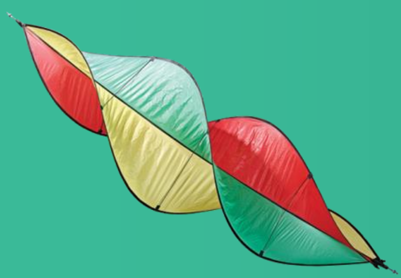
KU LEUVEN

TECHNOLOGIECAMPUS DE NAYER SINT-KATELIJNE-WAVER

# How to reach top accuracy for a visual pedestrian detection warning system from a car?




Floris De Smedt[1], Steven Puttemans[2] *(presenting author)*, Toon Goedemé[2]

[1] Robovision, Antwerpen, Belgium

[2] EAVISE Research Group, KU Leuven, Campus De Nayer, Sint-Katelijne-Waver, Belgium

# Introduction

- **Detection accuracy**
  - "How accurate can each pedestrian be localised?"
    → Accurate localisation of pedestrians is required!
  - "Reasonable setting?": ≥50px and ≥65% visible

- **System test**
  - "Is there an alarm when a pedestrian is in front of the car?"
    → Accurate localisation is **not** required
  - Independent of pedestrian size and occlusion level

# Dataset: KAIST

- 6 sets for training, 6 sets for testing (videos)
- 95.000 VGA image pairs (Color + LWIR) fully annotated
- 103.128 annotations
- 1.182 unique pedestrians
- LWIR is very beneficial during night conditions due to limited color information

