Appendix: Augmentation of Virtual Agents in Real Crowd Videos

Experimental Results on Pedestrian Detection/Tracking and Navigable Area Placement

Yalım Doğan · Serkan Demirci · Uğur Güdükbay · Hamdi Dibeklioğlu

We implemented the pedestrian detection and tracking part using Python programming language and OpenCV libraries. We evaluated our tracker on a video (PETS09-S2L1) from PETS dataset [1] and the videos we recorded. We use the multi-object tracking metrics explained in [2] to evaluate our pedestrian tracker. Precision is defined as the fraction of relevant detections (TPs) to all detections (TPs + FPs) and recall is defined as the fraction relevant detections (TPs) to all relevant detections (TPs + FNs) where TP is a true positive and FN is a false negative. Multi-object tracking accuracy (MOTA) [3],[4] is another metric used to evaluate the tracker's performance. MOTA combines three sources of errors: false negatives, false positives, and mismatch errors (MMEs):

$$MOTA = 1 - \frac{\sum_t (\mathrm{FN}_t + \mathrm{FP}_t + \mathrm{MME}_t)}{\sum_t \mathrm{GT}_t}, \qquad (1)$$

where t is the frame index and GT is the number of ground truth objects. MOTA can be negative if the number of errors made by the tracker is more than the number of all objects in the scene [5]. Mismatch error occurs if the tracker assigns different identification numbers to the same object. In our context, false negatives cause more noticeable errors in the simulation than false positives and mismatches. For our application, MOTA is not a useful metric because the three sources of errors are weighted equally. For our purposes, FNs are more important than the other two sources of errors, i.e., FPs and MMEs, as explained in the next paragraph. Lastly, the multi-object tracking precision (MOTP) metric measures the average dissimilarity between the true hypothesis and corresponding ground truth objects. This metric measures the localization precision of the tracker, which is affected by the accuracy of the associated detector. In our application, high localization errors decrease the quality of the agent projection. The more closer the MOTP values to zero, the better the tracker localization precision.

Y. Doğan · S. Demirci · U. Güdükbay (✉) · H. Dibeklioğlu
Department of Computer Engineering, Bilkent 06800 Ankara, Turkey
email: gudukbay@cs.bilkent.edu.tr.

Table 1 shows the tracker scores. In our application, false positives (FPs), detections that do not contain a pedestrian, cause agents to avoid empty areas during navigation. On the other hand, undetected pedestrians, i.e., false negatives (FNs), cause more noticeable errors because undetected pedestrians are not taken into account during collision detection and augmented agents sail through the undetected pedestrians in the video. Therefore, for our application, a tracker that favors recall score is more well-suited than the one that favors precision score. We adjusted the parameters of our detector and tracker to obtain high recall.

For each video, the position and orientation of the camera are manually adjusted to match the navigable area. The navigable areas are generated as object files (.obj) using 3D modeling software, such as Blender [6]. The navigable areas are imported as meshes with standard material properties.

For the augmented pedestrians to navigate on, the navigation mesh for the navigable area is generated right after the navigable area is imported into the real video. The lower left corner of the mesh is placed at the world origin. The mesh is visible to the user when it is loaded, but its visibility can be toggled afterward. Figure 1 shows the placements and their accuracies.

Figure 2 shows example detections for the tested videos. Additional agents can be seen projected onto the environment, which is caused by false detections (see Figure 2, last row, right part); however, it does not disturb the realism of simulation much. They can be considered as invisible obstacles for artificial agents. Multiple pedestrians projected onto the same location, which occurs in cases when the detection stage returns multiple bounding boxes for the same pedestrian, does not disturb the realism either.

Table 1: The quantitative results of our pedestrian tracker. The experiments were performed on a personal computer with Intel®Core^{TM}i7-4500U CPU @1.8 GHz, 8 GB RAM and NVIDIA 740M.

Videos	Resolution	Recall (%)	Precision (%)	MOTA (%)	MOTP	Frame processing time (sec)
PETS09-S2L1	768×576 @15.0fps	81.8	76.8	55.5	0.314	1.96
Video 1	1280×720 @30fps	18.1	41.4	-7.8	0.400	4.10
Video 2	1920×1080 @23.976fps	44.8	65.8	21.0	0.333	5.20

