

Quante camere e dove: altezze sfalsate, viste simultanee minime e soggetto che esce dalla scena

Anelli da 3 a 12 camere monocromatiche IR in una stanza di 6×6 m

SIMONE TORTORELLA

Ottobre 2026

SOMMARIO

Si studia quante camere servono e dove metterle per un volume di $3,5 \times 3,5 \times 2$ m in una stanza di 6×6 m, con marker sferici IR e camere monocromatiche. Si separano due domande che spesso si confondono: la qualità della **calibrazione** e la qualità della **ricostruzione** dei marker sull'atleta. La calibrazione con terna e T riesce già con 3 camere e nel volume coperto non migliora in modo sistematico con il numero di camere (p95 tra 0,3 e 0,45 mm da 3 a 12 camere). Il numero di camere conta invece per la ricostruzione: la quota di marker visti da almeno 3 camere passa dal 25% con 4 camere al 91% con 8 e al 96% con 10, poi si appiattisce; il richiamo dei marker ricostruiti segue la stessa curva (97,0%, 98,7%, 99,7% con 8, 10, 12 camere). Il guadagno converge quindi tra 8 e 10 camere. Camere sfalsate in altezza (due o tre file) riducono l'errore di ricostruzione del 10–20% rispetto a una sola fila alta; il vantaggio viene soprattutto dalle camere basse. Ogni immagine dell'oggetto di calibrazione deve essere vista da almeno 2 camere contemporaneamente (con 1 la calibrazione fallisce), meglio 3. Se il soggetto esce dal volume la calibrazione non peggiora: i frame in cui l'oggetto è visto da meno di 2 camere semplicemente non contribuiscono.

Parole chiave: numero di camere · posizionamento · altezze sfalsate · occlusione · viste simultanee

1. Introduzione

Nei sistemi a marker la domanda "quante camere?" ha due risposte diverse a seconda di cosa si misura. Per calibrare servono abbastanza viste dell'oggetto di calibrazione da vincolare ogni camera e collegarla alle altre. Per ricostruire un atleta servono abbastanza camere da vedere ogni marker da almeno due (meglio tre) direzioni nonostante il corpo che lo copre. Qui le due domande sono studiate separatamente con lo stesso simulatore: stanza di 6×6 m, volume di $3,5 \times 3,5 \times 2$ m, camere sulle pareti a distanza uniforme lungo il perimetro, puntate al centro del volume, con l'ottica più stretta che lo inquadra (sempre 120° in questa stanza).

Le altezze provate sono: tutte a 2,2 m ("una fila"), tutte a 1,0 m ("bassa"), alternate 2,4/1,0 m ("due file") e a rotazione 2,4/1,6/0,8 m ("tre file"). L'ipotesi di partenza è che **camere sfalsate in altezza** riducano le occlusioni e migliorino la geometria della triangolazione.

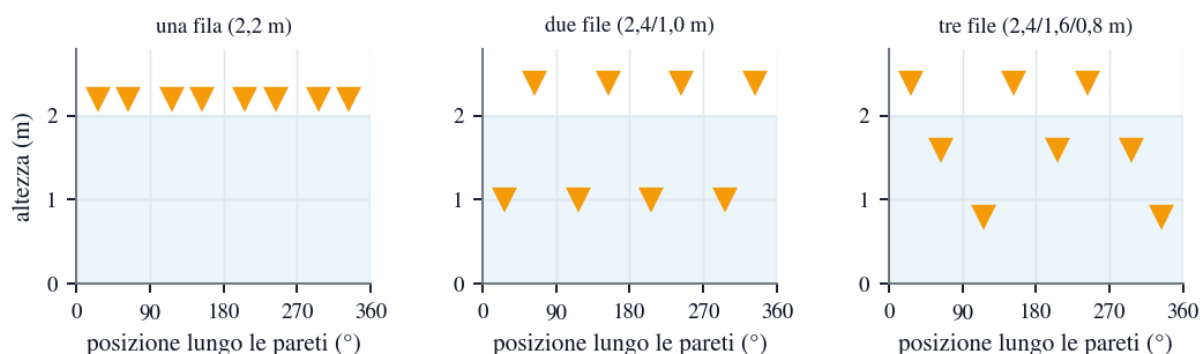


Figura 1. Disposizione di 8 camere nelle tre configurazioni di altezza, "srotolata" lungo il perimetro della stanza; la fascia azzurra è l'altezza del volume di misura (0–2 m).

