La Comunicazione - Note, Recensioni e Notizie n. 64

Anno 2021

Analisi dell'incertezza dei giunti scheletrici rilasciati dal sistema Azure Kinect Body Tracking SDK

Uncertainty analysis of the skeleton joints released by the Azure Kinect Body Tracking SDK

Laura Romeo♦, Roberto Marani□, Anna Gina Perri♦

♦ Dipartimento di Ingegneria Elettrica e dell'Informazione (DEI), Politecnico di Bari, Italia

□ Istituto di Sistemi e Tecnologie Industriali Intelligenti per il Manifatturiero Avanzato (STIIMA), Consiglio Nazionale delle Ricerche (CNR), Bari, Italia

Sommario

Il controllo in tempo reale di robot collaborativi, conosciuti anche con il termine cobot, è fondamentale negli ambienti industriali, in quanto permette di ridurre eventuali rischi per i lavoratori, migliorando la loro sicurezza. Le attività che il cobot controlla richiedono dati in input riguardo l'ambiente circostante, al fine di effettuare mappature del luogo in cui si trova e reagire in modo opportuno ad eventi imprevedibili, come possono essere le azioni umane. In particolare, in questo caso, avere una stima sulla posizione dell'operatore aiuta il cobot a meglio prevedere eventuali movimenti improvvisi dell'uomo. Tale scopo può essere raggiunto attraverso l'uso di telecamere RGB-D, il cui output vene processato attraverso sistemi di body tracking in modo da ottenere una stima della posizione dell'uomo in tempo reale. L'obiettivo di questo articolo è quello di analizzare le performance della telecamera Microsoft Azure Kinect RGB-D, insieme alla libreria di body-tracking ad essa associata. È stato realizzato un modello che tiene conto delle diverse condizioni ambientali al contorno durante l'acquisizione delle immagini, in modo da valutare l'incertezza sulla stima dei giunti del corpo a seconda delle condizioni di luce, della presenza di occlusioni, della risoluzione della telecamera, e della distanza tra umano e telecamera. I risultati di tale analisi hanno provato la necessità di saper gestire l'incertezza nel controllo dei cobot che lavorano a stretto contatto con gli umani.

Abstract

Real-time control of cooperative robots, or cobots, in industrial environments is a mandatory task to reduce the risk for workers by improving their safety. The task of cobot control always requires input data about the surroundings to enable planning procedures and proper reactions to unpredictable events, such as human actions. In this case, the exact position of the humans can be easily inferred from RGB-D cameras, whose output can be processed by body tracking modules to produce exact pose estimations in real-time. This paper experimentally explores the performance of the affordable Microsoft Azure Kinect RGB-D camera and its body-tracking library. A parametric analysis of the uncertainty of the estimation of the skeleton joints is performed by changing the ambient light conditions, the presence of occlusions, the infrared camera resolution, and the human-camera distance. The output of this investigation proves the need for uncertainty management in the control of cobots working with humans.

1 – Introduzione

Fin dalla nascita dei primi robot industriali, sviluppare sistemi per il loro controllo in tempo reale è sempre stato un aspetto fondamentale per l'implementazione di manipolatori e piattaforme mobili. Controllare un robot è un task complesso, in quanto tale necessità nasce come conseguenza del fatto che il robot effettua azioni in risposta a due tipologie di attività: prevedibile e imprevedibile. Nel primo caso, il controllo del robot è focalizzato a pianificare un set di azioni per raggiungere un determinato scopo, cercando la migliore soluzione per garantire la massima efficienza sotto tutti gli aspetti. Nel secondo caso, invece, gli eventi imprevedibili potrebbero scatenare diverse reazioni, che comprendono anche il fermo del robot, o una ripianificazione delle attività [1], [2].

Negli ultimi tempi, la rapida crescita di tecnologie atte al controllo dei robot ha permesso la nascita di una nuova generazione di robot industriali, denominati robot collaborativi, o cobot, i quali hanno la possibilità di cooperare con gli umani in celle di lavoro condivise [3]. Ovviamente, dalla presenza simultanea di cobot e operatori in uno spazio di lavoro condiviso

nasce la necessità di ridefinire gli standard di sicurezza [4]. In particolare, i requisiti di sicurezza non sono limitati ai soli limiti meccanici previsti nei cobot, o allo spazio di lavoro condiviso, ma risulta necessario estenderne il concetto in modo da migliorare la capacità di controllo del cobot stesso. Tale controllo deve necessariamente essere in tempo reale, in modo da garantire alta flessibilità e riconfigurabilità, adeguandosi alle dinamiche richieste dall'ambiente di lavoro. Una pianificazione preliminare del controllo del cobot non è sufficiente a garantire il raggiungimento dello scopo finale e, al tempo stesso, la sicurezza dell'operatore. Per questo motivo, il cobot ha bisogno di percepire l'ambiente circostante, in modo da conoscere tutte le possibili interazioni ed eventuali collisioni con l'operatore, andando così a garantire la sicurezza dell'umano senza la necessità di adottare sistemi di protezione esterni [5].