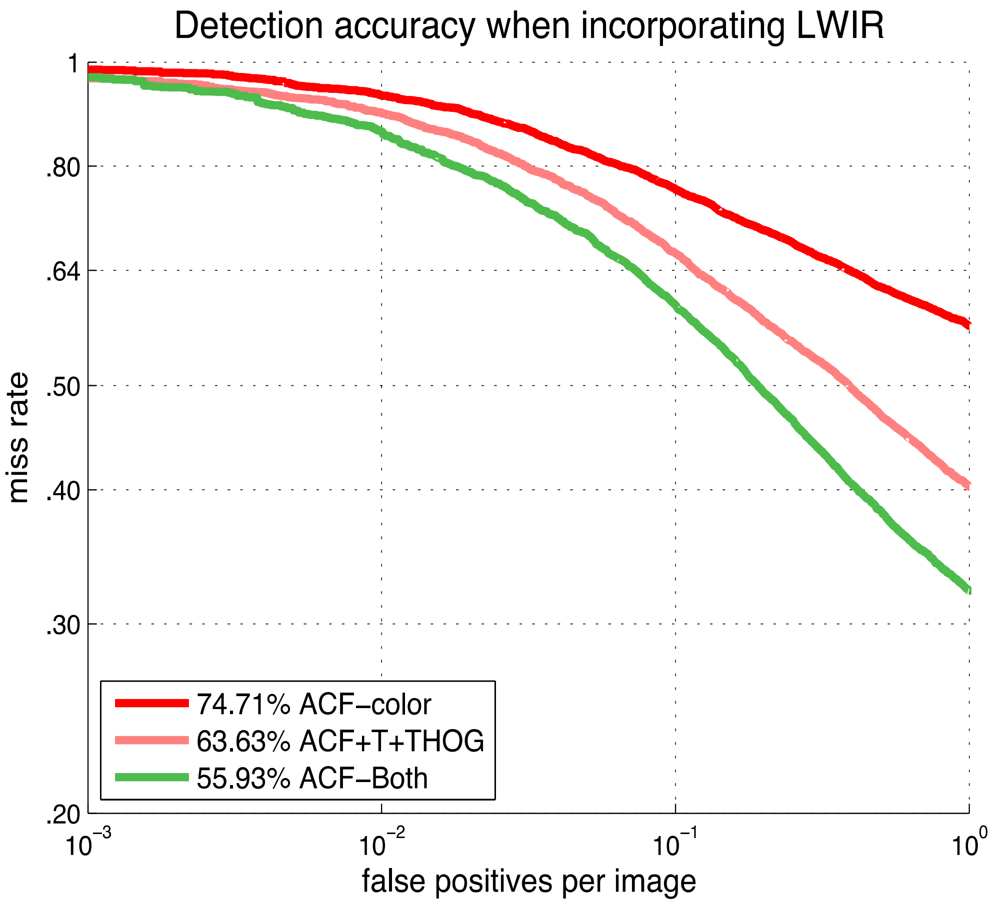
# Approach

- Own C++ implementation of the ACF (Aggregate Channel Features) detector by Piotr Dollar. [1]
- Study techniques that have shown to be beneficial on other pedestrian detection benchmarks (Caltech, INRIA,…)
  1. Combine color and LWIR
  2. ACF+ versus ACF
  3. Influence of selecting the training set
  4. *Amount of training data -> only in paper*
  5. *Influence of the model size -> only in paper*
  6. Using convolution masks to extend the features pool
  7. Influence of a ground constraint

[1] Fast feature pyramid for object detection, P. Dollár et al, PAMI2014

# 1. Combine color and LWIR

- ACF color channels
  - 3 color channels (LUV)
  - 6 gradient orientations
  - 1 gradient magnitude
- LWIR channels
  - 1 intensity
  - 6 gradient orientations
  - 1 gradient magnitude




→ Large improvement *(green)* over state-of-the-art *(red/pink)* [2]

→ Trained on both day and night images

[2] Multispectral pedestrian detection: benchmark dataset and baseline, Soonmin S. et al, CVPR2015

