



UNIVERSITÀ DEGLI STUDI DI SALERNO

Dipartimento di Informatica

Corso di Laurea Magistrale in Informatica

TESI DI LAUREA

# Pre-processing dei dati video per la gait analysis di pazienti con schizofrenia

RELATORI

Prof. Rita Francese

Dott. Laura De Santis

Università degli Studi di Salerno

CANDIDATO

**Giovanni Buonincontri**

Matricola: 0522500852

“I think the easiest thing in life is to give up on whatever you know, whatever you’re focused on, or whatever you’re hopeful for. The hardest thing is to continue to keep … hold on.”

---

—Lewis Hamilton

**Alla parte di me che dubita,**  
perché mi tiene vigile;  
**alla parte che teme,**  
perché mi rende umano;  
**alla parte che insiste,**  
perché mi spinge oltre;  
**ai sogni,** che danno senso a ogni fatica.

**A Monica,** metà ed orizzonte,  
per tutto ciò che siamo e che saremo.

## **Abstract**

Le alterazioni del cammino nei pazienti schizofrenici rappresentano un importante indicatore clinico che può essere analizzato tramite tecnologie di motion capture. Tuttavia, l'estrazione di dati affidabili da sistemi non specialistici presenta numerose sfide tecniche che limitano l'applicabilità di queste metodologie in ambito clinico. Questo lavoro di tesi implementa uno strumento avanzato per l'elaborazione e la correzione dei dati video acquisiti tramite telecamere RGB-D Intel RealSense, con l'obiettivo di ottenere parametri cinematici affidabili per la gait analysis di pazienti schizofrenici. Lo scopo di questa tesi è stato quello di sviluppare un algoritmo completo che permetta di estrarre, validare e correggere dati di movimento tridimensionali da registrazioni video, superando le limitazioni legate all'affidabilità dei keypoint estratti tramite tecniche di visione artificiale. L'approccio proposto integra l'estrazione di keypoint tramite MediaPipe, il calcolo preciso della profondità con tecniche robuste per la gestione dell'incertezza, e l'applicazione di filtri Savitzky-Golay adattivi che intervengono selettivamente in presenza di inconsistenze nei dati. La valutazione sperimentale si è concentrata sull'analisi del miglioramento della qualità dei dati post-processamento, con particolare attenzione alla riduzione del jitter e all'incremento degli score di affidabilità. Questo lavoro contribuisce al campo della gait analysis fornendo strumenti metodologici e computazionali che permettono di ottenere dati di movimento più affidabili da sistemi a basso costo.

---

# Indice

---

<b>Elenco delle Figure</b>	<b>iv</b>
<b>Elenco delle Tabelle</b>	<b>vi</b>
<b>Elenco degli Algoritmi</b>	<b>vii</b>
<b>1 Introduzione</b>	<b>1</b>
1.1 Panoramica sui disturbi motori nella schizofrenia . . . . .	1
1.2 Importanza della gait analysis in ambito clinico . . . . .	3
1.3 Problematiche dei sistemi di motion capture tradizionali vs. soluzioni low-cost . . . . .	6
1.4 Obiettivi e contributi della tesi . . . . .	9
1.5 Struttura della Tesi . . . . .	13
<b>2 Stato dell'arte</b>	<b>15</b>
2.1 Letteratura sulla gait analysis in pazienti psichiatrici . . . . .	15
2.2 Tecnologie di motion capture: da sistemi professionali a RGB-D . . .	18
2.3 MediaPipe e tecniche di pose estimation . . . . .	22
2.4 Metodi di correzione e filtraggio dei dati di movimento . . . . .	26
2.5 Il Contesto di Riferimento: il Progetto SPECTRA . . . . .	29

<b>3 Metodologia</b>	<b>32</b>
3.1 Architettura e Setup Sperimentale . . . . .	32
3.1.1 Architettura Hardware e Setup di Acquisizione . . . . .	32
3.1.2 Architettura Software e Pipeline di Elaborazione . . . . .	35
3.1.3 Procedura di Acquisizione Dati . . . . .	37
3.1.4 Considerazioni Etiche . . . . .	38
<b>4 Implementazione</b>	<b>39</b>
4.1 Ambiente di Sviluppo e Tecnologie Software . . . . .	39
4.2 Panoramica della Pipeline di Elaborazione . . . . .	41
4.3 Dall'Immagine ai Dati Grezzi: Estrazione e Ricostruzione 3D . . . . .	43
4.3.1 Stima della Posa 2D tramite MediaPipe . . . . .	43
4.3.2 Ricostruzione 3D e Gestione dell'Incertezza . . . . .	45
4.4 Valutazione della Qualità del Dato: Score di Affidabilità Temporale .	49
4.5 Correzione del Segnale: Filtraggio Adattivo Savitzky-Golay . . . . .	51
4.6 Gestione dell'Output e Visualizzazione dei Risultati . . . . .	53
4.6.1 Salvataggio dei Dati Quantitativi in Formato JSON . . . . .	54
4.6.2 Generazione degli Output Visivi per la Validazione . . . . .	55
<b>5 Risultati Sperimentali</b>	<b>57</b>
5.1 Presentazione del Caso di Studio e Protocollo Applicato . . . . .	57
5.2 Validazione Qualitativa della Correzione del Segnale . . . . .	58
5.2.1 Caso 1: Analisi di un Frame a Basso Jitter (Frame 142) . . . . .	58
5.2.2 Caso 2: Analisi di un Frame ad Alto Jitter (Frame 185) . . . . .	60
5.2.3 Caso 3: Analisi di un Frame a Bassa Incertezza (Frame 546) .	62
5.2.4 Analisi della Traiettoria Temporale di un Keypoint . . . . .	62
5.3 Valutazione Quantitativa dell'Efficacia del Filtraggio . . . . .	65
5.3.1 Metrica di Valutazione del Jitter . . . . .	65
5.3.2 Risultati della Riduzione del Jitter . . . . .	65
5.4 Discussione dei Risultati e Risposta alle Domande di Ricerca . . . . .	66
5.4.1 Risposta alle Domande di Ricerca . . . . .	67

<b>6 Conclusioni</b>	<b>69</b>
6.1 Vantaggi e Limitazioni dell'Approccio . . . . .	69
6.1.1 Vantaggi e Punti di Forza . . . . .	70
6.1.2 Limitazioni e Aspetti Critici . . . . .	71
6.2 Applicabilità Clinica e Integrazione nel Progetto SPECTRA . . . . .	72
6.3 Sviluppi Futuri . . . . .	73
<b>Bibliografia</b>	<b>76</b>

---

## Elenco delle figure

---

1.1	Suddivisione del ciclo del passo nella fase di appoggio e nella fase di oscillazione . . . . .	5
2.1	Esempio di Analisi del cammino con l'utilizzo del Motion Capture .	19
2.2	Esempio di landmark per mani e corpo di MediaPipe . . . . .	23
2.3	Esempio di frame video e corrispondente risultato di rilevazione della HPE. Le linee verdi rappresentano il movimento effettivo, le rosse il modello testato. . . . .	25
3.1	Configurazione del sistema di acquisizione video impiegato per la registrazione delle sessioni sperimentali con i pazienti. Sono visibili le videocamere, i cavalletti e il tappeto di camminata. . . . .	34
3.2	Riassunto della metodologia utilizzata. . . . .	36
5.1	Confronto visivo per il frame 142 (basso jitter). Non si notano differenze significative tra lo scheletro grezzo (a) e quello filtrato (b), indicando che il filtro ha correttamente preservato il dato originale. . . . .	59
5.2	Confronto visivo per il frame 185 (alto jitter). Lo scheletro grezzo (a) presenta una posa anatomicamente incoerente. Lo scheletro filtrato (b) mostra una posa corretta e stabile. . . . .	61

5.3 Confronto visivo per il frame 546 (bassa incertezza). La prossimità del soggetto alla telecamera risulta in dati grezzi di alta qualità, che vengono correttamente preservati dal filtro adattivo. . . . .	63
5.4 Andamento temporale della coordinata Y della caviglia sinistra ( <i>id: 27</i> ) per 31 frame consecutivi. La linea blu (segnaletico grezzo) mostra un evidente jitter. La linea rossa (segnaletico filtrato) segue il trend generale del movimento, ma rimuove efficacemente le oscillazioni ad alta frequenza. . . . .	64

