

A calibration object for anonymous IR markers: from the white threshold to the optimal shape

Retro-reflective spherical markers, monochrome cameras, threshold detection

SIMONE TORTORELLA

October 2026

ABSTRACT

We study the object used to calibrate a system of 8 monochrome cameras with an infrared illuminator and retro-reflective spherical markers, whose centres are estimated as centroids of the pixels above a white threshold. A pixel-level model of the marker (retro-reflective profile, PSF, motion blur, noise, compression) shows that with a dark IR background the centroid error is about 0.011 px for an 8 px marker at a 40% threshold, and that the dominant error is not noise but partial occlusion (0.82 px when a third of the disc is covered). In full simulations (a person moving the object, 8 cameras, bundle adjustment) we compare resolution (HD, Full HD, 4K), frame rate (30–120 fps), marker diameter (10–50 mm), six object designs and the static terna. The marker should be the smallest the resolution allows, about 3 px at the median distance: 30 mm in HD, 20 mm in Full HD, 10 mm in 4K. To recognise the object from anonymous points we define an identification margin and optimise it: a 4-marker T with the junction 14 cm off-centre and a 45 cm stem maximises the margin; 5–6 marker variants are not more robust. The construction tolerance of the static terna is the main limit of the absolute accuracy: 0.25 mm of tolerance takes the p95 from 0.68 to 1.62 mm.

Keywords: multi-camera calibration · infrared markers · blob centroid · threshold · T-wand · unlabelled marker identification

1. Introduction

Marker-based motion analysis systems use monochrome cameras with a ring of infrared LEDs around the lens and a filter that passes only infrared light. The markers are spheres covered with retro-reflective material: they send the light back to the camera and appear in the image as almost white discs on an almost black background. The centre of each marker is estimated with a white threshold and the centroid of the pixels above it. The cameras do not know which marker a blob belongs to: they only see anonymous points.

The calibration proposed in a previous study uses a 3D terna standing still at the origin (it defines origin, axes and scale) and a T-wand moved by a person through the volume (it constrains intrinsics, extrinsics and delays through a sparse bundle adjustment). Here we answer a more concrete question: **what the object should look like**, and which marker and camera parameters really matter. Every test is a simulation with known ground truth: the estimated calibration is always compared with the true one.

- §2 pixel model of the marker and of threshold detection;
- §3 from pixel to millimetre: resolution, fps and marker diameter in full calibrations;
- §4 recognising the object from anonymous points: identification margin and its optimisation;
- §5 the static terna; §6 the ideal object.

2. The marker as the camera sees it

A retro-reflective sphere of apparent diameter d pixels, lit by the coaxial ring, has an almost flat intensity profile in the image that drops near the rim. We model it as

$$I(r) = b + g \left(1 - \frac{r^2}{R^2} \right)^{1/4}, \quad r < R = d/2 \quad (1)$$

with background $b = 0.05$ and gain $g = 1.4$ (the centre saturates at 255). The image is $8\times$ super-sampled, convolved with the lens PSF (Gaussian, $\sigma = 0.7$ px) and with the motion blur of the exposure, integrated over the pixels, with photon noise (6000 e⁻ at white) and read noise (2 DN), quantised to 8 bits. The detector applies a threshold τ (fraction of white), finds connected components and computes their weighted centroid:

$$\hat{\mathbf{u}} = \frac{\sum_{i: I_i \geq \tau} (I_i - \tau) \mathbf{u}_i}{\sum_{i: I_i \geq \tau} (I_i - \tau)} \quad (2)$$

