Detecting Dynamic Objects with Multi-View Background Subtraction

Raúl Díaz, Sam Hallman, Charless C. Fowlkes
Computer Science Department, University of California, Irvine
{rdiazgar,shallman,fowlkes}@ics.uci.edu

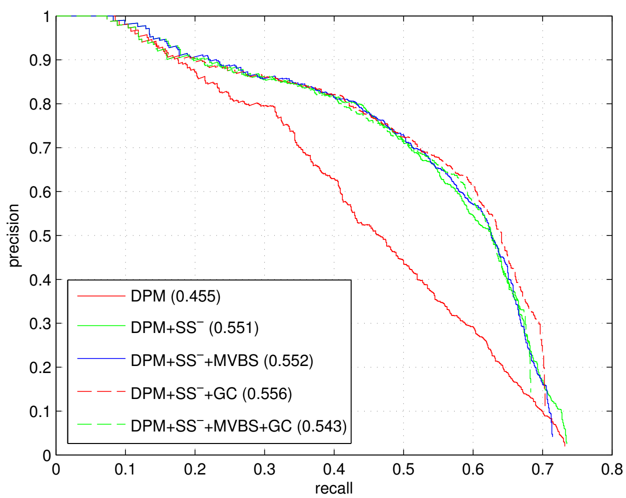


Figure 1: Precision-Recall for pedestrian detection with scene specific DPM detectors. **DPM** is the baseline Deformable Parts Model template detector trained on the PASCAL dataset. **+SS$^-$** is trained using scene-specific negative instances mined in an unsupervised manner from images of Notre Dame. **+GC** prunes detections where the bottom of the detection appears above the horizon based on the camera pose estimated using SfM. **+MVBS** prunes detections whose bounding box contain more than 20% estimated background pixels based on multi-view matching. The scene-specific model performs significantly better than the baseline. Multi-view background-subtraction and geometric consistency both provide additional gains in detector precision although the gains are less than in the case of the DT template detector.

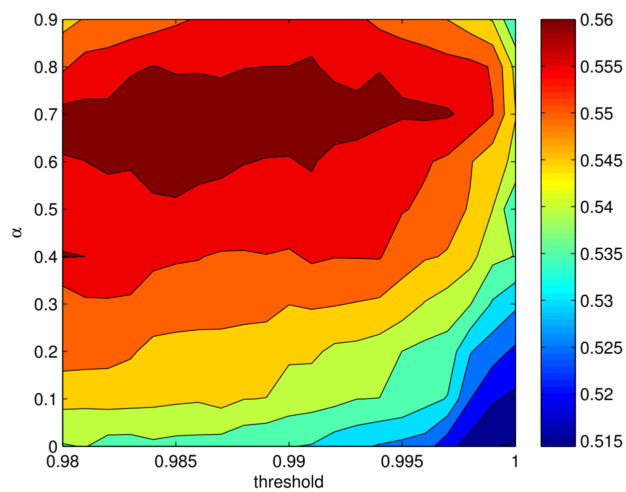


Figure 2: Dependence of detection performance on matching parameters. α indicates the patch matching threshold value used for construction of the background mask. For a given choice of α we also vary the bounding box pruning threshold $q(i) > \tau$ which determines how many candidate bounding boxes are retained. Since the range of usable thresholds τ depends strongly on α, the x-axis "threshold" in the figure corresponds to different settings of this threshold but parameterized by the percentage of true positive training windows kept given a certain background threshold. In general the performance is quite insensitive to the exact choice of α and τ.