---

## **Elenco delle tabelle**

---

2.1	Confronto specifiche telecamere RGB-D . . . . .	20
3.1	Caratteristiche tecniche principali della telecamera Intel RealSense D455f . . . . .	33
3.2	Configurazione dei parametri del modello MediaPipe Holistic . . . . .	36
5.1	Confronto quantitativo del jitter prima e dopo l'applicazione del filtro adattivo. I valori rappresentano la deviazione standard della magnitudine della velocità del keypoint (in unità pixel/frame). Una riduzione maggiore indica un miglioramento più marcato. . . . .	66

---

## **Elenco degli Algoritmi**

---

1	Estrazione dei Keypoint 2D Significativi . . . . .	44
2	Stima Robusta della Profondità . . . . .	46
3	Calcolo dell'Errore di Profondità . . . . .	47
4	Calcolo dello Score di Affidabilità del Frame (Jitter Score) . . . . .	50
5	Filtraggio Adattivo Savitzky-Golay . . . . .	52
6	Disegno dello Scheletro su Immagine . . . . .	56

# CAPITOLO 1

---

## Introduzione

---

### 1.1 Panoramica sui disturbi motori nella schizofrenia

La schizofrenia è universalmente riconosciuta come un disturbo psichiatrico complesso e multisfaccettato, le cui manifestazioni cliniche si estendono ben oltre i sintomi psicotici positivi, quali deliri e allucinazioni, che ne costituiscono spesso l'aspetto più eclatante [1]. Un dominio sintomatologico di cruciale importanza, sia dal punto di vista teorico che clinico, è rappresentato dalle alterazioni della sfera motoria e del comportamento non verbale [1, 2]. Storicamente, l'osservazione clinica attenta ha permesso di identificare una vasta gamma di disturbi del movimento nei pazienti affetti da schizofrenia. Questi possono variare da sottili anomalie della coordinazione, della fluidità e dell'armonia motoria, fino a manifestazioni più evidenti e talvolta invalidanti come la catatonìa, le stereotipie, i manierismi, o la presenza di movimenti involontari [1, 2].

È fondamentale sottolineare che tali segni motori non sono relegati a fasi tardive della malattia né sono esclusivamente interpretabili come effetti collaterali dei trattamenti farmacologici antipsicotici, sebbene questi possano indubbiamente modularne o indurne alcuni. Evidenze crescenti suggeriscono infatti che i disturbi motori possano essere presenti già nelle fasi prodromiche della schizofrenia, e addirittura essere

considerati come possibili marcatori precoci di rischio per lo sviluppo di psicosi [1, 3]. Questa precoce comparsa depone a favore di una loro natura intrinseca al disturbo, riflettendo possibili disfunzioni neurobiologiche e neuroevolutive sottostanti [1, 2].

Particolare rilievo assume, nel quadro dei disturbi motori, la marcata riduzione dell'espressività non verbale, che include la mimica facciale, la gestualità, la postura e il movimento corporeo generale [1, 4]. Questa ipomimia e la generale carenza espressiva sono considerate componenti centrali e distintive dei sintomi negativi della schizofrenia [1]. I sintomi negativi, che comprendono anche abulia, apatia, anedonia e alogia, determinano un deficit generalizzato nel comportamento diretto a uno scopo (goal-directed behavior) e si ipotizza possano essere correlati a disfunzioni nei meccanismi cerebrali che sottendono la motivazione, il movimento e gli stati emotivi [1, 5]. La mancanza di espressività emotiva, in particolare, si è dimostrata uno dei marcatori più consistenti e affidabili nella ricerca sull'emotività nella schizofrenia [1, 6]. Tale deficit nell'espressività motoria e non verbale può avere un impatto profondamente negativo sulla capacità dei pazienti di stabilire e mantenere relazioni interpersonali, sulla loro competenza sociale e, di conseguenza, sulla loro qualità di vita complessiva [1].

Oltre alla ridotta espressività, la ricerca ha documentato alterazioni motorie più specifiche. Ad esempio, pazienti con schizofrenia mostrano frequentemente deficit nei movimenti oculari di inseguimento lento (smooth-pursuit eye movement) e nell'integrazione del movimento visivo [1, 7]. Sono state riportate anche problematiche relative al controllo posturale e all'equilibrio (postural sway) [1, 8], nonché alterazioni nei pattern di esplorazione motoria di ambienti non familiari [1, 9]. Anche la pianificazione psicomotoria sembra essere compromessa fin dalle fasi iniziali della malattia [1, 10]. Un altro aspetto rilevante è la "lentezza" (slowness) cognitiva e motoria, identificata come un sintomo chiave dei disturbi dello spettro schizofrenico [1, 11]. Questa lentezza motoria può essa stessa essere un indicatore della gravità della schizofrenia, data la sua stretta associazione con i sintomi negativi [1].

In questo scenario, l'analisi quantitativa e oggettiva del movimento, e in particolare del cammino (gait analysis), emerge come uno strumento di grande potenziale. Il cammino è un'attività motoria complessa e automatica, ma al contempo sensibile a disfunzioni del sistema nervoso centrale. Le alterazioni della deambulazione sono state

ampiamente studiate in diverse patologie neurologiche [12] e potrebbero, per analogia e per le disfunzioni neuromotorie intrinseche alla schizofrenia, rappresentare un ambito di indagine fruttuoso anche in questo contesto psichiatrico. La possibilità di quantificare oggettivamente parametri del cammino potrebbe fornire biomarker utili per una migliore caratterizzazione dei disturbi motori, per il monitoraggio della progressione della malattia e per la valutazione della risposta a specifici interventi terapeutici mirati anche alla componente motoria e negativa della schizofrenia [1].

## 1.2 Importanza della gait analysis in ambito clinico

L’analisi del cammino, comunemente nota come *gait analysis*, rappresenta una metodologia diagnostica e di ricerca di crescente rilievo nel panorama clinico-scientifico [12]. Il suo scopo è analizzare in dettaglio i meccanismi del movimento umano, in particolare della camminata e della corsa, al fine di diagnosticare e trattare un’ampia gamma di anomalie dell’andatura, lesioni o disturbi neurologici. In ambito clinico, viene eseguita da specialisti come fisioterapisti o biomeccanici, che effettuano un’attenta osservazione dell’individuo mentre cammina o corre per valutare la postura, l’allineamento degli arti e l’attivazione muscolare. Il valore intrinseco della gait analysis risiede nella sua capacità di trasformare questa osservazione, altrimenti qualitativa e soggetta a variabilità inter-osservatore, in un insieme di dati numerici oggettivi, precisi, ripetibili e confrontabili nel tempo [12]. Le modalità con cui un individuo cammina sono infatti un riflesso complesso dell’integrità e della funzionalità del sistema muscolo-scheletrico, nervoso e cardiovascolare; di conseguenza, le alterazioni del pattern deambulatorio possono costituire un’importante spia, rivelando informazioni cruciali sulla qualità della vita, sullo stato di salute generale e sull’evoluzione di diverse condizioni patologiche [12].

L’applicazione della gait analysis in ambito clinico è vasta e trasversale a numerose discipline mediche. Essa si è dimostrata particolarmente preziosa nella diagnosi precoce, nel monitoraggio della progressione e nella valutazione dell’efficacia degli interventi terapeutici in un ampio spettro di disturbi neurologici, quali la malattia di Parkinson, la sclerosi multipla, le atassie cerebellari e le sequele motorie derivanti da ictus cerebrale [12]. Anche in campo ortopedico, la gait analysis è fondamentale per

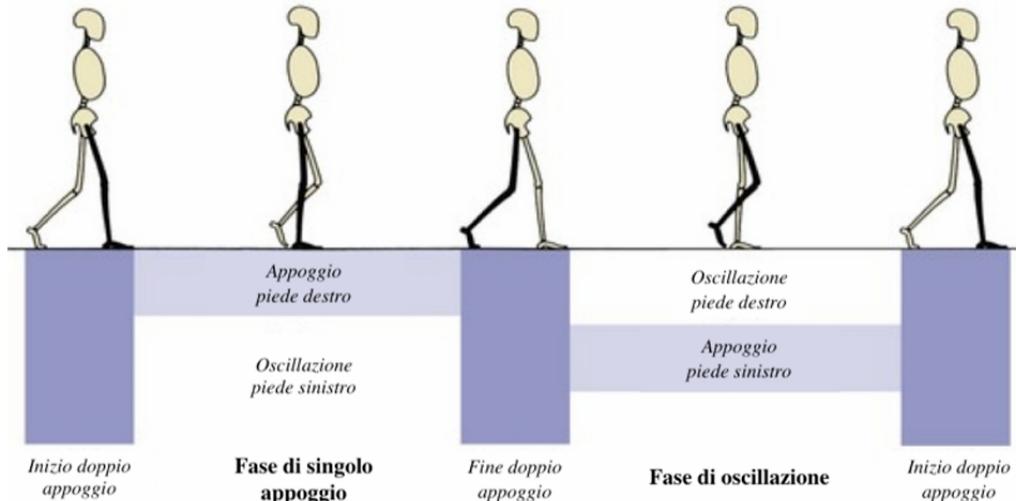
la valutazione funzionale pre e post-chirurgica (ad esempio, in caso di artroprotesi d'anca o di ginocchio), per la progettazione di ortesi e protesi personalizzate e per la gestione di patologie da sovraccarico. In ambito geriatrico, l'analisi del cammino è uno strumento chiave per identificare i soggetti a rischio di caduta, una delle principali cause di morbilità e mortalità nell'anziano, e per monitorare l'impatto delle malattie croniche sulla mobilità [12]. Persino in patologie sistemiche come le cardiopatie, il cammino può risultare significativamente alterato, e la sua analisi può fornire indicatori prognostici aggiuntivi [12, 13].

