3D Crowd Counting via Multi-View Fusion with 3D Gaussian Kernels Supplemental

Qi Zhang, Antoni B. Chan
Department of Computer Science, City University of Hong Kong
Tat Chee 83, Kowloon Tong, Hong Kong SAR, China
qzhang364-c@my.cityu.edu.hk, abchan@cityu.edu.hk

Single-view examples

In Fig. 1, we show more examples showing the 3 situations that will limit the single-view counting methods' performance for large and wide scenes. From top to bottom, the 3 rows show: limited field-of-view, low resolution and occlusions. It can be observed from the first row that a single camera cannot cover all the people in a wide scene, such as the football match stadium, or the street shot with a mobile phone. From the second row, the people who are far away from the single camera are in quite low resolution, and can be hard to distinguish from other objects. In the third row, there are many obstacles in the scenes that occlude a lot of people, such as the flags, buildings, the advertisement board, or the people near the camera. Therefore, under these 3 situations, results for single-view counting on these large and wide scenes will be inaccurate.

Datasets and 3D prediction results

Examples from the 3 multi-view counting datasets are presented in Fig. 2. More information about the 3 multi-view datasets can be found in Zhang and Chan 2019.

The 3D crowd counting ground-truth and the corresponding prediction results of the proposed method, and the 2D density map ground-truth and prediction of MVMS of the 3 multi-view datasets examples are shown in Figs. 3, 4 and 5. The 3D density maps of PETS2009 are thresholded at 5e-3 (values below this are not shown), while DukeMTMC and CityStreet are thresholded by 1e-3. All the 3D ground-truth and predictions are presented from 3 view angles (front, top and bottom). The 3D crowd density maps can be observed at an arbitrary view angle, which may contribute to better scene understanding, generation or vizualization.

References

Zhang, Q., and Chan, A. B. 2019. Wide-area crowd counting via ground-plane density maps and multi-view fusion cnns. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 8297–8306.

Figure 1: The examples for the 3 situations that limit the single-view counting performance on large and wide scenes: limited field-of-view, low-resolution due to long distance to the camera, and severe occlusion. The images are from ShanghaiTech (left) and UCF_QNRF (right) datasets.

