

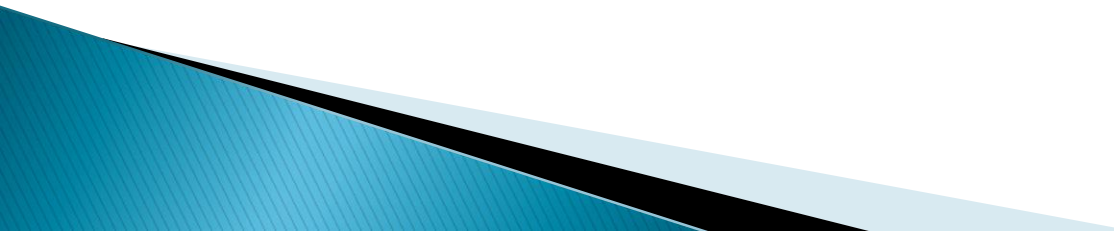
Histogram of Oriented Gradient for Human Detection

Farhad Fallah

11/17/2015



Outline

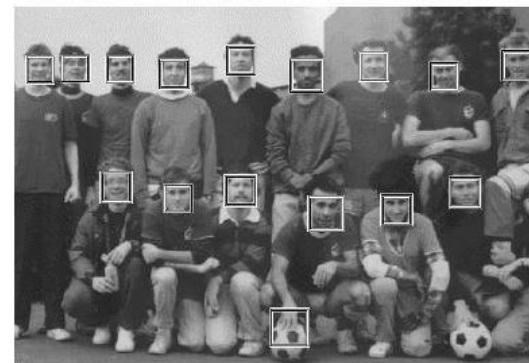
- ▶ Previous descriptors
 - ▶ Introduction to the HOG descriptor
 - ▶ Describing the algorithm and result of each parts
 - ▶ MIT and INRIA datasets
 - ▶ Conclusion
- 

Previous descriptors

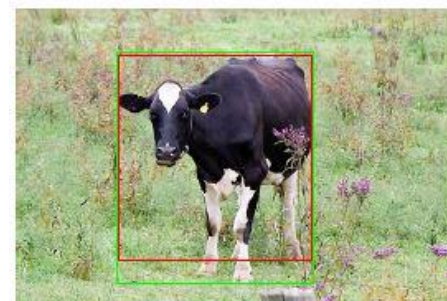
- Shape matching and Object recognition (Serge Belongie)



- Rapid Object Detection using a Boosted Cascade of Simple Features. (Paul Viola)



- Selective Search for Object Recognition (J.R.R. Uijlings)



Histogram of Oriented Gradient

- ▶ A feature descriptor for object detection
- ▶ Mostly use for human detection
- ▶ Compare to existing edge and gradient based descriptor, it performs significantly better.

Challenges for human detection



Various poses



Variable appearance



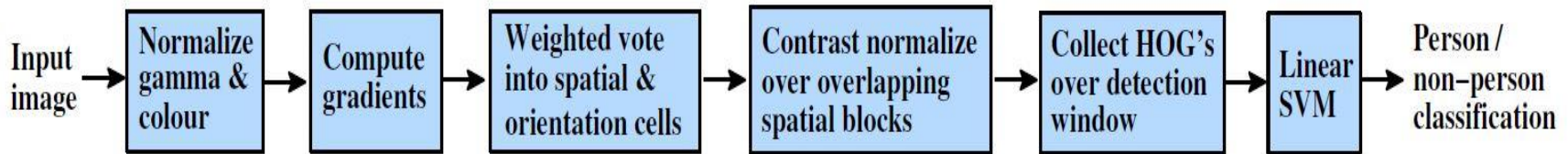
Complex background



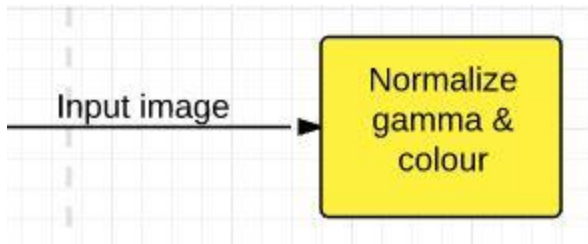
Unconstrained illumination



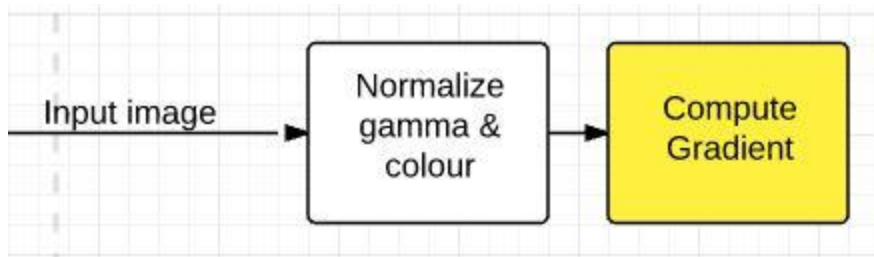
An overview of the object detection chain



- Evaluating several input pixels in gray scale, RGB and LAB color space to recognize which one has better performance (Optional)
- Divide the image window into blocks and cells to compute the gradient on each cell.
- Make a histogram of gradient direction of each cell.
- Grouping cells into large blocks and then apply the contrast normalization on each block.
- Concatenating all the histograms to a 1-D feature vector matrix and training a SVM classifier to find the positive and negative images.



- ▶ The input to the system is an image.
- ▶ Tested with RGB, LAB and gray scale color space images
 - RGB and LAB have better result compare to gray scale.
- ▶ Gamma compression: A non linear operation used to code and decode luminance.
- ▶ Using square root gamma compression of each channel improved performance.
- ▶ The Log compression doesn't have a good performance.



- ▶ The gradients computed using Gaussian smoothing followed by one of the several discrete derivative masks for computing gradients.
- ▶ The simple 1-D $[-1, 0, 1]$ mask at $\sigma = 0$ works best.
- ▶ For color images, pick the color channel with the highest gradient magnitude for each pixel.