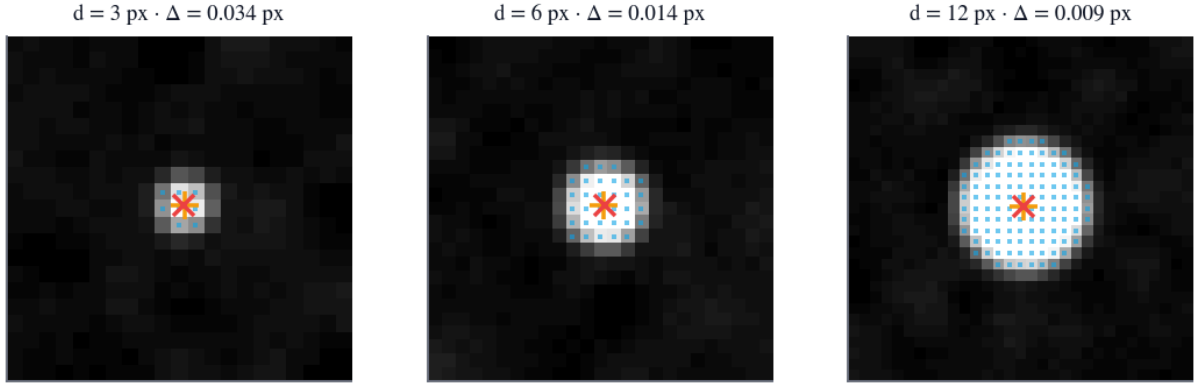


Figure 1. Simulated spheres of 3, 6 and 12 px. Blue squares: pixels above the 40% threshold. Yellow cross: true centre; red cross: estimated centroid.

Comment. Even the 3 px marker has an error of a few hundredths of a pixel: with a dark background and many photons the centroid is very precise. The problem, as we will see, is not noise.

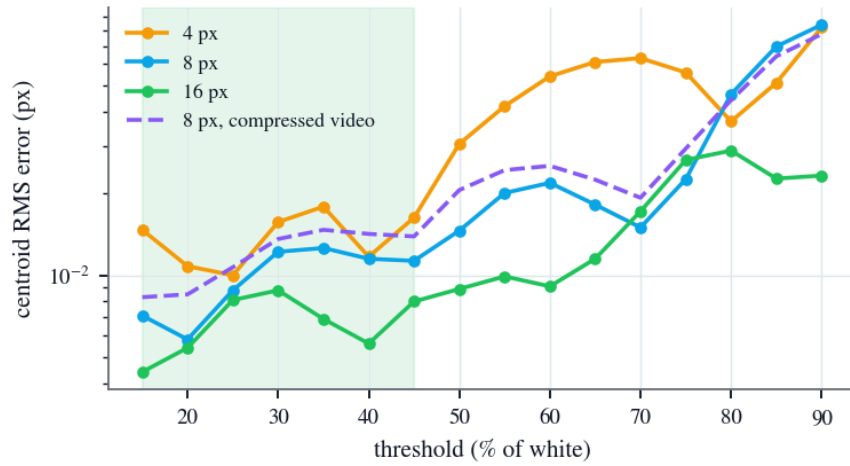


Figure 2. Centroid error (log scale) versus threshold, for 4, 8 and 16 px markers; dashed: 8 px marker with compressed video (quality 80). All markers are detected up to an 85% threshold.

Comment. With a dark IR background the best threshold is low: between 15% and 45% the error is minimal (0.011 px at 8 px and a 40% threshold). A higher threshold throws away the rim of the disc, which carries information: above 75% the error grows quickly (0.046 px at 80%). Compression costs little (0.014 px) when the video quality is ≥ 70 . We choose a 40% threshold: high enough to cut weak IR reflections (sunlight, shiny surfaces), low enough to use the whole disc.