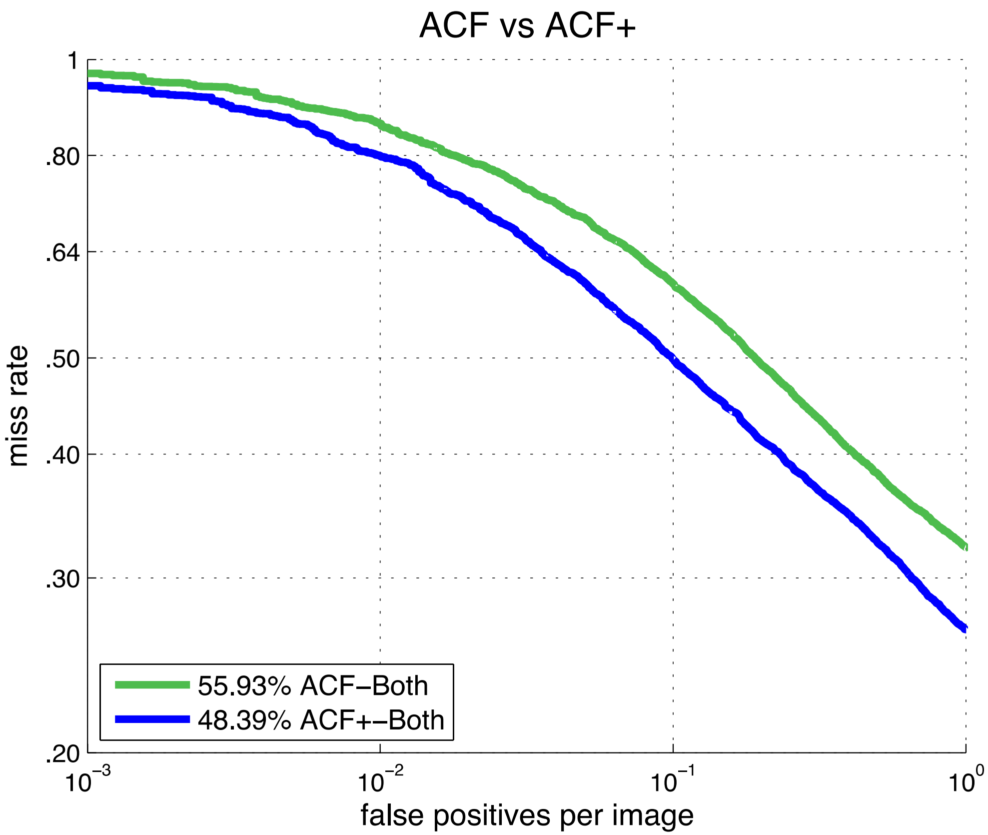
# 2. ACF+ versus ACF

## ACF

- 2.048 weak classifiers
- Depth-2 decision trees
- 5.000 negatives
- 10.000 accumulated negatives

## ACF+ [3]

- 4.096 weak classifiers
- Depth-5 decision trees
- 25.000 negatives
- 50.000 accumulated negatives




→ **7,5%** drop in miss-rate!

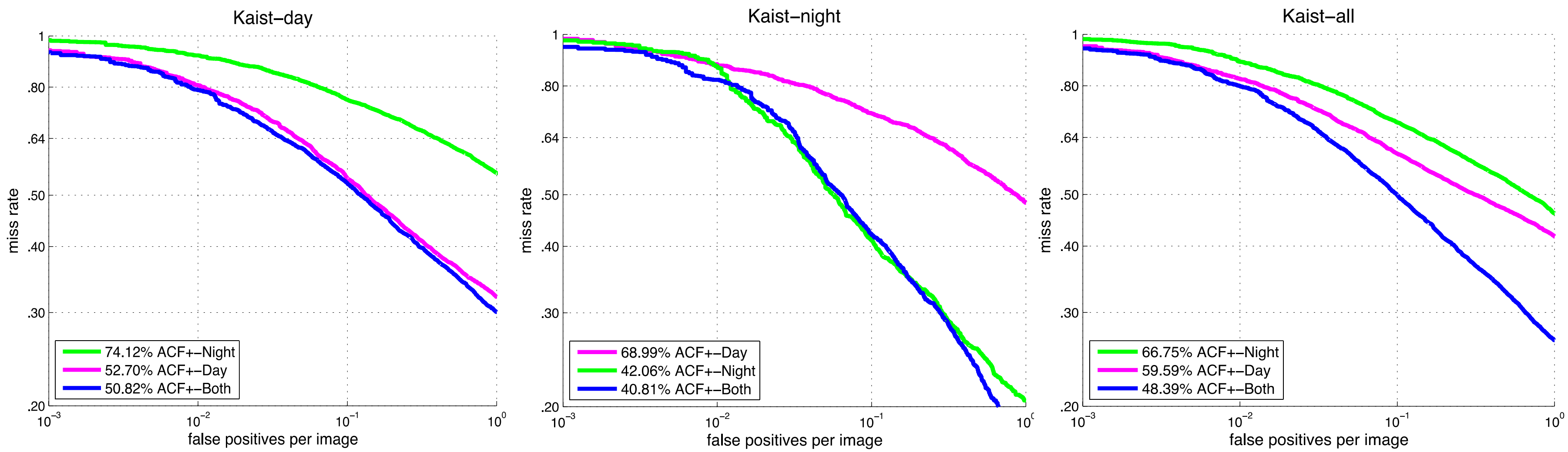
[3] Local decorrelation for improved pedestrian detection, W. Nam, ANIPS2014

