

Oggetto di calibrazione per marker IR anonimi: dalla soglia di bianco alla forma ottimale

Marker sferici retroriflettenti, camere monocromatiche, rilevamento a soglia

SIMONE TORTORELLA

Ottobre 2026

SOMMARIO

Si studia l'oggetto con cui calibrare un sistema di 8 camere monocromatiche con illuminatore infrarosso e marker sferici retroriflettenti, i cui centri sono stimati come baricentri dei pixel sopra una soglia di bianco. Un modello a livello di pixel del marker (profilo retroriflettente, PSF, mosso, rumore, compressione) mostra che con sfondo IR scuro l'errore del baricentro è di circa 0,011 px per un marker di 8 px con soglia al 40%, e che l'errore dominante non è il rumore ma l'occlusione parziale (0,82 px quando un terzo del disco è coperto). In simulazioni complete (persona che muove l'oggetto, 8 camere, bundle adjustment) si confrontano risoluzione (HD, Full HD, 4K), frequenza (30–120 fps), diametro dei marker (10–50 mm), sei disegni dell'oggetto e la terna statica. Il marker deve essere il più piccolo che la risoluzione consente, circa 3 px alla distanza mediana: 30 mm in HD, 20 mm in Full HD, 10 mm in 4K. Per riconoscere l'oggetto da punti anonimi si definisce un margine di identificazione e lo si ottimizza: la T a 4 marker con giunzione spostata di 14 cm e gambo di 45 cm massimizza il margine; le varianti a 5–6 marker non sono più robuste. La tolleranza di costruzione della terna statica è il limite principale dell'accuratezza assoluta: 0,25 mm di tolleranza portano il p95 da 0,68 a 1,62 mm.

Parole chiave: calibrazione multi-camera · marker infrarossi · baricentro di blob · soglia · bacchetta a T · identificazione di marker anonimi

1. Introduzione

I sistemi di analisi del movimento a marker usano camere monocromatiche con un anello di LED infrarossi attorno all'obiettivo e un filtro che lascia passare solo l'infrarosso. I marker sono sfere rivestite di materiale retroriflettente: rimandano la luce verso la camera e nell'immagine appaiono come dischi quasi bianchi su fondo quasi nero. Il centro di ogni marker si stima con una soglia di bianco e il baricentro dei pixel che la superano. Le camere non sanno a quale marker appartiene ciascun blob: vedono solo punti anonimi.

La calibrazione proposta in uno studio precedente usa una terna 3D ferma nell'origine (definisce origine, assi e scala) e una bacchetta a T mossa da una persona nel volume (vincola intrinseci, estrinseci e ritardi con un bundle adjustment sparso). Qui si risponde a una domanda più concreta: **come deve essere fatto l'oggetto**, e quali parametri del marker e della camera contano davvero. Tutte le prove sono simulazioni con verità nota: si confronta sempre la calibrazione stimata con quella vera.

- §2 modello a pixel del marker e del rilevamento a soglia;
- §3 dal pixel al millimetro: risoluzione, fps e diametro del marker in calibrazioni complete;
- §4 riconoscere l'oggetto da punti anonimi: margine di identificazione e sua ottimizzazione;
- §5 la terna statica; §6 l'oggetto ideale.

2. Il marker visto dalla camera

Una sfera retroriflettente di diametro apparente d pixel, illuminata dall'anello coassiale, ha nell'immagine un profilo di intensità quasi piatto che cala vicino al bordo. Lo modelliamo come

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

con fondo $b = 0,05$ e guadagno $g = 1,4$ (il centro satura a 255). L'immagine è sovra-campionata 8×, convoluta con la PSF dell'obiettivo (gaussiana, $\sigma = 0,7$ px) e con il mosso dell'esposizione, integrata sui pixel, con rumore fotonico (6000 e⁻ al bianco) e di lettura (2 DN), quantizzata a 8 bit. Il rilevatore applica una soglia τ (frazione del bianco), cerca le componenti connesse e ne calcola il baricentro pesato:

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

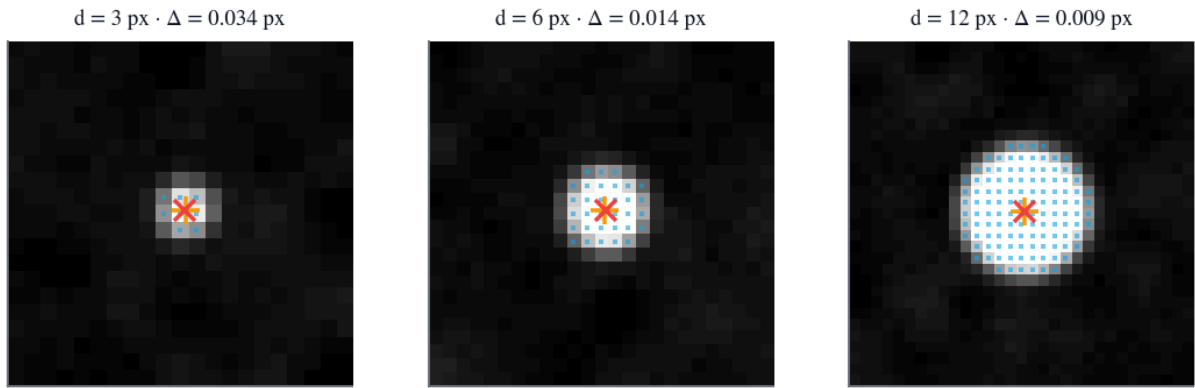


Figura 1. Sfere simulate di 3, 6 e 12 px. Quadratini azzurri: pixel sopra la soglia del 40%. Croce gialla: centro vero; croce rossa: baricentro stimato.

Commento. Anche il marker da 3 px ha un errore di pochi centesimi di pixel: con sfondo scuro e molti fotoni il baricentro è molto preciso. Il problema, come vedremo, non è il rumore.

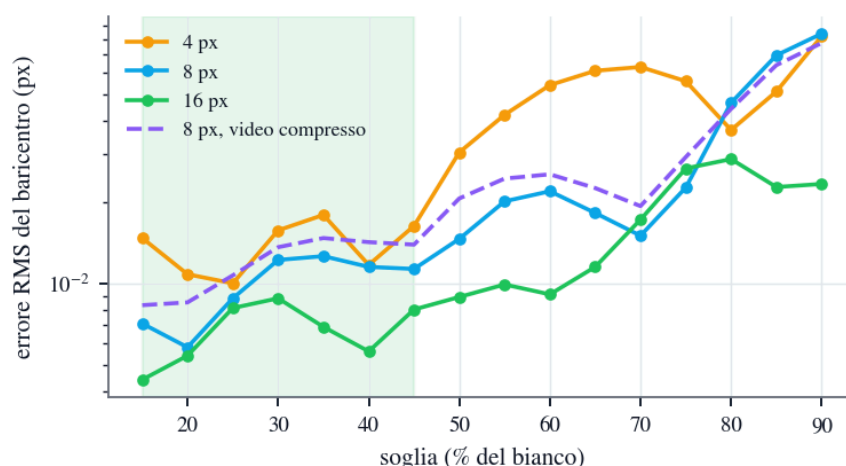


Figura 2. Errore del baricentro (scala log) al variare della soglia, per marker di 4, 8 e 16 px; tratteggio: marker di 8 px con video compresso (qualità 80). Tutti i marker sono rilevati fino all'85% di soglia.

Commento. Con fondo IR scuro la soglia ottimale è bassa: tra 15% e 45% l'errore è minimo (0,011 px a 8 px e soglia 40%). Alzare la soglia butta via il bordo del disco, che porta informazione: sopra il 75% l'errore cresce rapidamente (0,046 px all'80%). La compressione costa poco (0,014 px) se il video ha qualità ≥ 70 . Si sceglie una soglia del 40%: abbastanza alta da tagliare i riflessi IR deboli (luce solare, superfici lucide), abbastanza bassa da usare tutto il disco.