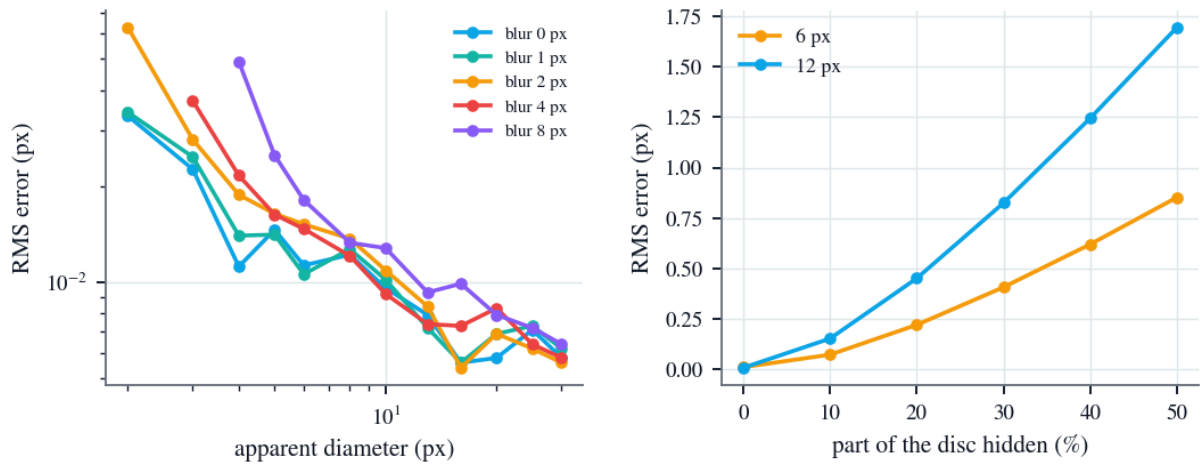


Figure 3. Left: centroid error versus apparent diameter and blur (40% threshold). Right: error when a chord hides part of the disc (an arm, the wand's rod, another marker).

Comment. Blur matters little: for markers of at least 6 px even an 8 px streak keeps the error below 0.03 px, because the centre of the streak is still the centre of the marker when the motion is uniform during the exposure; it only weighs on the smallest markers. Partial occlusion, instead, moves the centroid towards the visible part: with a third of the disc hidden the error is 0.41 px at 6 px and 0.82 px at 12 px, i.e. **proportional to the diameter**, and 30–100 times the noise. In millimetres this error is proportional to the physical size of the sphere: it is the first reason to prefer small markers and to discard deformed blobs (§3 and Paper 3).

Two nearby markers merge into one blob when the distance between their centres falls below about 1.05 apparent diameters (simulation, 40% threshold). Finally, an error component the model does not contain (residual of the distortion model, rolling shutter, vibration, non-uniform material) is represented in the following simulations by a term $\sigma_0 = 0.08$ px added in quadrature; Paper 3 studies the sensitivity to it.

3. From pixel to millimetre: resolution, fps, diameter

Each full simulation has 8 cameras on the walls of a 6×6 m room, at two alternating heights (2.4 and 1.0 m), with the narrowest lens that frames the 3.5×3.5×2 m volume (here always 120°, focal length 554 px at Full HD). A person walks for 20 s moving the T; the bundle adjustment uses 400 frames; one frame in five is kept aside for validation. The main metric is the p95 of the 3D error inside the part of the volume crossed by the T ("covered volume"), triangulating exact observations with the estimated cameras. The apparent diameter of a marker is $d = fD/Z$: a 30 mm sphere at 3.6 m (median distance) measures

Table 1. Apparent diameter (px) at 3.6 m (median) with a 120° lens; last column: 4K at the 95th percentile of the distance.

Diameter	HD	Full HD	4K	4K at 5.1 m
10 mm	1.0	1.5	3.1	2.2
14 mm	1.4	2.2	4.3	3.0
20 mm	2.1	3.1	6.2	4.3
30 mm	3.1	4.6	9.3	6.5
40 mm	4.1	6.2	12.4	8.6
50 mm	5.2	7.7	15.5	10.8

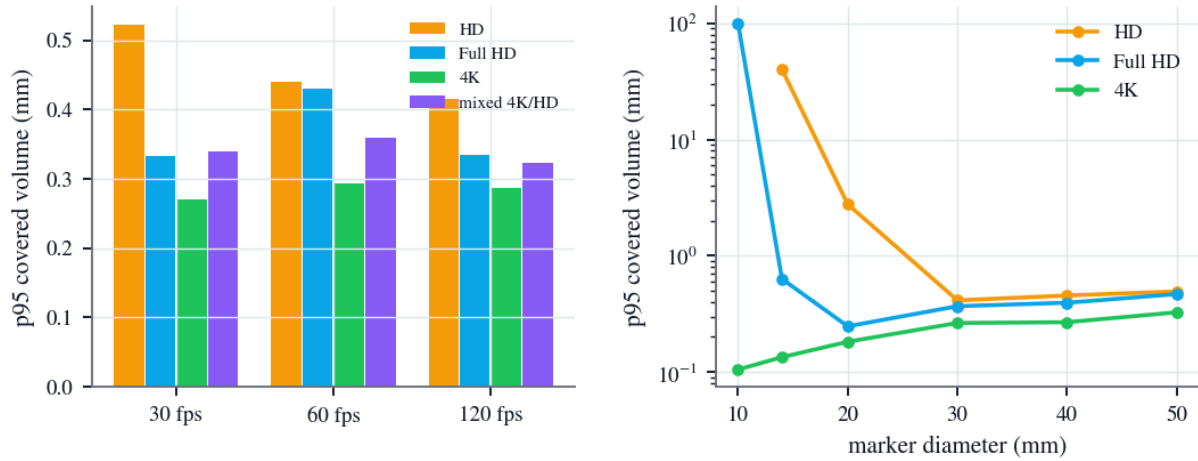


Figure 4. Left: p95 by resolution and frame rate (30 mm markers; exposure 1/(2·fps); median of 5 sessions). Right: p95 versus marker diameter for the three resolutions (60 fps). Missing values: calibration failed.