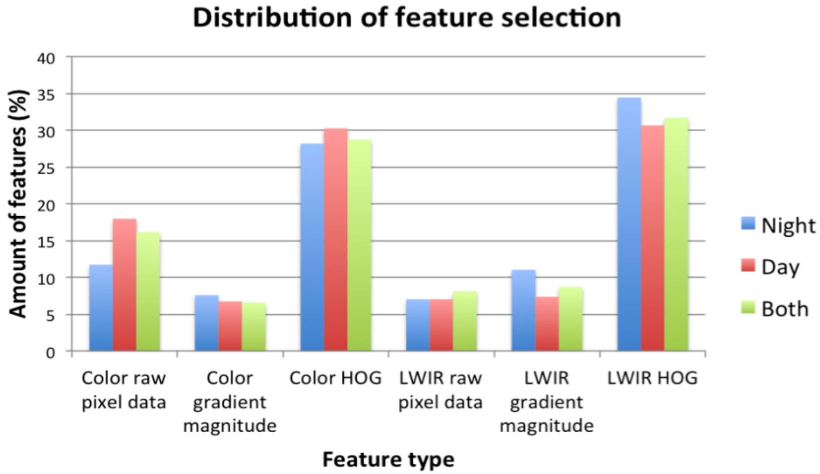
# 3. Influence of selecting the training set

- The goal is to learn a decision surface between pedestrians and background training samples.
- Having training conditions similar to the evaluation conditions is beneficial for each trained object detector. [4]
  - Selected features from feature pool are optimized for specific situation
  - Can we use a separate day and night model?
- We compare
  - Only day images for training
  - Only night images for training
  - Both day and night images for training

[4] Ten years of pedestrian detection, what have we learned?, R. Benenson et al, ECCV 2014 WS

# 3. Influence of selecting the training set

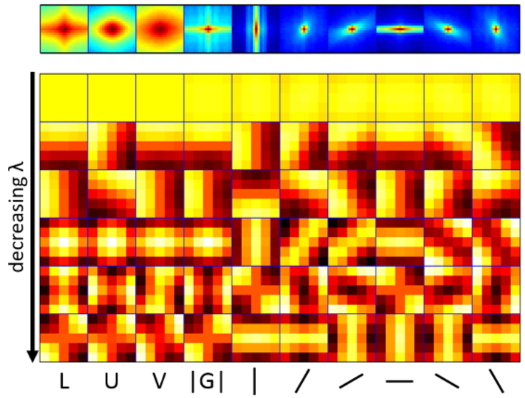


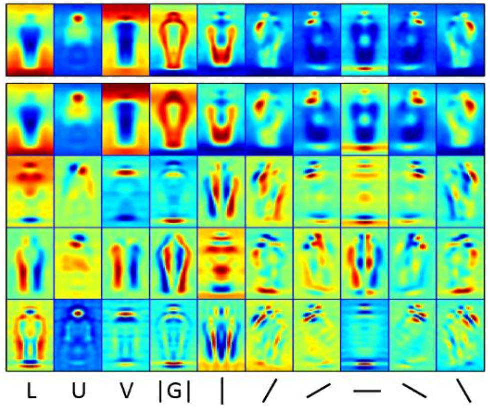




- No remarkable improvement by training for a specific condition *(green/pink)* over using both conditions *(blue)* at the same time.
- Using the combined training set is the best in all conditions.

# 6. Using convolution masks

- Convolve each feature channel with convolution masks
- LDCF: 4 convolution filters = 40 channels
- Filtered channel features: [5]
  - 61 convolution filters
  - State-of-the-art detection results
  - Very slow!
- Rotated channel features: [6]
  - 9 convolution filters
  - 6 x faster
  - 1% miss-rate increase in accuracy