I cobot che lavorano a distanza dagli umani possono eseguire il loro lavoro più velocemente, aumentando la produzione, per cui le informazioni in real-time sono importanti non solo per motivi di sicurezza, ma anche per migliorare l'efficienza del cobot stesso. Per il riconoscimento della posizione e del volume occupato dagli umani, ci si avvale spesso della tecnica del body tracking [6], il quale ha come obiettivo l'identificazione degli utenti all'interno della linea di vista della telecamera, andando a segmentarli in parti significative, riordinandoli in array di giunti, che vanno a creare la rappresentazione di uno scheletro. In ogni caso, i sistemi di body-tracking richiedono hardware e software appropriati.

Negli ultimi anni, sono stati proposti diversi moduli basati su sistemi di visione [7] in grado di attuare tecnologie di motion capture, come ad esempio il Vicon Motion System. Qui, un sistema di telecamere cattura un'area di interesse e, attraverso l'uso di marker riflettenti opportunamente posizionati, è in grado di visualizzare gli umani presenti nella scena e rappresentarli in giunti scheletrici. La triangolazione dei raggi ottici permette la ricostruzione della posizione 3D dei marker, e quindi dei giunti, con alta accuratezza [8]. Sfortunatamente, questa soluzione è spesso inattuabile in contesti industriali, a causa dei costi e del tempo richiesto per configurare il setup. Una soluzione alternativa al sistema Vicon sono le telecamere RGB-D, che sono in grado di acquisire le immagini RGB e la rappresentazione 3D

della scena inquadrata. Questo tipo di telecamere possono avere le seguenti configurazioni [9]:

- Stereovisione attiva: due immagini acquisite simultaneamente da due diversi punti di vista sono comparate in modo da trovare eventuali differenze, che dipendono dalla distanza tra la telecamera e il target ripreso. Le telecamere Intel RealSense implementano questo principio, ottenendo prestazioni migliori in termini di accuratezza [10]. Tuttavia, l'incertezza dei modelli 3D cresce sensibilmente per distanze maggiori di 1 m.
- Luce strutturata: viene proiettato un pattern di riferimento sul target, la cui forma è recuperata secondo eventuali deformazioni. Una telecamera molto usata, che si avvale di questo principio, è la Microsoft Kinect V1. In ogni caso, l'accuratezza sulla stima di profondità decresce esponenzialmente con l'aumentare della distanza tra target e telecamera [11], andando quindi a limitare il range di applicabilità di tale sensore.
- Tempo di Volo (Time of Flight, ToF): la distanza tra target e telecamera è misurata attraverso il tempo richiesto ad un impulso luminoso per colpire il target e riflettere la luce sull'emettitore. La profondità viene stimata con una accuratezza nettamente migliore rispetto ad altri sensori. È stato dimostrato, infatti, che la Microsoft Kinect V2, nata nel 2014, risulta la scelta migliore per la gesture recognition [12].

Recentemente, la produzione della Kinect V2 è stata interrotta in favore della Azure Kinect DK (Development Kit), la quale offre una accuratezza migliore rispetto ad altre telecamere RGB-D che attuano la tecnologia ToF [13]. Queste maggiori prestazioni permettono alla Microsoft Azure Kinect DK di essere considerata come la miglior soluzione per il body tracking da interno, il quale può essere usato in diversi scenari, dall'industria videoludica [15] al l'industria manifatturiera.

La qualità del body tracking non dipende soltanto dall'accuratezza e dalla risoluzione dei dati 3D, ma anche dagli algoritmi del body tracking stesso. A titolo di esempio, il rilascio della Microsoft Kinect è stato accompagnato dal rilascio di valide librerie per il body tracking, che si basano sull'uso di algoritmi di "decision forests" addestrati su determinati dataset [15] per il

riconoscimento di 20 giunti dello scheletro umano, che sono stati poi incrementati a 25 con l'utilizzo di nuove librerie Windows SDK (Software Develpment Kit) per Kinect [16]. Ad oggi, il numero dei giunti riconosciuti è arrivato a 32, attraverso l'uso di reti neurali e sistemi di deep learning [17], dando vita all'Azure Kinect Body Tracking SDK [18], il quale ha aggiunto ulteriori dettagli alla mappatura dei giunti, aumentando ad esempio il numero di giunti del viso.

La disponibilità delle ultime tecnologie hardware e software per il body tracking ha portato la comunità scientifica ad analizzare i movimenti dell'umano attraverso, ad esempio, l'andatura della camminata [19]. Albert et al. [13] hanno valutato i risultati del body tracking per la Kinect V2 e per l'Azure Kinect comparandoli con quelli del Vicon durante l'esecuzione di test dinamici, analizzando come le diverse tipologie di hardware e di algoritmi basati sul deep learning per il body tracking possono migliorare il tracciamento dei giunti umani su diversi soggetti. In ogni caso, è necessario anche analizzare come condizioni ambientali, occlusioni, risoluzione della telecamera, e distanza del soggetto possono alterare l'incertezza dell'estrazione dei giunti. Questo aspetto è di fondamentale importanza, soprattutto nel momento in cui il sistema di body tracking viene utilizzato per il controllo in tempo reale di un cobot che condivide lo stesso spazio di lavoro con un operatore umano.