Commento. Con una sola fila tutte le camere guardano il volume dall'alto con angoli simili: i piedi e la parte bassa del corpo sono visti di scorcio. Con due o tre file metà delle camere vede il volume quasi orizzontalmente.

2. Calibrazione e numero di camere

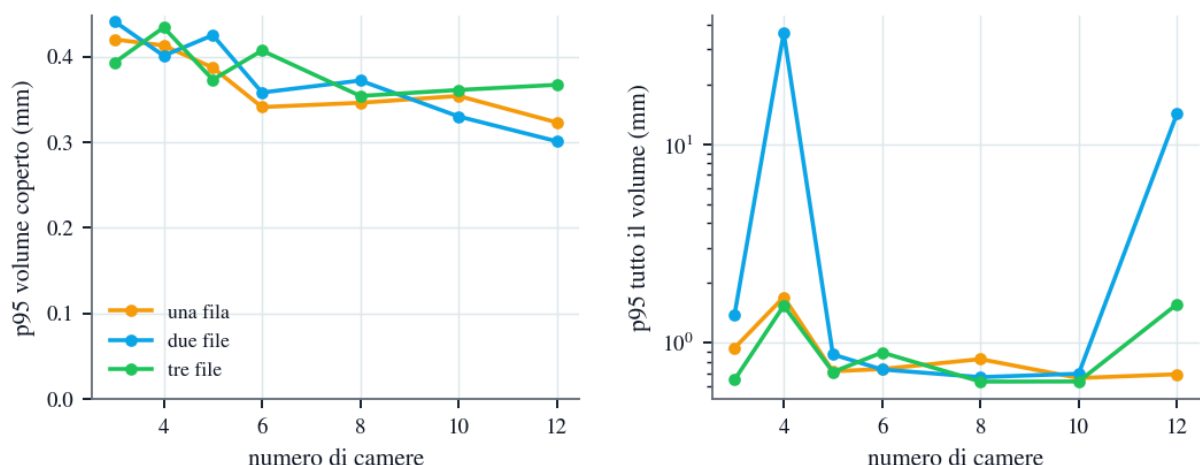


Figura 2. Errore di calibrazione (mediana di 6 sessioni da 20 s) in funzione del numero di camere e delle altezze. Sinistra: dove la T è passata; destra: su tutto il volume di 3,5×3,5×2 m.

Commento. Nel volume coperto l'errore resta tra 0,3 e 0,45 mm da 3 a 12 camere, senza una tendenza chiara: ogni camera è calibrata dalle stesse osservazioni della T e aggiungere camere non migliora le altre (errore massimo di posizione delle camere 1,2–1,9 mm, di rotazione 0,02–0,04°, per ogni numero di camere). Sul volume intero compaiono invece picchi di decine di millimetri con 4 e 12 camere su due file (oltre 5 mm in 100% e 100% delle sessioni). Non sono camere calibrate male: sono punti agli angoli del volume che una camera vede nell'angolo della sua immagine (raggio normalizzato 1,35) mentre la T, in quella camera, non è mai andata oltre 0,9–1,0. Lì la distorsione dell'ottica da 120° è estrapolata e il raggio è sbagliato. La conclusione è che **il numero di camere non è un parametro della calibrazione**: conta quanto la T copre il volume e **l'immagine di ogni camera fino agli angoli** (Paper 3).

3. Ricostruzione: quante camere vedono ogni marker

Per la ricostruzione si usa la scena del Paper 4: una persona che cammina nel volume con 30 marker da 14 mm sul corpo, occlusi dal corpo stesso (modello a capsule), osservati con rumore di 0,15 px e riflessi casuali. Si conta quanti marker sono visti da almeno 2 e 3 camere e quanti l'algoritmo di triangolazione per marker anonimi ricostruisce (richiamo), con il supporto minimo di 3 viste e il tracciamento.

