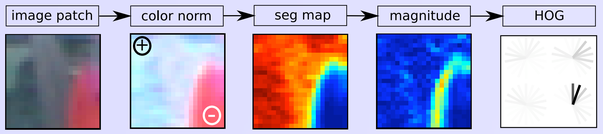
Implicit Color Segmentation Features for Pedestrian and Object Detection

Patrick Ott and Mark Everingham {ott | me}@comp.leeds.ac.uk

UNIVERSITY OF LEEDS

Contribution

We address the problem of pedestrian detection in still images, proposing a novel feature extraction scheme 'CHOG' which computes implicit 'soft segmentations' of image regions into foreground and background. The method, incorporated in a sliding window framework, uses HOG features to characterize the local segmentations.



Soft Segmentation

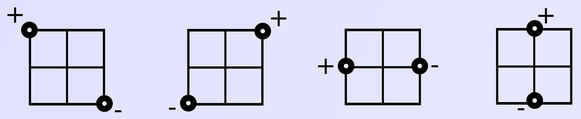
Given reference points $\mathbf{p}_1, \mathbf{p}_2$ hypothesized as lying on foreground and background respectively we compute a projection maximizing the separation between foreground and background:

$$\hat{\mathbf{w}} \propto (\mathbf{m}_2 - \mathbf{m}_1)$$

where $\mathbf{m}_1$ and $\mathbf{m}_2$ are the image means in the neighborhood of the reference points. This is the Fisher discriminant assuming the covariance in each region is the identity matrix. Projecting the original image pixels $\{\mathbf{x}\}$ by $\hat{\mathbf{w}}^T\mathbf{x}$ gives a 'soft segmentation' image S. This soft segmentation is described by a HOG descriptor, requiring the gradients of S. Denoting the original image I and the projected image S the gradients are:

$$\frac{\partial}{\partial x} S = \hat{\mathbf{w}}^T \left\langle \frac{\partial}{\partial x} I_R, \frac{\partial}{\partial x} I_G, \frac{\partial}{\partial x} I_B \right\rangle$$

Since the projection is linear, the gradients of the soft segmentation for any pair of reference points can be computed without needing to apply the convolution operator to the projected image.



We define $\mathbf{p}_1$ $(-)$ and $\mathbf{p}_2$ $(+)$ relative to individual HOG blocks, thus adapting to local foreground/background color distributions within a window.

[1] P. Ott, M. Everingham Implicit Color Segmentation Features for Pedestrian and Object Detection. *ICCV 2009*

Color Normalization

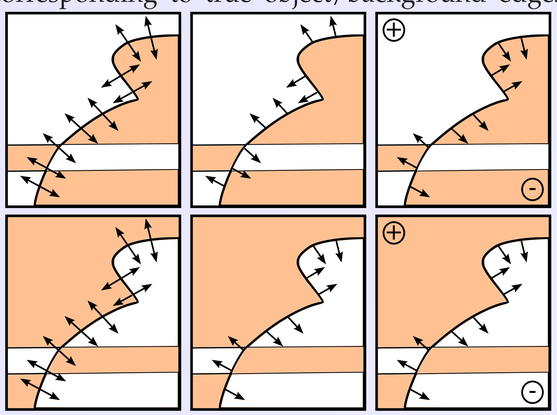
When using RGB pixels to form the soft segmentation, a significant component of the discriminant projections is difference in intensity. We therefore use an intensity-invariant color normalization. Prior to computation of projections and gradients, the RGB image pixels $\mathbf{x} = \langle x_r, x_g, x_b \rangle$ are normalized thus:

$$\mathbf{x}' = \frac{\mathbf{x}}{\max(x_r, x_g, x_b) + \epsilon}$$

Columns from left to right: (i) original image; (ii) gradient magnitude of original image; (iii) color normalized image; (iv) gradient magnitude of normalized image.

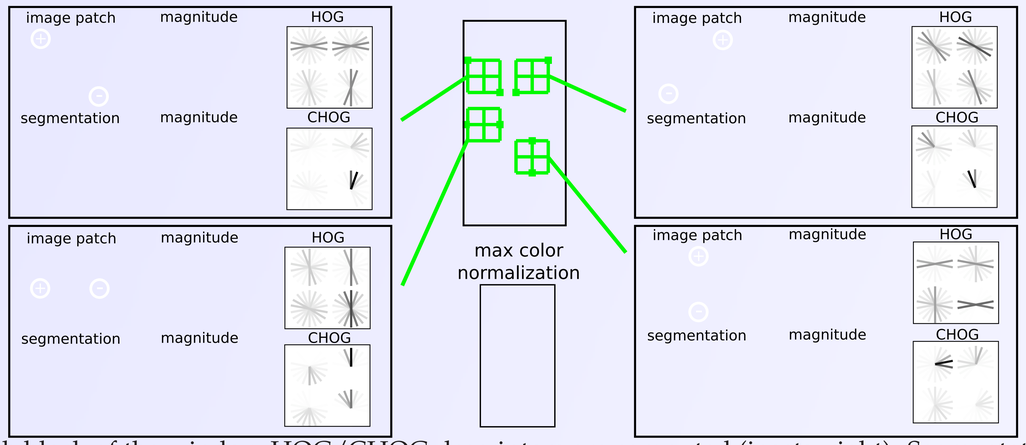
Discussion: Gradient Sign

Given locally coherent background intensity CHOG captures coherence of gradient sign in contrast to HOG which cannot distinguish between gradient orientations 180 degrees apart. As a result CHOG is able to distinguish between noisy edges in background regions and coherent gradients corresponding to true object/background edges.