Questo articolo ha come obiettivo quello di fare luce sull'alterazione dell'incertezza su acquisizioni quasi-statiche di corpi umani. È stato realizzato un setup sperimentale composto da un'Azure Kinect e dal sistema di body tracking ad essa associato, il quale si occupa di estrarre i giunti dell'utente nella scena. Sono state fatte diverse acquisizioni andando a cambiare parametri intrinseci ed estrinseci. Tutti i dati sono stati processati in modo da valutare l'incertezza in ogni condizione, evidenziando le condizioni di lavoro peggiori che potrebbero significativamente alterare i dati in uscita dall'Azure Kinect Body Tracking SDK. L'articolo è strutturato come segue: nel paragrafo 2 vengono definite le acquisizioni effettuate e il setup sperimentale realizzato; nel paragrafo 3 vengono presentati e commentati i risultati ottenuti; infine, il paragrafo 4 delinea le conclusioni.

2 – Materiali e Metodi

Come precedentemente citato, Albert et al. [13] hanno presentato uno studio sulle prestazioni dell'Azure Kinect e del body tracking ad essa associato, focalizzandosi sull'analisi della camminata di diversi soggetti. Hanno successivamente effettuato una comparazione con il sistema Vicon, attraverso l'analisi di media e deviazione standard delle distanze euclidee tra i giunti 3D estratti dal sensore Kinect e dal sistema Vicon. Nel presente lavoro, invece, ci si focalizzerà sulla deviazione standard, strettamente connessa alla misura dell'incertezza, calcolata considerando i seguenti parametri:

- Parametri intrinseci: risoluzione del sensore di profondità.
- Parametri estrinseci: condizioni di luce ambientali, occlusioni del corpo, distanza tra telecamera e soggetto.

Di seguito viene presentato il setup proposto, e la procedura di processing per la valutazione del sistema di body tracking.

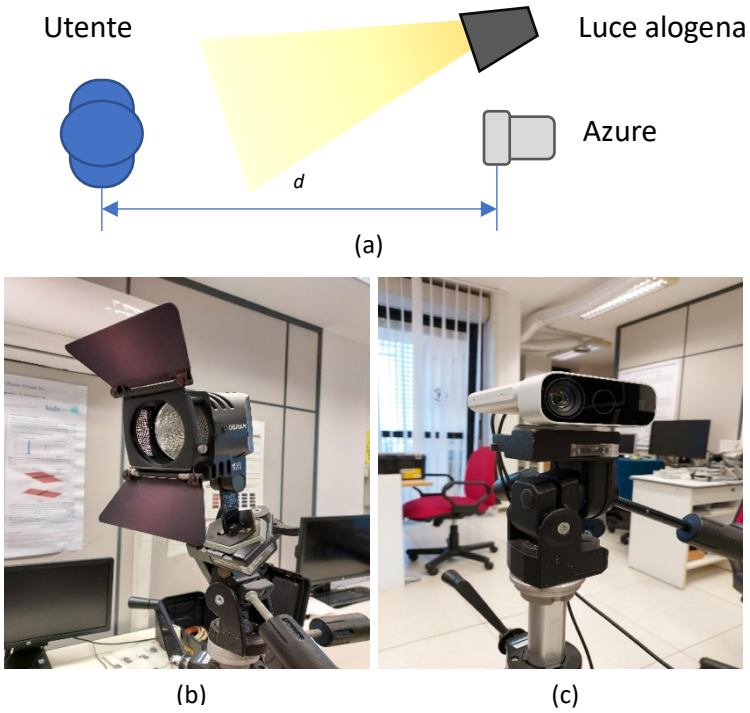

Figura 1 - (a) Rappresentazione del setup sperimentale composto da **(b)** una luce alogena e **(c)** una telecamera Azure Kinect. Il parametro *d* rappresenta la distanza tra la telecamera e l'utente.

2.1 – Definizione del Setup

Il lavoro presentato è stato realizzato avvalendosi dello schema sperimentale in Figura 1. Il sensore Azure Kinect è stato posizionato ad una distanza dall'utente, pari a *d*, distanza che verrà variata in un range che va da 1 a 3 m, in step di 1 m. Per illuminare la scena, si è scelta una luce alogena di potenza 300 W. Le condizioni di luce ambientali sono state decise considerando la fonte di luce descritta come accesa o spenta. In particolare, quando la luce è accesa, la luminosità E_v ad 1 m dalla fonte luminosa è di 1750 lux, mentre tale valore scende a 10 lux quando la luce è spenta. In entrambi i casi, l'esposizione della telecamera è stata configurata in modalità *auto*, in modo che il tempo di esposizione può essere al massimo uguale all'inverso del framerate della telecamera. In Figura 2, è possibile osservare le due condizioni di luce utilizzate nelle acquisizioni sperimentali. Sebbene la condizione di scarsa fonte luminosa non sia realistica in contesti ambientali, è stato deciso comunque di analizzare tale parametro in modo da valutare l'Azure Kinect anche in condizioni limite.

(a)

(b)

Figura 2 - Confronto tra le condizioni di luce proposte: **(a)** E_v = 1750 lux; **(b)** E_v = 10 lux.

Come definito precedentemente, i giunti vengono calcolati considerando il corpo dell'utente perfettamente visibile dalla telecamera (*Occl = w/o*), o con un ostacolo opaco che occlude la parte inferiore del corpo dell'utente (*Occl = w/*). Inoltre, la Azure Kinect prevede due modalità per il sensore di profondità, wide (*Res = W*) e narrow (*Res = N*), le quali differiscono per quanto riguarda il campo di vista (120°×120° e 75°×65°, rispettivamente) e per la risoluzione di profondità (512×512 e 640×576, rispettivamente). Entrambe le configurazioni sono state considerate negli esperimenti.