Comment. Resolution matters more than frame rate: 4K takes the p95 to about 0.29 mm against 0.44 mm for HD (60 fps). With a hand-held T the blur is small even at 30 fps, so frame rate matters little for the calibration (it matters for synchronisation, preliminary study). The diameter has an optimum: below about 3 px at the median distance (Table 1) far markers disappear and the calibration degrades or fails (HD below 30 mm, Full HD below 14–20 mm); above it the error grows again with the diameter because partial occlusion and merging of nearby blobs weigh more. In 4K even 10–14 mm work and give the lowest error (0.10 mm).

4. Recognising the object from anonymous points

In every image the program has to say which blob is which marker. The object is planar: its image is a projective transform of the drawing, well approximated by an affine one when the object is small compared with the distance. For an assignment π of the blobs $\mathbf{u}_i$ to the drawing markers $\mathbf{x}_{\pi(i)}$, the normalised residual of the best affine fit is

$$r(\pi) = \frac{1}{s} \min_{A, \mathbf{t}} \sqrt{\frac{1}{k} \sum_{i=1}^k \|A \mathbf{x}_{\pi(i)} + \mathbf{t} - \mathbf{u}_i\|^2}, \quad s = \text{RMS spread of the blobs} \quad (3)$$

An affine map has 6 degrees of freedom: with 3 markers every permutation has zero residual and there is nothing to check; with a symmetric drawing the left/right reflection is an exact affine map too. We define the **identification margin** as the difference between the best residual of a wrong labelling and that of the right one, at the 10th percentile over 200 viewpoints with up to 70° of tilt:

$$m = Q_{10\%} \left[\min_{\pi \neq \pi^*} r(\pi) - r(\pi^*) \right] \quad (4)$$

The margin is dimensionless (a fraction of the object's size in the image): the labelling stays safe as long as the centroid noise, divided by the object's size in pixels, is much smaller than m . With the hypothesis of a ghost blob (reflection) we also consider labellings that drop one marker.

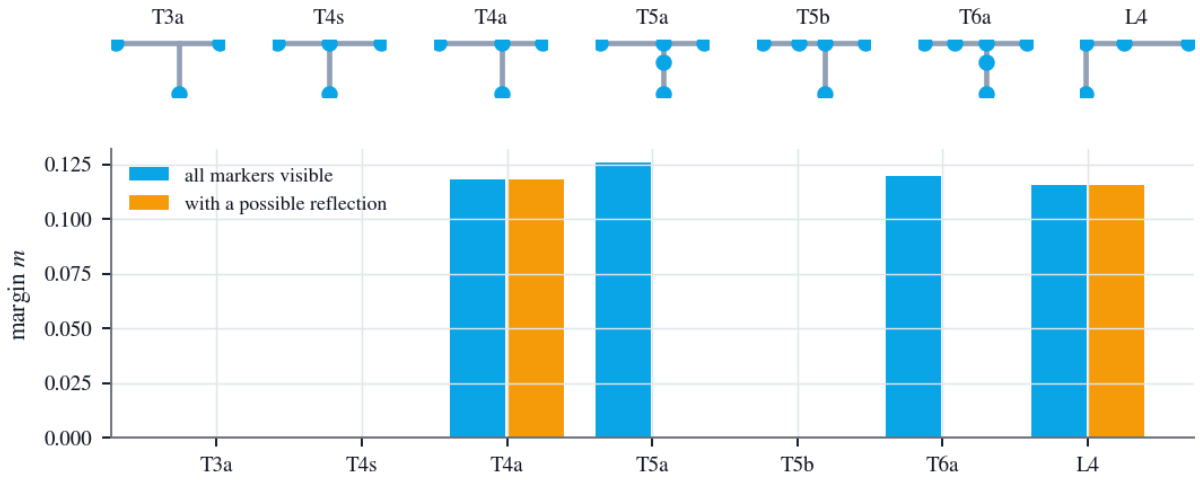


Figure 5. The seven designs (0.75 m bar) and their identification margin. T3a: 3 markers; T4s/T4a: symmetric/asymmetric 4-marker T; T5a: T4a with a marker in the stem; T5b: 4 collinear markers on the bar; T6a: T5a with one more bar marker; L4: 4-marker L.