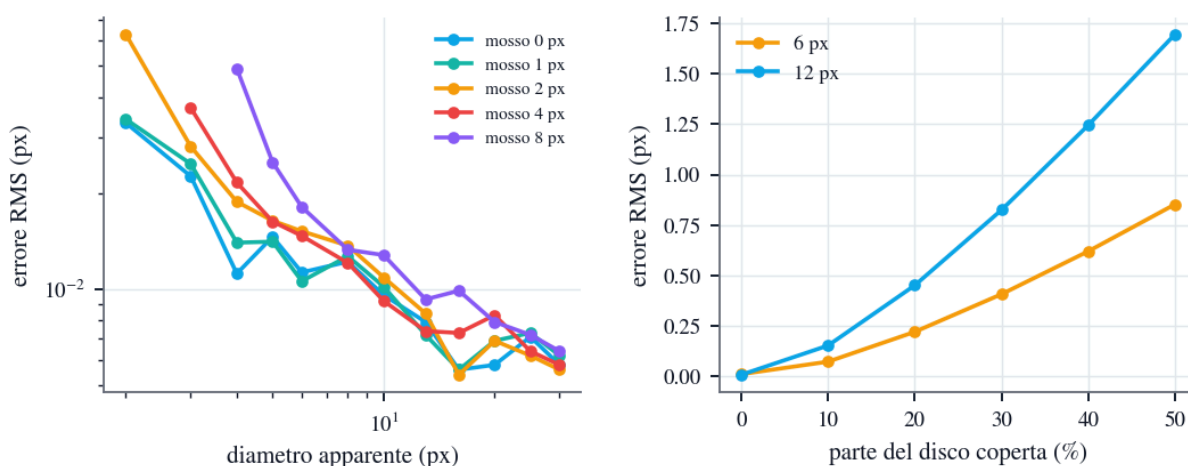


Figura 3. Sinistra: errore del baricentro in funzione del diametro apparente e del mosso (soglia 40%). Destra: errore quando una corda copre parte del disco (un braccio, il bastone della T, un altro marker).

Commento. Il mosso conta poco: per marker di almeno 6 px anche una scia di 8 px lascia l'errore sotto 0,03 px, perché il centro della scia resta il centro del marker se il moto è uniforme durante l'esposizione; pesa solo sui marker più piccoli. L'occlusione parziale invece sposta il baricentro verso la parte visibile: con un terzo del disco coperto l'errore è 0,41 px a 6 px e 0,82 px a 12 px, cioè **proporzionale al diametro**, e 30–100 volte il rumore. In millimetri questo errore è proporzionale al diametro fisico della sfera: è il primo motivo per preferire marker piccoli e per scartare i blob deformati (§3 e Paper 3).

Due marker vicini si fondono in un solo blob quando la distanza tra i centri scende sotto circa 1,05 diametri apparenti (simulazione, soglia 40%). Infine, una componente d'errore che il modello non contiene (residuo del modello di distorsione, rolling shutter, vibrazioni, non uniformità del materiale) è rappresentata nelle simulazioni successive da un termine $\sigma_0 = 0,08$ px aggiunto in quadratura; il Paper 3 ne studia la sensibilità.

3. Dal pixel al millimetro: risoluzione, fps, diametro

Ogni simulazione completa ha 8 camere sulle pareti di una stanza di 6×6 m, a due altezze alternate (2,4 e 1,0 m), con l'ottica più stretta che inquadra il volume di 3,5×3,5×2 m (qui sempre 120°, focale 554 px in Full HD). Una persona cammina per 20 s muovendo la T; il bundle adjustment usa 400 frame; un frame su cinque è tenuto da parte come verifica. La metrica principale è il p95 dell'errore 3D nella parte del volume attraversata dalla T ("volume coperto"), triangolando osservazioni esatte con le camere stimate. Il diametro apparente di un marker è $d = fD/Z$: una sfera da 30 mm a 3,6 m (distanza mediana) è di

Tabella 1. Diametro apparente (px) a 3,6 m (mediana) con ottica da 120°; ultima colonna: 4K al 95° percentile della distanza.