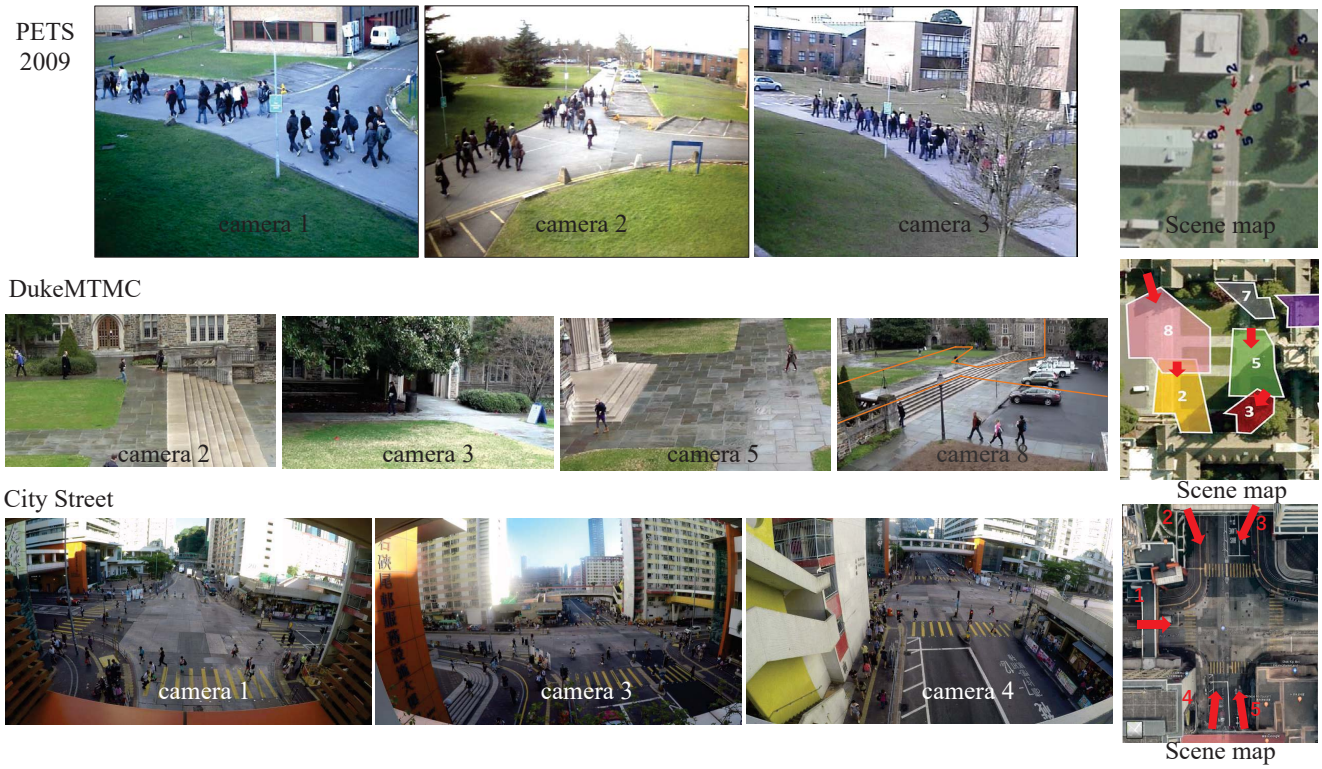

Figure 2: The multi-view datasets examples: PETS2009, DukeMTMC and CityStreet.

