

How many cameras and where: staggered heights, minimum simultaneous views and a subject leaving the scene

Rings of 3 to 12 IR monochrome cameras in a 6×6 m room

SIMONE TORTORELLA

October 2026

ABSTRACT

We study how many cameras are needed and where to place them for a $3.5 \times 3.5 \times 2$ m volume in a 6×6 m room, with IR spherical markers and monochrome cameras. We separate two questions that are often mixed up: the quality of the **calibration** and the quality of the **reconstruction** of the markers on the athlete. Calibration with terna and T already works with 3 cameras and inside the covered volume it does not improve systematically with the number of cameras (p95 between 0.3 and 0.45 mm from 3 to 12 cameras). The number of cameras matters for reconstruction: the share of markers seen by at least 3 cameras goes from 25% with 4 cameras to 91% with 8 and 96% with 10, then flattens; the recall of reconstructed markers follows the same curve (97.0%, 98.7%, 99.7% with 8, 10, 12 cameras). The gain therefore converges between 8 and 10 cameras. Cameras staggered in height (two or three rows) reduce the reconstruction error by 10–20% compared with a single high row; the advantage comes mostly from the low cameras. Every image of the calibration object must be seen by at least 2 cameras at the same time (with 1 the calibration fails), better 3. If the subject leaves the volume the calibration does not get worse: frames in which the object is seen by fewer than 2 cameras simply do not contribute.

Keywords: number of cameras · placement · staggered heights · occlusion · simultaneous views

1. Introduction

In marker systems the question "how many cameras?" has two different answers depending on what is measured. To calibrate, there must be enough views of the calibration object to constrain every camera and link it to the others. To reconstruct an athlete, there must be enough cameras to see every marker from at least two (better three) directions despite the body covering it. Here the two questions are studied separately with the same simulator: 6×6 m room, $3.5 \times 3.5 \times 2$ m volume, cameras on the walls evenly spaced along the perimeter, aimed at the centre of the volume, with the narrowest lens that frames it (always 120° in this room).

The heights tested are: all at 2.2 m ("one row"), all at 1.0 m ("low"), alternating 2.4/1.0 m ("two rows") and rotating 2.4/1.6/0.8 m ("three rows"). The starting hypothesis is that **cameras staggered in height** reduce occlusion and improve the triangulation geometry.

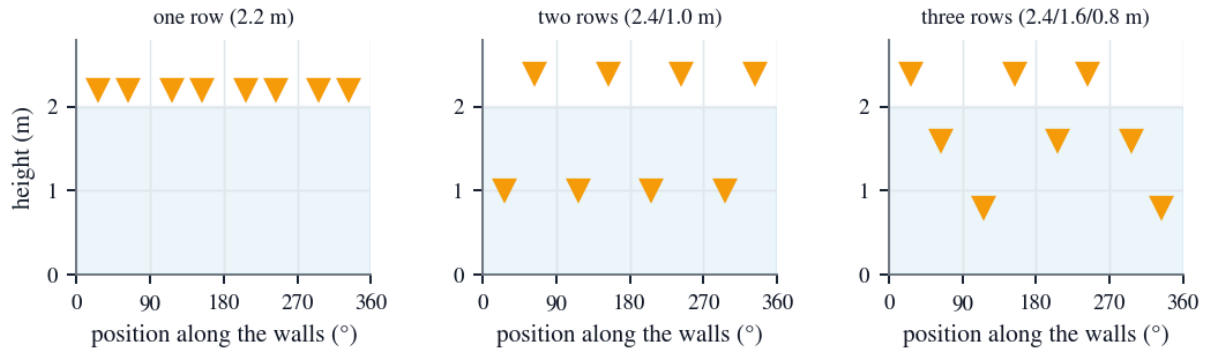


Figure 1. Layout of 8 cameras in the three height configurations, "unrolled" along the room perimeter; the blue band is the height of the measurement volume (0–2 m).