La forza della gait analysis risiede nella sua capacità di fornire misurazioni oggettive che superano i limiti delle tradizionali scale di valutazione clinica [12]. Per ottenere dati quantitativi e dettagliati, l'osservazione clinica è supportata da strumenti tecnologici avanzati. Tra questi, i sistemi di motion capture stereofotogrammetrici, che utilizzano un insieme di più telecamere emettitrici di luce e marker riflettenti posizionati sul corpo del paziente, sono tra i più utilizzati per creare una rappresentazione 3D del movimento. In aggiunta alle telecamere, vengono spesso impiegate le pedane di forza, dispositivi posizionati a terra che misurano la direzione, l'intensità e la tempistica delle forze di reazione al suolo prodotte durante il movimento, analizzando i pattern motori attraverso l'applicazione della cinematica alla dinamica.

L'unità fondamentale di analisi è il **ciclo del passo** (*gait cycle*), definito come l'intervallo di tempo che intercorre tra due contatti iniziali consecutivi dello stesso piede con il suolo. Il ciclo del passo è convenzionalmente suddiviso in due fasi principali:

- **Fase di appoggio (Stance Phase):** È il periodo in cui il piede è a contatto con il terreno. Inizia con il contatto del tallone (*heel strike*) e termina quando le dita dello stesso piede si staccano dal suolo (*toe-off*). Questa fase costituisce circa il 60% dell'intero ciclo del passo in una deambulazione normale ed è ulteriormente suddivisa in sottofasi come la risposta al carico, l'appoggio intermedio, l'appoggio terminale e il pre-pendolamento.
- **Fase di oscillazione (Swing Phase):** Rappresenta il periodo in cui l'arto di riferimento non è a contatto con il suolo, ma oscilla in avanti. Inizia subito dopo il distacco delle dita e termina con il successivo contatto del tallone. Questa

fase occupa circa il 40% del ciclo del passo e include le sottofasi di oscillazione iniziale, intermedia e terminale.



**Figura 1.1:** Suddivisione del ciclo del passo nella fase di appoggio e nella fase di oscillazione

Dall’analisi di queste fasi è possibile estrarre un’ampia gamma di parametri. I **parametri spazio-temporali** sono tra i più comuni e includono: la velocità del cammino, la cadenza, la lunghezza del passo e del semipasso, la durata percentuale delle fasi di stance, swing e di doppio appoggio, e le oscillazioni del corpo. Questa ricchezza di dati quantitativi, altrimenti non accessibili, riduce significativamente il margine di errore e fornisce una base solida per decisioni cliniche informate, consentendo di personalizzare i trattamenti e di valutarne l’efficacia in modo rigoroso [12].

Nel contesto specifico dei disturbi psichiatrici, come la schizofrenia, l’applicazione della gait analysis è un campo di indagine più recente ma di grande promessa. Data la già discussa presenza di disturbi motori intrinseci e la frequente compromissione della coordinazione e dell’espressività motoria in questi pazienti [1, 2], l’analisi oggettiva del cammino potrebbe offrire un biomarker comportamentale non invasivo e facilmente misurabile. La “lentezza motoria”, ad esempio, è stata identificata come un aspetto centrale dei sintomi negativi e un possibile indicatore della severità della schizofrenia [1, 11]. La gait analysis potrebbe quindi contribuire a quantificare oggettivamente tali deficit, a correlarli con altri aspetti della psicopatologia, a monitorare la

risposta a interventi farmacologici o riabilitativi specifici, e a stratificare i pazienti in sottogruppi con profili motori distinti. L'esigenza di metodi affidabili e convenienti per misurare il "deficit di espressività", che include la ritardatezza psicomotoria, è stata sottolineata come cruciale nella ricerca sulla schizofrenia [1, 14]. La gait analysis, specialmente con l'avvento di tecnologie più accessibili, si candida a rispondere a questa esigenza, fornendo uno strumento per approfondire la comprensione delle basi neurobiologiche dei disturbi motori in psichiatria e per migliorare l'assessment clinico.

### **1.3 Problematiche dei sistemi di motion capture tradizionali vs. soluzioni low-cost**

L'analisi quantitativa del movimento umano, e in particolare del cammino, si è storicamente avvalsa di sistemi di motion capture (MoCap) considerati il *gold standard* in termini di accuratezza e precisione [15]. Questi sistemi, prevalentemente basati su tecnologia ottica, impiegano un array di telecamere specializzate (spesso a infrarossi) per tracciare la posizione tridimensionale di marker riflettenti passivi applicati su specifici punti di repere anatomici del corpo del soggetto [15, 12]. Sistemi di fascia alta, come quelli prodotti da Vicon o OptiTrack (nelle sue configurazioni più avanzate), sono in grado di fornire dati cinematici con un'elevatissima risoluzione spaziale e temporale. Studi di validazione hanno dimostrato che tali sistemi possono raggiungere errori RMS inferiori al millimetro in condizioni statiche e leggermente superiori, ma comunque molto contenuti (fino a circa 4.2 mm in alcuni casi), durante test dinamici complessi [15, 16]. Questa precisione è il risultato di hardware sofisticato, che include sensori di immagine ad alta risoluzione (CCD o CMOS), illuminatori potenti, e software avanzati per la calibrazione del volume di cattura, il tracciamento dei marker e la ricostruzione 3D delle traiettorie [15]. Tali sistemi permettono un'analisi dettagliata e simultanea di molteplici parametri del movimento, offrendo una base dati robusta per la ricerca biomeccanica e clinica [12].

Nonostante l'indubbia superiorità in termini di accuratezza, i sistemi di motion capture tradizionali presentano una serie di problematiche significative che ne ostano

### *1.3 – Problematiche dei sistemi di motion capture tradizionali vs. soluzioni low-cost*

colano una diffusione capillare, specialmente in contesti clinici routinari o in studi su larga scala. Il primo e più evidente ostacolo è rappresentato dal costo estremamente elevato, che può facilmente raggiungere le centinaia di migliaia di euro per l'acquisto dell'hardware, del software e per la successiva manutenzione [15]. In secondo luogo, tali sistemi richiedono la predisposizione di ambienti di laboratorio dedicati e controllati, con specifiche caratteristiche di illuminazione e spazio, limitando drasticamente la possibilità di effettuare misurazioni in contesti ecologici o durante lo svolgimento delle normali attività quotidiane del paziente [12, 15]. La gestione di questi laboratori implica inoltre la presenza di personale altamente specializzato e formato per le complesse procedure di calibrazione del sistema, per l'applicazione meticolosa dei marker sul soggetto (un processo che può essere lungo e talvolta invasivo o fastidioso per il paziente), per l'acquisizione dei dati e per la loro successiva, spesso laboriosa, elaborazione [12, 15]. Queste limitazioni rendono i sistemi tradizionali poco pratici per uno screening su vasta scala o per un monitoraggio continuo e a lungo termine.