HOG unsigned HOG signed CHOG signed

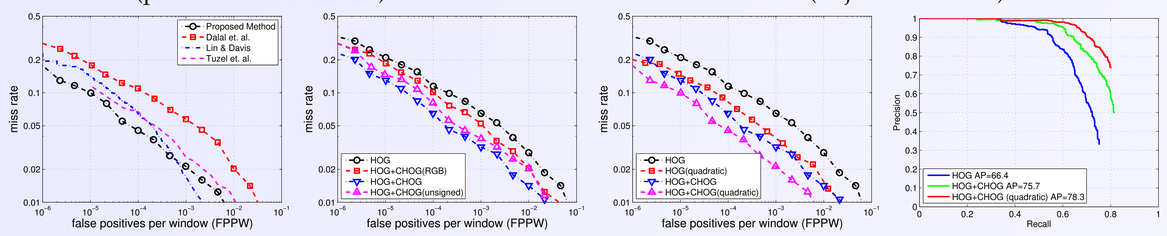
Discussion: Segmentation Maps



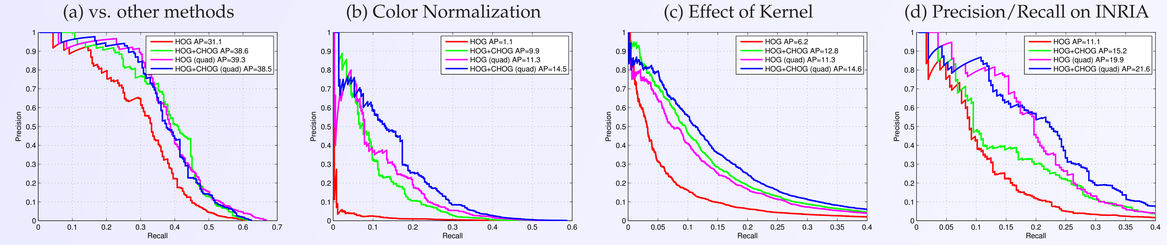
For each block of the window HOG/CHOG descriptors are computed (insets, right). Segmentation maps for CHOG features are shown (insets, bottom left). Note how in the top left and top right insets CHOG gives a strong response to the boundary of the coat, and suppresses responses to shading and background clutter. Overall, CHOG features emphasize the edges of the pedestrian while attenuating shading on the clothing and clutter edges in the background.

Experiments

We concatenate CHOG and HOG features into one vector and use a SVM with linear and quadratic kernel as a classifier. Bootstrapping is used to extract 'hard' negatives. We evaluate the method on the INRIAPerson dataset (pedestrian detection) and the PASCAL VOC 2010 dataset (object detection).



(a) vs. other methods (b) Color Normalization (c) Effect of Kernel (d) Precision/Recall on INRIA



(e) PASCAL Bus (f) PASCAL Sheep (g) PASCAL Person (h) PASCAL Motorbike

At 10^{-4} FPPW (a) miss rate improves by $\sim 30\%$ *c.f.* the next best method; (b) signed gradients account for $\sim 30\%$ relative reduction in miss rate; (c) a quadratic kernel improves miss rate substantially; (d) we improve Average Precision by $\sim 18\%$; (e)–(h) substantial improvements for various VOC 2010 classes.

	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	pers	mbike	plant	sheep	sofa	train	tv
HOG	13.5	22.4	0.1	0.1	7.8	22.6	12.5	0.2	1.3	1.3	0.9	0.2	2.1	5.6	11.5	2.0	0.8	0.7	1.4	21.5
HOG+CHOG	21.3	22.8	0.4	0.6	9.1	24.7	15.4	1.6	3.0	2.1	0.4	1.4	8.6	11.8	18.6	3.1	9.6	1.4	6.4	23.7
HOG (quad)	**29.0**	23.5	2.8	**1.1**	**11.2**	27.8	**19.4**	2.5	4.4	3.6	**2.3**	3.7	**12.8**	10.3	19.0	3.6	7.9	**2.6**	12.9	22.8
H+C (quad)	24.6	**23.8**	**3.0**	1.0	10.6	**28.7**	19.0	**3.1**	**5.4**	**4.9**	**1.7**	**3.9**	12.1	**13.4**	**21.4**	**4.0**	**12.6**	2.3	**14.0**	**24.3**

We show substantial improvements in Average Precision for all VOC 2010 classes. Combining HOG and CHOG improves results for 19 of 20 classes. Utilizing a quadratic kernel generally improves Average Precision further.