

Supplemental Material

Robust Multi-Resolution Pedestrian Detection in Traffic Scenes

Junjie Yan Xucong Zhang Zhen Lei Shengcai Liao Stan Z. Li
 Center for Biometrics and Security Research & National Laboratory of Pattern Recognition
 Institute of Automation, Chinese Academy of Sciences, China
 {jjyan, xc Zhang, zlei, scliao, szli}@nlpr.ia.ac.cn

1. The Detailed Optimization Procedure for Multi-Task DPM

In section 3.2, we omit some derivations for the space limitation. More details are provided here. The coordinate descent procedure is motivated by [4].

1.1. Optimize W_a and w_s

When P_H and P_L are fixed, we use the following derivations to transform the multi-task problem to be a standard latent SVM problem for optimizing W_a and w_s . To combine f_{I_H} and f_{I_L} , we denote $P_H P_H^T + P_L P_L^T$ as A , and $A^{\frac{1}{2}} W_a$ as \widetilde{W}_a . Here A is a symmetric nonnegative definite matrix, and its square root is denoted as $A^{\frac{1}{2}}$. We use resolution aware transformations, for high resolution samples, $\widetilde{\Phi}_a(I_n, L_n^*)$ stands for $A^{-\frac{1}{2}} P_H \Phi_a(I_n, L_n^*)$, and for low resolution samples, $\widetilde{\Phi}_a(I_n, L_n^*)$ stands for $A^{-\frac{1}{2}} P_L \Phi_a(I_n, L_n^*)$. With these notions, the optimization problem becomes to be:

$$\arg \min_{\widetilde{W}_a, w_s} \frac{1}{2} \|\widetilde{W}_a\|_F^2 + \frac{1}{2} w_s^T w_s + C \sum_{N_H+N_L} \max_{L_n^*} [0, 1 - y_n(\text{Tr}(\widetilde{W}_a^T \widetilde{\Phi}_a(I_n, L_n^*)) + w_s^T \phi_s(L_n^*))] \quad (1)$$

Eq. 1 equals to:

$$\arg \min_{\widetilde{W}_a, w_s} \frac{1}{2} \text{Vec}(\widetilde{W}_a)^T \text{Vec}(\widetilde{W}_a) + \frac{1}{2} w_s^T w_s + C \sum_{N_H+N_L} \max_{L_n^*} [0, 1 - y_n(\text{Vec}(\widetilde{W}_a)^T \text{Vec}(\widetilde{\Phi}_a(I_n, L_n^*)) + w_s^T \phi_s(L_n^*))] \quad (2)$$

where $\text{Vec}(\cdot)$ is used to reshape a matrix to a column vector. Denoting $[\text{vec}(\widetilde{W}_a); w_s]$ as w , and $[\text{Vec}(\widetilde{\Phi}_a(I_n, L_n^*)); \phi_s(L_n^*)]$ as $\phi(I_n, L_n^*)$, the Eq. 2 is:

$$\arg \min_w \frac{1}{2} w^T w + C \sum_{N_H+N_L} \max_{L_n^*} [0, 1 - y_n(w^T \phi(I_n, L_n^*))] \quad (3)$$

which is a standard Latent SVM problem discussed in [2]. It takes the part locations as latent variable. In solving the problem with latent variable, we use the solver in [2], which follows the EM-like procedure, that infers the latent part location and optimizes the detection model iteratively.

1.2. Optimize P_H and P_L

When W_a and w_s are fixed, We transform the optimization problem to a standard SVM problem. Since f_{I_H} , f_{I_L} are independent with the fixed $\{W_a, w_s\}$ and they are of the same form, here we only show how to optimize f_{I_H} . Denoting $W_a W_a^T$ as A , $A^{\frac{1}{2}} P_H$ as \widetilde{P}_H , and $A^{-\frac{1}{2}} W_a \Phi_a(I_{H_n}, L_n^*)^T$ as $\widetilde{\Phi}_a(I_{H_n}, L_n^*)$, the optimization problem becomes to be:

$$\arg \min_{\widetilde{P}_H} \frac{1}{2} \|\widetilde{P}_H\|_F^2 + C \sum_{N_H} \max [0, 1 - y_n(\text{Tr}(\widetilde{P}_H^T \widetilde{\Phi}_a(I_{H_n}, L_n^*)) + w_s^T \phi_s(L_n^*))] \quad (4)$$

which equals to:

$$\arg \min_{\widetilde{P}_H} \frac{1}{2} \text{Vec}(\widetilde{P}_H)^T \text{Vec}(\widetilde{P}_H) + C \sum_{N_H} \max[0, 1 - y_n(\text{Vec}(\widetilde{P}_H)^T \text{Vec}(\widetilde{\Phi}_a(I_{H_n}, L_n^*)) + w_s^T \phi_s(L_n^*))] \quad (5)$$