Diametro	HD	Full HD	4K	4K a 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

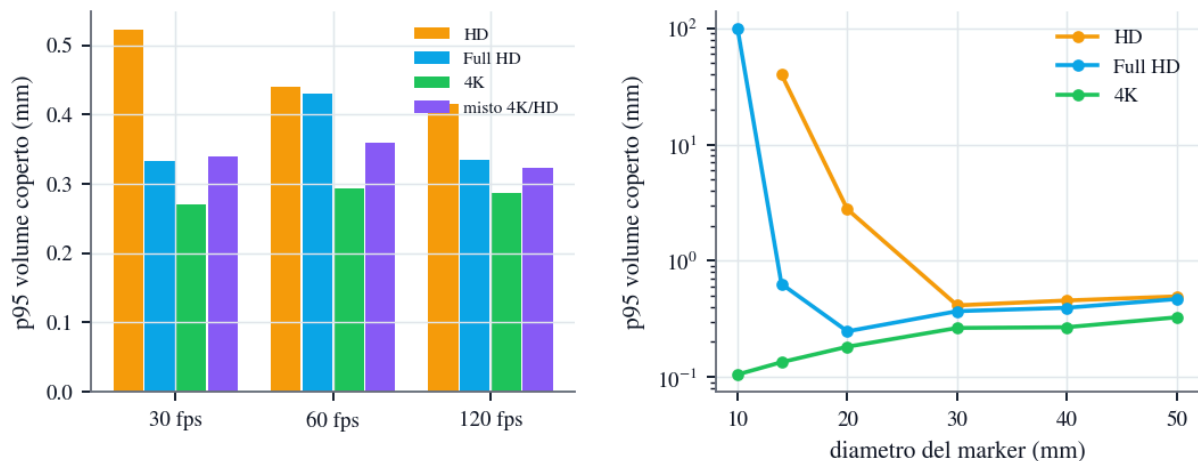


Figura 4. Sinistra: p95 per risoluzione e frequenza (marker da 30 mm; esposizione 1/(2-fps); mediana di 5 sessioni). Destra: p95 in funzione del diametro dei marker per le tre risoluzioni (60 fps). Valori mancanti: calibrazione non riuscita.

Commento. La risoluzione conta più della frequenza: il 4K porta il p95 a circa 0,29 mm contro 0,44 mm dell'HD (60 fps). Con la T mossa a mano il mosso è piccolo anche a 30 fps, quindi la frequenza conta poco per la calibrazione (conta per la sincronizzazione, studio preliminare). Il diametro ha un ottimo: sotto circa 3 px alla distanza mediana (Tabella 1) i marker lontani spariscono e la calibrazione degrada o fallisce (HD sotto 30 mm, Full HD sotto 14–20 mm); sopra, l'errore cresce di nuovo con il diametro perché l'occlusione parziale e la fusione di blob vicini pesano di più. In 4K anche 10–14 mm funzionano e danno l'errore più basso (0,10 mm).

4. Riconoscere l'oggetto da punti anonimi

In ogni immagine il programma deve dire quale blob è quale marker. L'oggetto è piano: la sua immagine è una trasformazione proiettiva del disegno, ben approssimata da una affine quando l'oggetto è piccolo rispetto alla distanza. Per un'assegnazione π dei blob $\mathbf{u}_i$ ai marker $\mathbf{x}_{\pi(i)}$ del disegno, il residuo normalizzato del miglior fit affine è