-1	1
----	---

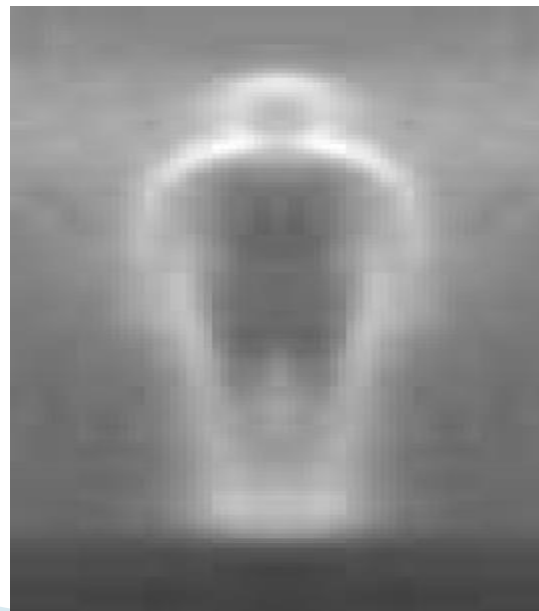
Uncentred

-1	0	1
----	---	---

Centered

1	-8	0	8	-1
---	----	---	---	----

Cubic-corrected



Body edges

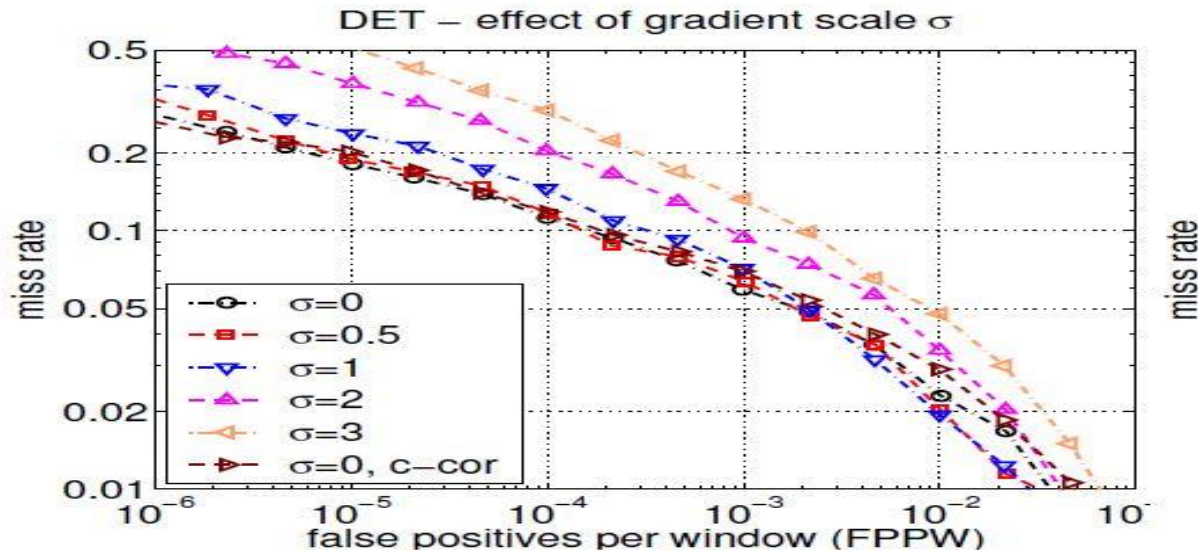
0	1
-1	0

Diagonal

-1	0	1
-2	0	2
-1	0	1

Sobel

Result of changing gradient scale Ω



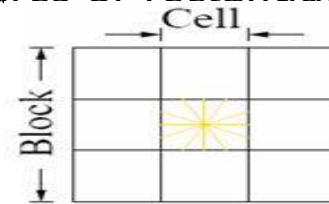
- To quantify the detector performance the Detection Error Tradeoff (DET) curves are plotted on a log-log scale. (The DET is a graph to plot false reject rate vs. false accept rate)
- The vertical axis is “miss rate” ($\frac{FalseNeg}{TruePos+FalseNeg}$)
- The Horizontal axis is False Positives Per Window (FPPW)
- DET plots allow small probability to be distinguished more easily.
- The miss rate at 10^{-4} FPPW as a reference point for results.
- In these diagrams, Lower values are better.

Block and cell

- ▶ For the next level, we should divide the image window into grids.
- ▶ There are two classes of block geometries that will divide the image window into grids.

❑ **R-HOG:** Will partition the image into grids of squares or rectangular. It is represented by three factors:

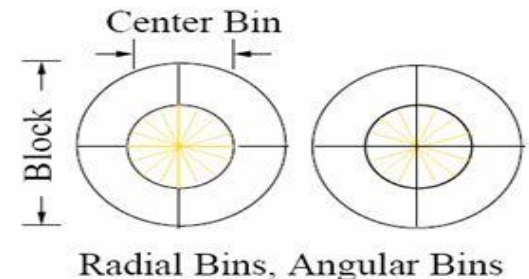
- Number of cells per block.
- Number of pixels per cell.
- Number of channel per cell histogram.



❑ **C-HOG:** Circular blocks partitioned into cells in log-polar mode. One of the blocks is a single circular central cell and the other one With an angularly-divided central cell. These block are represented by:

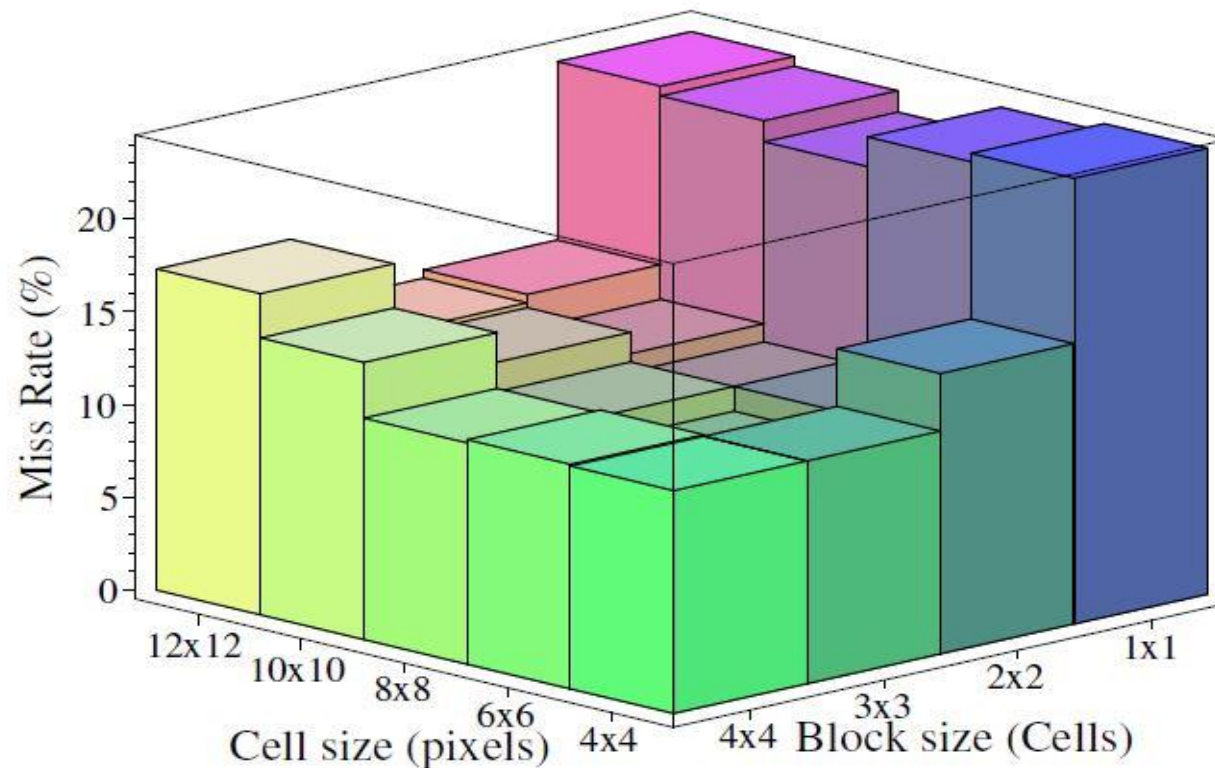
- Number of angular and radial bins.
- the radius of the center bin.
- The expansion factor for subsequent radii.