Comment. Three-marker and symmetric designs have zero margin, as expected. T4a and L4 have a margin of ~0.118. Adding markers does not help: T5a and T6a have the same margin when all markers are visible, but become ambiguous as soon as a reflection is allowed (margin ≤ 0), because dropping one marker leaves subsets with three collinear points that fit wrong labellings well. T5b, with four points on the same line, is ambiguous even without reflections: the distance ratios along a line are too similar in the two directions.

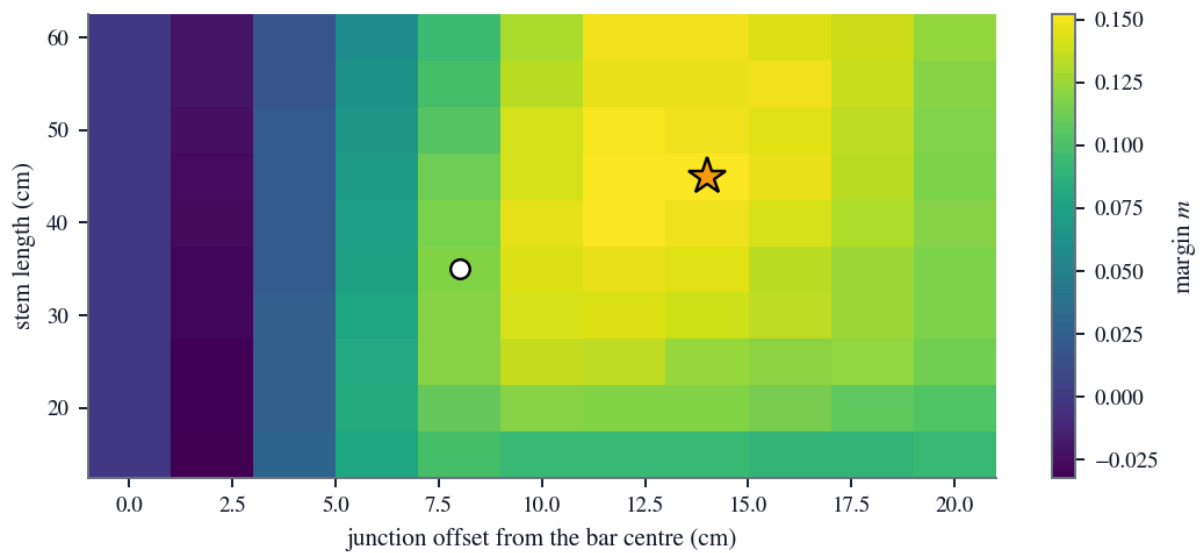


Figure 6. Identification margin of the 4-marker T (0.75 m bar) versus junction offset and stem length. Star: optimum; circle: the T of the preliminary study (8 cm, 35 cm).

Comment. The margin grows with the asymmetry up to about 14 cm and with the stem length up to 45 cm, then flattens: beyond that the T becomes awkward to handle with no gain. The optimum (0.152) is 28% above the T of the preliminary study (0.119). Short stems make the T almost collinear and reduce the margin; offsets beyond 16–18 cm bring the junction close to the end of the bar and another symmetry (the L) becomes possible.

