Sparsity Potentials for Detecting Objects with the Hough Transform

Nima Razavi[1]
nrazavi@vision.ee.ethz.ch

Nima Sedaghat Alvar[2]
n.sedaghat@gmail.com

Juergen Gall[3]
juergen.gall@tue.mpg.de

Luc van Gool[1,4]
luc.vangool@esat.kuleuven.be

[1]Computer Vision Laboratory
ETH Zurich, Switzerland

[2]Machine Vision and Intelligence Lab
Sharif University of Technology, Iran

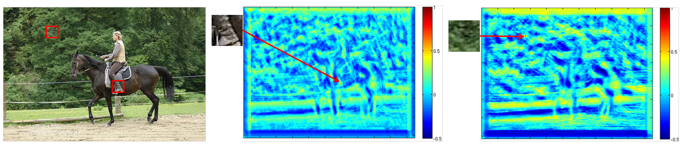
[3]Perceiving Systems Department
MPI for Intelligent Systems, Germany

[4]IBBT/ESAT-PSI
K.U. Leuven, Belgium

Hough transform based object detectors divide an object into a number of patches and combine them using a shape model. For efficient combination of patches into the shape model, the individual patches are assumed to be independent of one another. Although this independence assumption is key for fast inference, it requires the individual patches to have a high discriminative power in predicting the class and location of objects. In this paper, we make the following two observations:

- the similarity in appearance of patches in a neighborhood of a central patch exhibit different sparsity values when the central patch appears on an object as opposed to a background region.
- the codebook entries associated with texture or simple edge patterns are consistently less sparse in their neighborhood as opposed to entries which are associated to more complex patterns (see Fig. 1).

Based on these observations, we argue that the sparsity of the appearance of a patch in its neighborhood can be a very powerful measure to increase the discriminative power of a local patch and incorporate it as a sparsity potential for object detection.



(a) (b) (c)

Figure 1: The patches in an image exhibit different sparsity values. While the self-similarity of a non-texture patch (a) to its neighboring patches is low, the patch on the tree (b) is less sparse and much more similar to its neighborhood. Base on this observation, we introduce a measure which captures the *sparseness* of a patch within its neighborhood and incorporate it as a "sparsity potential" for object detection.

We base our sparsity or distinctiveness measure on self-similarity. Let us assume that we have a metric that measures the similarity of a patch f_i with all patches in its neighborhood, $\{f_n | n \in \mathcal{N}^i\}$, e.g. by Normalized Cross Correlation as in Fig. 1. Further, we assume that the returned similarity is normalized to be in $[0,1]$ with 1 representing the most similar and 0 the most dissimilar patch. In this case, one is getting a real valued self-similarity vector $\mathbf{ss}_i = (ss_1, \ldots, ss_{|\mathcal{N}^i|})$ where each element ss_n records the normalized similarity of f_n to f_i.

The sparsity of the self-similarity vector $\mathbf{ss}_i$ can be measured in many different ways, e.g., by using entropy or various vector norms. In this work, we use the L1-norm,

$$||\mathbf{ss}_i||_1 = \sum_{n \in \mathcal{N}^i} |ss_n| \tag{1}$$

which is both simple and fast to calculate.

For detecting objects, we incorporate the sparsity measure by training a classifier for each code-word and object class. For training the sparsity classifiers, first a set of features on the validation set, both on objects as well as background, are extracted and are assigned to one or more codebook entries ω_j. Given a neighborhood function $\mathcal{N}^i$, the sparsity measure of every feature f_i is calculated. Next, for each ω_j and class label c, these sparsity measures are collected and used to learn a simple threshold θ_{c,ω_j}. These thresholds are then used to estimate class probability $p(c|\omega_j, \mathcal{N}^i)$ as

$$p(c|\omega_j, \mathcal{N}^i) \propto \begin{cases} p(c|\omega_j) & \text{if} \quad ||\mathbf{ss}_i||_1 \leq \theta_{c,\omega_j} \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

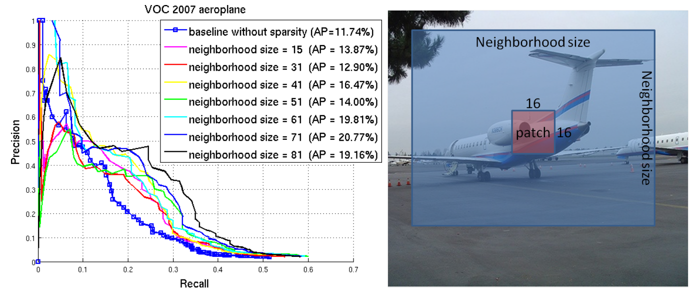


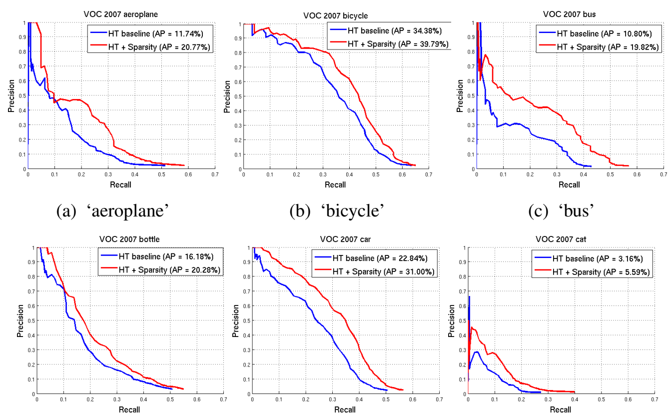
Figure 2: This figure evaluates the effect of the neighborhood size used for calculating the sparsity on the accuracy of the detector. The performance comparison of our Hough Forest baseline [1] with and without sparsity measure potentials is shown. As can be seen, the proposed sparsity potential improves the accuracy. The performance tends to increase with the window size until it saturates at around 71 pixels, almost doubling the average precision (AP) compared to the baseline.The sparsity potential is calculated on a square neighborhood of every 16×16 patch.

where $p(c|\omega_j)$ is the class probability estimated at the codebook entry ω_j.

Using the sparsity measure as a single dimensional feature, the thresholds are learned such as to separate the positive and negatives with the best classification accuracy with zero false negatives on the training data.

The evaluation is carried out on the PASCAL VOC 2007 dataset. Our experiments confirm the benefit of using the proposed sparsity potential for object detection increase the mean average precision (mAP) of our Hough transform baseline [1] from 14.82 to 20.68. Example Precision/Recall curves for some categories of the VOC'07 dataset is shown in Fig. 3.

In the future, it would be interesting to use the sparsity potentials in a multi-class setup to also discriminate classes from one another. Since the self-similar patches tend to belong to the same label, it would be also interesting to incorporate their sparsity as a higher order potential for image segmentation.



(a) 'aeroplane' (b) 'bicycle' (c) 'bus'

(d) 'bottle' (e) 'car' (f) 'cat'

Figure 3: The precision recall curves for some categories of the PASCAL VOC 2007. As can be seen, the proposed sparsity potentials substantially improve the detection performance. The average precision (AP) is calculated by the integral under the curve.

[1] J. J. Gall, A. Yao, N. Razavi, L. Van Gool, and V. Lempitsky. Hough forests for object detection, tracking, and action recognition. *TPAMI*, 33(11):2188–2202, 2011.