- ▶ Each of the blocks in the image contains pixel cells.



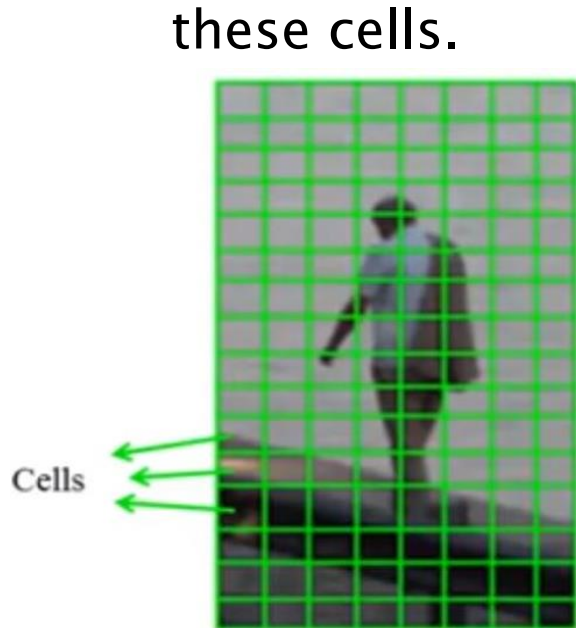
The effects of cells and blocks

- For human detection, 3*3 cell blocks of 6*6 pixel cells perform best by 10.4% of miss rate at 10^{-4} .
- Totally, 2*2 and 3*3 block size work best. (Paper default is 2*2 block size of 8*8 cell size).

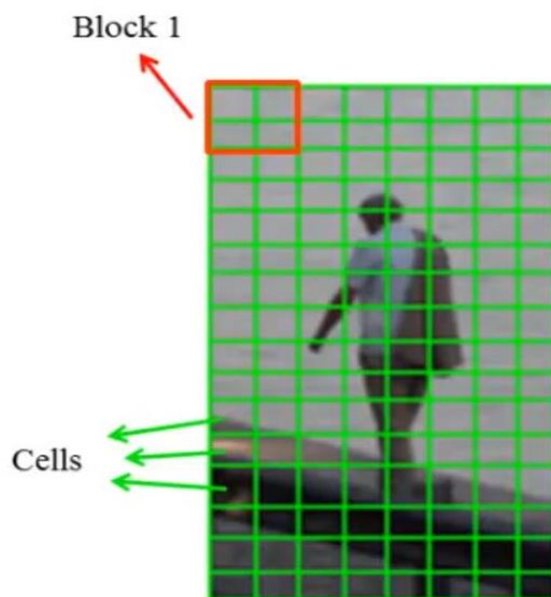


The properties of the Dalal default detector

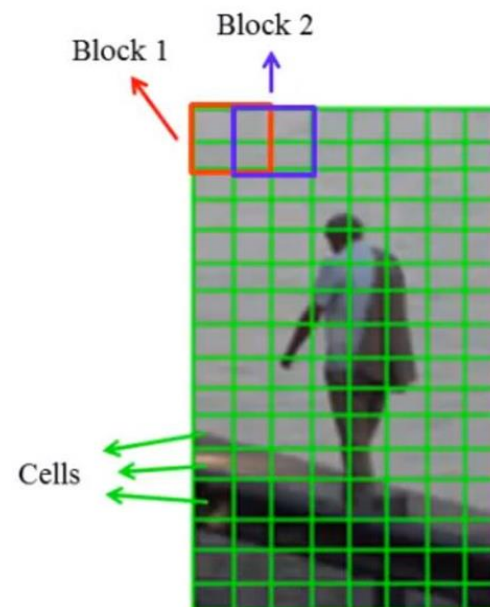
- ▶ The default detector has 64×128 detection window.
- ▶ The cells are 8×8 pixels, and the blocks are 2×2 cells.
- ▶ Each block will have 4 cells, each cell is 8×8 pixels. Also blocks should overlap each other by 50%.
- ▶ For the next level, we should divide the image window into these cells.



Each cell is 8×8 pixel

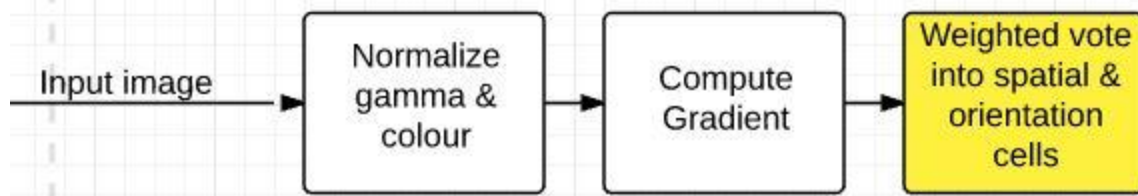


Each block is 16×16 pixel (2×2 cells)

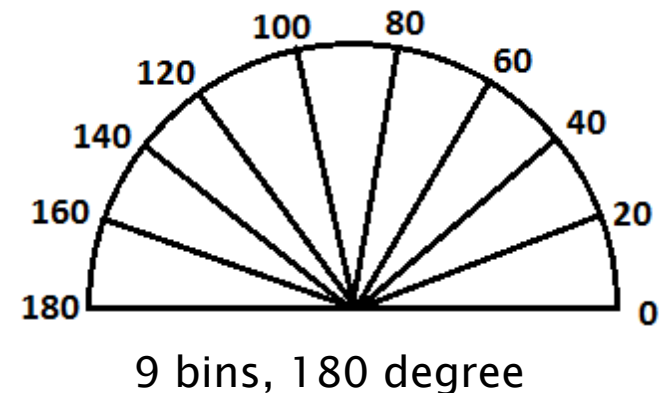
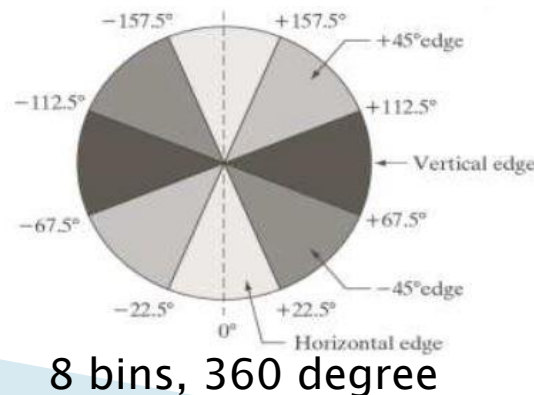


50% overlapping

- Will have $7 \times 15 = 105$ blocks totally.