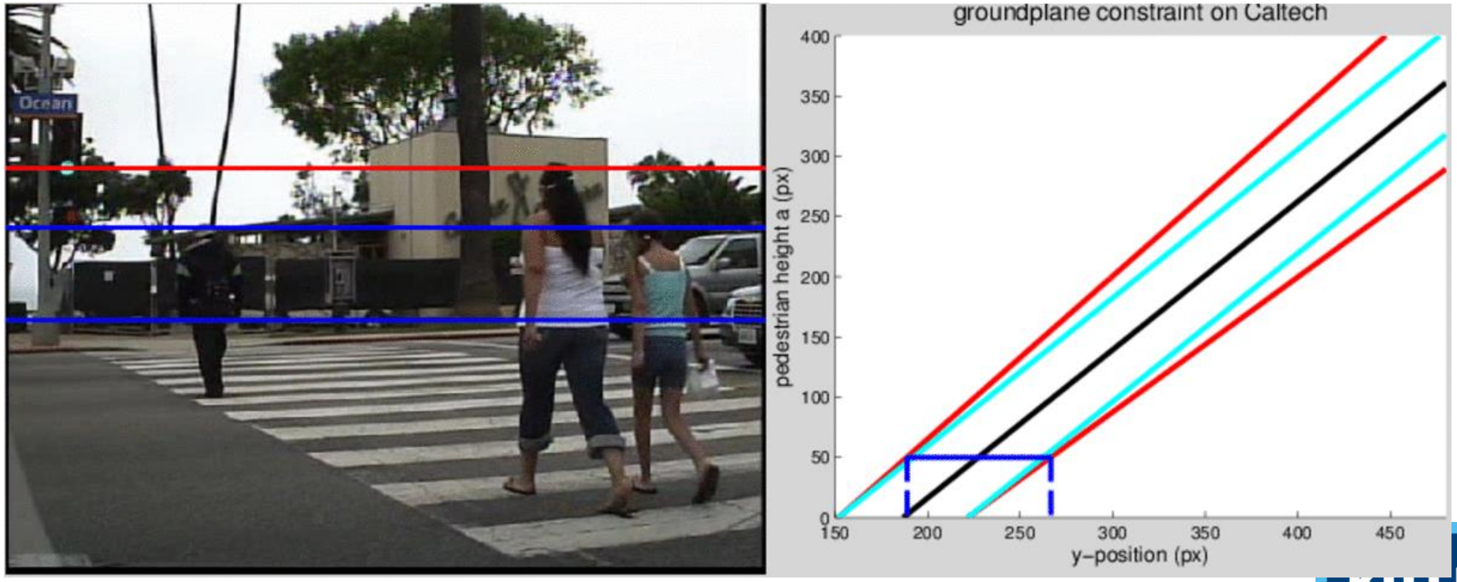


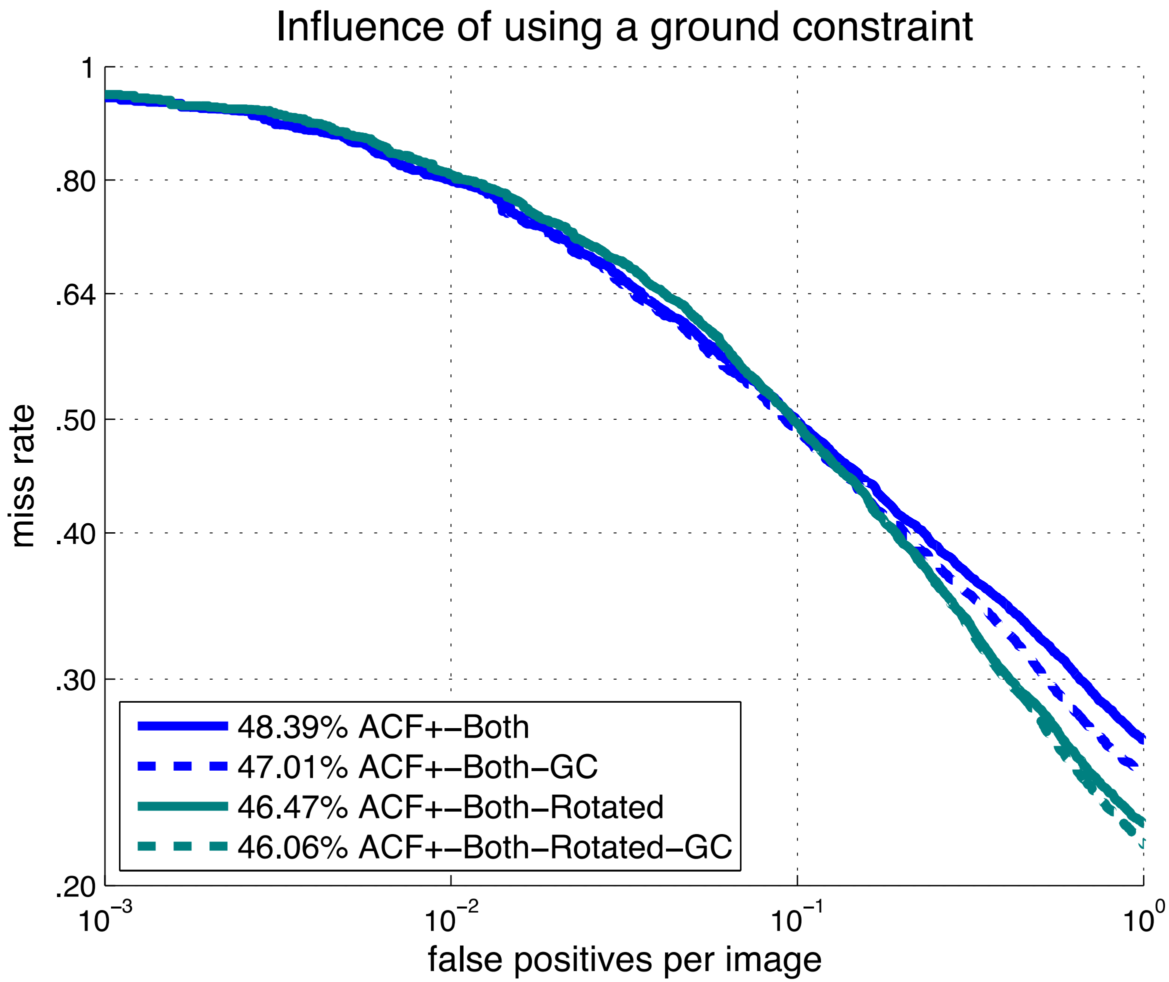

[5] Filtered channel features for pedestrian detection, S. Zhang et al, CVPR2015
[6] How far are we from solving pedestrian detection?, S. Zhang et al, CVPR2016

# 7. Influence of a ground constraint

- Each pedestrian size (height) can only be found in a limited range of y-positions inside the image.
  - Fit a relation between annotations and position in the image.