In risposta a queste criticità, gli ultimi anni hanno visto un'importante evoluzione tecnologica verso lo sviluppo e la diffusione di soluzioni low-cost per il motion capture [12, 15]. Queste alternative mirano a democratizzare l'accesso all'analisi quantitativa del movimento, offrendo strumenti con un rapporto costo-efficacia più vantaggioso. Tra queste soluzioni emergenti, un ruolo di primo piano è ricoperto dalle telecamere RGB-D (Red Green Blue - Depth), come il sensore Kinect di Microsoft o le telecamere Intel RealSense, capaci di fornire simultaneamente un'immagine a colori e una mappa di profondità della scena [12]. Parallelamente, si sono sviluppati sistemi markerless, che non richiedono l'applicazione di marker sul corpo, ma si affidano ad algoritmi di computer vision e machine learning per stimare la posa scheletrica del soggetto direttamente dalle immagini video (2D o 3D) [12]. I vantaggi principali di queste tecnologie low-cost sono evidenti: un costo di acquisto drasticamente inferiore (spesso nell'ordine di poche centinaia di euro), una maggiore portabilità che ne consente l'utilizzo anche al di fuori di laboratori specializzati, e una potenziale maggiore facilità d'uso, riducendo la necessità di personale iper-specializzato [15, 12]. Uno studio comparativo condotto da Thewlis e collaboratori (2011) tra un sistema OptiTrack di fascia media (considerato low-cost rispetto al Vicon) e un sistema Vicon high-end ha messo in luce come, pur persistendo una superiorità del sistema Vicon

in termini di accuratezza lineare assoluta, anche il sistema low-cost fosse in grado di fornire risultati comparabili, con errori percentuali assoluti inferiori all'1% nei test di accuratezza lineare e differenze relativamente contenute nelle cinematiche angolari durante l'analisi del cammino [15]. In particolare, per molti parametri cinematici del cammino, le differenze tra i due sistemi non superavano i pochi gradi, sebbene discrepanze maggiori potessero emergere per rotazioni sul piano trasverso, notoriamente più sensibili agli artefatti da movimento dei tessuti molli e al posizionamento dei marker (o alla stima della posa) [15].

Tuttavia, l'adozione di soluzioni low-cost non è esente da sfide. La principale limitazione risiede in una minore accuratezza intrinseca rispetto ai sistemi high-end [12]. I dati grezzi forniti da sensori come le telecamere RGB-D possono essere affetti da un rumore maggiore, da una risoluzione spaziale e temporale inferiore, e da artefatti specifici legati alla tecnologia di misurazione della profondità (ad esempio, sensibilità a superfici riflettenti o trasparenti, limitazioni nel range di misurazione, occlusioni) [12] (si veda Tabella 2 in [12] per un confronto delle tecnologie di misurazione della profondità). Alcuni sistemi low-cost possono presentare volumi di cattura più ristretti e frequenze di campionamento massime inferiori rispetto ai sistemi tradizionali [15]. Inoltre, la stima markerless della posa, pur eliminando l'onere dell'applicazione dei marker, introduce una complessità algoritmica significativa e può portare a una minore precisione nella localizzazione dei giunti scheletrici, specialmente in condizioni di illuminazione non ottimali o con movimenti rapidi e complessi. Di conseguenza, i dati ottenuti da sistemi low-cost richiedono spesso procedure di elaborazione e correzione più sofisticate per poter essere utilizzati in applicazioni cliniche che necessitano di un elevato grado di affidabilità. La gestione dell'incertezza nella stima della profondità, il filtraggio del rumore, la correzione degli errori di tracciamento e la validazione dei keypoint estratti diventano passaggi cruciali per migliorare la qualità dei dati cinematici [12]. È proprio in questo ambito che si colloca la necessità di sviluppare pipeline di elaborazione dedicate, come quella proposta nella presente tesi, al fine di massimizzare il potenziale informativo delle tecnologie low-cost per l'analisi del cammino in contesti clinici.

## 1.4 Obiettivi e contributi della tesi

Le sezioni precedenti hanno delineato un quadro in cui emerge con chiarezza, da un lato, la significativa presenza e rilevanza clinica dei disturbi motori nella schizofrenia [1, 2] e, dall’altro, il valore crescente della gait analysis come strumento di valutazione oggettiva e quantitativa in molteplici contesti clinici [12]. È stato altresì evidenziato il divario esistente tra i sistemi di motion capture tradizionali, altamente accurati ma costosi e complessi, e le più recenti soluzioni low-cost, promettenti in termini di accessibilità e portabilità, ma che presentano intrinseche sfide legate all’affidabilità e alla qualità del dato grezzo [15, 12]. In questo scenario, la necessità di sviluppare metodologie avanzate per l’elaborazione e la correzione dei dati provenienti da sistemi a basso costo diventa imperativa per poterne sfruttare appieno le potenzialità in ambito clinico.

Il presente lavoro di tesi si inserisce in tale contesto con l’obiettivo primario di progettare, implementare e validare uno strumento software specificamente concepito per l’elaborazione e la correzione accurata dei dati video tridimensionali acquisiti mediante telecamere RGB-D, con particolare riferimento al modello Intel RealSense. La finalità ultima di tale strumento è quella di estrarre parametri cinematici del cammino che siano sufficientemente affidabili e robusti da poter essere impiegati per la gait analysis in pazienti affetti da schizofrenia, superando le limitazioni tecniche tipiche dei sistemi non specialistici.

Per investigare sistematicamente il problema e guidare lo sviluppo, il presente lavoro si propone di rispondere alle seguenti domande di ricerca (Research Questions - RQs):

**RQ1** È possibile, attraverso una pipeline di elaborazione software dedicata, mitigare le imprecisioni intrinseche di un sensore RGB-D a basso costo per ottenere dati cinematici del cammino con un’affidabilità sufficiente per l’analisi clinica?

**RQ2** In che misura un approccio di filtraggio adattivo, basato su filtri di Savitzky-Golay i cui parametri sono modulati dalla qualità del segnale, è efficace nel ridurre il rumore (jitter) delle traiettorie dei keypoint, preservando al contempo le caratteristiche cinematiche significative del movimento?

**RQ3** È possibile definire e implementare una metrica quantitativa oggettiva, come uno score di affidabilità basato sulla coerenza temporale, per validare l'efficacia della pipeline di correzione e per identificare automaticamente i segmenti di dati meno affidabili?

Per raggiungere tale macro-obiettivo, la ricerca si articola nei seguenti obiettivi specifici, che rispecchiano fedelmente l'architettura dello strumento software implementato:

- **Sviluppo di una pipeline di elaborazione automatizzata per dati RealSense:**  
Creare un processo integrato che, partendo da un file di registrazione grezzo ('.bag') di una telecamera Intel RealSense, gestisca l'intero flusso di lavoro. Questo include l'estrazione sincronizzata dei frame di colore e di profondità e la loro preparazione per le fasi successive.
- **Estrazione e ricostruzione 3D robusta dei keypoint scheletrici:**
  - Utilizzare la libreria **MediaPipe Holistic** per l'estrazione automatica dei 33 landmark principali del corpo umano dai frame video a colori, configurandola per massimizzare l'accuratezza e la sensibilità del rilevamento anche in condizioni non ottimali.
  - Implementare una strategia di fusione dei dati di profondità che superi la semplice lettura del singolo pixel. Per ogni keypoint 2D, la profondità viene stimata calcolando la **media dei valori validi in un intorno spaziale** (una finestra 5x5) e, in caso di dati mancanti, sfruttando una **cronologia temporale** degli ultimi valori di profondità noti per quel medesimo keypoint, garantendo così una maggiore continuità e robustezza del dato 3D.
  - Quantificare l'incertezza della misura di profondità per ogni keypoint, attraverso l'implementazione di un modello di errore che calcola l'**errore RMS teorico in millimetri**, basandosi sui parametri intrinseci della telecamera (distanza focale, baseline) estratti dal file di registrazione.
- **Sviluppo di un sistema di validazione della qualità dei dati e di correzione adattiva:**

- Definire e implementare una metrica quantitativa di affidabilità per ogni frame. Tale metrica valuta la coerenza temporale del movimento analizzando la variabilità delle distanze 3D inter-frame dei keypoint e penalizzando attivamente le stime affette da profondità nulla o bassa visibilità, fornendo un indicatore immediato della qualità del dato grezzo.
- Implementare un **filtro di Savitzky-Golay adattivo** per la riduzione del jitter e degli artefatti di movimento. A differenza di un approccio fisso, questo filtro modula dinamicamente i suoi parametri (dimensione della finestra e ordine polinomiale) in base a metriche di qualità del segnale, come la visibilità media del keypoint e la varianza del suo movimento. Il filtraggio viene applicato selettivamente solo alle coordinate che superano una determinata soglia di variabilità, preservando così i movimenti fluidi e correggendo solo le sequenze problematiche.

