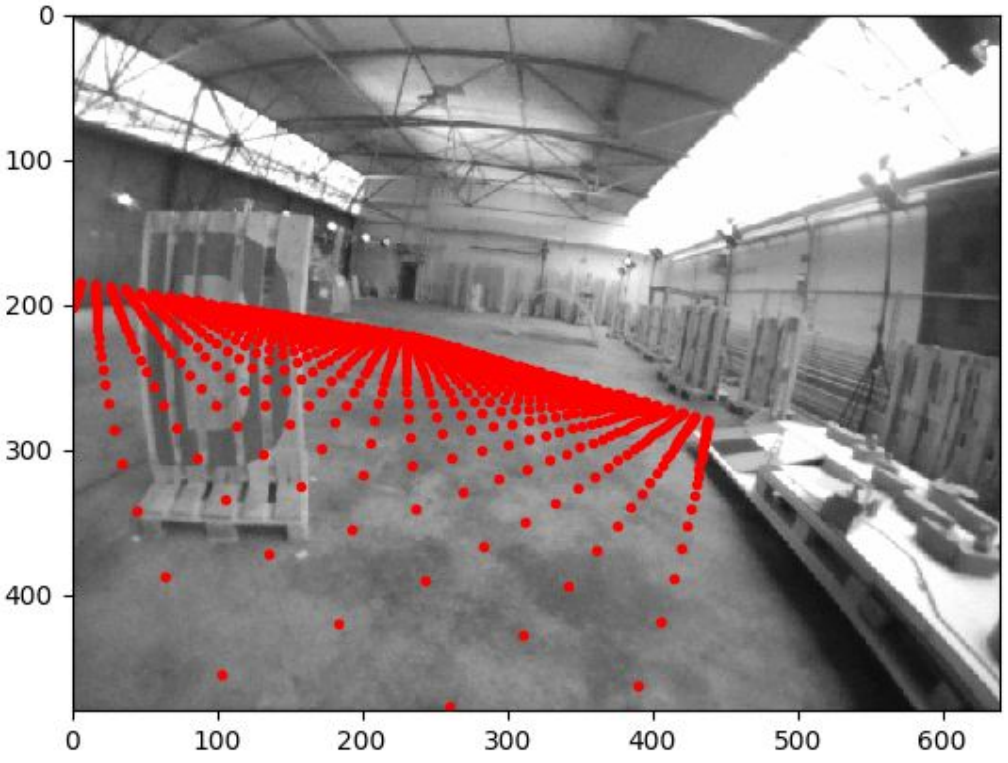
Report on a rotation issue with the original UZH FPV dataset ground truth

Titus Cieslewski, Giovanni Cioffi and Davide Scaramuzza, **Nov. 10th, 2020**

We have recently discovered that there is an issue with the rotations provided by the ground truth in the UZH FPV dataset. To visualize this rotation error, we introduce virtual 3D points that are supposed to be very close to the hangar floor and project them into the camera image according to the pose predicted by the ground truth (code). This is the result:



A frame from sequence indoor forward 5 with the virtual points projected according to the ground truth. The points should be coplanar with the ground but visibly are not. See also the video below.

indoor forward 5	ground truth video

As can be seen, the virtual points are sometimes not properly aligned with the floor in the images. More subtly, the center of expansion does sometimes not seem consistent between images and virtual points on the ground, showing also an error in yaw.

What does this mean for results (paper & competitions)?

Evaluations that only consider position (such as ATE after alignment) are not affected (see our previous evaluation: method and results)! However, evaluations that also take into account rotations are wrong and should be re-evaluated. For example, the trajectory evaluation method we have used for the UZH FPV competitions relies on aligning sub-trajectories using the first pose of the sub-trajectory between the estimate and ground truth.

What's next?

1. We re-evaluate the UZH FPV competition results using a **temporary** metric that approximates the proper evaluation metric using only ground truth positions, not ground truth rotations (see below).
2. We temporarily remove rotation from the ground truth of all publicly available data.
3. We implement a new method for ground truth pose estimation which also incorporates visual information in the optimization problem. The previous ground truth was based on the fusion of IMU and Leica position measurements only. Once we can thoroughly test the new ground truth we will release it and re-evaluate the competition submissions to get **final** results.

Temporary competition re-evaluations

Without going into too many details, the RPE evaluation metric we use relies on evaluating drift for sub-trajectories of different lengths. To do this, each sub-trajectory estimate is aligned to the corresponding ground-truth trajectory by aligning the first pose in the sub-trajectory. Since rotations are erroneous in our ground truth, this results in an exaggerated evaluation of drift.

To approximate this metric temporarily without relying on rotation estimates, we instead align each sub-trajectory based on the first 10% (distance-wise, not timewise) worth of positions. Since gravity is observable, we only optimize position and yaw. This results in the **temporary** results below. This method is for now also implemented in the automatic website evaluation script, so the results on the website temporarily also use this metric. Note that in this temporary evaluation, only position drift can be evaluated.

Once we re-calculate the ground truth with visual error terms, the submissions will be re-evaluated using the original evaluation metric and the ranking updated.

red = results not at the full frame rate
Note: these are not the final results! They should, however, be closer to the final results than our previous evaluation.

New rank	Name	T err% (temporary)	T err% (previous)	Comp.
1	Lenovo	2.009	7.005	IROS20

2	VCU	2.896	6.891	IROS20
3	Open Vins 1	3.092	7.023	IROS19
4	LARVIO	3.100	6.919	IROS20
5	OKVIS 2.0	3.182	7.148	ICRA20
6	MEGVII	3.198	6.819	IROS20
7	Moerwald-Leica	3.212	7.034	IROS19
8	OSU ETHZ	3.496	7.277	ICRA20
9	Open Vins 2	3.502	7.198	ICRA20
10	Basalt	3.593	7.494	IROS20
11	Usenko-TUM	4.554	7.778	IROS19
12	Rahman-USC	4.623	36.048	IROS19
13	Xin Zhang	6.716	9.140	IROS20
14	adjustmentteam	9.899	11.896	IROS19
15	QuetzalC++	33.751	34.273	IROS20