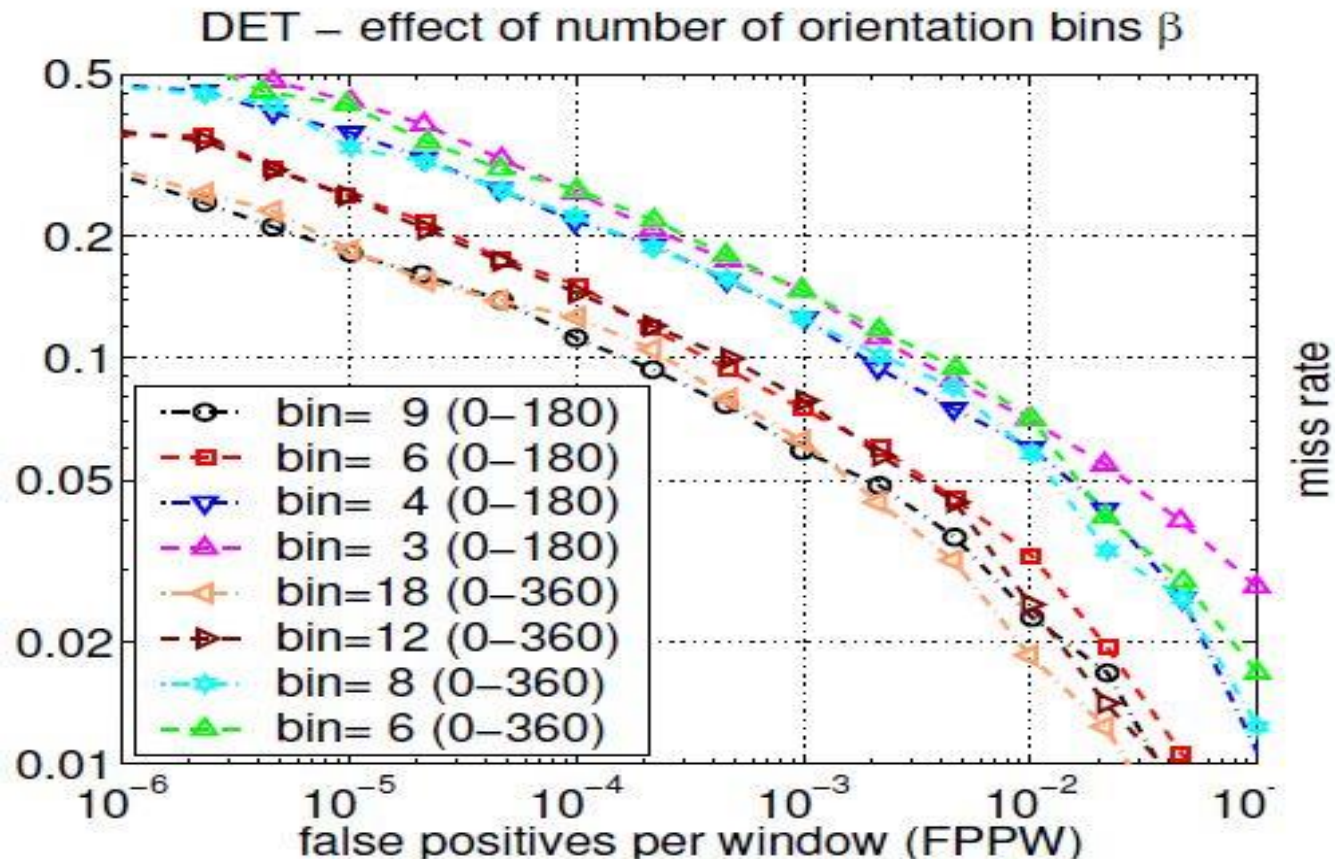


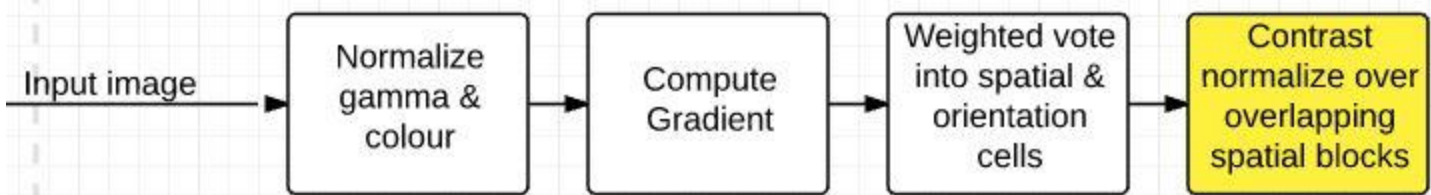
- ▶ The histogram of the gradient direction should be computed.
- ▶ Quantize the gradient orientation into 9 bins in 0 to 180 degree.(the same as what we did for 8 bins in the homework)
- ▶ Each pixel within the cell casts a weighted vote for an orientation-based histogram channel based on the values found in the gradient computation.
- ▶ The vote is a function of the gradient magnitude.
- ▶ During the overlapping, each cell contributes more than once to the final descriptor.
- ▶ We could weighted a vote with respect to its gradient magnitude. (if it's a strong edge or weak edge)



Effects of the changing in the number of bins

- Increasing the number of orientation bins increases performance significantly up to about 9 bins spaced over 0–180 degree



