Exploring Prior Knowledge for Pedestrian Detection

Yi Yang[1]
yangyi@nlpr.ia.ac.cn
Zhenhua Wang[2]
wzh@ntu.edu.sg
Fuchao Wu[1]
fcwu@nlpr.ia.ac.cn

[1] National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences
Beijing, China

[2] Rapid-Rich Object Search Lab
School of EEE, Nanyang Technological University
Singapore

Pedestrian detection is a classical and hot issue in object detection. Many approaches have been proposed in this area. However, it remains a challenging problem due to the variances in lighting conditions, scene structures, clothes, view angles, postures, scales, occlusions, *etc*.

Previous survey [1] has summarized that using better features plays an important role in improving detection quality. In addition, prior knowledge has shown good success in designing haar-like features for pedestrian detection [4]. Inspired by it, our work aims to integrate more prior knowledge into the design of features to enhance performance of pedestrian detection. By observing the pedestrian samples, we have discovered several important priors that are always ignored by previous methods, *e.g.* the symmetry of human body and the differences among different channels. We therefore utilize these priors to design two kinds of features: 1) symmetric features which capture the difference between two local symmetric regions, and 2) cross-channel features which capture the difference between two different channels of the same region. Figure 1 gives some visual examples of these two features. To the best of our knowledge, we are the first to use symmetric and cross-channel priors in designing features for pedestrian detection. By integrating these prior information into feature design, our detector achieves state-of-the-art performance.

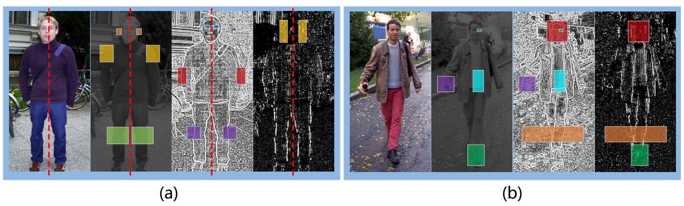

Figure 1: (a) Symmetric features. Rectangles with the same color in the same channel represent one symmetric feature. (b) Cross-channel features. Rectangles with the same color in different channels represent one cross-channel feature.

Symmetric Features: A pedestrian body shape model can be obtained by computing an average edge map based on gradient magnitudes extracted from a large number of training samples [4]. The shape model seems like a fairly standard silhouette of a person standing facing the front, with hands falling naturally on body sides and feet keeping naturally together. It is worth to mention that, this fairly standing body silhouette is symmetrical about the vertical line at the center of the model, as the red dashed lines shown in Figure 1. This symmetry should be a good prior for designing effective features.

Based on this observation, we design a kind of symmetric features to capture the symmetric prior. For convenient implementation and efficient computation, we constrain our feature templates to be rectangles. More specifically, we define the symmetric features to be second-order rectangle features which are composed of two separate local symmetric rectangular regions. The two local symmetric regions share the same size and should be in the same channel. See Figure 1(a) for some examples. The feature values can be effectively computed as the difference between the responses of these two symmetric regions by using integral image.

To enrich the symmetric information, we actually take consideration of 3 more symmetry axes with 0°, 45°, and 135°symmetry angles (see Figure 2) in our implementation. In addition, as the whole human body is not symmetrical about these three additional axes, we actually divide the model region into 4 subregions and generate symmetric rectangles in these subregions to capture the local symmetric information.

Cross-channel Features: In recent pedestrian detectors with multiple channels, *e.g.* HOG+LUV channels, different channels contain different kinds of information. Previous methods mainly use features that

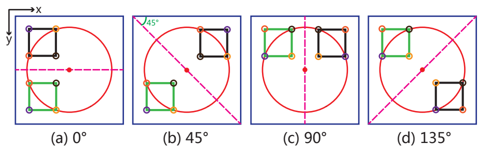

Figure 2: Some simple examples of symmetric rectangles.

capture the information in a single channel, *e.g.*, `ChnFtrs` only takes responses of rectangles in one channel at a time. Such features inevitably lose information cross different channels. To deal with this problem, we propose a kind of cross-channel features to capture such valuable information by comparing the responses of the same rectangle in different channels. Note that the cross-channel features are also second-order features. Differently, the two rectangles share the same size and position but locate in different channels. The difference of responses between these two rectangles is recorded as feature value. To make the values of different channels comparable, it is required to normalize all the channels before computing such features. See Figure 1(b) for some examples.

Generating Feature Pool: The details for generating the proposed symmetric and cross-channel features are described as follows. Firstly, we denote a rectangle as a 4 dimensional vector $\mathbf{r} = (x,y,w,h)$ in which x, y are the coordinates of the top-left vertex and w, h are the width and height respectively. Then, the valid rectangle set $\mathcal{R}$ is defined as the set of all possible rectangles that are inside of the model region and larger than a predefined area threshold S. In an usual image coordinate system, *e.g.* with the origin being the top-left vertex of the model region, the positive x-axis pointing right and the positive y-axis pointing down, such a set can be represented as:

$$\mathcal{R} = \{\mathbf{r_i} : x_i \leq W - w_i, y_i \leq H - h_i, w_i \leq W, h_i \leq H, w_i \times h_i \geq S, x_i, y_i, w_i, h_i \in \mathbb{N}\} \quad (1)$$

where W and H are the width and height of the model region respectively. With the valid rectangle set, we can further generate a pool of rectangular templates. For symmetric features, we select two symmetric rectangles $\mathbf{r_i}$ and $\mathbf{r_j}$ from $\mathcal{R}$ and randomly generate their channel index. For cross-channel features, we select one rectangle $\mathbf{r_i} \in \mathcal{R}$ and randomly generate two different channel indexes.

Experiments: We conduct our experiments on two public benchmark data sets: the INRIA [2] and Caltech-USA [3] pedestrian data sets. For Caltech-USA data set, we conduct a dense sampling of the training data (every 4 frames) and report our results on the "reasonable", "scale" and "occlusion" subsets of the test set. Our detector achieves three best performances among all the tested detectors and six best performances among the detectors without using motion features.

[1] Rodrigo Benenson, Mohamed Omran, Jan Hosang, and Bernt Schiele. Ten years of pedestrian detection, what have we learned? *In ECCV, CVRSUAD workshop*, 2014.

[2] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.

[3] Piotr Dollár, Christian Wojek, Bernt Schiele, and Pietro Perona. Pedestrian detection: A benchmark. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 304–311. IEEE, 2009.

[4] Shanshan Zhang, Armin B. Cremers, and Christian Bauckhage. Informed haar-like features improve pedestrian detection. *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 947 – 954, 2014.