viola & Jones [2001] face detector

Our contributions



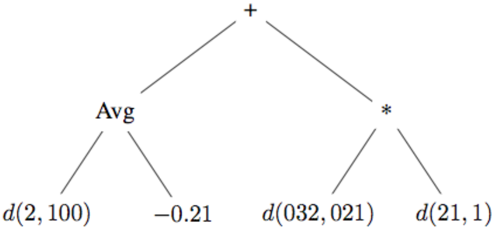
- Employ GP to induce the base classifiers
- Use quad-tree-based features instead of Haar wavelets

Feature definition

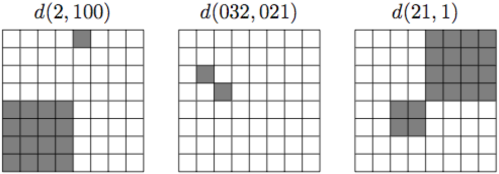
- A quad tree stacked over 32×32 input window
  - Tree nodes correspond one-to-one to rectangular image regions (*tiles*).
  - The nodes at consecutive depths correspond to 16×16, 8×8, 4×4, and 2×2 tiles; there are, 4, 16, 64, and 256 of them $\Longrightarrow$ total of 340 tiles.
  - Each tile uniquely identified by *quad key* – a variable-length sequence of quaternary digits.
- Feature $d(m, n)$ = difference between mean brightness values in *two* tiles identified by m and n.
- Total number of features: $340 \times 340 = 115,600$
- A clever trick (integral image) makes extraction of such features very effective (4x memory access + 3 subtractions).

Exemplary tree and accessed features

An exemplary GP tree (base classifier):



The tiles accessed by particular features (16x16 grid not shown):



The Experiment: Tasks and Settings

Parameter	Setting
Individual	5 GP trees
Population size	1024
Population initialization	standard ramped half-and-half
Selection	tournament (7)
Crossover	tree-swap, probability 0.9
Mutation	subtree-replacement, probability 0.1
Elitism	no
Tree depth limit	10
Number of generations	100

- Evolutionary Computation in Java, ECJ
- Runs repeated 5 times

Fitness function

- F-measure = the harmonic mean of *precision* p and *recall* r (sensitivity)

$$fitness = F_{measure} = \frac{2pr}{p+r},$$

$$p = \frac{TP}{TP+FP}, \; r = \frac{TP}{TP+FN}$$

The Experiment: Data

- 33 true-color satellite images of spatial resolution 0.2m/pixel

The Experiment: Data

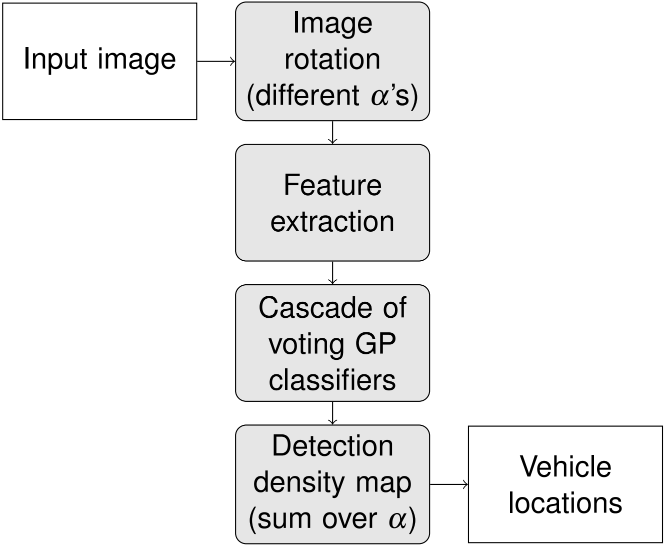
- Different environments: urban, rural, parking lots, bridges, etc
- 4 to 378 cars per image, 22×9 pixels on average
- Training example = 32x32 window of the original image
  - positive example: window centered on a vertically aligned car
  - negative example: any window non overlapping with any car
- Training set: 659 cars extracted from 24 training images,
- Testing set: 635 cars extracted from 9 testing images,
- All images converted to grayscale for further analysis

Learning progress

Cascade node	TP	TN	FP	FN
1	374	11865	135	285
2	368	87	48	6
3	367	18	30	1
4	367	5	25	0
5	367	8	17	0
Total	367	11983	17	292

On the training data, this detector attains precision $p = 0.956$, recall $r = 0.557$, and F-measure of 0.704.

Evaluation on test images



Evaluation on test images

Detection density map (DDM):

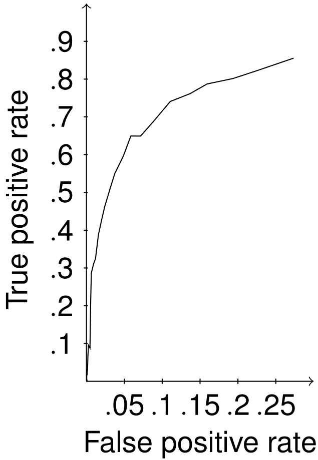
- aggregated over 8 rotated versions of the test images (every 22.5 degrees)
- each detection increases the belief in vehicle presence at the particular location and its surroundings (Gaussian distribution with $\sigma = 2.6$)
- local maxima in DDM with belief values greater than t lead to detections

Detections

Detections

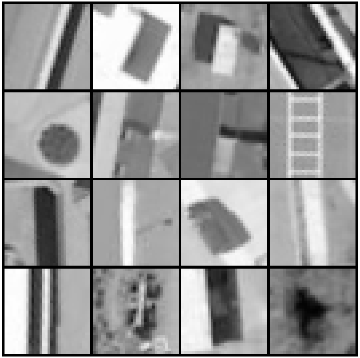
ROC curve

- Obtained by varying the *t* threshold of the DDM
- Detection within a true vehicle contour counted as TP, otherwise FP



Some false positives and false negatives

FP



FN

Summary

- Decent performance on a challenging task using simple features
- No contextual information used (particularly road/street locations)
- Further work: different aggregation schemes, colour, multiobjective evolutionary search for precision and recall