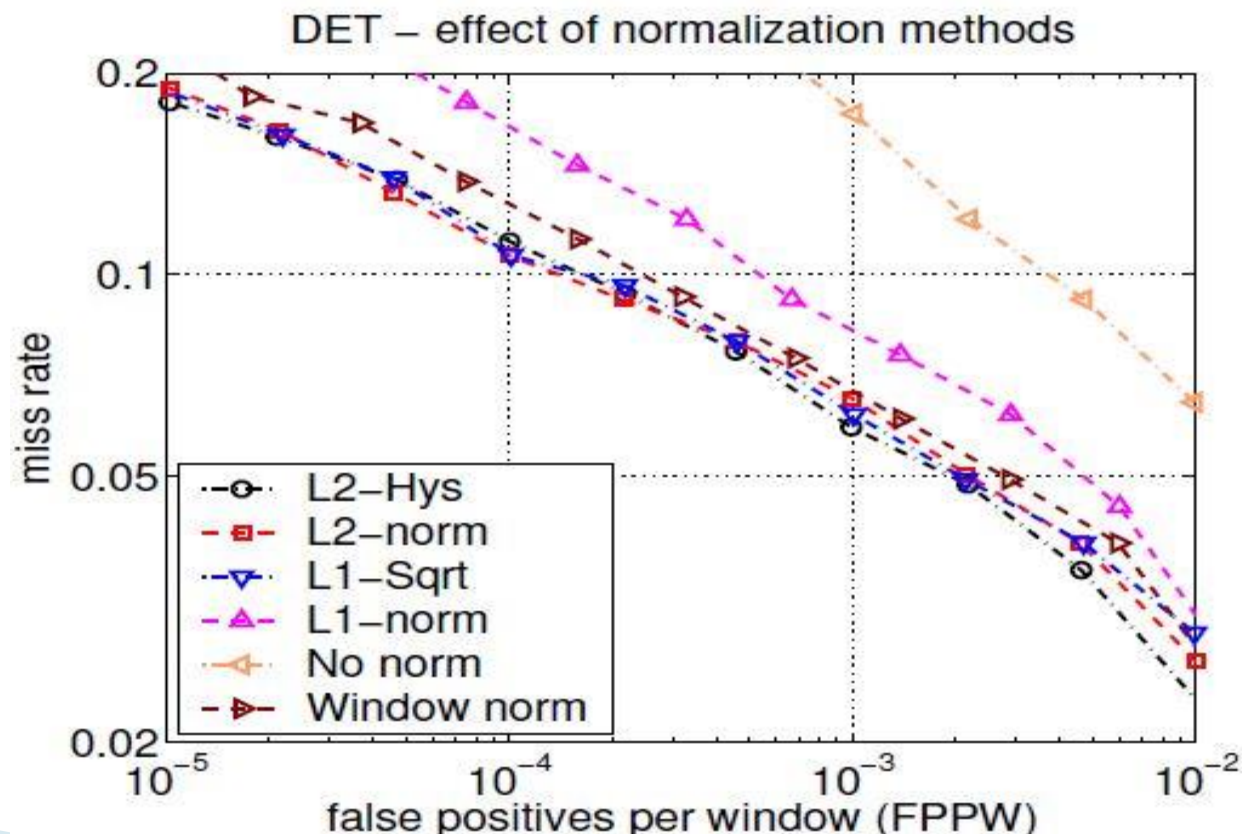


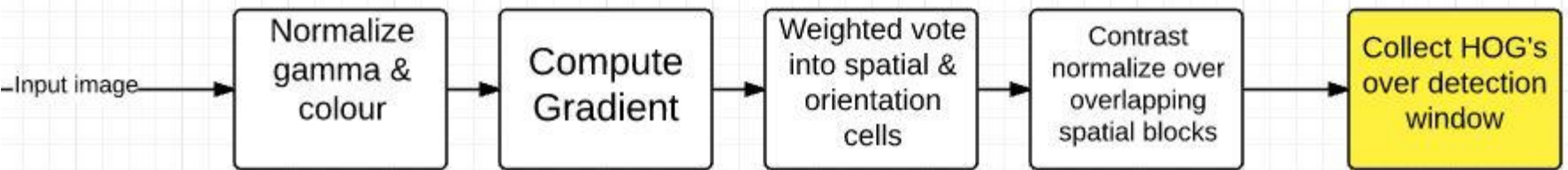
- ▶ Gradient strengths vary over a wide range owing to local variations in illumination and foreground-background contrast, so effective local contrast normalization turns out to be essential for good performance.
- ▶ In a global normalization changes of pixels values are based on the range of values of the entire image, however, in Local a global normalization values changes based on a local patch of the image.
- ▶ They tried 4 different kinds of normalization.
- ▶ Let (v) be the block to be normalized and (ϵ) be a small constant.

- ❖ L2-norm \longrightarrow $L2 - norm : v \longrightarrow v / \sqrt{\|v\|_2^2 + \epsilon^2}$
- ❖ L1-norm \longrightarrow $L2 - norm : v \longrightarrow v / \sqrt{\|v\|_2^2 + \epsilon^2}$
- ❖ L2-hys \longrightarrow L2-norm followed by clipping (Limiting the maximum values)
- ❖ L1-sqrt \longrightarrow $L1 - sqrt : v \longrightarrow \sqrt{v / (\|v\|_1 + \epsilon)}$

Effects of different normalization method

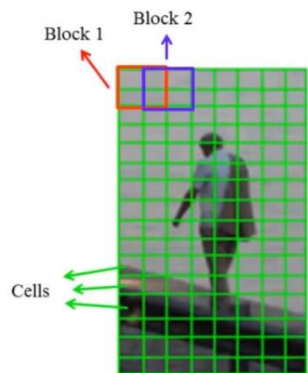
- All methods showed very significant improvement over the non-normalized data. The best methods are L2-norm and L1-sqrt.





- ▶ Now we have extracted all the histogram of directions (105 blocks). In this part, we should concatenate all of them to a 1-D matrix. (each block histogram has 9 dimensions)

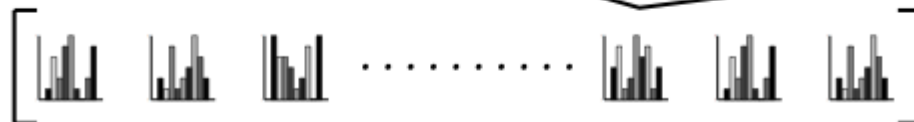
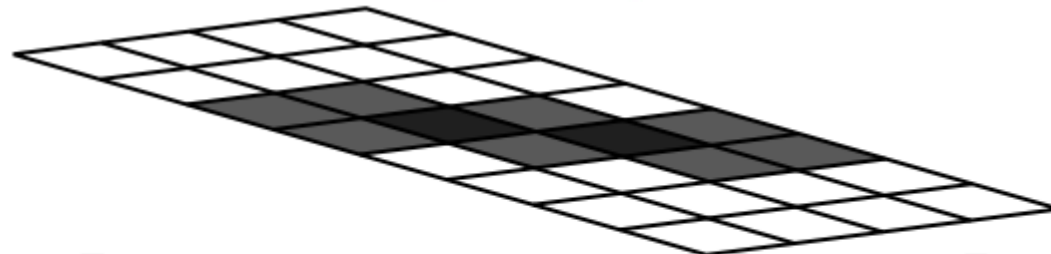
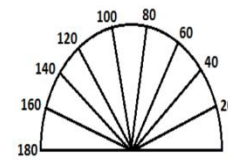
- ▶ Number of Feature vectors = $(15 \times 7) \times 9 \times 4 = 3780$.