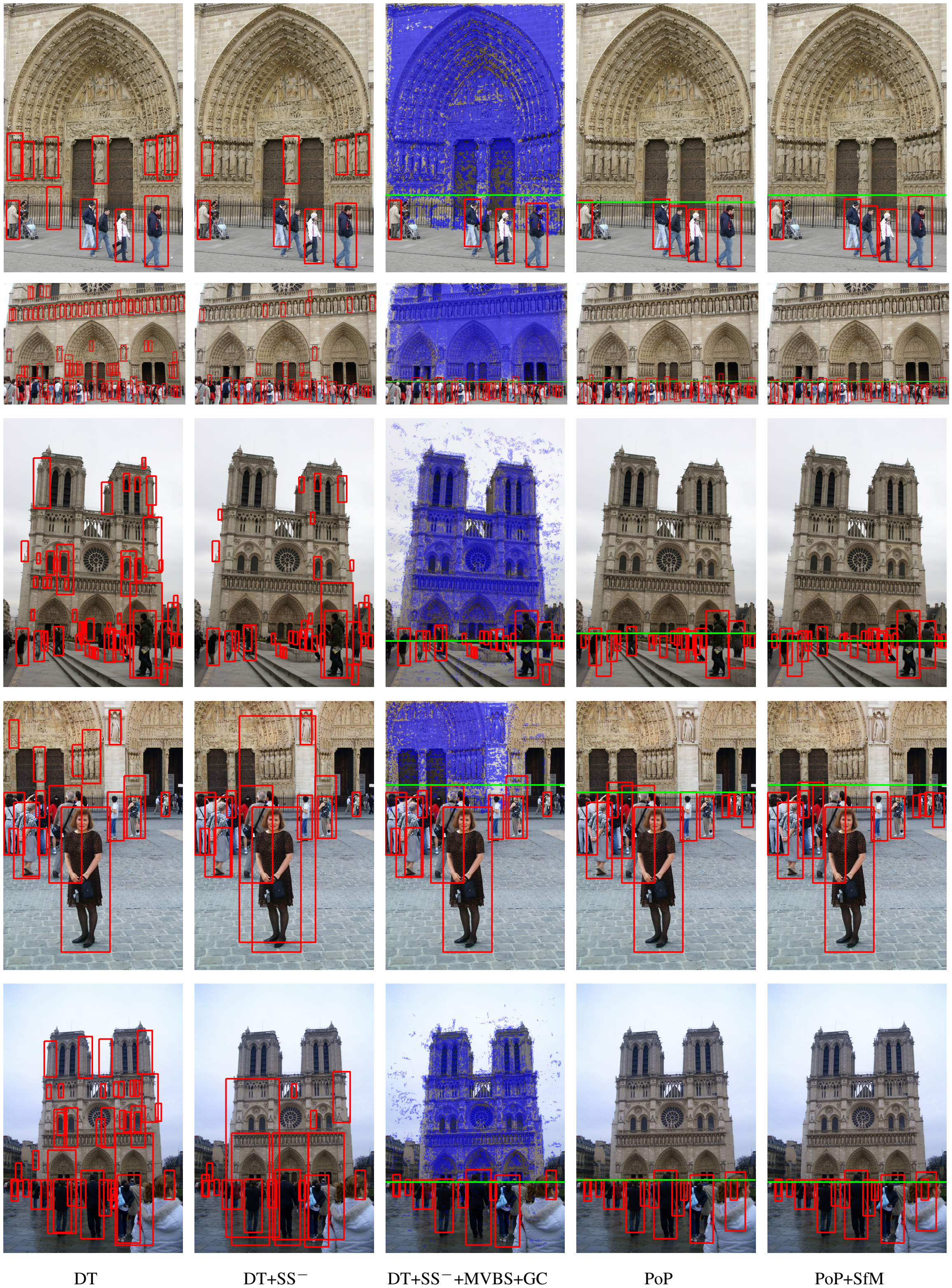