In this step, we don't take L_n^* as latent variable any more, and infer them by the detection model parameterized by $\{P_H^{k-1}, P_L^{k-1}, W_a^k, w_s^k\}$. Then w_s and $\phi_s(L_n^*)$ are fixed, so that $w_s \phi_s(L_n^*)$ can be taken as a real value. We denote $[\text{Vec}(\widetilde{P}_H); 1]$ as w , and $[\text{Vec}(\widetilde{\Phi}_a(I_{H_n}, L_n^*)); w_s^T \Phi_s(L_n^*)]$ as $\phi(I, L_n^*)$. It equals to the standard SVM problem:

$$\arg \min_w \frac{1}{2} w^T w + C \sum_{N_H} \max[0, 1 - y_n(w^T \phi(I_n, L_n^*))] \quad (6)$$

1.3. Coordinate Descent Procedure

Data: The annotated high resolution samples (taller than 80 pixels in our paper) I_H , and annotated low resolution samples (30-80 pixels high) I_L .

1. **Result:** The DPM detector parameterized by W_a and w_s in the mapped common subspace, the resolution aware transformations P_H, P_L .
- 2 Calculate PCA on the HOG features of high and low resolution data, and take the first n_d eigenvectors as the initial values for P_H^0 and P_L^0 , respectively;
- 3 **for** $k \leftarrow 1$ **to** K **do**
- 4 Fix P_H^{k-1} and P_L^{k-1} to optimize W_a^k and w_s^k with Eq. 1;
- 5 Infer the L_n^* for every training samples with the parameters $\{P_H^{k-1}, P_L^{k-1}, W_a^k, w_s^k\}$ by the resolution aware detection model;
- 6 Fix W_a^k and w_s^k to update P_H^k and P_L^k by Eq. 4, respectively;
- 7 **end**

Algorithm 1: The coordinate descent procedure for Multi-Task DPM

1.4. More Resolution Partitions

Two resolution partitions are used in the paper, but extending the proposed model to more resolution partitions is straightforward. For R partitions, the objective function of MT-DPM can be written as:

$$\arg \min_{W_a, w_s, \mathcal{P}} \frac{1}{2} w_s^T w_s + \sum_{r=1}^R f_{I_r}(W_a, w_s, P_r) \quad (7)$$

where $\mathcal{P} = \{P_1, P_2, \dots, P_R\}$ and the $f_{I_r}(W_a, w_s, P_r)$ is:

$$f_{I_r}(W_a, w_s, P_r) = \frac{1}{2} \|P_r^T W_a\|_F^2 + C \sum_{N_r} \max[0, 1 - y_n(\text{Tr}(W_a^T P_r \Phi_a(I_{r_n}, L_n^*)) + w_s^T \phi_s(L_n^*))] \quad (8)$$

2. Quantitative Results on Caltech Pedestrian Testing Data

The curves are generated under the standard experimental settings advised in [1]. Results of other methods except for the proposed methods can be found at the website of the Caltech Pedestrian Benchmark:

http://www.vision.caltech.edu/Image_Datasets/CaltechPedestrians/files/USA-test-plots.pdf. Although the proposed method are mainly for multi-resolution handling, we find the dramatic performance improvements on all the 9 sub-experiments. The improvements on "Medium scale", "Typical aspects ratios", "No-occlusion", "Partial-occlusion" and "Heavy-occlusion" are more than 9%, and more than 6% on "Overall", "Atypical aspect ratios". Some methods achieved good performance on "Near scale", but our method (MT-DPM+Context) improves 3% further. For the pedestrians shorter than 30 pixels, all the methods performed poorly for the limit useful information and our methods improve 2% in this condition.