Effettuando diverse acquisizioni con tutte le combinazioni dei parametri presentati, si otterranno 24 video dalla Azure Kinect, che verranno poi processati dall'Azure Body Tracking SDK (v 1.0.1), in modo da ottenere 24 set di skeleton rappresentativi di un singolo utente, con 32 giunti, i cui indici sono mappati in [18]. Tutti i video hanno una durata di 60 s ed un framerate di 15 fps (frame per second). Per una migliore comprensione delle sperimentazioni effettuate, gli skeleton sono rinominati nel presente articolo come *Sk*(*Res*, *Occl*, E_v, *d*). Ad esempio, *Sk*(N, w/o, 1750, 2) si riferirà ad una acquisizione eseguita con risoluzione narrow (640×576 pixels), senza occlusioni, con alta illuminazione, e con una distanza utente-telecamera di 2 m, come mostrato in Figura 3.

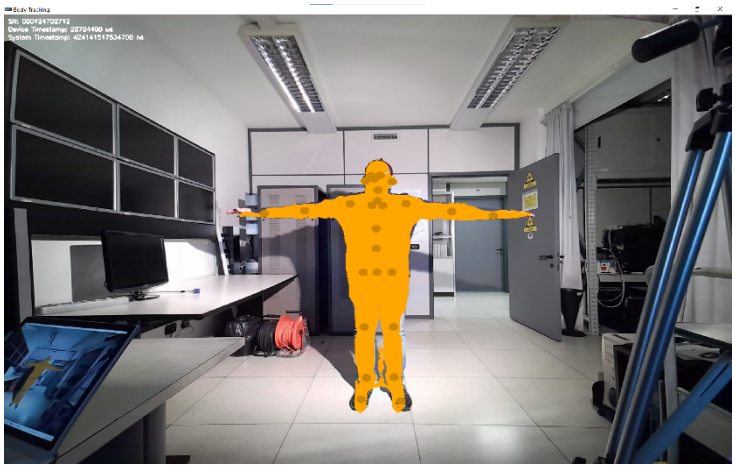

Figura 3 - Risultato del Body Tracking dell'acquisizione Sk(N, w/o, 1750, 2). I punti arancioni rappresentano la posizione stimata dei giunti dello skeleton nell'immagine 2D.

2.2 – Fase di Processing

Tutte le acquisizioni effettuate rilasciano come output la rappresentazione del corpo umano sotto forma di giunti scheletrici *Sk*, per un totale di 32 giunti, le cui posizioni 3D sono J[j,t] = (x_1[j,t], x_2[j,t], x_3[j,t]), dove j = 0, …, 31 rappresenta l'indice dei giunti, e t = 1, …, T rappresenta l'indice dei campioni nel tempo. Il sistema di riferimento (x_1, x_2, x_3) è allineato alle coordinate della telecamera definite in [18], dove T = 900 è il risultato dato da acquisizioni della durata di 60 s, a 15 fps.

Come indicato nei paragrafi precedenti, gli esperimenti proposti hanno come scopo la valutazione della misura dell'incertezza. Nei video acquisiti, l'utente è in piedi e fermo di fronte alla telecamera, tenendo le braccia aperte ed i piedi uniti. L'utente mantiene la sua posizione mentre la telecamera acquisisce dati per 60 s. Tali dati danno come risultato le posizioni dei giunti, che vengono acquisite in 3D. In Figura 4 è possibile visualizzare lo scatter plot dei giunti acquisiti, nel tempo, dalla configurazione *Sk*(N, w/o, 1750, 1).

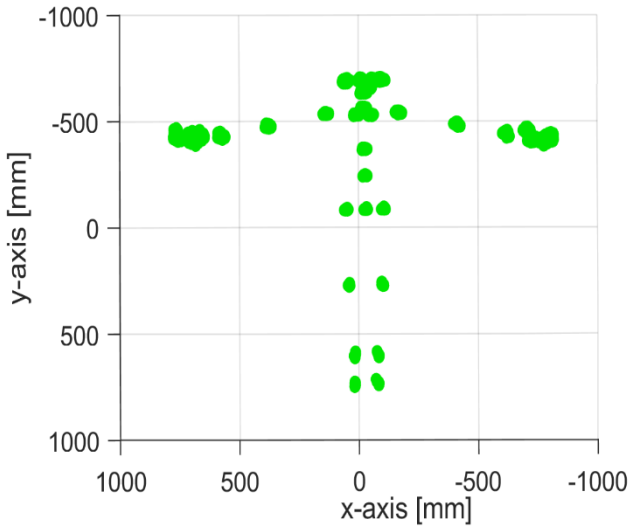

Figura 4 - Giunti dello skeleton in 2D acquisito da Sk(N, w/o, 1750, 1).