  - This leads to a strong reduction of the object search space compared to a full multiscale sliding window detection.
  - Allows approximately a 4x speed-up.



# 7. Influence of a ground constraint




- Limited accuracy benefit
- Both in case of convolution filters and ground constraint
- However still a large speed-up when using the ground constraint during processing

# System test experiments

- Required breaking distance (rule of thumb)

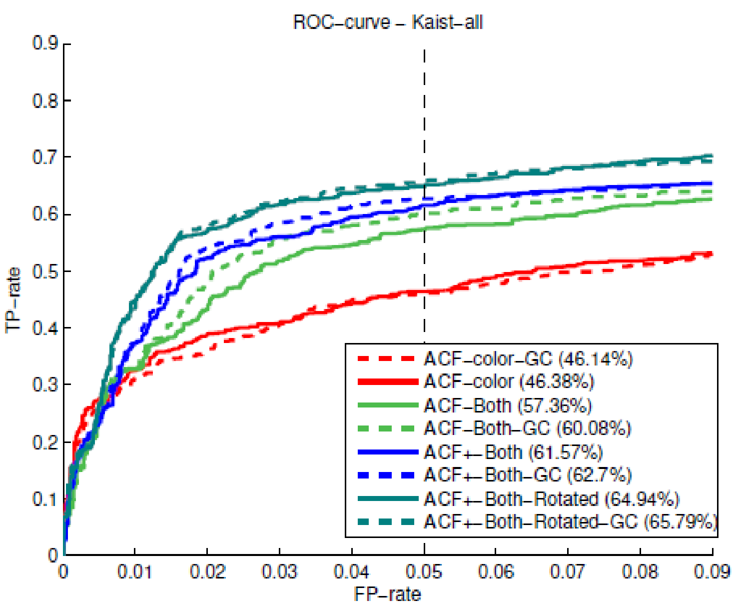
$$BD_{dry} = \frac{(\frac{v}{10})^2}{2}$$