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

Un'affine ha 6 gradi di libertà: con 3 marker ogni permutazione ha residuo nullo e non c'è nulla da verificare; con un disegno simmetrico la riflessione sinistra/destra è anch'essa un'affine esatta. Definiamo il **margin** di identificazione come la differenza tra il miglior residuo di un'etichettatura sbagliata e quello dell'etichettatura giusta, al 10° percentile su 200 punti di vista con inclinazione fino a 70°:

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

Il margine è adimensionale (frazione della dimensione dell'oggetto nell'immagine): l'etichettatura resta sicura finché il rumore dei baricentri, diviso la dimensione dell'oggetto in pixel, è molto più piccolo di m . Con l'ipotesi di un blob fantasma (riflesso) si considerano anche le etichettature che scartano un marker.

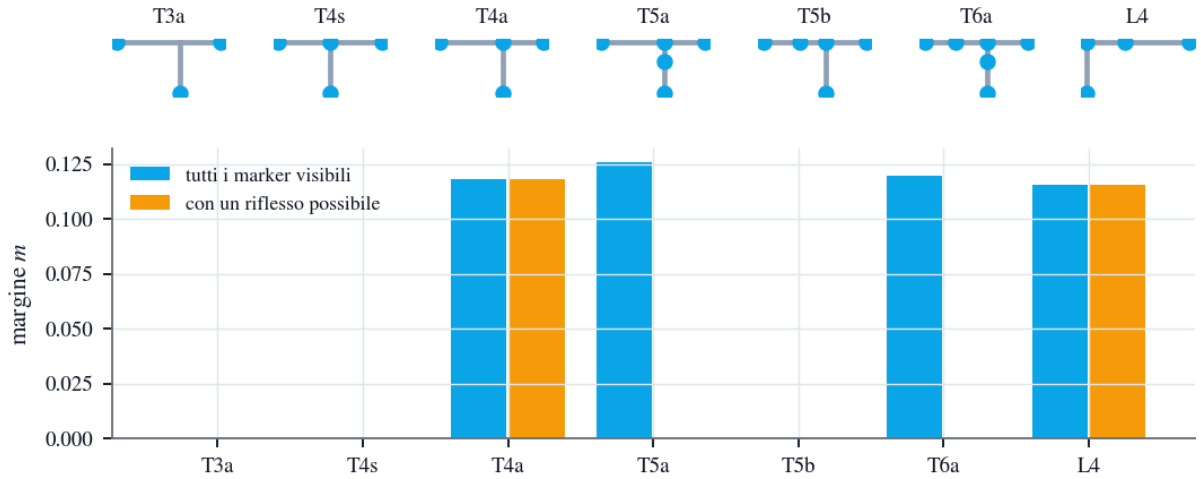


Figura 5. I sette disegni (barra di 0,75 m) e il loro margine di identificazione. T3a: 3 marker; T4s/T4a: T a 4 simmetrica/asimmetrica; T5a: T4a con un marker nel gambo; T5b: 4 marker allineati sulla barra; T6a: T5a con un marker in più sulla barra; L4: L a 4 marker.

Commento. I disegni a 3 marker e quelli simmetrici hanno margine nullo, come previsto. T4a e L4 hanno margine ~0,118. Aggiungere marker non aiuta: T5a e T6a hanno lo stesso margine quando tutti i marker si vedono, ma diventano ambigui appena si ammette un riflesso (margine ≤ 0), perché togliendo un marker restano sottoinsiemi con tre punti allineati che si adattano bene a etichettature sbagliate. T5b, con quattro punti sulla stessa retta, è ambiguo anche senza riflessi: i rapporti tra le distanze lungo una retta sono troppo simili nei due versi.