Nonostante lo sforzo fatto dall'utente per mantenere la sua posizione, è naturale che il corpo abbia fluttuazioni, determinando le condizioni quasi-statiche delle sperimentazioni. Tale fenomeno è particolarmente evidente nei giunti periferici, come ad esempio le mani. Per

risolvere tale problema, è stato calcolato il valore medio delle distanze euclidee di ogni coordinata di giunto dal centroide corrispondente, il quale è stato posizionato come C[j,t] = ($x_{c,1}$[j,t], $x_{c,2}$[j,t], $x_{c,3}$[j,t]), dove:

Equazione 1. Equazione del centroide.

$$x_{c,i}[j,t] = \frac{1}{2N+1}\sum_{p=t-N}^{t+N} x_i[j,p] \qquad i = 1,2,3$$

Dopo N campioni, questa informazione rappresenta il risultato di una media mobile, calcolata su una finestra di 2N+1 campioni, centrata intorno al t-esimo campione d'interesse. Bisogna tener conto che la media mobile è anche calcolata considerando i limiti dei vettori di input J[j,t], precisamente a $t < N + 1$ e $t > T - N$. In questi casi, la lunghezza della finestra è limitata opportunamente con le entrate di J[j,t]. Nelle Figure 5 e 6 è possibile osservare i risultati della media mobile sulle coordinate del centroide del torso (j = 1, SPINE_NAVAL in 0) e della mano destra (j = 8, HAND_LEFT in 0). Durante tutti gli esperimenti, N è impostato a 15, che corrisponde ad una lunghezza della finestra di circa 2 s a 15 fps.

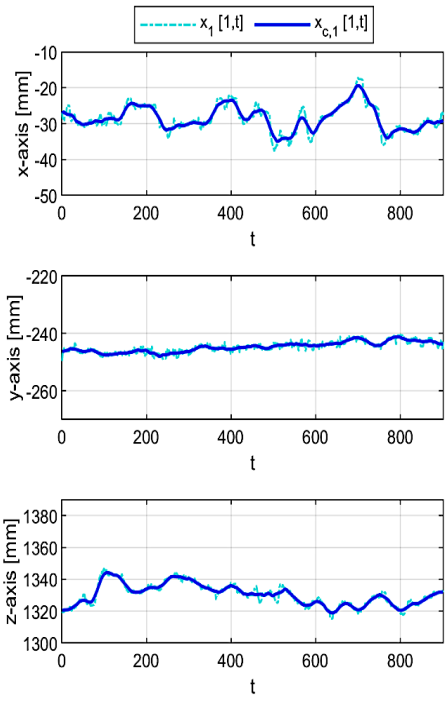

Figura 5 - Confronto tra le posizioni del giunto del torso (linea azzurra) e la posizione del centroide corrispondente (linea blu) nel tempo. L'acquisizione è Sk(N, w/o, 1750, 1).

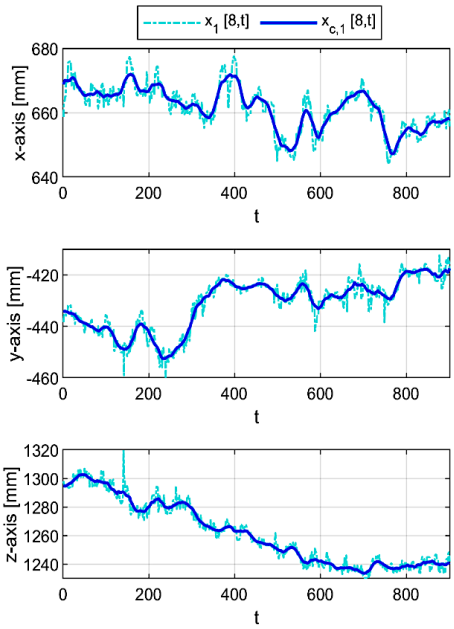

Figura 6 - Confronto tra le posizioni del giunto della mano sinistra (linea azzurra) e la posizione del centroide corrispondente (linea blu) nel tempo. L'acquisizione è Sk(N, w/o, 1750, 1).

Come ci si poteva aspettare, i grafici in Figura 5 e 6 mostrano che le fluttuazioni del corpo condizionano maggiormante i giunti delle mani rispetto al giunto del torso, il quale rimane comunque in una posizione statica. In ogni caso, la componente relativa alle fluttuazioni del corpo a causa della condizione quasi-statica delle acquisizioni verrà ignorato andando a calcolare l'errore quadratico ($SE[j,t]$) come segue:

Equazione 2. Calcolo del valore di SE.

$$SE[j,t] = \sum_{i=1}^{3} (x_i[j,t] - x_{c,i}[j,t])^2$$

Allo stesso modo, la distanza euclidea potrà essere calcolata da *SE*[*j*,*t*] e dalla media nel tempo dei campioni, rilasciando quindi il valore medio della distanza (Mean Distance Error, MDE[*j*]) del *j*-esimo giunto:

Equazione 3. Calcolo del valore di MDE.

$$MDE[j] = \frac{1}{T}\sum_{t=1}^{T}\sqrt{SEj,t]}$$

Un esempio del calcolo della MDE[*j*] da *Sk*(N, w/o, 1750, 1) è mostrato in Figura 7. In questo caso, è possibile notare che i più alti valori di MDE corrispondono a quelli delle mani (*j* = 8, HAND_LEFT e *j* = 15, HAND_RIGHT in 0), dei pollici (*j* = 10, THUMB_LEFT e *j* = 17, THUMB_RIGHT in 0), e delle dita (*j* = 9, HANDTIP_LEFT e *j* = 16, HANDTIP_RIGHT in 0). In questi esperimenti, le dita delle mani ed i pollici non sono di interesse, in quanto il loro utilizzo è tipicamente necessario nella gesture recognition. L'analisi della loro affidabilità non è quindi tra gli scopi del presente lavoro, il quale invece si focalizza nella segmentazione delle persone per il controllo in real-time e in sicurezza dei cobot. Per questo motivo, le successive analisi si concentreranno soltanto su quattro giunti rappresentativi: la testa (*j* = 26, HEAD in 0), il bacino (*j* = 0, PELVIS in 0), la mano sinistra (*j* = 8, HAND_LEFT in 0), ed il piede destro (*j* = 25, FOOT_RIGHT in 0). I risultati ottenuti per i giunti destri o sinistri sono replicabili anche per le parti opposte del corpo.