Comment. With a single row all cameras look at the volume from above at similar angles: feet and lower body are seen at a grazing angle. With two or three rows half of the cameras see the volume almost horizontally.

2. Calibration and number of cameras

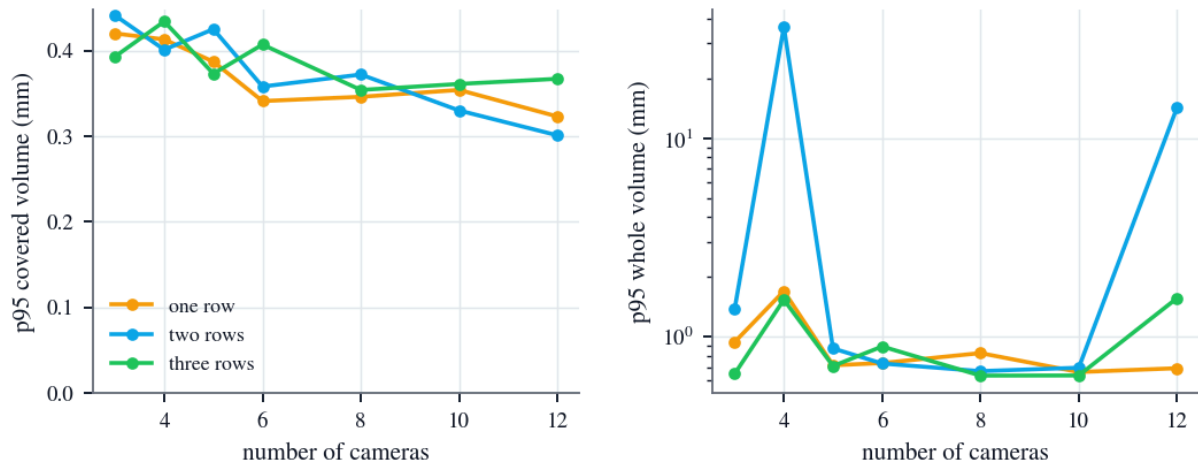


Figure 2. Calibration error (median of 6 sessions of 20 s) versus number of cameras and heights. Left: where the T went; right: over the whole 3.5×3.5×2 m volume.

Comment. Inside the covered volume the error stays between 0.3 and 0.45 mm from 3 to 12 cameras, with no clear trend: each camera is calibrated by the same observations of the T and adding cameras does not improve the others (maximum camera position error 1.2–1.9 mm, rotation 0.02–0.04°, for every number of cameras). Over the whole volume, instead, peaks of tens of millimetres appear with 4 and 12 cameras on two rows (above 5 mm in 100% and 100% of the sessions). These are not badly calibrated cameras: they are points at the corners of the volume that a camera sees in the corner of its image (normalised radius 1.35) while the T, in that camera, never went beyond 0.9–1.0. There the distortion of the 120° lens is extrapolated and the ray is wrong. The conclusion is that **the number of cameras is not a calibration parameter**: what matters is how well the T covers the volume and **every camera's image up to the corners** (Paper 3).

3. Reconstruction: how many cameras see each marker

For reconstruction we use the scene of Paper 4: a person walking through the volume with 30 14 mm markers on the body, hidden by the body itself (capsule model), observed with 0.15 px noise and random reflections. We count how many markers are seen by at least 2 and 3 cameras and how many the triangulation algorithm for anonymous markers reconstructs (recall), with a minimum support of 3 views and tracking.