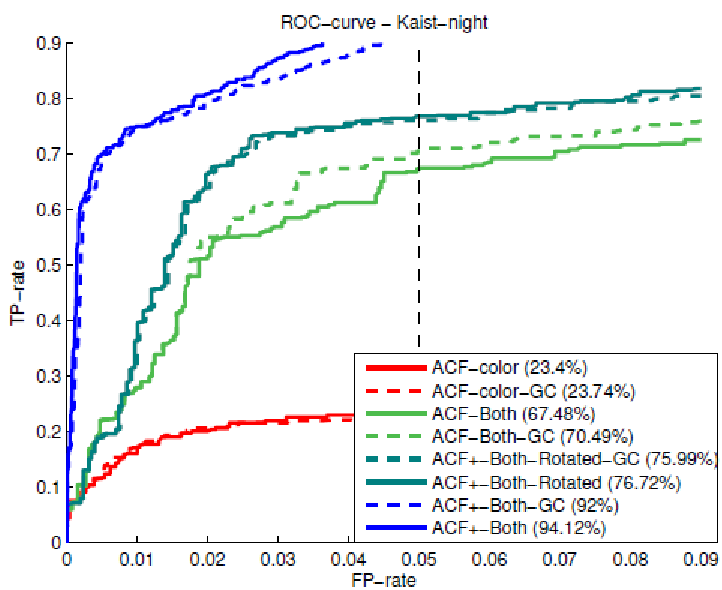
$$BD_{wet} = BD_{dry} \times 1.5$$

- Speed = 50km/h (13,9m/s) & distance(pedestrian,car) = 20m
  - DRY: 12.5m <-> WET: 18.75m
  - Remaining distance: DRY 7,5m <-> WET 1,25m
  - @13,9m/s this equals: DRY 1,85 sec <-> WET 11,12 sec
  - In this time you need minimal 1 frame
    → required processing speed = **1.85 fps**
    → required processing speed = **11.12 fps**
- Pedestrian size **75px** at 20m
- 5% FP-rate is acceptable *[Hoedemaeker et al, Foundation research and traffic security 2010]*

# System test experiments

| Technique | ≥ 50px (fps) | ≥ 75px (fps) | ≥ 100px (fps) | TPR (all) | TPR GC (all) | TPR (night) | TPR GC (night) |
|---|---|---|---|---|---|---|---|
| ACF - color | 10.73 | 18.8 | 26.03 | 46.38% | 46.14% | 23.4% | 23.74% |
| ACF - Both | 9.51 | 11.81 | 21.13 | 57.37% | 60.08% | 67.48% | 70.49% |
| ACF+ - Both | 8.75 | 10.43 | 19.28 | 61.57% | 62.70% | **94.12%** | 92% |
| ACF+ - Rot Both | 0.875 | 1.39 | 1.91 | 64,94% | **65.79%** | 76.72% | 75.99% |




- Required speed under dry conditions (1.85 fps) can be reached taking into account the 4x speed up of the ground plane constraint
- Night conditions are the hardest for the driver but we reach a high performance of 94,12%.

# Conclusion

- We proposed using current state-of-the-art pedestrian detectors as a warning system for car drivers.
- We used a system test as validation:
  - An alarm should be generated if pedestrians are too close to the car
  - Independent of the amount of occlusion
- Shown a drastic accuracy improvement over the state-of-the-art on the KAIST dataset by study different techniques to improve detection quality.
- Reached top accuracy in night conditions, where the system is most useful.

# Thank you for your attention!

Contact:
- floris.desmedt@robovision.eu
- toon.goedeme@kuleuven.be

More info:
- http://www.eavise.be