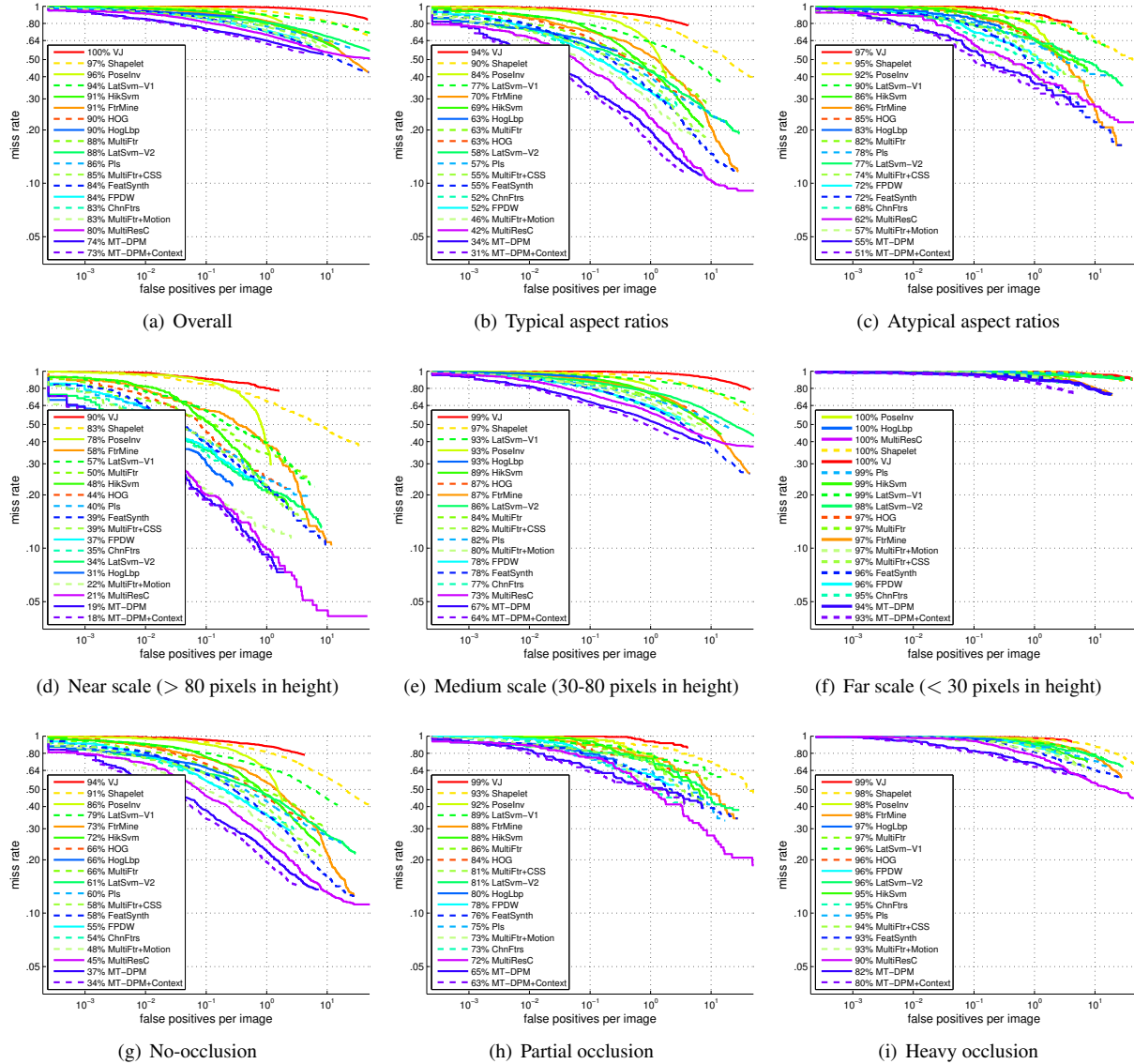


Figure 1. Miss rates versus false positive per-image curves of different conditions on the Caltech Pedestrian testing data. Lower curves mean better performance. (a) Overall performance. (b-c) Performance w.r.t. aspect ratio (computed for un-occluded pedestrians taller than 50 pixels). (d-f): Performance w.r.t. scale (computed for un-occluded pedestrians). (g-i): Performance under varying levels of occlusion (computed for pedestrians taller than 50 pixels).

3. Qualitative Results

The qualitative results are demonstrated in the video attached. In “QualitativeComparisons-set07-view000.avi”, we show the ground truth and the detections of our proposed “MT-DPM+ Context” on set07-view000. The detections are evaluated at every 30th frame, following the protocol advised in [1]. For comparison, we show the results of two state-of-the-art methods: “Multi-ResC” [3] and “MultiFtr+Motion” [5]. We also run a continuous video and show it in “QualitativeDetetionResult-s.avi”.

References

[1] P. Dollár, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection: An evaluation of the state of the art. *TPAMI*, 2012.

- [2] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *TPAMI*, 2010. 1
- [3] D. Park, D. Ramanan, and C. Fowlkes. Multiresolution models for object detection. *ECCV*, 2010. 3
- [4] H. Pirsiavash and D. Ramanan. Steerable part models. In *CVPR*. IEEE, 2012. 1
- [5] S. Walk, N. Majer, K. Schindler, and B. Schiele. New features and insights for pedestrian detection. In *CVPR*. IEEE, 2010. 3