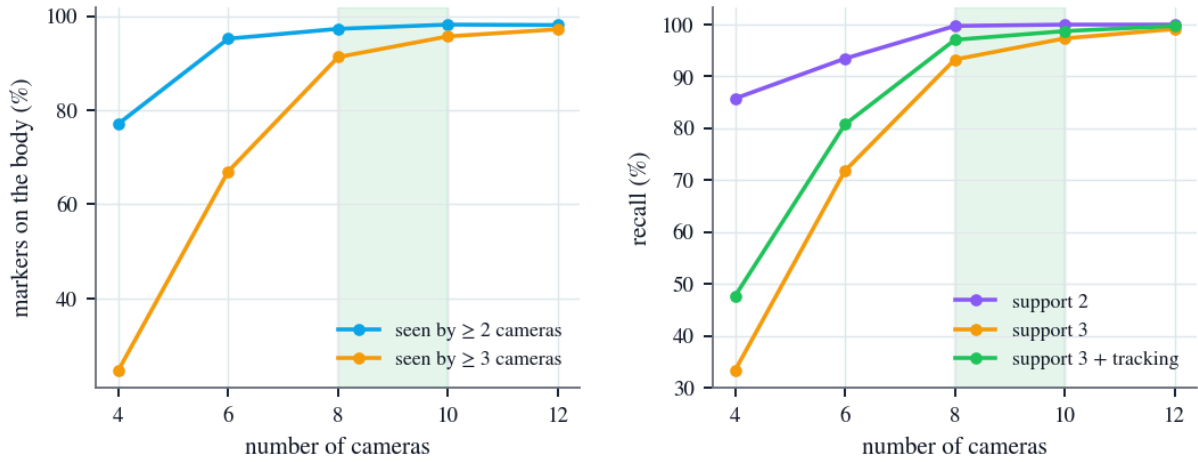


Figure 3. Left: share of body markers seen by at least 2 and 3 cameras. Right: recall of the 3D reconstruction of anonymous markers (Paper 4) versus number of cameras (two height rows).

Comment. Here the number of cameras matters. With 4 cameras only 25% of the markers are seen by 3 cameras and the recall with support 3 is 33%; with 8 cameras it goes up to 91% and 97.0%; with 10 to 96% and 98.7%; with 12 it gains less than one point. **The gain converges between 8 and 10 cameras:** beyond that, each extra camera adds about 1% of markers and costs bandwidth and computation (the algorithm's time grows with the square of the number of cameras, Paper 4). The practical minimum for a moving athlete is 6 cameras (recall 81% with tracking).

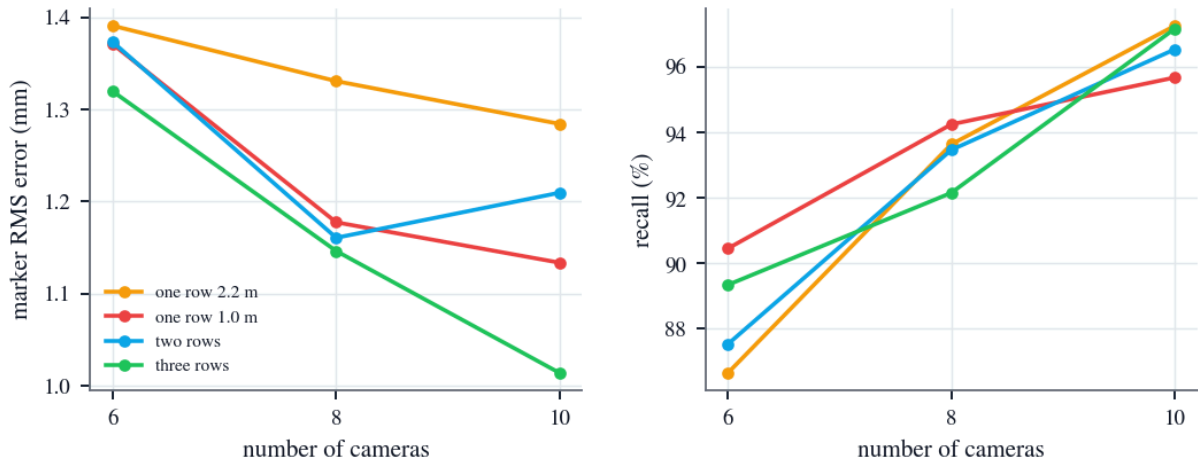


Figure 4. Effect of the heights on the reconstruction of body markers (median of 4 sequences, support 3 with tracking).

