The proposed descriptors are reminiscent of edge orientation histograms [, SIFT descriptors and shape contexts, but they are computed on a dense grid of uniformly spaced cells and they use overlapping local contrast normalizations for improved performance.

Test case is “pedestrian detection”

METHOD:
The basic idea is that local object appearance and shape can often be characterized rather well by the distribution of local intensity gradients or edge directions, even without precise knowledge of the corresponding gradient or edge positions. In practice this is implemented by dividing the image window into small spatial regions (“cells”), for each cell accumulating a local 1-D histogram of gradient directions or edge orientations over the pixels of the cell. The combined histogram entries form the representation.

Contrast normalisation can be done by accumulating a measure of local histogram “energy” over somewhat larger spatial regions (“blocks”) and using the results to normalize all of the cells in the block. We will refer to the normalized descriptor blocks as Histogram of Oriented Gradient (HOG) descriptors.

Method reached maturity when combined with local spatial histogramming and normalization in Lowe’s Scale Invariant Feature Transformation (SIFT) approach to wide baseline image matching, in which it provides the underlying image patch descriptor for matching scaleinvariant keypoints.

IMPLEMENTATION:
INTRO:
default detector which has the following properties, described below: RGB colour space with no gamma correction; [−1, 0, 1] gradient filter with no smoothing; linear gradient voting into 9 orientation bins in 0◦–180◦; 16×16 pixel blocks of four 8×8 pixel cells; Gaussian spatial window with σ = 8pixel; L2-Hys (Lowe-style clipped L2 norm) block normalization; block spacing stride of 8 pixels (hence 4-fold coverage of each cell); 64×128 detection window;

GAMMA/COLOUR NORM.:
RGB and LAB colour spaces give comparable results, but restricting to grayscale reduces performance by 1.5% at 10^{-4} FPPW (False positives per window). Square root gamma compression of each colour channel improves performance at low FPPW (by 1% at 10^{-4} FPPW) but log compression is too strong and worsens it by 2% at 10^{-4} FPPW

GRADIENT COMPUTATION:
Several smoothing scales were tested includingσ=0 (none). Masks tested included various 1-D point derivatives as well as 3×3 Sobel masks Simple 1-D masks at σ=0 work best. Using larger masks always seems to decrease performance, and smoothing damages it significantly. Using uncentred derivative masks also decreases performance, presumably because orientation estimation suffers as a result of the x and y filters being based at different centres.

SPATIAL/ORIENTATION BINNING:
Each pixel calculates a weighted vote for an edge orientation histogram channel based on the orientation of the gradient element centred on it, and the votes are accumulated into orientation bins over local spatial regions that we call cells. Cells can be either rectangular or radial (log-polar sectors). The orientation bins are evenly spaced over 0◦– 180◦ (“unsigned” gradient) or 0◦–360◦ (“signed” gradient). To reduce aliasing, votes are interpolated bilinearly between the neighbouring bin centres in both orientation and position. The vote is a function of the gradient magnitude at the pixel, either the magnitude itself, its square, its square root, or a clipped form of the magnitude representing soft presence/absence of an edge at the pixel. In practice, using the magnitude itself gives the best results.

RECTANGULAR HOG:
They are computed in dense grids at a single scale without dominant orientation alignment and used as part of a larger code vector that implicitly encodes spatial position relative to the detection window, whereas SIFT’s are computed at a sparse set of scale-invariant key points, rotated to align their dominant orientations, and used individually. SIFT’s are optimized for sparse wide baseline matching, R-HOG’s for dense robust coding of spatial form. We usually use square R-HOG’s, i.e. ς×ς grids of η×η pixel cells each containing β orientation bins, where ς, η, β are parameters.

CIRCLULAR HOG:
Circular block (C-HOG) descriptors are reminiscent of Shape Contexts except that, crucially, each spatial cell contains a stack of gradient-weighted orientation cells instead of a single orientation-independent

edgepresence count. The log-polar grid was originally suggested by the idea that it would allow fine coding of nearby structure to be combined with coarser coding of wider context, and the fact that the transformation from the visual field to the V1 cortex in primates is logarithmic. However small descriptors with very few radial bins turn out to give the best performance, so in practice there is little inhomogeneity or context. It is probably better to think of C-HOG's simply as an advanced form of centre-surround coding.

BLOCK NORMALISATION:
Let **v** be the unnormalized descriptor vector, **v**k be its k-norm for k=1, 2, and be a small constant.
The schemes are: (a) L2-norm, **v** → **v**/**v**22 + 2; (b) L2-Hys, L2-norm followed by clipping (limiting the maximum values of **v** to 0.2) and renormalizing;(c) L1-norm, **v** → **v**/(**v**1 +); and (d) L1-sqrt, L1-norm followed by square root **v** → **v**/(**v**1 +), which amounts to treating the descriptor vectors as probability distributions and using the Bhattacharya distance between them. L2-Hys, L2-norm and L1-sqrt all perform equally well, while simple L1-norm reduces performance by 5%, and omitting normalization entirely reduces it by 27%, at 10−4 FPPW.

CENTRE NORMALISATION:
Alternative centre-surround style cell normalization scheme, in which the image is tiled with a grid of cells and for each cell the total energy in the cell and its surrounding region (summed over orientations and pooled using Gaussian weighting) is used to normalize the cell. However this decreases performance relative to the corresponding block based scheme