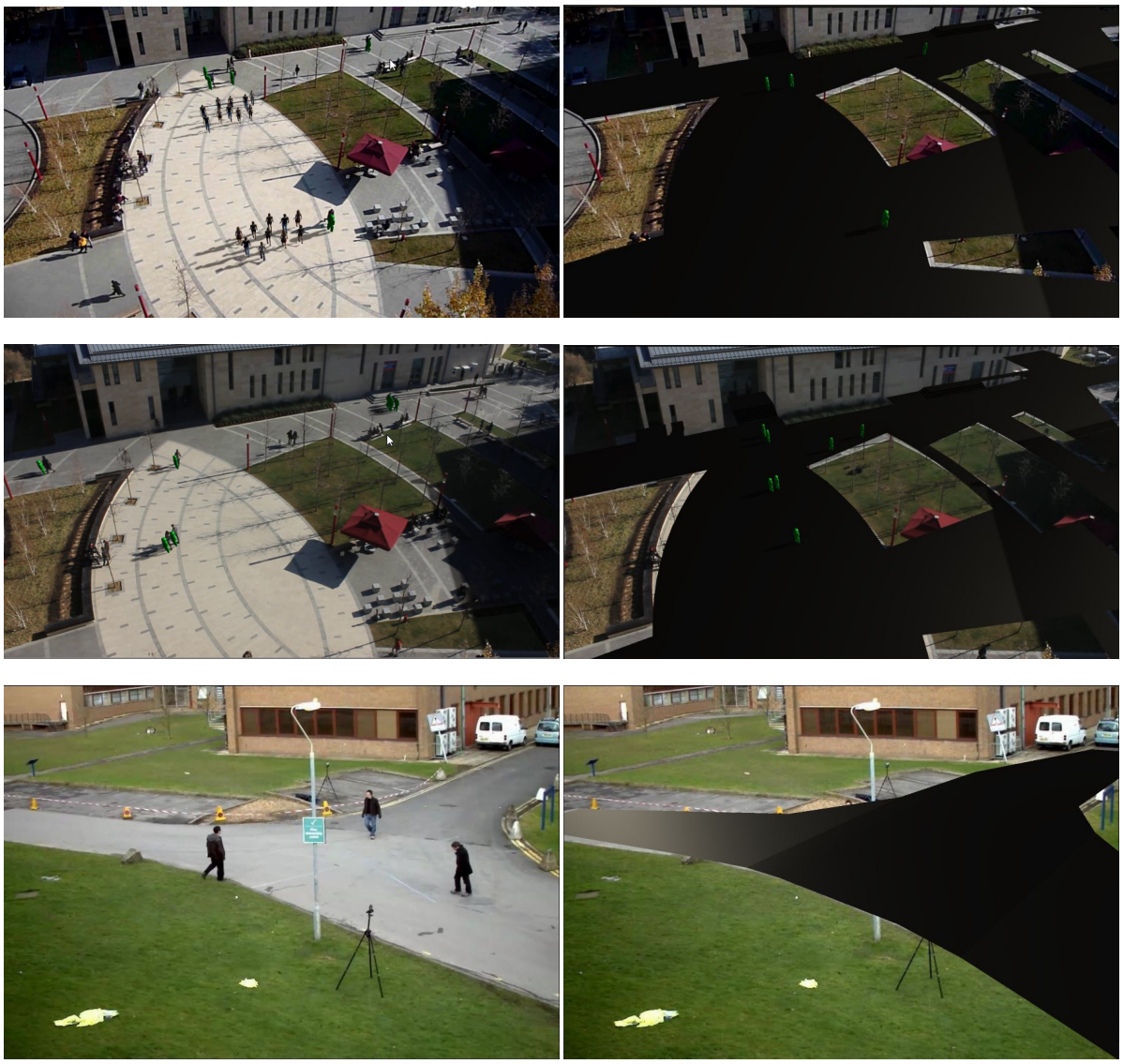

Fig. 1: The placed navigable area can be seen on the right, where a screenshot from the related video can be seen on the left. The accuracy of navigable areas is important as an augmented agent's traversal on non-navigable areas would degrade realism of the resulting video.