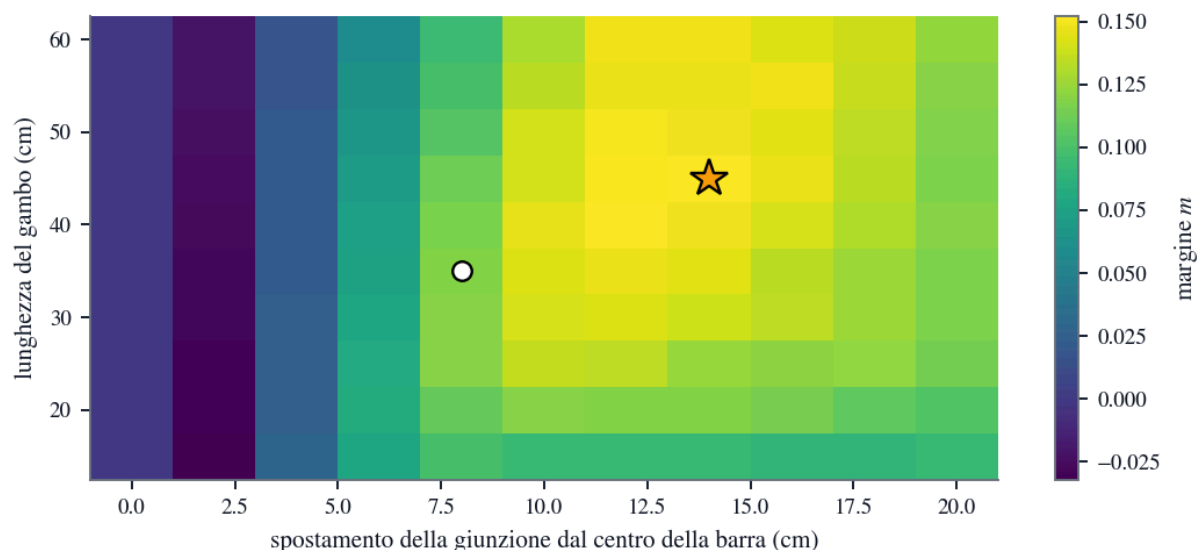


Figura 6. Margine di identificazione della T a 4 marker (barra 0,75 m) in funzione dello spostamento della giunzione e della lunghezza del gambo. Stella: ottimo; cerchio: la T dello studio preliminare (8 cm, 35 cm).

Commento. Il margine cresce con l'asimmetria fino a circa 14 cm e con la lunghezza del gambo fino a 45 cm, poi si appiattisce: oltre, la T diventa scomoda da maneggiare senza guadagni. L'ottimo (0,152) è il 28% sopra la T dello studio preliminare (0,119). Gambi corti rendono la T quasi collineare e riducono il margine; spostamenti oltre 16–18 cm avvicinano la giunzione all'estremo della barra e un'altra simmetria (la L) diventa possibile.

