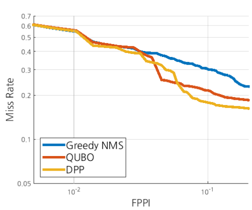
Individualness and Determinantal Point Processes for Pedestrian Detection: Supplementary Material

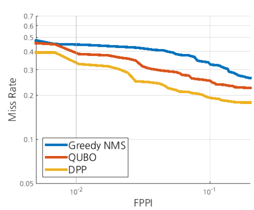
Donghoon Lee[1], Geonho Cha[1], Ming-Hsuan Yang[2], and Songhwai Oh[1]

[1]Electrical and Computer Engineering, Seoul National University, Korea
[2]Electrical Engineering and Computer Science, University of California at Merced
{donghoon.lee, geonho.cha}@cpslab.snu.ac.kr, mhyang@ucmerced.edu,
songhwai@snu.ac.kr

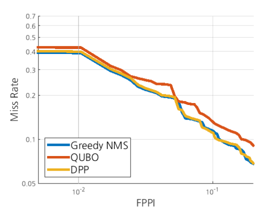
In this supplementary material, we describe experimental results which were not included in the paper due to the page limitation. Comparisons of NMS, QUBO, and the proposed method are provided in Section 1. A sensitivity analysis of parameters is reported in Section 2. Section 3 shows the effectiveness of the proposed quality and similarity term design by comparing it to other models.



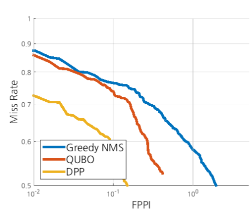
(a) INRIA, DPM



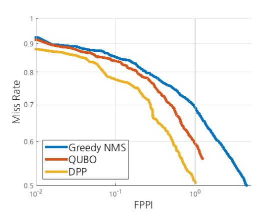
(b) INRIA, ACF



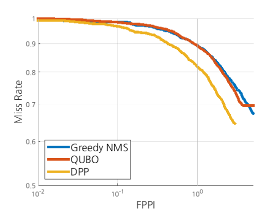
(c) INRIA, faster RCNN



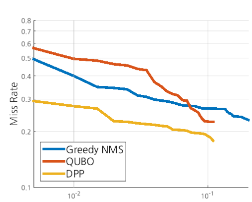
(d) PETS, DPM



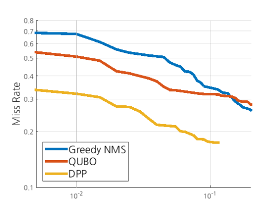
(e) PETS, ACF



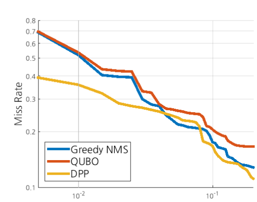
(f) PETS, faster RCNN



(g) Terrace, DPM



(h) Terrace, ACF



(i) Terrace, faster RCNN

Fig. 1. DET curves for all combinations of datasets and detectors.

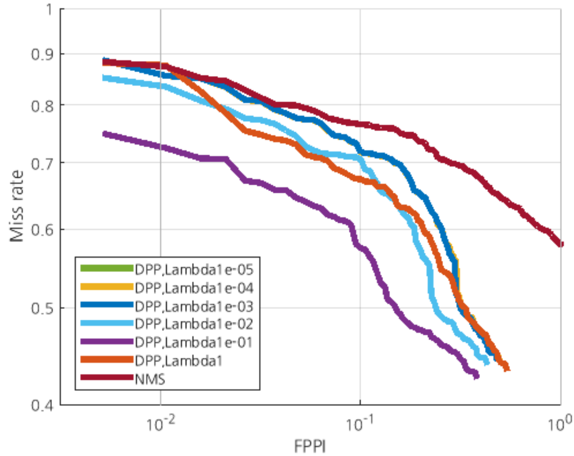


Fig. 2. Sensitivity of analysis λ using the DPM detector and PETS 2009 dataset.

1 DET curves

For each dataset and a pedestrian detector, we compare the performance of greedy NMS, QUBO, and the proposed method based on a DPP. As shown in Figure 1, the proposed method works favorably against other methods for all datasets and detectors.

2 Sensitivity analysis

We use a DPM detector [1] and the PETS 2009 dataset [2] to analyze the sensitivity of parameters in the proposed algorithm. We analyze λ in (4), σ in (6), and w in (12) of the manuscript, and present the results in Figure 2, 3, and 4.

Sensitivity of λ. The results show that λ, which penalizes large detection boxes, is an important parameter. Too large or too small penalties make (4) less effective. We use $\lambda = 0.1$ which gives the best accuracy. Interestingly, the value also gives the best fit to Figure 3(b) in the paper which demonstrates that (4) is a reasonable model. Note that green and yellow lines are almost overlapped by a blue line ($\lambda = 1e - 03$).

Sensitivity of σ. This parameter is a standard deviation of pedestrian height when it is modeled using Gaussian as in (6). We define $\sigma = \widetilde{h}r$ where $\widetilde{h}$ is the expected height as defined in (5) and $0 < r \leq 1$ is a ratio. The best accuracy is obtained around $r = 0.2$. We can see that the prior information becomes less effective as the standard deviation gets bigger.