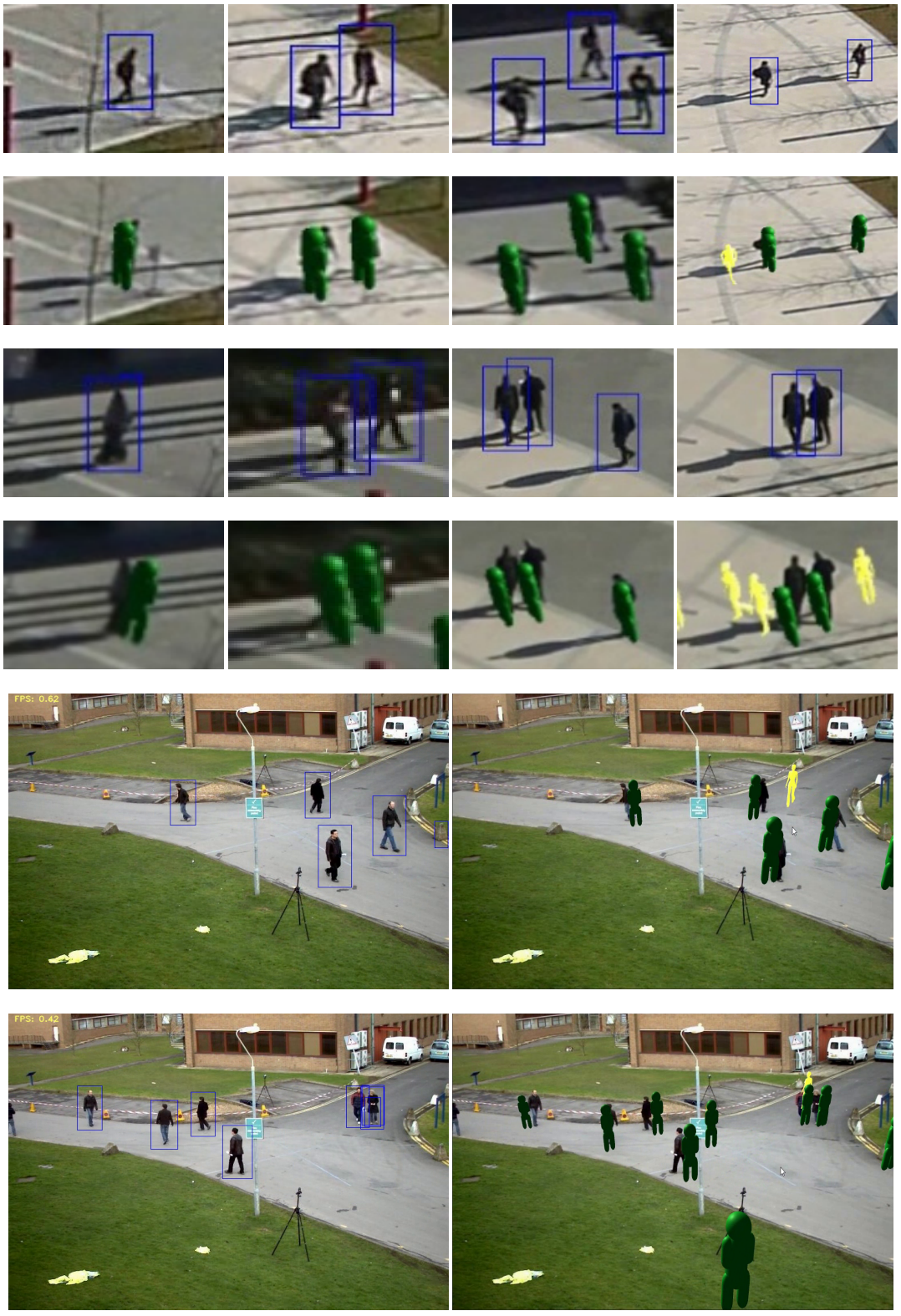

Fig. 2: Pedestrian detection and projection results for our videos (top four rows) and PETS09-S2L1 (bottom 2 rows). Yellow models are artificial agents we controlled during the simulation and green models are dummies that represent the location of the agent projection for each tracking result.

References

1. J. Ferryman and A. Ellis, "PETS2010: dataset and challenge," in *Proceedings of the 7th IEEE International Conference on Advanced Video and Signal Based Surveillance*, Aug 2010, pp. 143–150.
2. L. Leal-Taixé, A. Milan, I. D. Reid, S. Roth, and K. Schindler, "MOTChallenge 2015: Towards a benchmark for multi-target tracking," *CoRR*, vol. abs/1504.01942, 2015. [Online]. Available: http://arxiv.org/abs/1504.01942
3. R. Stiefelhagen, K. Bernardin, R. Bowers, J. Garofolo, D. Mostefa, and P. Soundararajan, "The CLEAR 2006 evaluation," in *Proceedings of the International Evaluation Workshop on Classification of Events, Activities and Relationships*. Springer, 2006, pp. 1–44.
4. K. Bernardin and R. Stiefelhagen, "Evaluating multiple object tracking performance: the CLEAR MOT metrics," *Journal on Image and Video Processing*, vol. 2008, p. 1, 2008.
5. L. Leal-Taixé, I. D. Reid, S. Roth, and K. Schindler, "MOT16: A benchmark for multi-object tracking," *CoRR*, vol. arXiv:1603.00831v2, 2016. [Online]. Available: https://arxiv.org/pdf/1603.00831.pdf
6. Blender Foundation, "Blender," 2002. [Online]. Available: http://www.blender.org