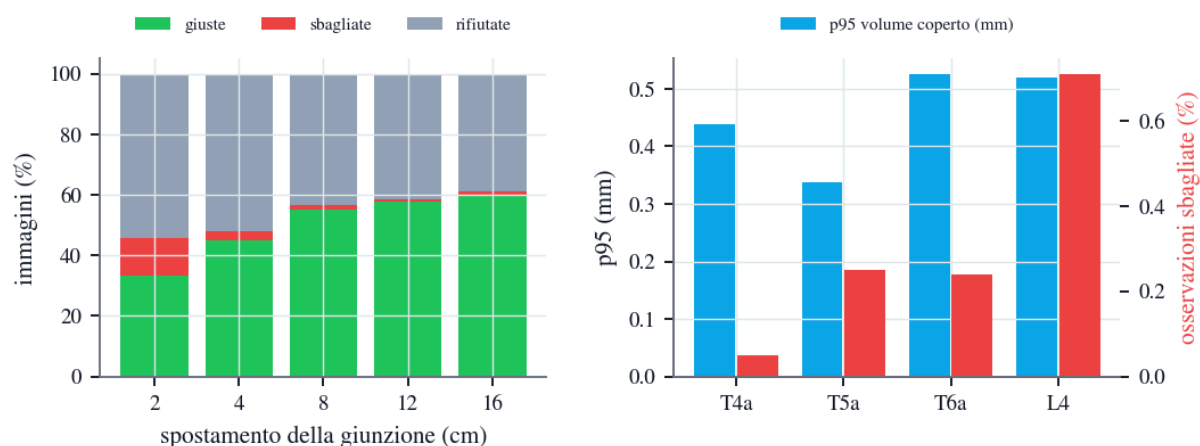


Figura 7. Sinistra: esito dell'etichettatura per immagine (fit affine con rifiuto, 10% di riflessi) al variare dell'asimmetria della T4. Destra: calibrazione completa con le etichette dell'algoritmo (fit affine, filtro, ri-etichettatura multi-vista) per quattro disegni.

Commento. Con 2 cm di spostamento il 12,6% delle immagini è etichettato male; da 8 cm in su si scende all'1–1,4%. Il rifiuto (~40%) non è un problema: le immagini rifiutate vengono recuperate dalla ri-etichettatura multi-vista. In calibrazione la T4a lascia 0,05% di osservazioni sbagliate; T5a e T6a rifiutano l'80% delle immagini e lasciano 0,25% di errori, L4 il 0,71%: il disegno a 4 marker è il più robusto.

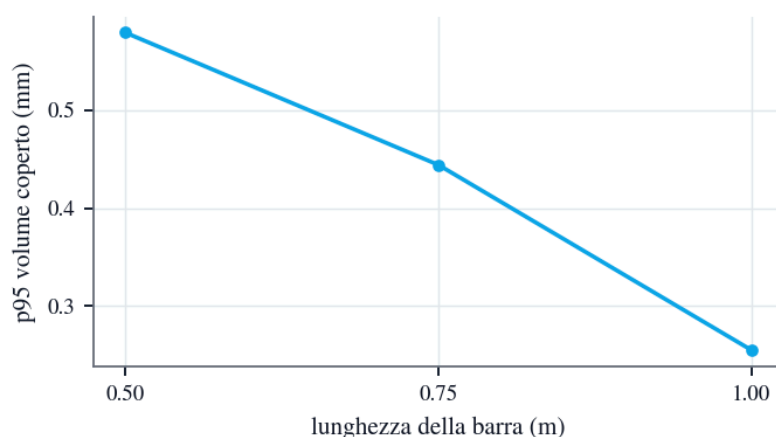


Figura 8. p95 in funzione della lunghezza della barra della T4a (gambo pari a metà barra).

Commento. Una barra più lunga vincola meglio scala e intrinseci: da 0,5 a 1,0 m il p95 passa da 0,58 a 0,26 mm. Oltre 1 m la T diventa difficile da muovere nel volume e da tenere dentro l'inquadratura delle camere vicine; 0,75–1,0 m è il compromesso.

5. La terna statica

La terna ferma nell'origine fissa il sistema di riferimento e dà la stima iniziale delle camere (PnP). Si confrontano una L piana (due bracci sul pavimento) e una terna 3D (tre bracci ortogonali), con bracci da 0,3 a 1,0 m e marker da 40 mm, e l'effetto di una tolleranza di costruzione σ_T (scarto casuale di ogni marker dalla posizione di progetto).