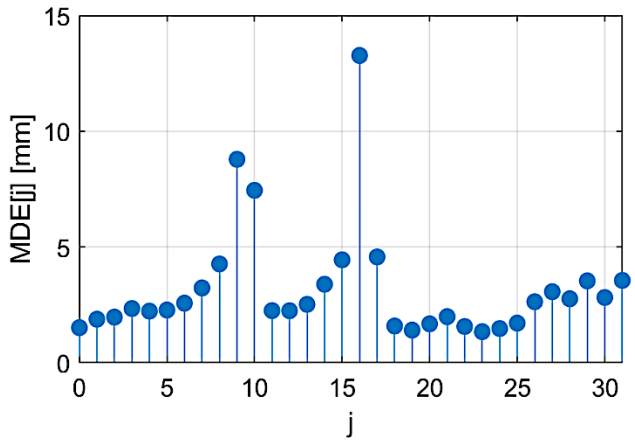

Figura 7 - MDE dei 32 giunti estratti dall'Azure Kinect e la rispettiva libreria di body tracking, dall'acquisizione *Sk(N, w/o, 1750, 1)*. L'indice dei giunti corrisponde a quello in [18].

3 – Analisi dei Risultati

Come precedentemente specificato, i video acquisiti forniscono informazioni circa le posizioni in 3D dei giunti dello scheletro ad ogni frame, mentre l'utente rimane nella posizione indicata per 60 s. Le 24 acquisizioni sono state analizzate considerando 4 giunti significativi: la testa, il bacino, la mano sinistra e il piede destro.

In Tabella 1 sono riportati i valori di MDE per i quattro giunti considerati, prendendo in esame le acquisizioni effettuate senza occlusioni, al variare della risoluzione di profondità della telecamera (*Res*), della luce ambientale (E_v), e della distanza *d* tra soggetto e telecamera.

Tabella 1 - Valore di MDE della testa, del bacino, della mano sinistra e del piede destro al cambiare delle condizioni di input. In tutti i casi, le acquisizioni sono state eseguite senza occlusioni. I valori sono espressi in millimetri.

Condizioni di input	Testa	Bacino	Mano sinistra	Piede destro
Sk(N,w/o,10,1)	2.49	1.71	4.53	1.83
Sk(N,w/o,10,2)	2.35	1.70	6.66	1.54
Sk(N,w/o,10,3)	2.94	1.77	8.83	2.66
Sk(W,w/o,10,1)	2.67	2.04	10.46	1.15
Sk(W,w/o,10,2)	3.03	2.86	9.83	2.82
Sk(W,w/o,10,3)	6.92	3.90	20.97	7.73
Sk(N,w/o,1750,1)	2.63	1.51	4.27	1.71
Sk(N,w/o,1750,2)	2.66	1.51	8.82	1.82
Sk(N,w/o,1750,3)	4.74	2.42	17.63	3.57
Sk(W,w/o,1750,1)	3.21	2.73	12.99	6.67
Sk(W,w/o,1750,2)	7.70	4.84	21.51	10.65
Sk(W,w/o,1750,3)	16.63	7.08	35.84	9.38

I risultati mostrano chiaramente che il valore di MDE cresce al crescere della distanza. Tale andamento è in realtà comprensibile, considerando che all'aumentare della distanza, la risoluzione della stima della profondità diminuisce.

Come ci si poteva aspettare, i dati presenti in Tabella 1 confermano che, tra i quattro giunti, la mano sinistra è stimata con il più alto valore di MDE in tutte le acquisizioni, a prescindere dalle condizioni di input. Questo risultato è ulteriormente confermato se si analizzano i valori medi di MDE dei giunti presi in considerazione, i quali sono pari a 4.83, 2.84, 13.53 e 4.29 mm, rispettivamente per testa, bacino, mano sinistra e piede destro. Inoltre, la Tabella 1 mostra anche che la scelta della risoluzione di profondità della telecamera è fondamentale nel giudicare le prestazioni della Kinect Azure. In particolare, l'impostazione wide (*Res* = W) aumenta il livello d'incertezza rispetto all'impostazione narrow (*Res* = N). I valori medi di MDE dei giunti rappresentativi della testa, del bacino, della mano sinistra e del piede destro, riferiti all'impostazione wide (*Res* = W), sono rispettivamente 2.06, 2.17, 2.28 e 2.96 volte più alti rispetto all'impostazione narrow (*Res* = N). I risultati riportati evidenziano anche che l'utilizzo di una fonte luminosa comporta un aumento del valore di MDE. Tale valore è inoltre più cospicuo quando la distanza tra utente e telecamera aumenta. Considerando l'intensità luminosa pari a E_v = 1750 lux, i valori di MDE per i giunti di testa, bacino, mano sinistra e piede destro sono rispettivamente, in media, 1.66, 1.33, 1.57 e 2.37 volte maggiori rispetto ai valori ottenuti con E_v = 10 lux. Ciò è probabilmente dovuto alla modalità di illuminazione utilizzata, nel setup sperimentale che punta direttamente sull'utente. L'illuminazione diretta, infatti, genera come conseguenza una mappa di profondità con maggior contributo di rumore, andando quindi ad aumentare i valori di MDE. Una luce meno intensa ma più diffusa, invece, va a limitare il rumore acquisito dal sensore di profondità, per cui l'incertezza dei giunti rimane relativamente bassa.