- **Validazione e output per l'analisi cinematica:**

- Estrarre e salvare in formato JSON strutturato sia i dati cinematici grezzi che quelli filtrati, arricchiti con metadati quali l'errore di profondità calcolato e gli score di affidabilità, al fine di renderli pronti per una successiva e dettagliata analisi dei parametri del cammino (es. lunghezza e velocità del passo, angoli articolari).
- Produrre output visivi per la validazione qualitativa dell'efficacia del filtraggio, inclusa la generazione di video che mostrano lo scheletro originale, quello filtrato e un confronto fianco a fianco o sovrapposto, permettendo una valutazione immediata della riduzione del jitter e del miglioramento della fluidità delle traiettorie.

I principali contributi scientifici e tecnologici di questo lavoro di tesi possono essere così sintetizzati:

1. **Contributo Metodologico:** La proposizione di un approccio integrato e innovativo per l'elaborazione dei dati di movimento 3D da telecamere RGB-D. Questo approccio si distingue per la combinazione unica di una stima della profondità resa più robusta da una media spaziale e una memoria temporale e un sistema

di valutazione quantitativa della qualità del dato tramite un reliability score basato sul jitter, infine un **filtraggio di Savitzky-Golay adattivo e condizionale**, che interviene solo dove e quando necessario. Tale metodologia rappresenta un passo avanti significativo per superare le limitazioni di affidabilità dei sensori low-cost.

2. **Contributo Computazionale:** Lo sviluppo di uno strumento software completo e automatizzato, che implementa l'intera pipeline descritta. Questo strumento fornisce una soluzione pratica e riproducibile, potenzialmente trasferibile ad altri contesti di ricerca o clinici che necessitano di un'analisi del movimento affidabile partendo da tecnologie accessibili.
3. **Contributo Applicativo:** La dimostrazione pratica che, tramite un'elaborazione software mirata e sofisticata, è possibile elevare la qualità dei dati provenienti da sistemi RGB-D low-cost a un livello tale da renderli idonei per applicazioni cliniche esigenti come la gait analysis in pazienti schizofrenici. Questo lavoro facilita concretamente l'introduzione di valutazioni oggettive della motricità nella routine clinica psichiatrica, in linea con la crescente necessità di disporre di marker quantitativi per i deficit funzionali [1].
4. **Avanzamento della Conoscenza:** La fornitura di un metodo quantitativo e di output visivi per valutare e dimostrare l'efficacia delle tecniche di correzione. Contribuendo alla letteratura con una metodologia specifica per il miglioramento dei dati da sensori low-cost, questo lavoro rafforza la validità di tali sistemi come alternative praticabili ai più costosi sistemi di motion capture tradizionali [15].

In sintesi, questa tesi si propone di colmare parte del divario tra le potenzialità offerte dai sistemi di motion capture a basso costo e le esigenze di affidabilità dei dati richieste dall'ambito clinico, con un'applicazione specifica e rilevante allo studio dei disturbi motori nella schizofrenia.

## 1.5 Struttura della Tesi

Il presente lavoro di tesi è articolato nei seguenti capitoli, che guidano il lettore dal contesto teorico fino all'implementazione pratica e alla discussione dei risultati:

**Capitolo 1 - Introduzione:** Introduce il contesto clinico dei disturbi motori nella schizofrenia e l'importanza della gait analysis. Vengono presentate le sfide delle tecnologie di motion capture e vengono definiti il problema di ricerca, gli obiettivi, le domande di ricerca e i contributi principali di questa tesi.

**Capitolo 2 - Stato dell'Arte:** Fornisce una rassegna completa della letteratura scientifica e tecnologica. Vengono analizzati gli studi esistenti sull'analisi del cammino in ambito psichiatrico, le diverse tecnologie di motion capture, le tecniche di stima della posa come MediaPipe e i metodi di filtraggio dei dati di movimento. Infine, si contestualizza il lavoro all'interno del più ampio progetto di ricerca SPECTRA.

**Capitolo 3 - Metodologia:** Descrive in dettaglio l'approccio metodologico adottato. Vengono presentate l'architettura generale del sistema, il setup sperimentale per l'acquisizione dei dati e il design della pipeline di elaborazione software, giustificando le scelte tecnologiche e algoritmiche.

**Capitolo 4 - Implementazione:** Approfondisce gli aspetti pratici e implementativi del software sviluppato. Vengono illustrati l'ambiente di sviluppo, le librerie utilizzate e, tramite l'uso di pseudocodice, vengono spiegate in dettaglio le funzioni chiave per l'estrazione dei dati, la ricostruzione 3D, la validazione della qualità e la correzione adattiva del segnale.

**Capitolo 5 - Risultati Sperimentali:** Presenta e analizza i risultati ottenuti dall'applicazione della pipeline sviluppata. Questa sezione includerà una valutazione quantitativa del miglioramento della qualità dei dati (es. riduzione del jitter, incremento degli score di affidabilità) e un'analisi qualitativa basata sugli output visivi.

**Capitolo 6 - Conclusioni:** Interpreta i risultati ottenuti nel contesto dello stato dell'arte e degli obiettivi iniziali. Vengono discussi i vantaggi e le limitazioni

dell'approccio proposto, la sua potenziale applicabilità in ambito clinico e le possibili direzioni per sviluppi futuri della ricerca.

# CAPITOLO 2

---

## Stato dell'arte

---

### 2.1 Letteratura sulla gait analysis in pazienti psichiatrici

Le alterazioni della sfera motoria costituiscono una manifestazione clinica di rilievo nei disturbi dello spettro schizofrenico, essendo osservabili nella maggior parte dei soggetti affetti [17]. Queste anomalie motorie genuine (Genuine Motor Abnormalities - GMA) possono interessare fino all'80% dei pazienti e sono state documentate anche nel 66% dei casi al primo episodio psicotico, prima dell'inizio di trattamenti antipsicotici. Inoltre, manifestazioni motorie più sfumate sono state rilevate in individui ad altissimo rischio di sviluppare la malattia (Ultra High Risk - UHR) e nei familiari di primo grado non affetti, suggerendo che le GMA possano rappresentare un indicatore di vulnerabilità neuroevolutiva e fungere da potenziale biomarcatore prognostico [17]. Nonostante la loro elevata prevalenza e il potenziale significato clinico, le anomalie motorie sono state storicamente sottostimate nella pratica clinica e la diagnostica psichiatrica tradizionale raramente include una valutazione quantitativa e standardizzata del comportamento motorio, specialmente per quanto riguarda il movimento dell'intero corpo [17].

La caratterizzazione delle GMA nella schizofrenia è complicata dalla notevole eterogeneità delle loro descrizioni e classificazioni, che tendono a variare in base

all’approccio concettuale e agli strumenti di valutazione impiegati dai ricercatori [17]. Ad esempio, alcune categorizzazioni distinguono tra segni neurologici sfumati (Neurological Soft Signs - NSS), movimenti involontari anomali (ipercinetici come discinesia e acatisia, o ipocinetici come il parkinsonismo) e fenomeni catatonici [17]. Altre classificazioni includono anche il rallentamento psicomotorio che impatta sia la motricità fine sia quella grossolana, come la deambulazione [17]. Le scale di valutazione tradizionali per le GMA, basandosi sull’osservazione clinica, presentano limiti intrinseci legati alla soggettività del giudizio, alla dipendenza dalla formazione del valutatore e a una ridotta sensibilità nel cogliere anomalie motorie subcliniche [17].

Nell’ambito delle GMA, i Neurological Soft Signs (NSS) hanno suscitato un interesse particolare. Gli NSS sono considerati possibili indicatori sia di tratto (predisposizione alla malattia) sia di stato (utili per monitorare l’evoluzione clinica) e sono stati correlati non solo con la sintomatologia psicopatologica della schizofrenia, ma anche con compromissioni significative del funzionamento cognitivo (ad es., attenzione, velocità di elaborazione, memoria) e sociale dei pazienti [17]. Tra le diverse manifestazioni di NSS, l’incoordinazione motoria è una delle più frequentemente documentate, evidenziata da difficoltà nell’esecuzione di movimenti rapidi alternati, nella marcia in tandem o nel test dito-naso. Alcuni studi hanno indicato che i deficit di coordinazione motoria e di coordinazione tra gli arti possono essere particolarmente utili nel discriminare tra soggetti ad alto rischio e controlli, o tra pazienti con schizofrenia e pazienti con disturbi dell’umore [17].