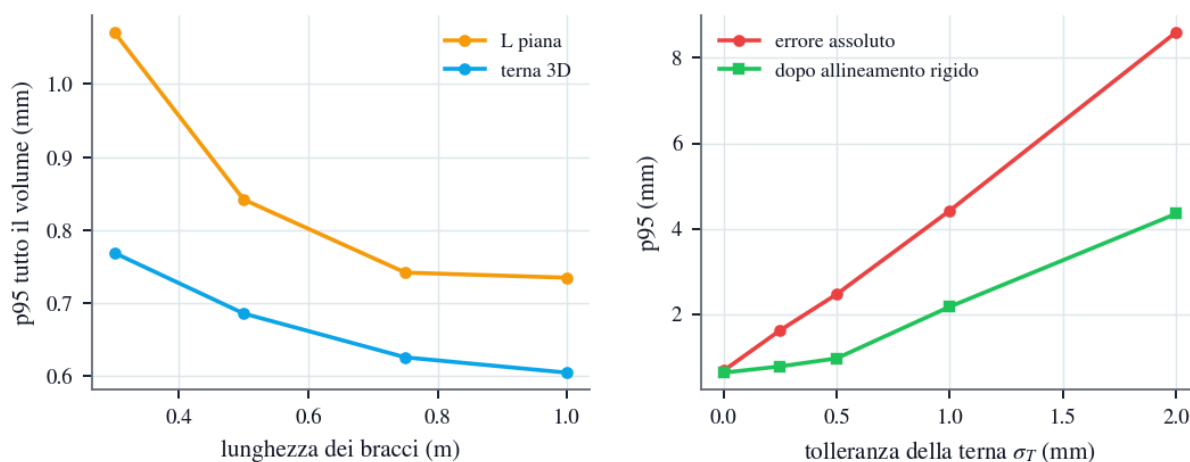


Figura 9. Sinistra: L piana contro terna 3D al variare della lunghezza dei bracci. Destra: effetto della tolleranza di costruzione della terna (bracci da 0,75 m); in verde l'errore che resta dopo il miglior allineamento rigido del volume ricostruito.

Commento. La terna 3D è migliore della L a ogni lunghezza (0,63 contro 0,74 mm a 0,75 m) e bracci più lunghi aiutano poco oltre 0,75 m. La tolleranza invece domina: ogni 0,25 mm di scarto aggiungono circa 1 mm di p95. Circa metà di questo errore è un **errore del sistema di riferimento** (rotazione e traslazione dell'intero volume): dopo l'allineamento rigido restano 0,78 mm con 0,25 mm di tolleranza e 2,17 mm con 1 mm; la scala resta corretta entro lo 0,03%. Per misure relative (angoli articolari, distanze) conta la metà rimanente; per misure assolute (posizione rispetto al pavimento o al treadmill) la terna va costruita entro 0,1–0,2 mm o misurata.

6. L'oggetto ideale e conclusioni

Oggetto ideale (marker IR, camere monocromatiche)

- Bacchetta a T con 4 sfere retroriflettenti, barra 0,75–1,0 m, giunzione e gambo spostati di 14 cm dal centro, gambo di 45 cm (margine 0,152).
- Diametro dei marker: il più piccolo che dia circa 3 px alla distanza mediana (Tabella 1): 30 mm in HD, 20 mm in Full HD, 10 mm in 4K (ottiche da 120° in una stanza di 6 m). Più grande non aiuta.
- Distanza tra marker ≥ 3 diametri: i blob si fondono sotto 1,05 diametri apparenti e a 70° di scorcio le distanze si riducono a un terzo.
- Soglia di bianco 40% (finestra utile 15–45%), baricentro pesato, scarto dei blob non circolari.
- Terna 3D a bracci da 0,75 m, sfere da 40 mm, costruita o misurata entro 0,1–0,2 mm.
- Non servono 5–6 marker né marker di dimensioni diverse: aumentano le ambiguità con i riflessi.

Il risultato meno intuitivo è che **marker più grandi non sono meglio**: oltre la soglia di rilevabilità l'errore dominante è l'occlusione parziale, che in millimetri è proporzionale al diametro della sfera. Il secondo è che il riconoscimento da punti anonimi si progetta: il margine di identificazione dà un criterio quantitativo per scegliere la forma, e premia una T a 4 marker ben asimmetrica rispetto a oggetti con più marker.

Limiti: il modello del marker non contiene materiale non uniforme, riflessi speculari sul bordo, rolling shutter; l'errore non modellato è riassunto in σ_0 . Le ottiche da 120° sono imposte da una stanza di 6 m con camere a parete; con ottiche più strette i diametri consigliati scalano con la focale.

Riferimenti

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.