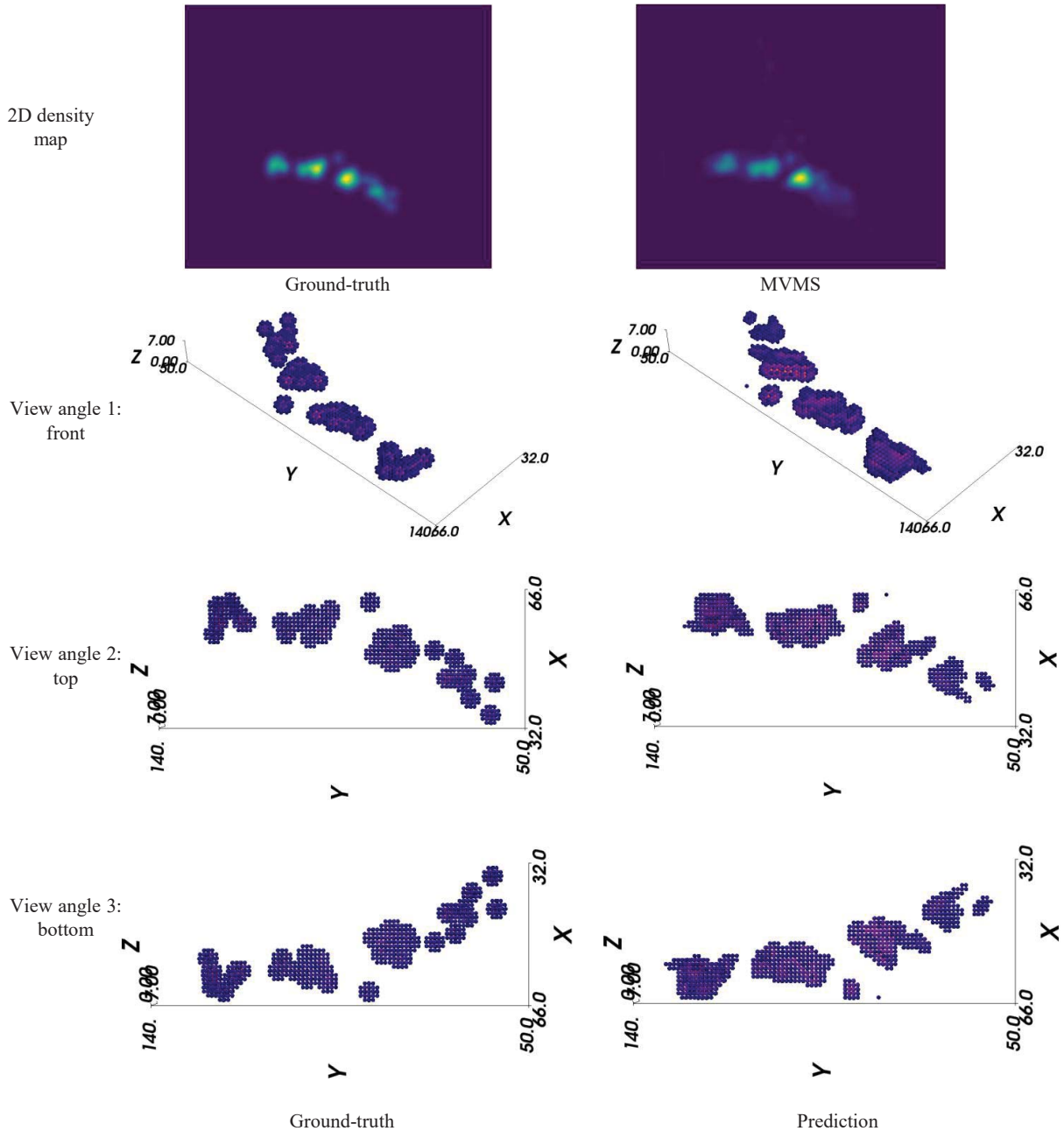

Figure 3: Example ground-truth and the prediction results on PETS2009 (shown from 3 view angles: front, top and bottom).

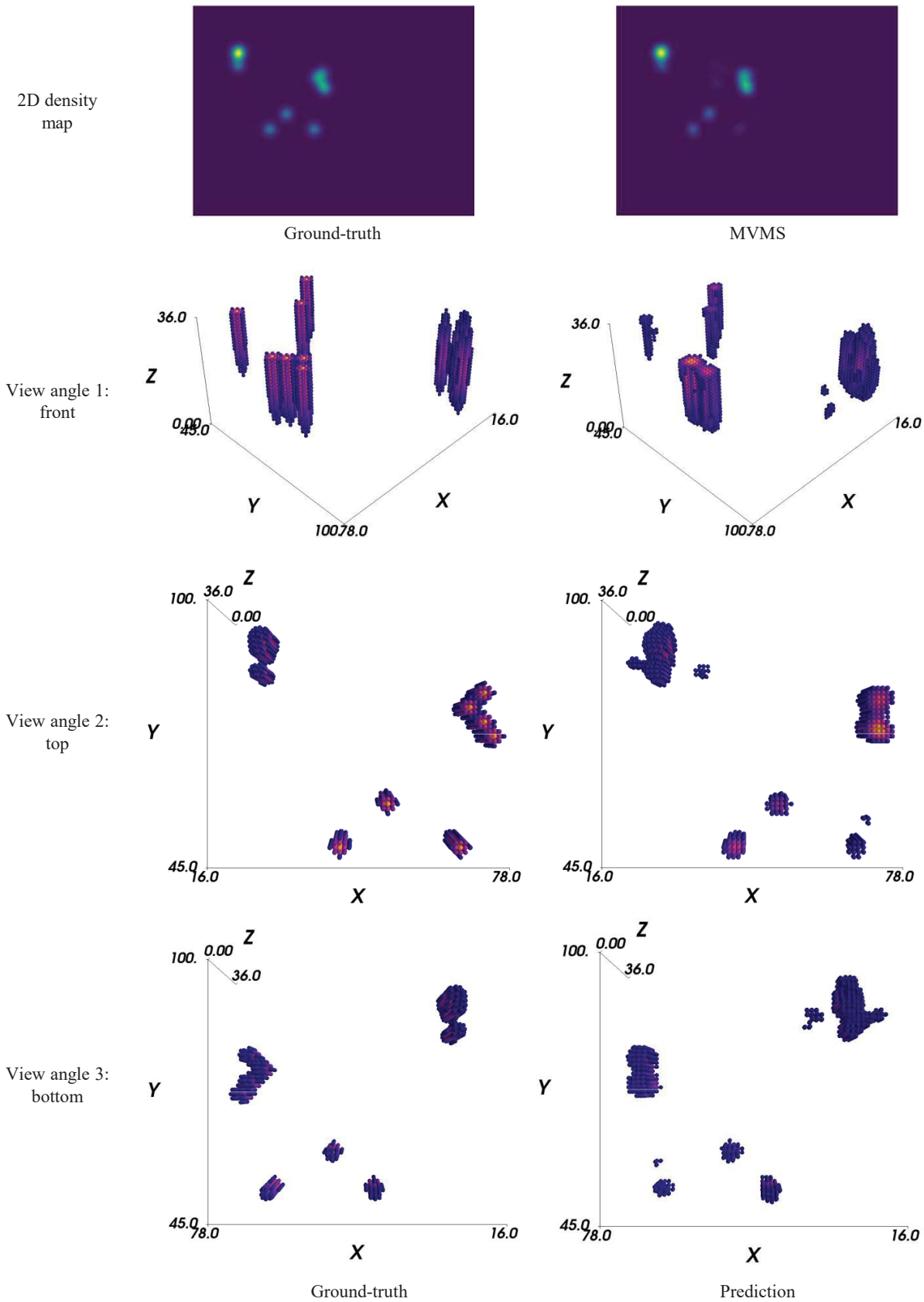

Figure 4: Example ground-truth and the prediction result on DukeMTMC (shown from 3 view angles: front, top and bottom).