Il riconoscimento dell’importanza di una valutazione oggettiva delle GMA ha spinto verso l’adozione di tecnologie moderne. Tuttavia, sebbene siano stati utilizzati strumenti come accelerometri, sensori di pressione plantare o actigrafi, la maggior parte di queste indagini si è concentrata sulla motricità fine o sui movimenti degli arti superiori. Gli studi dedicati specificamente al movimento dell’intero corpo, e in particolare all’analisi quantitativa e dettagliata del cammino (gait analysis), sono stati meno numerosi [17]. Quando la deambulazione è stata oggetto di studio, l’analisi si è spesso limitata a un numero ristretto di variabili, come la lunghezza del passo e la cadenza, o a misure molto generali dell’attività motoria, tralasciando un’esplorazione approfondita delle qualità cinematiche e dinamiche del movimento [17].

L’analisi del cammino mediante sistemi di motion capture (MoCap) rappresenta l’approccio metodologico più accurato e completo per lo studio della biomeccanica del movimento [18, 17]. Sebbene questa tecnologia sia ampiamente utilizzata in altri contesti clinici e di ricerca, la sua applicazione in ambito psichiatrico per analizzare pattern di movimento con rilevanza diagnostica è stata più sporadica. Uno degli studi pionieristici citati nella letteratura psichiatrica è quello condotto su pazienti con depressione maggiore, i quali, rispetto ai controlli, mostravano una velocità di cammino ridotta, una minore oscillazione delle braccia, un’attenuazione del movimento verticale del corpo, una postura del tronco più cifotica e un’accentuata oscillazione laterale [17].

Per quanto riguarda specificamente la schizofrenia, la ricerca sulla gait analysis ha iniziato a delineare alcune caratteristiche distintive. È stato riportato che i pazienti schizofrenici presentano disturbi nei parametri spazio-temporali del cammino, con una velocità di deambulazione generalmente inferiore a quella dei controlli sani. Tale riduzione della velocità sembra essere primariamente dovuta a una diminuzione della lunghezza del passo, piuttosto che a una significativa alterazione della cadenza [17]. Altri studi hanno investigato la variabilità del passo (stride-to-stride variability), riscontrando differenze significative tra pazienti e controlli soprattutto quando il compito di deambulazione veniva associato a un secondo compito cognitivo (condizione di dual-task), suggerendo possibili difficoltà nell’allocazione delle risorse attentive durante il movimento [17]. Anche aspetti posturali durante il cammino sono stati oggetto di indagine, con osservazioni relative, ad esempio, a una maggiore inclinazione anteriore della testa nei pazienti [17]. Parallelamente, studi basati sulla video-analisi hanno permesso di quantificare il movimento corporeo globale durante le interazioni sociali, evidenziando una correlazione tra la sua riduzione e l’intensità dei sintomi negativi, un nucleo sintomatologico centrale nella schizofrenia [1].

Un contributo recente e particolarmente dettagliato è fornito dallo studio di Martin e collaboratori (2022), che ha impiegato la tecnologia MoCap per un’analisi approfondita dei pattern di cammino dell’intero corpo in pazienti con schizofrenia. Questa ricerca ha identificato sedici specifici marcatori di movimento (Movement Markers - MM) che discriminano significativamente i pazienti dai controlli. Tali marcatori riguardano alterazioni della postura (testa più inclinata in avanti), della

velocità (ridotta), della regolarità del ciclo del passo (minore), dell'ampiezza e della fluidità delle oscillazioni dei segmenti corporei (ridotta oscillazione delle braccia, aumentata oscillazione laterale del tronco), della flessibilità articolare (ridotta) e dell'integrazione e coordinazione tra i diversi segmenti corporei e tra i due emilati del corpo [17]. Questi risultati sono risultati robusti anche dopo aver controllato per l'effetto dei farmaci antipsicotici e hanno mostrato correlazioni significative con la gravità dei Neurological Soft Signs, specialmente quelli relativi alla coordinazione motoria e all'integrazione sensoriale [17]. Nonostante questi promettenti avanzamenti, la valutazione strumentale e oggettiva del cammino e delle altre anomalie motorie non è ancora parte integrante della prassi clinica e diagnostica in psichiatria [17]. Rimane pertanto una significativa necessità di affinare le metodologie di analisi, di validare ulteriormente i marcatori di movimento identificati e di sviluppare strumenti che siano non solo accurati ma anche accessibili e di facile implementazione nei contesti clinici. L'efficacia della gait analysis nel migliorare la gestione clinica e le decisioni terapeutiche è ben documentata in diverse altre specialità mediche [18], fornendo un solido razionale per promuoverne l'applicazione anche nello studio e nel trattamento dei disturbi psichiatrici, inclusa la schizofrenia.

## **2.2 Tecnologie di motion capture: da sistemi professionali a RGB-D**

L'analisi quantitativa del movimento umano, e in particolare della deambulazione, si avvale di una varietà di tecnologie di motion capture (MoCap), che spaziano da sistemi professionali ad alta precisione a soluzioni più recenti, a basso costo e maggiormente accessibili [12, 15]. La scelta della tecnologia più appropriata dipende da molteplici fattori, tra cui il livello di accuratezza richiesto, il contesto applicativo (ricerca, clinica, riabilitazione domiciliare), il budget disponibile e la complessità del setup gestibile.

I sistemi di motion capture ottici basati su marker sono tradizionalmente considerati il *gold standard* per l'analisi biomeccanica del movimento in ambito di ricerca e in contesti clinici specializzati [15, 12]. Questi sistemi, come quelli prodotti da



**Figura 2.1:** Esempio di Analisi del cammino con l'utilizzo del Motion Capture

Vicon o OptiTrack (nelle loro configurazioni di fascia alta), utilizzano un array di telecamere multiple, tipicamente a infrarossi, per tracciare con elevata precisione la posizione tridimensionale di marker riflettenti passivi applicati su specifici punti di repere anatomici del corpo del soggetto [15, 12]. L'evoluzione di tali sistemi ha portato all'integrazione di sensori di immagine ad alta risoluzione (CCD o CMOS), comunicazione wireless e alte frequenze di campionamento, raggiungendo in test statici errori RMS inferiori al millimetro e, in test dinamici, errori comunque molto contenuti, sebbene possano aumentare fino a pochi millimetri a seconda del sistema e delle condizioni [15, 16]. Anche i software di gestione si sono evoluti, offrendo algoritmi di calibrazione e tracciamento migliorati, streaming dei dati in tempo reale e formati di output standardizzati come il C3D [15]. Nonostante l'accuratezza, i costi elevati di acquisto e manutenzione (che possono raggiungere centinaia di migliaia di euro), la necessità di laboratori dedicati e personale altamente specializzato per la calibrazione, l'applicazione dei marker e l'elaborazione dei dati, ne hanno limitato la diffusione capillare, specialmente al di fuori dei centri di ricerca avanzati [15, 12, 19].

In risposta a queste limitazioni, è emersa una nuova generazione di tecnologie di motion capture a basso costo, che mirano a rendere l'analisi del movimento più accessibile e fruibile in contesti diversificati [15, 19, 20]. Tra queste, le telecamere RGB-D (Red Green Blue - Depth) hanno guadagnato una notevole popolarità. Questi dispositivi, come il Microsoft Kinect (nelle sue varie versioni) e la gamma Intel RealSense (ad es. D435, D455, SR300), sono in grado di acquisire simultaneamente informazioni sul colore (immagine RGB) e sulla profondità della scena, permettendo di ricostruire un modello tridimensionale dell'ambiente e dei soggetti presenti [19, 20, 21]. Le tecnologie di misurazione della profondità impiegate variano: il Kinect V1

utilizzava la luce strutturata, il Kinect V2 la tecnologia Time-of-Flight (ToF), mentre le camere Intel RealSense spesso impiegano la stereoscopia attiva con proiezione di pattern infrarossi per migliorare l'accuratezza della stima della profondità [20, 19, 21, 22]. Le camere RealSense, ad esempio, sono composte da una telecamera IR, una telecamera RGB e un proiettore laser IR; la profondità viene calcolata confrontando le immagini catturate dai due sensori IR, con l'ausilio del pattern proiettato per migliorare la rilevazione in scene con poca texture [21, 22].