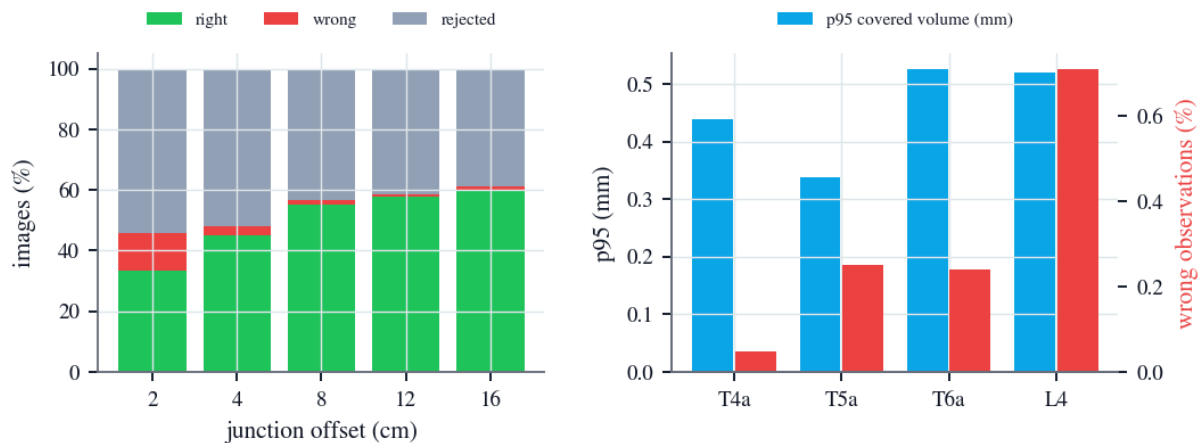


Figure 7. Left: per-image labelling outcome (affine fit with rejection, 10% reflections) versus the asymmetry of the T4. Right: full calibration with the algorithm's labels (affine fit, filter, multi-view relabelling) for four designs.

Comment. With a 2 cm offset 12.6% of the images are mislabelled; from 8 cm upwards it drops to 1–1.4%. Rejection (~40%) is not a problem: rejected images are recovered by multi-view relabelling. In calibration the T4a leaves 0.05% of wrong observations; T5a and T6a reject 80% of the images and leave 0.25% errors, L4 0.71%: the 4-marker design is the most robust.

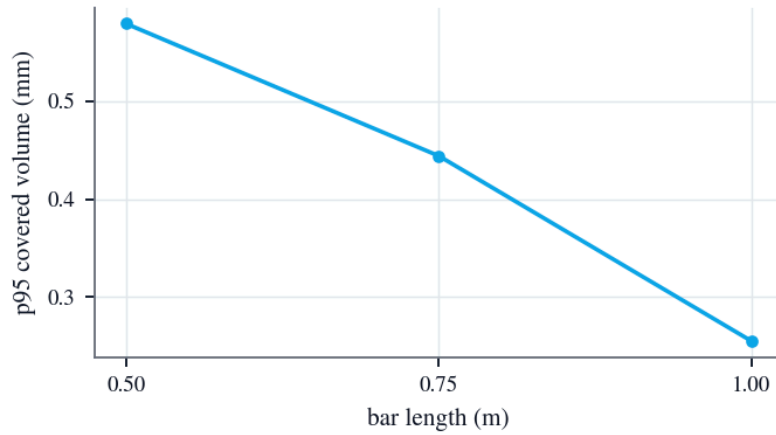


Figure 8. p95 versus the bar length of the T4a (stem half the bar).

Comment. A longer bar constrains scale and intrinsics better: from 0.5 to 1.0 m the p95 goes from 0.58 to 0.26 mm. Beyond 1 m the T becomes hard to move through the volume and to keep inside the frame of nearby cameras; 0.75–1.0 m is the compromise.

5. The static terna

The terna standing at the origin fixes the reference frame and gives the initial camera estimate (PnP). We compare a planar L (two arms on the floor) and a 3D terna (three orthogonal arms), with arms from 0.3 to 1.0 m and 40 mm markers, and the effect of a construction tolerance σ_T (random deviation of each marker from its design position).