L'analisi parametrica proposta ha come obiettivo il fornire informazioni riguardo quanto una occlusione parziale dell'utente possa alterare il valore di MDE dei giunti. Questo aspetto è di fondamentale importanza in quanto possibili occlusioni risultano essere molto comuni in ambienti industriali. Un operatore che svolge task in ambito manifatturiero, come ad esempio

l'assemblaggio di pezzi, può essere occluso dalla telecamera a causa di strumentazioni, scrivanie, nastri scorrevoli, o dal cobot stesso.

La Tabella 2 mostra i valori di MDE di tre giunti presi in considerazione, la testa, il bacino e la mano sinistra, estratti dallo skeleton acquisito mentre la parte inferiore del corpo è occlusa. Il giunto relativo al piede destro non viene considerato in questa analisi, in quanto è appunto occluso durante le acquisizioni.

Tabella 2 - MDE di testa, bacino e mano sinistra al cambiare delle condizioni di input. La stima del piede in questo caso non è applicabile, in quanto il giunto è occluso. In tutti i casi, le acquisizioni sono effettuate con la presenza di occlusioni nella parte inferiore del corpo dell'utente. I valori sono espressi in millimetri.

Condizioni di input	**Testa**	**Bacino**	**Mano sinistra**
Sk(N,w/,10,1)	1.82	1.36	8.08
Sk(N,w/,10,2)	2.24	1.71	8.43
Sk(N,w/,10,3)	4.93	2.03	10.01
Sk(W,w/,10,1)	2.24	1.79	5.21
Sk(W,w/,10,2)	4.59	2.88	13.25
Sk(W,w/,10,3)	30.23	15.65	67.15
Sk(N,w/,1750,1)	2.20	1.30	4.83
Sk(N,w/,1750,2)	2.86	1.89	6.53
Sk(N,w/,1750,3)	6.77	4.20	19.96
Sk(W,w/,1750,1)	3.67	2.75	9.52
Sk(W,w/,1750,2)	11.07	6.82	33.22
Sk(W,w/,1750,3)	29.04	8.61	53.85

Le osservazioni già effettuate per la Tabella 1 sono valide anche con riferimento alla Tabella 2.

Nello specifico:

- I valori di MDE aumentano all'aumentare della distanza tra soggetto e telecamera. Ciò è vero in tutte le condizioni proposte dal setup.
- *Res* = W fa aumentare i valori di MDE per i tre giunti considerati. Tali valori infatti sono triplicati rispetto all'impostazione *Res* = N.
- E_v = 1750 lux, ottenuta con illuminazione diretta, abbassa in generale le prestazioni del body tracking, con un aumento del valore di MDE, che risulta in media 1.44 volte maggiore rispetto al valore ottenuto per E_v = 10 lux. In ogni caso, nelle acquisizioni *Sk*(N, w/, E_v, 1) e *Sk*(W, w/, E_v, 3), con E_v = 10 lux, i valori di MDE sono comparabili con le acquisizioni corrispondenti a E_v = 1750 lux.
- La testa e il bacino sono caratterizzati da risultati nettamente migliori rispetto al quello rilasciato dal giunto della mano sinistra, il quale mostra i valori peggiori di MDE in tutte le condizioni di input.

Facendo un paragone tra Tabella 1 e 2, si può notare come i giunti mostrano alti valori di incertezza quando lo skeleton è parzialmente occluso. Per avere una stima dell'andamento dell'incertezza, si possono considerare le medie dei valori di MDE per ogni acquisizione considerata, sia occlusioni sia senza occlusioni. Il risultato di questa analisi mostra che le occlusioni vanno ad aumentare il valore di MDE di testa, bacino e mano sinistra rispettivamente del 75.37%, del 49.66% e del 47.86%. Ciò dimostra che tutti i giunti sono stimati con una maggiore incertezza quando il corpo è occluso, a prescindere da dove si trovi l'occlusione. In ogni caso, considerando tutti i parametri intrinseci ed estrinseci, i valori di MDE dei giunti considerati in questa analisi parametrica oscillano da un minimo di circa 1 mm ad un massimo di circa 53 mm, con un valore medio di 8 mm e una deviazione standard di 6 mm.

4 – Conclusioni

In questo articolo, è stata effettuata una analisi parametrica della misura dell'incertezza nell'algoritmo di body tracking proposto dalla Kinect Azure. Nello specifico, sono state valutate le prestazioni della Microsoft Azure Kinect nell'estrarre i giunti dallo skeleton al variare di parametri intrinseci ed estrinseci, ovvero la risoluzione della telecamera, l'illuminazione ambientale, la distanza tra soggetto e telecamera, e l'aggiunta di occlusioni sull'utente. I risultati dell'analisi hanno provato che:

- La stima dei giunti della mano soffre sempre dell'incertezza più alta
- Gli skeleton acquisiti con una risoluzione di profondità wide hanno sempre un livello d'incertezza maggiore rispetto a quelli acquisiti con una risoluzione di profondità narrow
- L'incertezza degli skeleton aumenta all'aumentare della distanza tra telecamera e soggetto.