**Tabella 2.1:** Confronto specifiche telecamere RGB-D .

Proprietà	Dispositivo			
	Kinect™ V1	Kinect™ V2	Intel	RealSense™ D435
Tecnologia	Luce Strutt.	Time-of-flight	Stereoscopia Att.	
Range PROF. (m)	0.8 - 4.0	0.5 - 4.5	0.2 - 4.5*	
Risoluzione	640x480	1920x1080	1920x1080	
FPS	30	30	30	
Risol. PROF.	320x240	512x424	1280x720	
FoV PROF.	57x43	70x60	85.2x58	
FPS PROF.	30	30	90	
N. Articolazioni	20	25	19	

I principali vantaggi delle telecamere RGB-D risiedono nel costo significativamente inferiore rispetto ai sistemi ottici tradizionali, nella maggiore portabilità e nella potenziale facilità d'uso, che non richiede necessariamente l'applicazione di marker sul corpo del soggetto, specialmente se abbinate ad algoritmi di stima della posa scheletrica (skeletal tracking) [19, 20]. Diversi studi hanno esplorato la validità di queste tecnologie per l'analisi del cammino e altri movimenti. Ad esempio, il Microsoft Kinect è stato testato in numerosi contesti clinici, mostrando un'accuratezza sufficiente per la valutazione di alcuni parametri del cammino e per applicazioni riabilitative, sebbene con prestazioni generalmente inferiori ai sistemi gold standard [20, 19]. La sua SDK (Software Development Kit) forniva moduli per il tracciamento scheletrico, facilitandone l'adozione da parte di ricercatori e sviluppatori [19]. Con la dismissione del Kinect da parte di Microsoft, le camere Intel RealSense sono emerse come una valida alternativa, offrendo funzionalità comparabili o superiori in termini di risoluzione, campo visivo e opzioni di tracciamento (facciale, gestuale, scheletrico) [19, 20]. Le camere RealSense come la D435 o la D455 sono state progettate per un'ampia gamma di applicazioni, inclusa la robotica e la visione artificiale, e offrono

SDK che permettono l’accesso ai flussi di dati grezzi (RGB, profondità, infrarosso) e a funzionalità di più alto livello [19, 21, 22]. La D435, ad esempio, utilizza la stereoscopia attiva e presenta un ampio campo visivo per la profondità, rendendola adatta a tracciare oggetti in movimento [19, 20]. Studi preliminari che hanno confrontato la RealSense D435 con il Kinect V2 per la misurazione della velocità del cammino hanno suggerito che questo dispositivo, piccolo ed economico, può essere utilizzato per misurare alcune variabili spazio-temporali, sebbene possano presentarsi problemi di rumore nei dati relativi agli arti distali (polsi e caviglie) a distanze maggiori [20]. La verifica dell’accuratezza della profondità fornita da queste camere è cruciale; ad esempio, esperimenti condotti con la RealSense D455 hanno mostrato un errore di misurazione della profondità inferiore all’1.5% a distanze di 1, 1.5 e 2 metri [21].

Nonostante i progressi, le tecnologie RGB-D presentano ancora delle sfide. L’accuratezza della stima della profondità e, di conseguenza, del tracciamento scheletrico 3D, può essere influenzata da vari fattori ambientali come le condizioni di illuminazione (specialmente per i sistemi basati solo su RGB o con proiettori IR meno potenti), la presenza di superfici riflettenti o trasparenti, e le occlusioni parziali o totali del corpo [19, 12]. Il campo visivo e il range ottimale di profondità sono altre considerazioni importanti: mentre le nuove camere D400 di Intel hanno un range effettivo che può superare i 10 metri in condizioni ideali, le prestazioni ottimali si ottengono solitamente a distanze più contenute [19, 20]. Inoltre, la risoluzione della mappa di profondità e la frequenza di campionamento, sebbene in miglioramento, possono essere inferiori a quelle dei sistemi professionali, limitando l’analisi di movimenti molto rapidi o fini [15, 19]. La robustezza degli algoritmi di skeletal tracking integrati negli SDK o forniti da librerie terze (come Nuitrack SDK menzionato per la RealSense D435 [20], o MediaPipe discusso più avanti) è fondamentale per ottenere dati cinematici affidabili, ma questi algoritmi possono ancora soffrire di instabilità, perdita di tracciamento o stime imprecise dei giunti, specialmente in posture complesse o non standard.

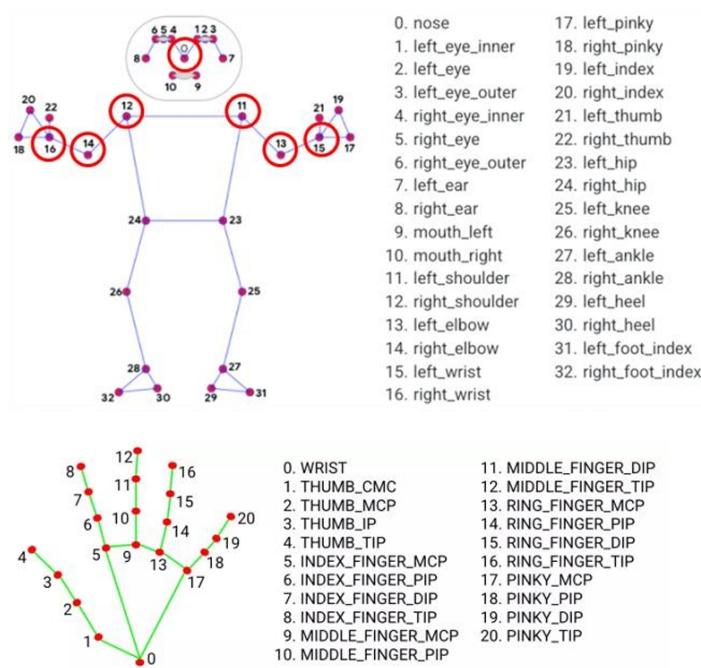
La comparazione diretta tra sistemi low-cost e high-end, come quella effettuata da Thewlis e collaboratori (2011) tra un sistema OptiTrack e un sistema Vicon, ha evidenziato che, sebbene i sistemi high-end mantengano una superiorità in termini di accuratezza lineare assoluta, anche le soluzioni low-cost possono raggiungere

prestazioni eccezionali, con errori percentuali assoluti inferiori all’1% in alcuni test lineari. Per quanto riguarda le cinematiche angolari durante il cammino, le differenze relative tra i due tipi di sistemi possono essere contenute per molti parametri, anche se discrepanze maggiori possono emergere per movimenti complessi o rotazioni su piani specifici (es. rotazione del ginocchio sul piano trasverso), che sono noti per essere sensibili al posizionamento dei marker e agli artefatti da movimento dei tessuti molli [15]. È importante notare che anche i sistemi low-cost presentano limitazioni come volumi di cattura potenzialmente più ridotti, frequenze di campionamento massime inferiori e software di gestione talvolta meno ottimizzati rispetto alle controparti professionali [15]. In conclusione, mentre i sistemi professionali rimangono il riferimento per studi che richiedono la massima precisione, le tecnologie RGB-D e altre soluzioni low-cost rappresentano un’alternativa promettente per ampliare l’accesso all’analisi del movimento, specialmente in contesti clinici e riabilitativi. La loro continua evoluzione, sia hardware che software, e lo sviluppo di metodologie di elaborazione dati sempre più sofisticate, come quelle che verranno discusse nelle sezioni successive, sono cruciali per superare le attuali limitazioni e validarne l’impiego per la misurazione affidabile degli outcome di salute [19].

## 2.3 MediaPipe e tecniche di pose estimation

La stima della posa umana (Human Pose Estimation - HPE) è un campo di ricerca della computer vision in rapida evoluzione, con l’obiettivo di localizzare le parti del corpo umano e determinarne la configurazione spaziale a partire da immagini o sequenze video [23, 24]. Questa tecnologia ha trovato applicazione in svariati settori, inclusi la videosorveglianza intelligente, l’assistenza medica e la riabilitazione, l’analisi del movimento sportivo, e l’interazione uomo-robot [23, 21]. L’output tipico di un sistema HPE è uno scheletro virtuale composto da un insieme di keypoint (o landmark) che rappresentano le principali articolazioni e punti di riferimento del corpo. Questi keypoint possono poi essere utilizzati per classificare posture, misurare angoli articolari, analizzare la correttezza di un movimento o riconoscere specifiche azioni [23, 24].