Number of Blocks

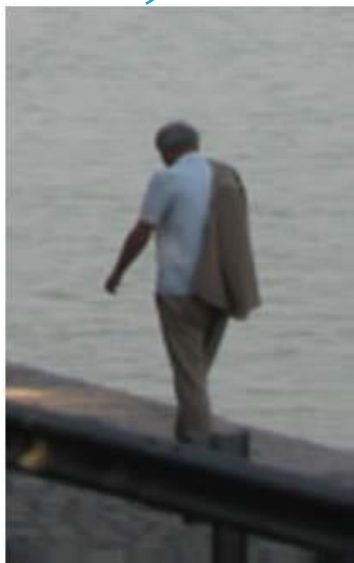
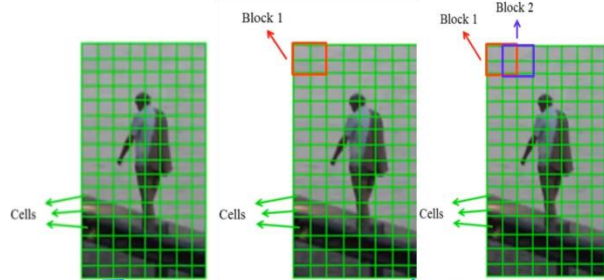
9 dimension

Number of normalization by neighboring cells

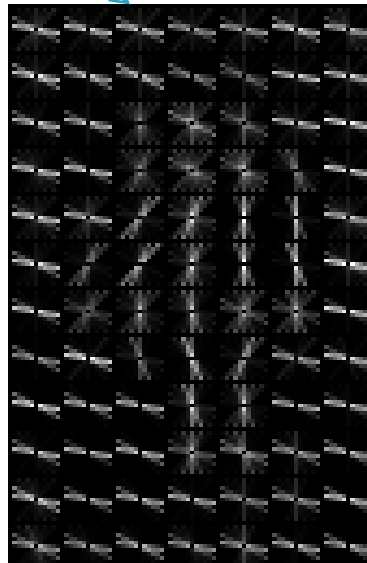


Examples

- Each block in the descriptor shows the histogram of orientation.
- Blocks corresponded with the arm of the person show the dominant direction of the edge



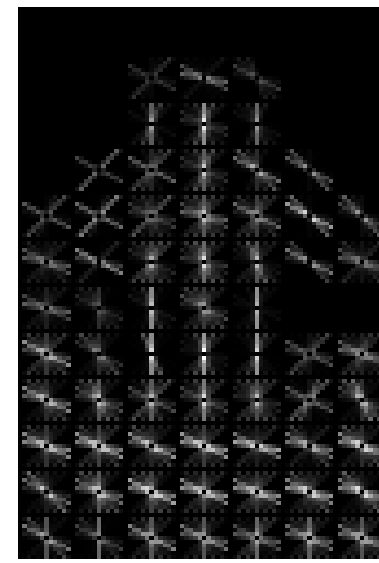
Human



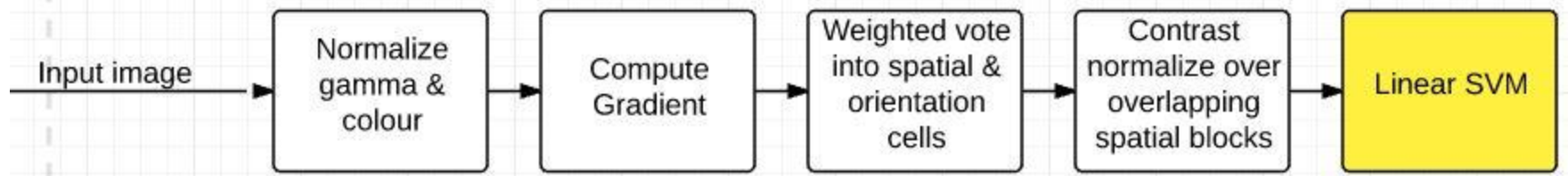
R-HOG descriptor



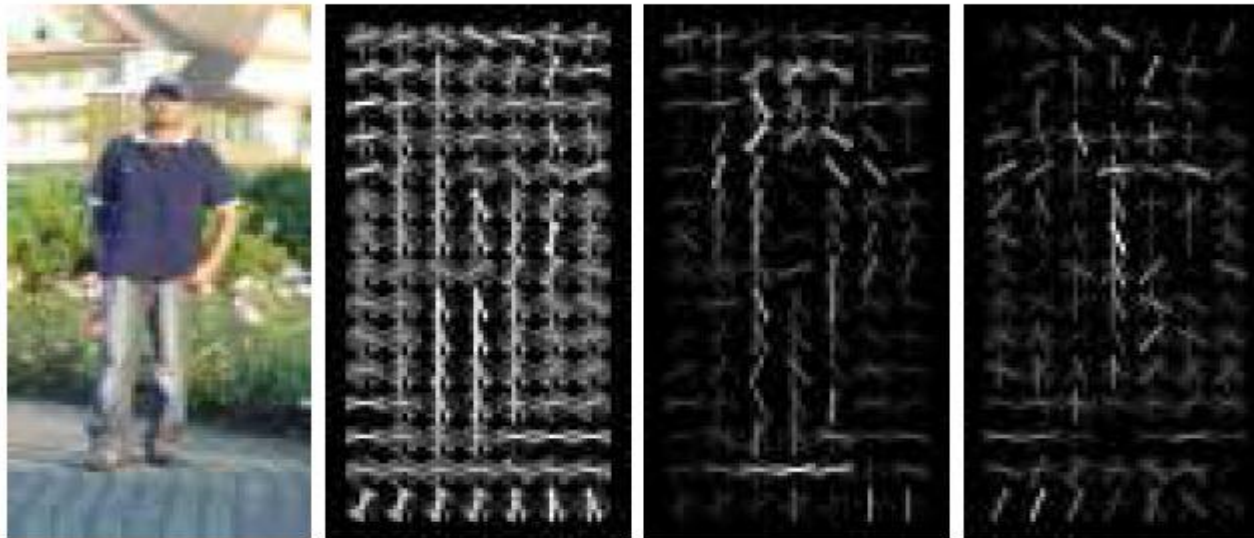
Human



R-HOG descriptor



- ▶ In this part, the HOG descriptors are fed into a recognition system based on SVM which will recognize the human from non-human.