Comment. The hypothesis is partly confirmed. With 8 cameras the high row gives 1.33 mm, two rows 1.16 mm, three rows 1.15 mm, the low row 1.18 mm; with 10 cameras three rows reach 1.01 mm against 1.28 mm for the high row. The gain (10–20%) comes from the low cameras: they see leg and foot markers that high cameras see at a grazing angle or hidden by the trunk, and their rays cross those of the high cameras at a wider angle, reducing the error along the vertical. An all-low row does almost as well on error, but loses arms above the head and markers hidden by equipment; staggering keeps the advantages of both. Recall, instead, changes little ($\pm 2\%$): the number of cameras matters more than their height.

4. How many cameras must see the object at once

The bundle adjustment links the cameras through the frames in which the T is seen by more than one camera. To measure the minimum needed, in each frame we keep at most k cameras (chosen at random among those seeing the T) and delete the observations of the others. With $k = 1$ each camera sees the T alone: intrinsics remain constrained (as with a checkerboard), but the cameras are linked only by the terna.

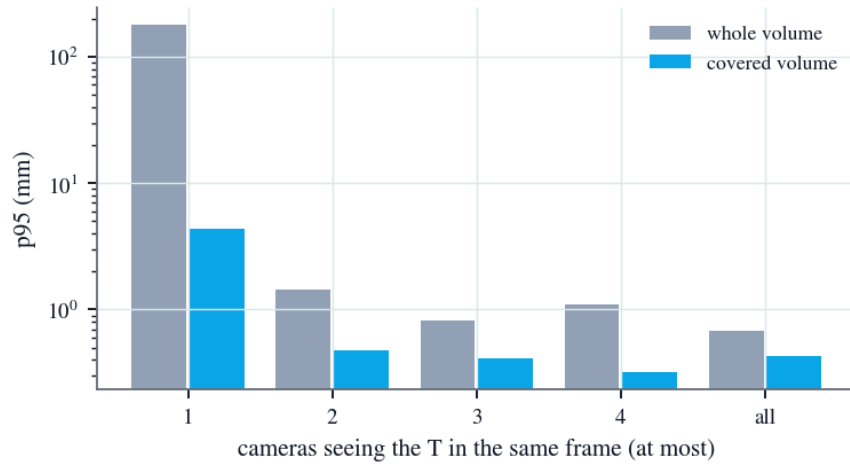


Figure 5. Calibration error when in each frame the T can be seen by at most k cameras (8 cameras, median of 5 sessions).

Comment. With a single camera per frame the calibration fails (179 mm over the volume); without shared views the relative position of the cameras depends only on the terna, seen from far away. With 2 cameras per frame it works (1.44 mm over the volume, 0.48 mm where the T went); with 3 it is already close to the case with all cameras (0.81 against 0.68 mm). **Minimum: 2 simultaneous views; recommended: 3.** In practice the T must pass where the cameras overlap, not only near the walls.

5. The subject leaves the scene

In reality whoever moves the T leaves the volume, gets close to a wall, walks in front of a camera. We simulate this by widening the person's path by a factor ρ (1 = always inside the volume; 2 = up to 0.5 m from the walls) and count in each frame how many cameras see at least 3 T markers.