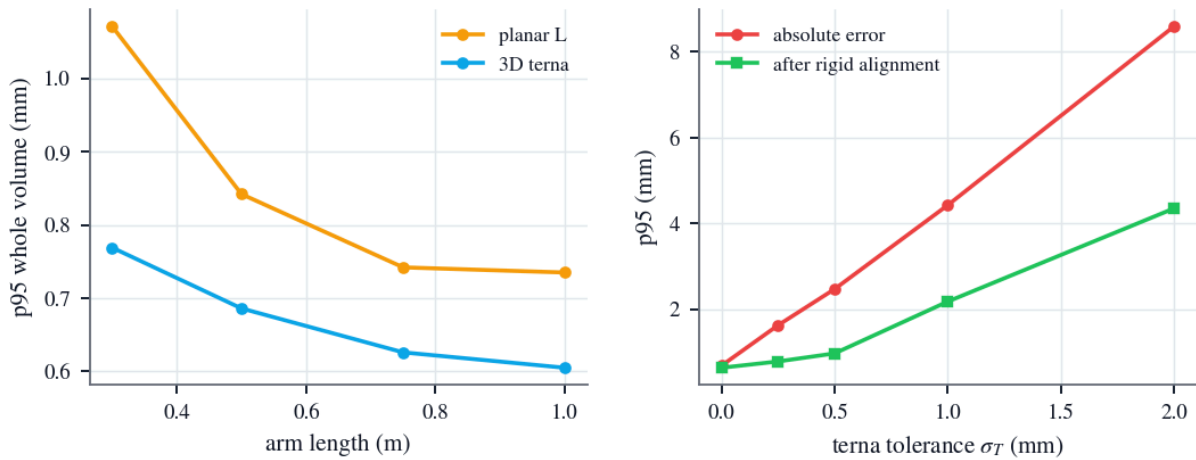


Figure 9. Left: planar L versus 3D terna for different arm lengths. Right: effect of the construction tolerance of the terna (0.75 m arms); green: the error left after the best rigid alignment of the reconstructed volume.

Comment. The 3D terna beats the L at every length (0.63 against 0.74 mm at 0.75 m) and longer arms help little beyond 0.75 m. Tolerance, instead, dominates: every 0.25 mm of deviation adds about 1 mm of p95. About half of this error is a **reference-frame error** (rotation and translation of the whole volume): after rigid alignment 0.78 mm remain with 0.25 mm of tolerance and 2.17 mm with 1 mm; the scale stays correct within 0.03%. For relative measurements (joint angles, distances) the remaining half matters; for absolute measurements (position relative to the floor or the treadmill) the terna must be built within 0.1–0.2 mm or measured.

6. The ideal object and conclusions

Ideal object (IR markers, monochrome cameras)

- T-wand with 4 retro-reflective spheres, 0,75–1,0 m bar, junction and stem 14 cm off-centre, 45 cm stem (margin 0.152).
- Marker diameter: the smallest that gives about 3 px at the median distance (Table 1): 30 mm in HD, 20 mm in Full HD, 10 mm in 4K (120° lenses in a 6 m room). Bigger does not help.
- Distance between markers ≥ 3 diameters: blobs merge below 1.05 apparent diameters and at a 70° grazing angle distances shrink to a third.
- White threshold 40% (useful window 15–45%), weighted centroid, rejection of non-circular blobs.
- 3D terna with 0.75 m arms, 40 mm spheres, built or measured within 0.1–0.2 mm.
- 5–6 markers or markers of different sizes are not needed: they increase ambiguity with reflections.

The least intuitive result is that **bigger markers are not better**: above the detectability limit the dominant error is partial occlusion, which in millimetres is proportional to the diameter of the sphere. The second is that recognition from anonymous points can be designed: the identification margin gives a quantitative criterion to choose the shape, and rewards a strongly asymmetric 4-marker T over objects with more markers.

Limits: the marker model does not contain non-uniform material, specular reflections at the rim, rolling shutter; the unmodelled error is summarised by σ_0 . The 120° lenses are imposed by a 6 m room with wall-mounted cameras; with narrower lenses the recommended diameters scale with the focal length.

References

1. Z. Zhang, *A flexible new technique for camera calibration*, IEEE Trans. PAMI 22(11), 2000.
2. R. Hartley, A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed., Cambridge University Press, 2004.
3. B. Triggs, P. McLauchlan, R. Hartley, A. Fitzgibbon, *Bundle adjustment — a modern synthesis*, Vision Algorithms: Theory and Practice, LNCS 1883, 2000.
4. M. R. Shortis, T. A. Clarke, T. Short, *A comparison of some techniques for the subpixel location of discrete target images*, Proc. SPIE 2350, Videometrics III, 1994.
5. T. Svoboda, D. Martinec, T. Pajdla, *A convenient multicamera self-calibration for virtual environments*, Presence 14(4), 2005.
6. W. Kabsch, *A solution for the best rotation to relate two sets of vectors*, Acta Crystallographica A32, 1976.