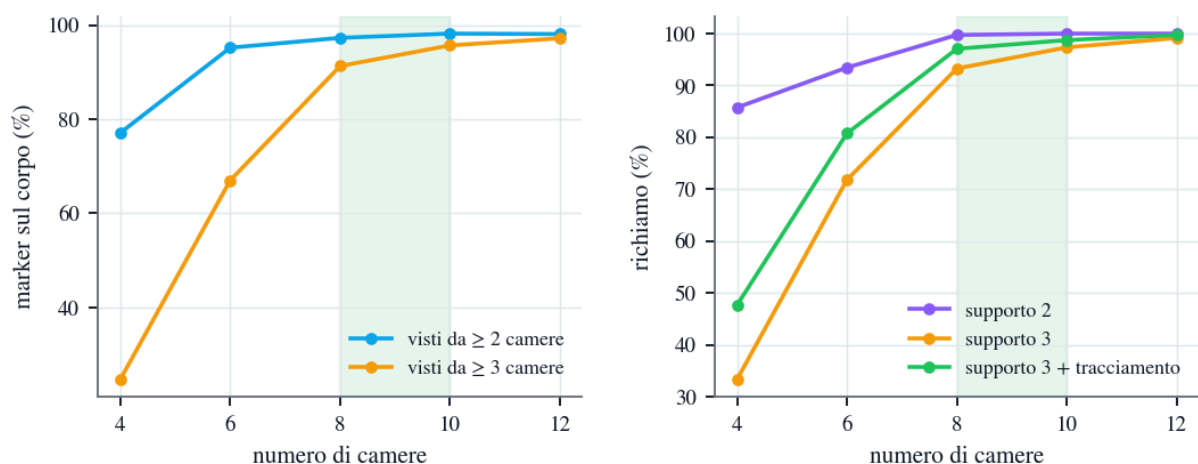


Figura 3. Sinistra: quota di marker del corpo visti da almeno 2 e 3 camere. Destra: richiamo della ricostruzione 3D di marker anonimi (Paper 4) in funzione del numero di camere (due file di altezza).

Commento. Qui il numero di camere conta. Con 4 camere solo il 25% dei marker è visto da 3 camere e il richiamo con supporto 3 è il 33%; con 8 camere si sale al 91% e 97,0%; con 10 al 96% e 98,7%; con 12 si guadagna meno di un punto. **Il guadagno converge tra 8 e 10 camere:** oltre, ogni camera in più aggiunge circa l'1% di marker e costa banda e calcolo (il tempo dell'algoritmo cresce come il quadrato del numero di camere, Paper 4). Il minimo pratico per un atleta in movimento è 6 camere (richiamo 81% con tracciamento).

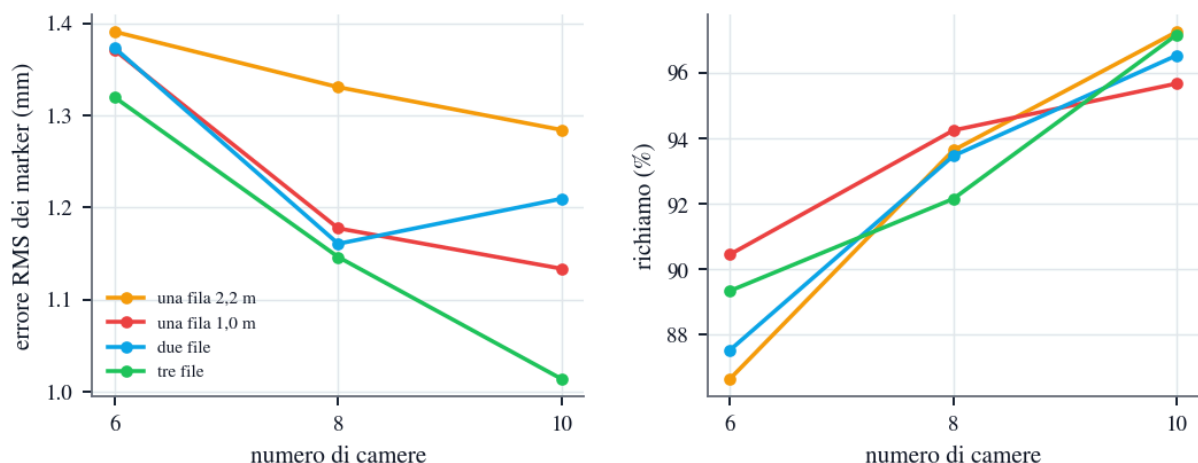


Figura 4. Effetto delle altezze sulla ricostruzione dei marker del corpo (mediana di 4 sequenze, supporto 3 con tracciamento).