Human

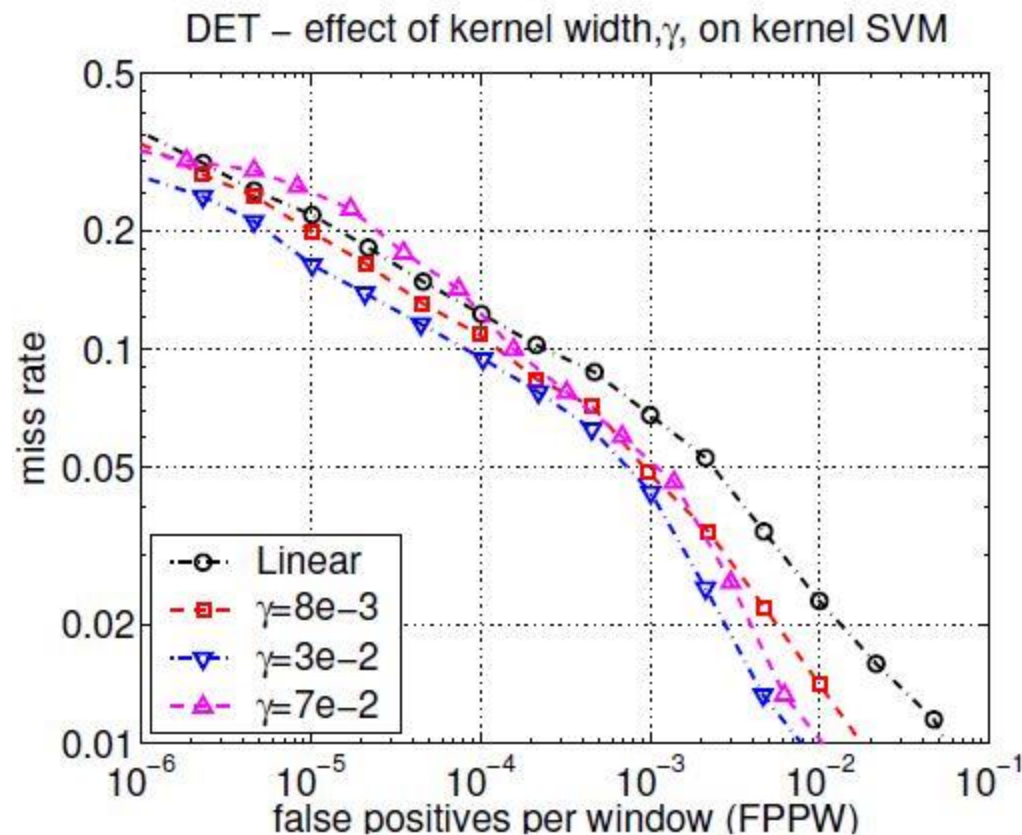
R-HOG descriptor

Weighted by
positive SVM
weight

Weighted by
negative SVM
weight

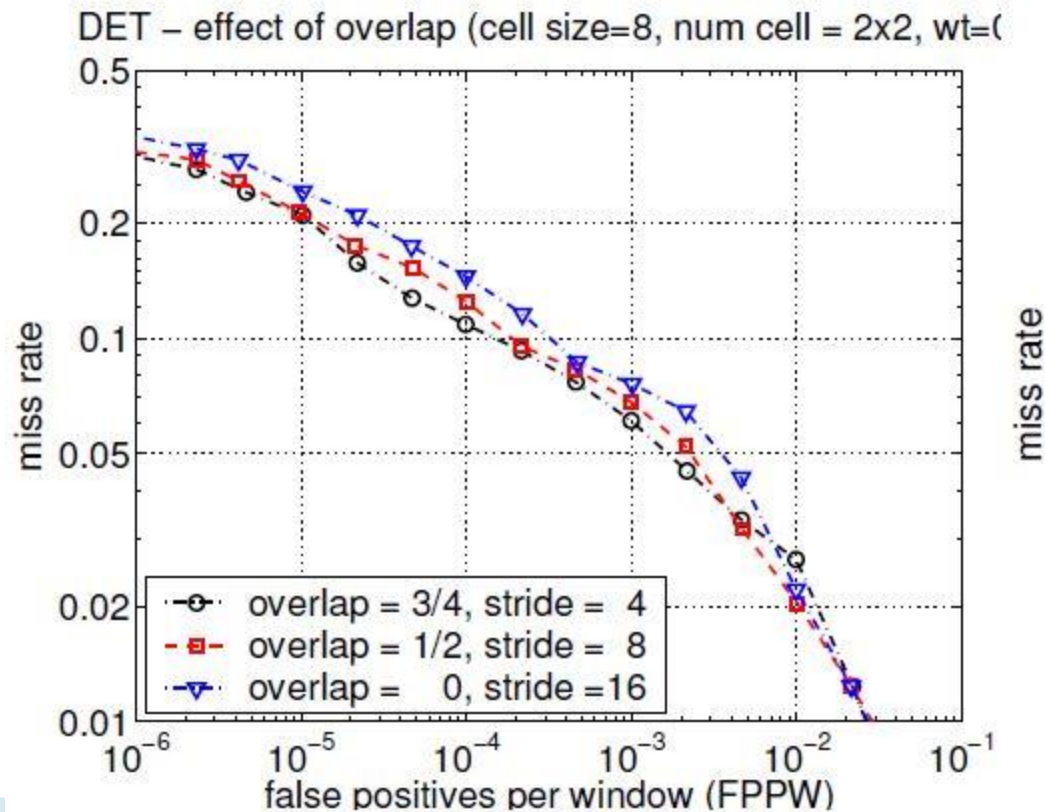
Effect of kernel width on the kernel SVM

- Using a Gaussian kernel SVM improves the performance by about 3%.