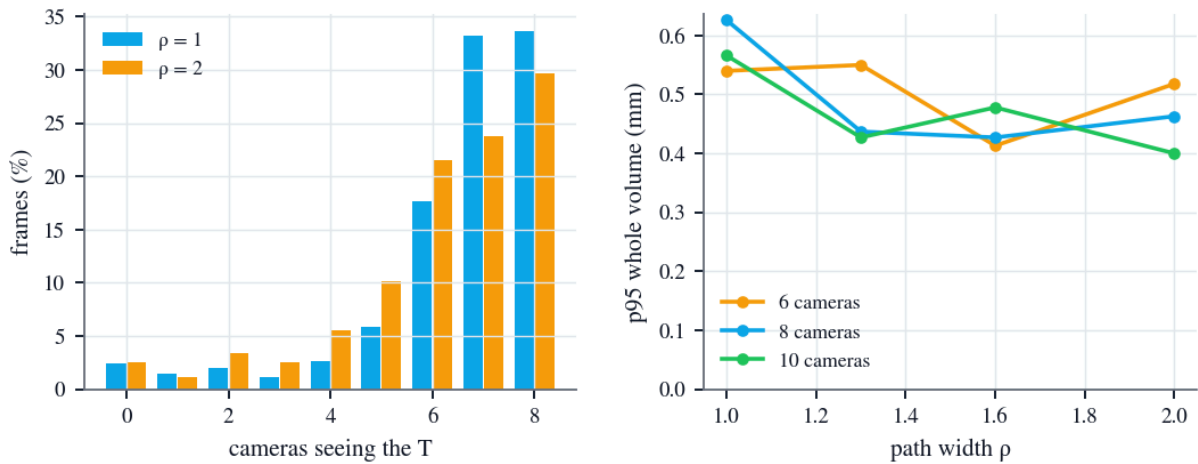


Figure 6. Left: distribution of the number of cameras seeing the T (8 cameras) with the path inside the volume ($\rho = 1$) and up to the walls ($\rho = 2$). Right: calibration error versus ρ .

Comment. Leaving the volume raises the share of frames where the T is seen by few cameras (≤ 4 cameras: from 10% to 15%), but the calibration does not get worse; it stays between 0.4 and 0.6 mm: frames with fewer than 2 views carry no information but do no harm, and a T passing near the walls widens the covered volume. Occlusions due to leaving the scene are therefore not a problem as long as the bundle adjustment uses only valid observations and the T crosses the centre of the volume in the rest of the session.

6. Orientation and mixed resolutions

Table 1. Camera orientation and rigs with different resolutions (median of the sessions).

Configuration	p95 (mm)	Note
8 landscape cameras	0,63	whole volume
8 portrait cameras	0,59	whole volume
8 Full HD cameras	0,43	covered volume, 60 fps
4 4K + 4 HD cameras, alternating	0,36	covered volume, 60 fps
8 4K cameras	0,29	covered volume, 60 fps

Portrait and landscape cameras are equivalent in an almost cubic volume. A mixed rig (half 4K, half HD) sits between the two homogeneous rigs: the bundle adjustment weights each observation by its expected error, so high-resolution cameras count more without special rules. Different frame rates between cameras instead require estimating a delay per camera (preliminary study) and were not simulated here.

7. Conclusions

Recommendations

- Calibration: 3 cameras are enough to calibrate; the number of cameras does not improve calibration, coverage of the volume and of every image's corners does.
- Reconstructing an athlete: minimum 6 cameras, optimum 8–10; beyond 10 the gain is < 1% of markers (98.7% → 99.7%).
- Heights: staggered on two rows (2.4 and 1.0 m) or three with 10 cameras or more: –10–20% error compared with a single high row; low cameras are needed for legs and feet.
- Simultaneous views of the T: minimum 2, recommended 3: move the T through the centre of the volume, where cameras overlap.
- The subject may leave the scene: frames with fewer than 2 views do not contribute and do not harm the calibration.

Limits: a single room type (6×6 m) and a fixed volume; wall-mounted cameras only; a simplified capsule body. In larger rooms lenses get narrower and markers bigger in pixels (Paper 1), but the reasoning on the number of views does not change.

References

1. R. Hartley, A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed., Cambridge University Press, 2004.
2. B. Triggs, P. McLauchlan, R. Hartley, A. Fitzgibbon, *Bundle adjustment — a modern synthesis*, LNCS 1883, 2000.
3. T. Svoboda, D. Martinec, T. Pajdla, *A convenient multicamera self-calibration for virtual environments*, Presence 14(4), 2005.
4. S. Tortorella, *Paper 1 — Oggetto di calibrazione per marker IR anonimi*, 2026; *Paper 4 — Ricostruzione 3D di marker anonimi*, 2026.