Sensitivity of w. This parameter is used to mix the appearance individualness and the spatial individualness. The results show that solely using the appearance

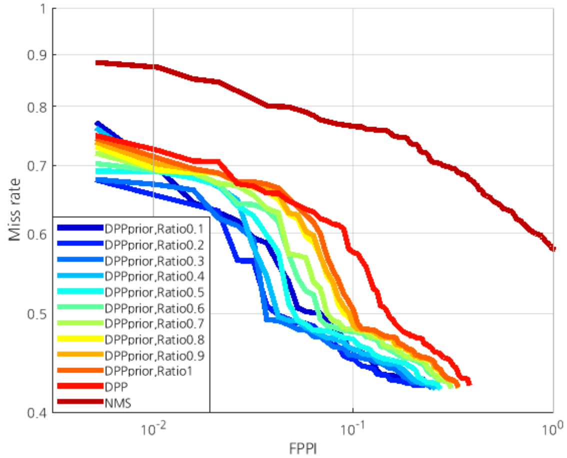


Fig. 3. Sensitivity of analysis σ using the DPM detector and PETS 2009 dataset.

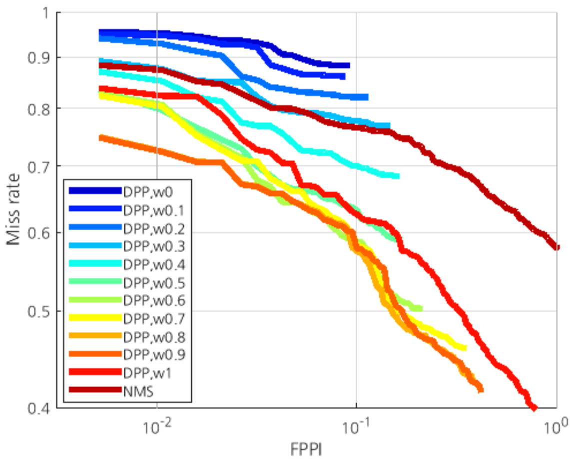


Fig. 4. Sensitivity of analysis w using the DPM detector and PETS 2009 dataset.

individualness is not robust as discussed in the paper. The best accuracy is achieved when $w = 0.8$. Also, note that the proposed algorithm performs better than NMS without the appearance individualness.

3 Effectiveness of the quality and similarity term design

In this section, we demonstrate the effectiveness of the proposed quality and similarity term design by comparing it to [3] which estimates human pose using a structured DPP. We begin by noting three differences between our work and [3]. First and obviously, the problem of detection and pose estimation is different. Therefore, the design of quality term and diversity feature should be changed which is the most critical part of a DPP. Second, they deal with people at similar scale without serious occlusions. On the other hand, our goal is to successfully handle significant scale variations and severe occlusions between pedestrians. Finally, their diversity feature generally does not have a unit norm. It biases their model which can degrade the performance [4]. Our formulation does not suffer from such biases.

Since it is not possible to apply [3] to the detection problem directly, we reformulate their design while preserving its concept. Please refer to [3] for the original design. The quality term is reformulated as follows:

$$q_i = \alpha(s_i^o)^\beta, \tag{1}$$

where α and β are constants and s_i^o is a raw detection score of detection i. The diversity feature is defined as follows:

$$\phi_i^r = f(\frac{dist(d_i, x_r)}{\sigma}), \tag{2}$$

where ϕ_i^r is the r-th element of ϕ_i, f is the standard normal density function, d_i is the i-th detection, x_r is the r-th reference point which is evenly spaced grid on the image, $dist(d_i, x_r)$ is the Euclidean distance between the detection i and the reference point x_r, and σ is a constant.

Figure 5 shows a detection result based on (1) and (2) using a DPM detector on the PETS 2009 dataset. We estimate the detection accuracy by varying the number of reference points from 4 by 2 to 80 by 40. The result shows that even if many number of reference points are used, the detection accuracy is saturated at equal or worse than NMS. It demonstrates that the above design is not effective to detect pedestrians.

References

1. Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part-based models. IEEE Transactions on Pattern Analysis and Machine Intelligence **32**(9) (2010) 1627–1645 2
2. Ferryman, J., Shahrokni, A.: PETS2009: Dataset and challenge. In: Winter-PETS. (2009) 2
3. Kulesza, A., Taskar, B.: Structured determinantal point processes. In: Advances in neural information processing systems. (2010) 1171–1179 3, 4, 5
4. Kulesza, A., Taskar, B.: Determinantal point processes for machine learning. arXiv preprint arXiv:1207.6083 (2012) 4

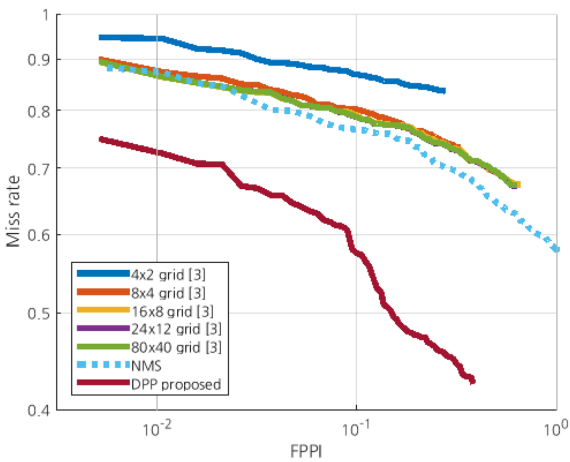


Fig. 5. Detection accuracy using [3].