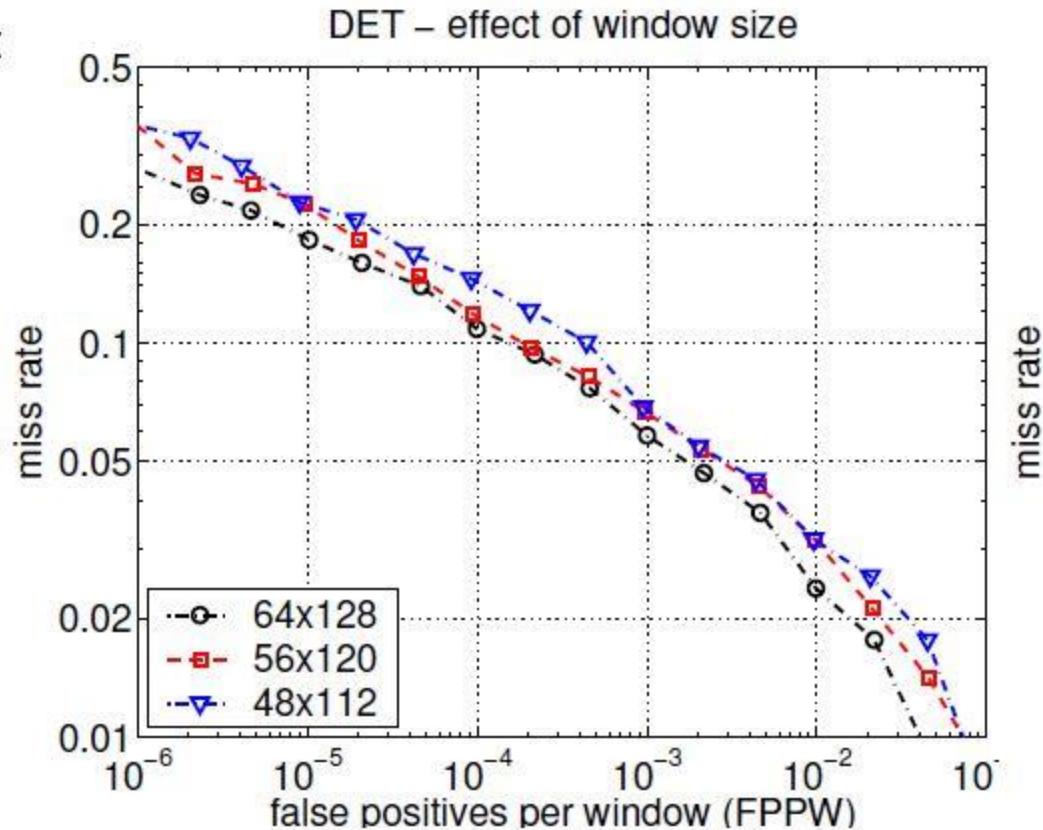
Effects of the overlapping

- Using overlapping descriptor blocks decreases the miss rate by around 5%.



Effect of the window size

- Reducing the 16 pixel margin around the 64×128 detection window decreases the performance about 3%.



Data set and methodology

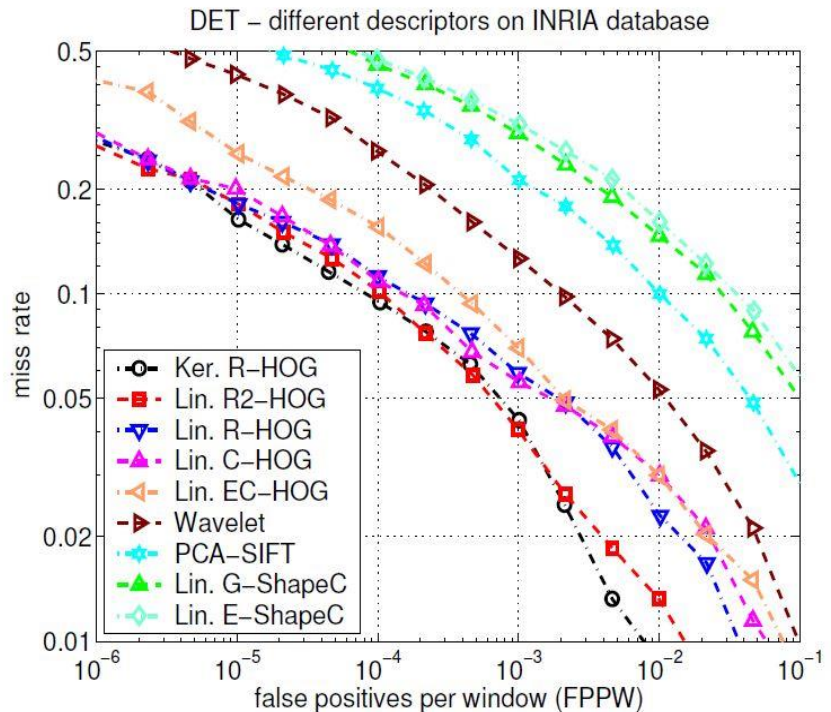
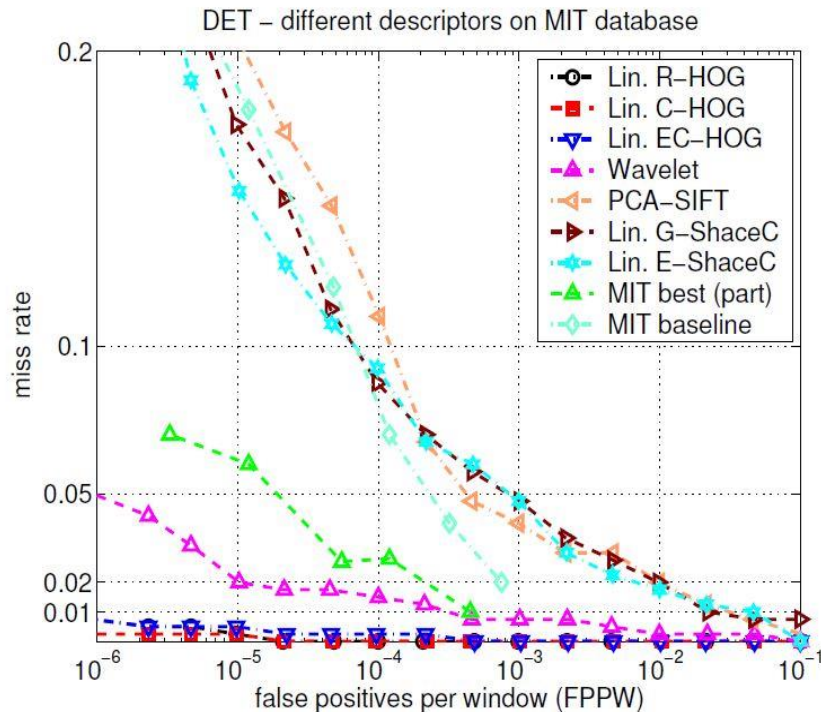
- The detector is tested on two data sets
 - MIT pedestrian (509 training and 200 test images)
 - INRA designed by authors (1805 64×128 images)
- The following images are from the INRA



Images of INRA data set

Applying descriptors on databases

- The HOG-based detectors greatly outperform the wavelet, PCA-SIFT and shape context.



Conclusion

- Very popular for the human detection.
 - Outperform previous works such as SIFT, shape context and wavelet.
 - Introducing a new and more challenging pedestrian database.
 - We should pay if we want to use SIFT detection due to being patent
-
- It has problems with occlusion.
 - Just to detect the entire body (using Deformable Parts Module)