Commento. L'ipotesi è confermata in parte. Con 8 camere la fila alta dà 1,33 mm, le due file 1,16 mm, le tre file 1,15 mm, la fila bassa 1,18 mm; con 10 camere le tre file arrivano a 1,01 mm contro 1,28 mm della fila alta. Il guadagno (10–20%) viene dalle camere basse: vedono i marker di gambe e piedi che quelle alte vedono di scorcio o coperti dal tronco, e i loro raggi incrociano quelli delle camere alte con un angolo più ampio, riducendo l'errore lungo la verticale. Una fila tutta bassa fa quasi altrettanto bene sull'errore, ma perde le braccia sopra la testa e i marker coperti dagli attrezzi; sfalsare tiene i vantaggi di entrambe. Il richiamo invece cambia poco ($\pm 2\%$): conta il numero di camere più della loro altezza.

4. Quante camere devono vedere l'oggetto insieme

Il bundle adjustment collega le camere attraverso i frame in cui la T è vista da più di una camera. Per misurare il minimo necessario, in ogni frame si lasciano al più k camere (scelte a caso tra quelle che vedono la T) e si cancellano le osservazioni delle altre. Con $k = 1$ ogni camera vede la T da sola: gli intrinseci restano vincolati (come con una scacchiera), ma le camere sono collegate solo dalla terna.

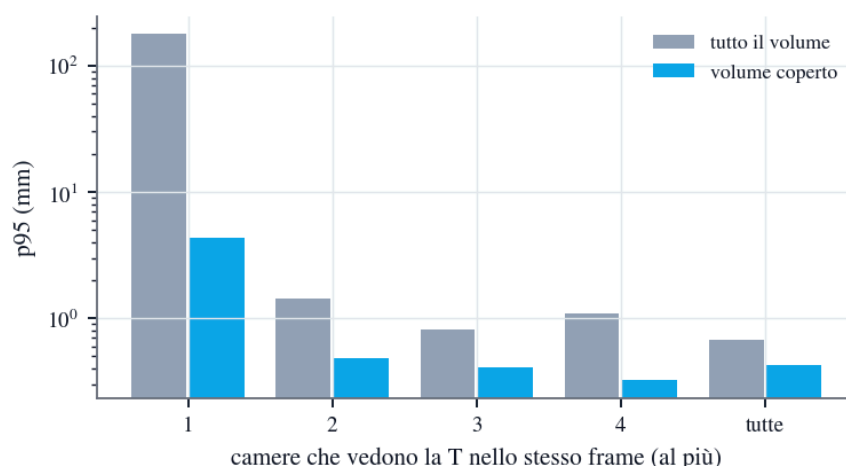


Figura 5. Errore di calibrazione quando in ogni frame la T può essere vista al più da k camere (8 camere, mediana di 5 sessioni).

Commento. Con una sola camera per frame la calibrazione fallisce (179 mm sul volume); senza viste condivise la posizione relativa delle camere dipende solo dalla terna, vista da lontano. Con 2 camere per frame funziona (1,44 mm sul volume, 0,48 mm dove passa la T); con 3 si è già vicini al caso con tutte le camere (0,81 contro 0,68 mm). **Minimo: 2 viste simultanee; consigliate: 3.** In pratica significa che la T deve passare dove le camere si sovrappongono, non solo vicino alle pareti.

5. Il soggetto esce dalla scena

Nella realtà chi muove la T esce dal volume, si avvicina a una parete, passa davanti a una camera. Si simula allargando il percorso della persona di un fattore ρ (1 = sempre nel volume; 2 = fino a 0,5 m dalle pareti) e si conta in ogni frame quante camere vedono almeno 3 marker della T.