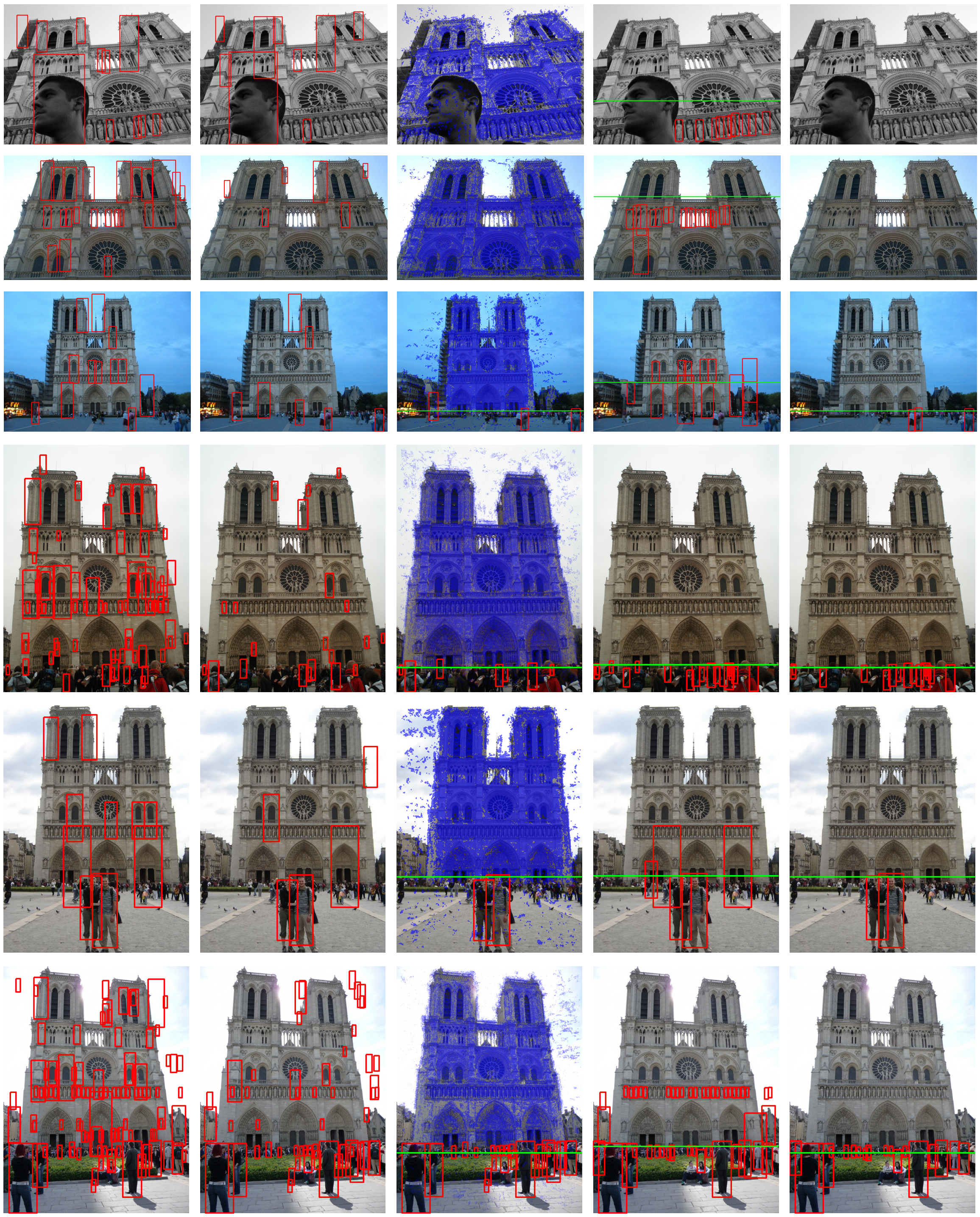