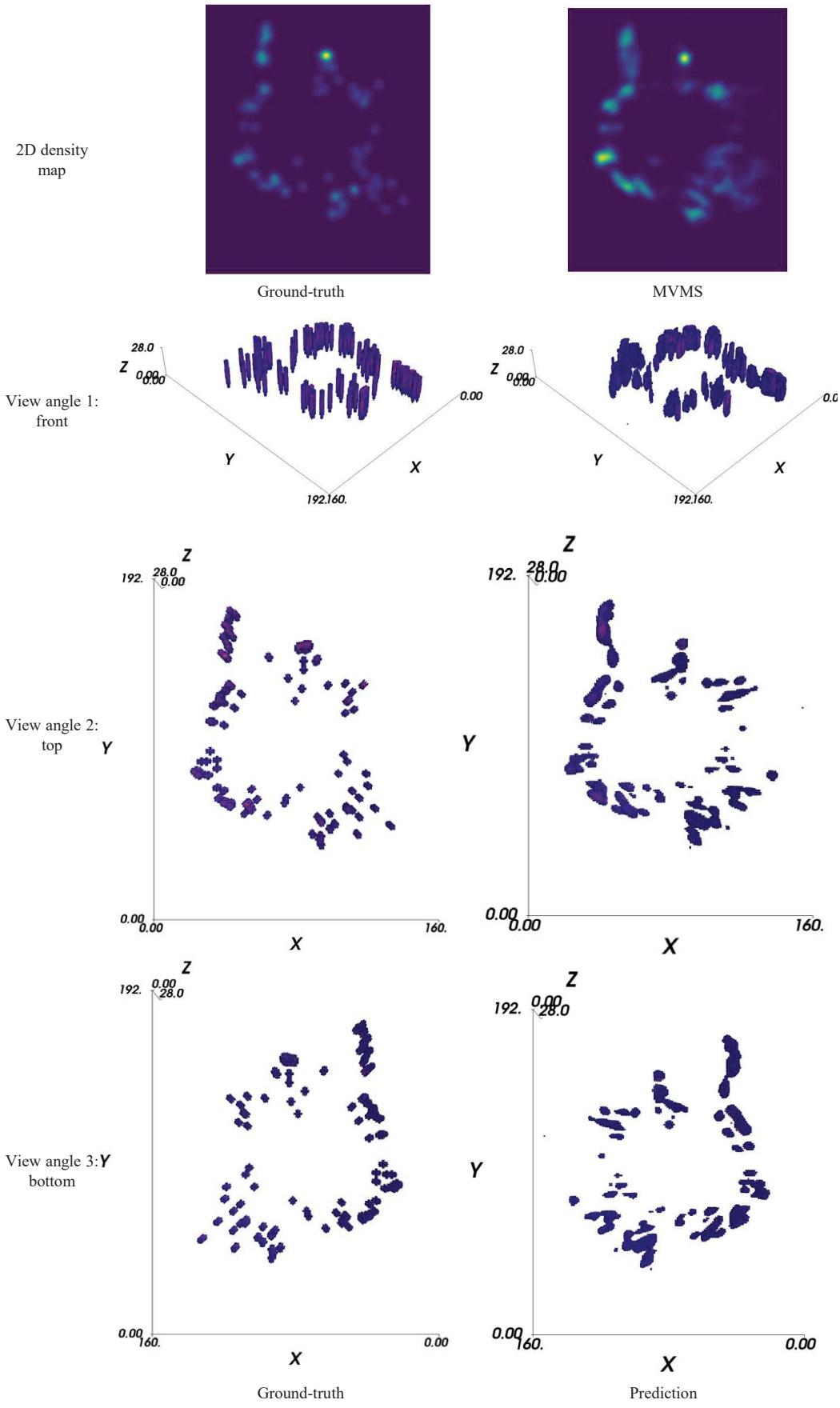

Figure 5: Example ground-truth and the prediction result on CityStreet (shown from 3 view angles: front, top and bottom).