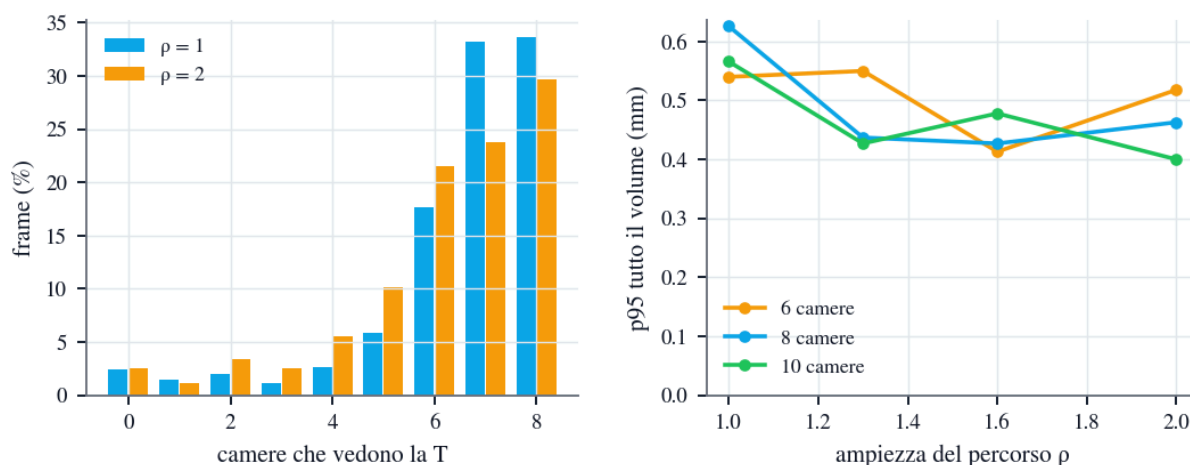


Figura 6. Sinistra: distribuzione del numero di camere che vedono la T (8 camere) con il percorso nel volume ($\rho = 1$) e fino alle pareti ($\rho = 2$). Destra: errore di calibrazione al variare di ρ .

Commento. Uscendo dal volume cresce la quota di frame in cui la T è vista da poche camere (≤ 4 camere: dal 10% al 15%), ma la calibrazione non peggiora, anzi resta tra 0,4 e 0,6 mm: i frame con meno di 2 viste non portano informazione ma non fanno danni, e la T che passa vicino alle pareti allarga il volume coperto. Le occlusioni dovute all'uscita di scena non sono quindi un problema se il bundle adjustment usa solo le osservazioni valide e se nel resto della sessione la T attraversa il centro del volume.

6. Orientamento e risoluzioni miste

Tabella 1. Orientamento delle camere e rig con risoluzioni diverse (mediana delle sessioni).

Configurazione	p95 (mm)	Nota
8 camere orizzontali	0,63	tutto il volume
8 camere verticali	0,59	tutto il volume
8 camere Full HD	0,43	volume coperto, 60 fps
4 camere 4K + 4 HD alternate	0,36	volume coperto, 60 fps
8 camere 4K	0,29	volume coperto, 60 fps

Camere verticali e orizzontali si equivalgono in un volume quasi cubico. Un rig misto (metà 4K, metà HD) sta tra i due rig omogenei: il bundle adjustment pesa ogni osservazione con il suo errore atteso, quindi le camere ad alta risoluzione contano di più senza bisogno di regole speciali. Frequenze diverse tra camere richiedono invece la stima del ritardo per camera (studio preliminare) e non sono state simulate qui.

7. Conclusioni

Raccomandazioni

- Calibrazione: 3 camere bastano a calibrare; il numero di camere non migliora la calibrazione, la copertura del volume e degli angoli di ogni immagine sì.
- Ricostruzione di un atleta: minimo 6 camere, ottimo 8–10; oltre 10 il guadagno è $< 1\%$ di marker ($98,7\% \rightarrow 99,7\%$).
- Altezze: sfalsate su due file (2,4 e 1,0 m) o tre con 10 camere o più: $-10-20\%$ di errore rispetto a una sola fila alta; servono camere basse per gambe e piedi.
- Viste simultanee della T: minimo 2, consigliate 3: muovere la T al centro del volume, dove le camere si sovrappongono.
- Il soggetto può uscire dalla scena: i frame con meno di 2 viste non contribuiscono e non danneggiano la calibrazione.

Limiti: un solo tipo di stanza (6×6 m) e un volume fisso; camere solo a parete; corpo semplificato a capsule. In stanze più grandi le ottiche si stringono e i marker diventano più grandi in pixel (Paper 1), ma il ragionamento sul numero di viste non cambia.

Riferimenti

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.