Figure 3: Example detector outputs at 50% recall. Scene specific training makes the detector better able to reject common distractors while MVBS can prune additional false positives. Putting Objects in Perspective (PoP) performance improves with accurate horizon estimation provided by structure from motion (SfM).



DT DT+SS$^-$ DT+SS$^-$+MVBS+GC PoP PoP+SfM

Figure 4: Example detector outputs at 50% recall. Scene specific training makes the detector better able to reject common distractors while MVBS can prune additional false positives. Putting Objects in Perspective (PoP) performance improves with accurate horizon estimation provided by structure from motion (SfM).

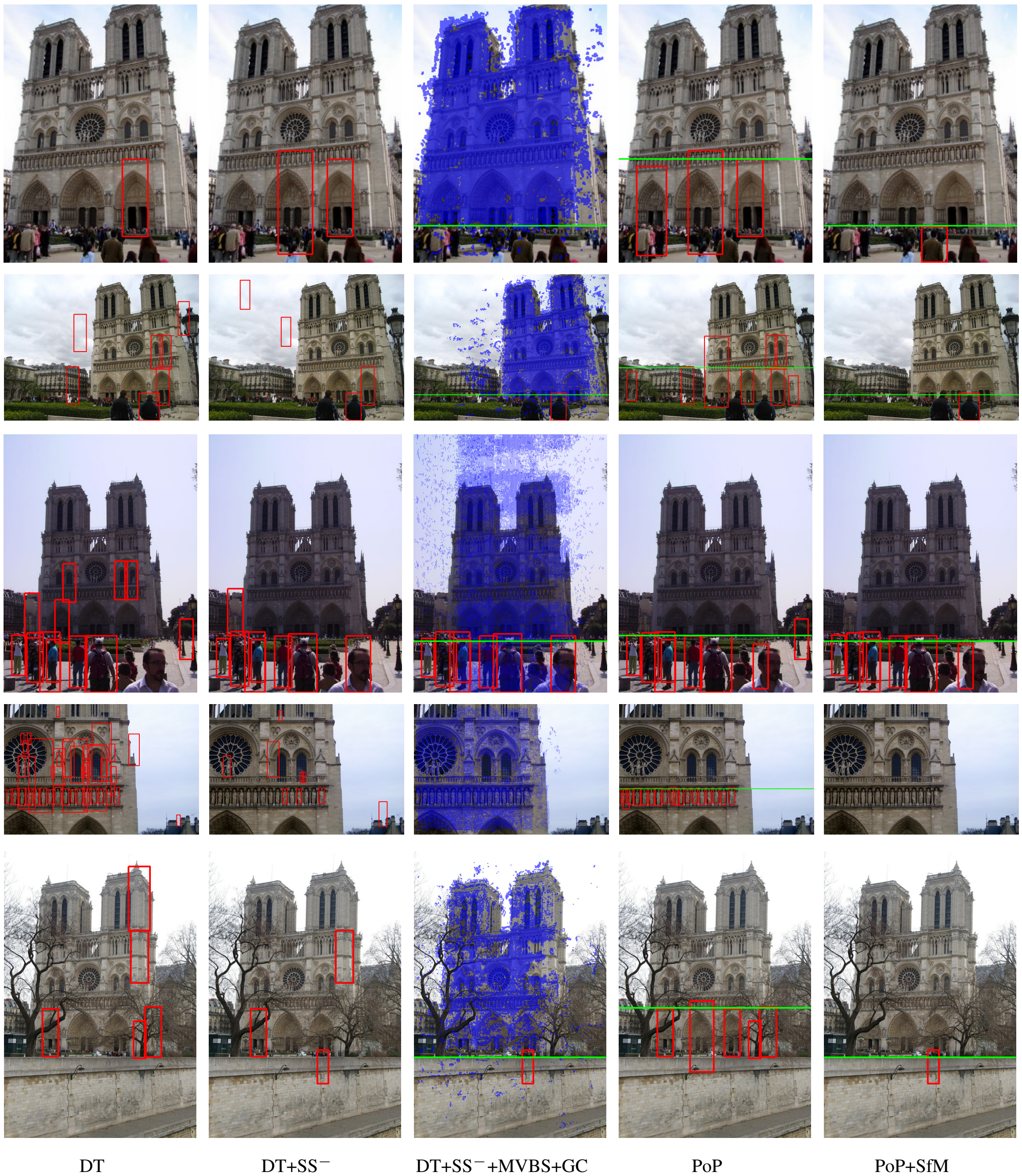


Figure 5: Example detector outputs at 50% recall. Scene specific training makes the detector better able to reject common distractors while MVBS can prune additional false positives. Putting Objects in Perspective (PoP) performance improves with accurate horizon estimation provided by structure from motion (SfM).