Inoltre, l'illuminazione diretta va ad alterare la mappa di profondità e, quindi, l'accuratezza dei giunti dello skeleton. Anche la presenza di occlusioni va ad aumentare l'incertezza su tutto lo skeleton, anche sui giunti che sono lontani dall'occlusione stessa.

La conoscenza dell'incertezza nell'estrazione dello skeleton durante il body tracking per analizzare le condizioni di lavoro in ambito industriale può essere di fondamentale importanza per migliorare la sicurezza nel controllo real-time dei cobot che operano in contemporanea con umani. Lavori futuri si focalizzeranno sull'analisi di ulteriori parametri, sia intrinseci che estrinseci, come ad esempio l'esposizione dell'immagine, e la presenza di più utenti in scena. Ulteriori esperimenti verranno proposti anche per considerare le condizioni dinamiche mentre il soggetto esegue task specifici.

5 – Bibliografia

[1] - Siciliano B., Khatib O., "Springer handbook of robotics". 2nd ed., Berlin Heidelberg, Germany: Springer-Verlag, 2016.

[2] - Tsarouchi P., Makris S., Chryssolouris G., "Human–robot interaction review and challenges on task planning and programming", *Int. J. Comput. Integr. Manufact.*, vol. 29, no. 8, pp. 916-931, 2016.

[3] - Colgate J. E., Edward J., Peshkin M. A., Wannasuphoprasit W., "Cobots: Robots for Collaboration with Human Operators", *Proc. ASME Int. Mechanical Engineering Congress and Exposition*, Atlanta, pp. 433–439, 1996.

[4] - International Organization for Standardization, ISO/TS 15066:2016 – Robots and Robotic Devices – Collaborative Robots, 2016.

[5] - Marvel J. A., Norcross R., "Implementing speed and separation monitoring in collaborative robot workcells", *Robot Comput Integr Manuf*, vol. 44, pp. 144–55, 2017.

[6] - Knoop S., Vacek S., Dillmann R., "Sensor fusion for 3D human body tracking with an articulated 3D body model", *Proc. IEEE Int. Conf. on Robotics and Automation* (ICRA), Orlando, USA, pp. 1686-1691, 2006.

[7] - Halme R. J., Lanz M., Kämäräinen J., Pieters R., Latokartano J., Hietanen A., "Review of vision-based safety systems for human-robot collaboration", *Procedia CIRP*, vol. 72, pp. 111-116, 2018.

[8] - Barker S., Craik R., Freedman W., Herrmann N., Hillstrom H., "Accuracy, reliability, and validity of a spatiotemporal gait analysis system", *Med. Eng. Phys.*, vol. 28, pp. 460–467, 2006.

[9] - Zollhöfer M., "Commodity RGB-D Sensors: Data Acquisition", in: Rosin P., Lai YK., Shao L., Liu Y. (eds) *RGB-D Image Analysis and Processing. Advances in Computer Vision and Pattern Recognition*. Cham, Switzerland: Springer Nature, pp. 3-13, 2019.

[10] - Carfagni M., Furferi R., Governi L., Santarelli C., Servi M., Uccheddu F., Volpe Y., "Metrological and critical characterization of the Intel D415 stereo depth camera", *Sensors*, vol. 19, no. 3, pp. 489, 2019.

[11] - Mallick T., Das P.P, Majumdar A.K., "Characterizations of noise in Kinect depth images: a review", *IEEE Sens. J.*, vol. 14, pp. 1731–1740, 2014.

[12] - Sarbolandi H., Lefloch D., Kolb A., "Kinect range sensing: Structured-light versus time-of-flight Kinect", *Comput. Vis. Image Und.*, vol. 139, pp. 1–20, 2015.

[13] - Albert J. A., Owolabi V., Gebel A., Brahms C. M., Granacher U., Arnrich B., "Evaluation of the pose tracking performance of the azure kinect and kinect v2 for gait analysis in comparison with a gold standard: A pilot study", *Sensors*, vol. 20, no. 18, pp. 5104, 2020.

[14] - Zhang M., Zhang Z., Chang Y., Aziz E.-S., Esche S., Chassapis C., "Recent developments in game-based virtual reality educational laboratories using the microsoft Kinect", *International Journal of Emerging Technologies in Learning (iJET)*, vol. 13, pp. 138-159, 2018.

[15] - Shotton J., Fitzgibbon A., Cook M., Sharp T., Finocchio M., Moore R., Kipman A., Blake A., "Real-time human pose recognition in parts from single depth images", *Proc. IEEE Int. Conf. Computer Vision And Pattern Recognition* (CVPR), Colorado Springs, USA, pp. 1297-1304, 2011.

[16] - https://www.microsoft.com/en-us/download/details.aspx?id=44561.

[17] - LeCun Y., Bengio Y., Hinton G., "Deep learning", *Nature*, vol. 521, no. 7553, pp. 436-444, 2015.

[18] - https://docs.microsoft.com/en-us/azure/kinect-dk/

[19] - Clark R. A., Mentiplay B. F., Hough E., Hua Y. H., "Three-dimensional cameras and skeleton pose tracking for physical function assessment: A review of uses, validity, current developments and Kinect alternatives"," *Gait Posture*, vol. 68, pp. 193–200, 2019.