Negli ultimi anni, con la crescente domanda di soluzioni HPE efficienti e facili da integrare, sono state sviluppate numerose librerie e algoritmi, molti dei quali basati su approcci di machine learning e deep learning [23, 24]. Tra queste, MediaPipe, un framework open-source sviluppato da Google, si è distinto per la sua capacità di fornire soluzioni di percezione multimodale (volto, mani, corpo) ottimizzate per l'esecuzione in tempo reale su diverse piattaforme, inclusi dispositivi mobili, desktop e web [21, 23, 24]. All'interno dell'ecosistema MediaPipe, la soluzione MediaPipe Pose è specificamente dedicata alla stima della posa del corpo umano.



**Figura 2.2:** Esempio di landmark per mani e corpo di MediaPipe

MediaPipe Pose utilizza un approccio basato su machine learning per inferire la posizione di 33 landmark tridimensionali del corpo a partire da un'immagine RGB in input [21, 23, 24]. Questi landmark includono punti chiave per il volto (es. naso, occhi, orecchie, bocca), il tronco (spalle, fianchi), gli arti superiori (gomiti, polsi, dita) e gli arti inferiori (ginocchia, caviglie, piedi) [21, 23]. La capacità di fornire coordinate 3D "real-world" (sebbene la terza dimensione, la profondità, sia stimata rispetto alla telecamera e possa avere limitazioni se non supportata da un sensore di profondità dedicato) e la sua efficienza computazionale ne hanno favorito l'adozione in diverse applicazioni [24, 23]. Il sistema è progettato per funzionare in tempo reale,

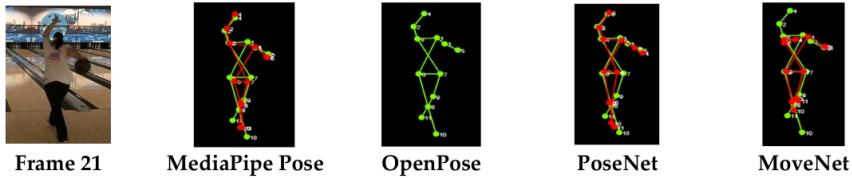
rendendolo adatto per applicazioni interattive [23, 25].

L’architettura di MediaPipe Pose, in particolare per la stima della posa di una singola persona, si basa su una pipeline a due stadi che impiega il modello BlazePose [21, 23, 25]. Nel primo stadio, un rilevatore (detector) identifica la regione di interesse (Region of Interest - ROI) contenente la persona all’interno del frame dell’immagine. Successivamente, un tracker (o predittore di landmark) utilizza questa ROI per stimare la posizione precisa dei 33 keypoint scheletrici. Nelle sequenze video, il rilevatore viene solitamente invocato nel primo frame (o quando il tracciamento viene perso) per localizzare la persona, mentre il tracker utilizza la ROI identificata per predire i keypoint nei frame successivi, ottimizzando così la velocità e la coerenza temporale. Se il tracker perde la persona, il rilevatore viene richiamato per ristabilire il tracciamento [21, 23, 25]. Questo approccio top-down (prima si localizza la persona, poi si stimano i suoi keypoint) è comune in molte soluzioni HPE [23]. Dal punto di vista della rete neurale sottostante, MediaPipe Pose utilizza reti convoluzionali (CNN) ottimizzate per l’efficienza [23].

La popolarità di MediaPipe Pose è dovuta a diversi fattori, tra cui la sua natura open-source, la facilità d’uso grazie a API ben documentate, e la sua efficienza computazionale che ne permette l’impiego anche su dispositivi con risorse limitate [24, 21]. È stato utilizzato in una varietà di applicazioni, come sistemi per la correzione della postura seduta, per la valutazione dell’equilibrio su tavolette proprietarie, o per il tracciamento di esercizi a corpo libero [23]. Ad esempio, nella collaborazione uomo-robot (Human-Robot Collaboration - HRC), la capacità di ottenere coordinate 3D affidabili dello scheletro umano è fondamentale per permettere ai robot di interagire in sicurezza con gli operatori, consegnare oggetti o evitare collisioni. In questo contesto, la combinazione di MediaPipe Pose per l’estrazione dello scheletro 2D dall’immagine RGB e di un sensore di profondità (come Intel RealSense) per ottenere la terza dimensione per ciascun keypoint è un approccio promettente per la stima 3D accurata [21].

Tuttavia, come per tutte le tecnologie di stima della posa basate su computer vision, anche MediaPipe Pose presenta delle limitazioni e la sua accuratezza può essere influenzata da diversi fattori. Studi comparativi tra diverse librerie HPE (come OpenPose, PoseNet, MoveNet e MediaPipe Pose) hanno cercato di valutarne le

prestazioni in vari scenari [23].



**Figura 2.3:** Esempio di frame video e corrispondente risultato di rilevazione della HPE. Le linee verdi rappresentano il movimento effettivo, le rosse il modello testato.

Chung et al. (2022) hanno condotto un'analisi comparativa utilizzando il dataset COCO (per immagini statiche) e Penn Action (per video), valutando le prestazioni tramite la metrica Percentage of Detected Joints (PDJ). Nei test su immagini statiche, MediaPipe Pose ha mostrato prestazioni inferiori rispetto ad altre librerie come MoveNet o OpenPose, con un numero maggiore di immagini in cui meno del 50% dei keypoint veniva rilevato correttamente [23]. Tuttavia, nei test su dataset video, MediaPipe Pose ha dimostrato prestazioni competitive, risultando leggermente inferiore a MoveNet in termini di PDJ medio complessivo, ma ottenendo il miglior punteggio in un numero maggiore di singole azioni. Le prestazioni di tutte le librerie HPE tendono a degradare in presenza di sfide comuni come l'auto-occlusione (parti del corpo che ne nascondono altre) o angolazioni della telecamera non ottimali [23].

Dill et al. (2023) hanno condotto una valutazione specifica dell'accuratezza di MediaPipe Pose (versione che fornisce coordinate 3D "real-world") per la stima di metriche rilevanti in esercizi di fisioterapia, confrontandola con un sistema MoCap inerziale (Xsens MTw Awinda) come ground truth. I risultati hanno indicato che l'accuratezza della stima della posa con MediaPipe è altamente dipendente dall'angolo di visuale della telecamera e dal tipo di esercizio eseguito. Sebbene si possa ottenere un'elevata accuratezza in condizioni ottimali (ad es. soggetto ben visibile, angolazione frontale o laterale a 90°), questa tende a diminuire rapidamente in condizioni meno favorevoli, come angolazioni intermedie o esercizi che comportano una maggiore auto-occlusione (ad es. esercizi a terra) [24]. Ad esempio, la stima di dimensioni corporee come l'ampiezza delle spalle o dei fianchi è risultata affetta da bias (dovuto alla mancanza di informazioni sulla taglia reale del soggetto nel modello MediaPipe) e da una notevole varianza, indicando una certa inconsistenza

del modello nel tempo e al variare dell’angolazione. Anche la stima degli angoli articolari (es. flessione del ginocchio) ha mostrato una dipendenza dall’angolo di visuale, con errori maggiori quando il movimento avviene su piani meno visibili dalla telecamera o quando gli angoli si avvicinano a valori estremi ( $0^\circ$  o  $180^\circ$ ) [24]. Gli autori suggeriscono che per migliorare l’accuratezza, specialmente in contesti clinici che richiedono elevata affidabilità, potrebbe essere necessario fusionare i dati da più telecamere o integrare modelli biomeccanici a priori per vincolare le stime delle proporzioni corporee e dei movimenti articolari [24]. Nonostante queste limitazioni, la facilità d’uso e l’efficienza di MediaPipe Pose lo rendono uno strumento prezioso per molte applicazioni, inclusa la potenziale integrazione con sensori di profondità per migliorare la stima 3D, come proposto nel presente lavoro di tesi per l’analisi del cammino.

## 2.4 Metodi di correzione e filtraggio dei dati di movimento

L’acquisizione di dati di movimento, specialmente tramite sistemi di motion capture low-cost o tecniche di stima della posa basate su computer vision, è intrinsecamente soggetta a diverse fonti di rumore e imprecisione [26, 27]. Il rumore può derivare dalle limitazioni hardware dei sensori, da fattori ambientali (come variazioni di illuminazione o occlusioni parziali), dalle approssimazioni degli algoritmi di stima della posa, o dalle fluttuazioni naturali del movimento umano stesso che possono non essere rilevanti per l’analisi specifica che si intende condurre [26, 27]. Ad esempio, durante la deambulazione, la posizione e l’orientamento del corpo del soggetto mostrano naturali fluttuazioni dovute all’alternanza dei passi, che possono essere considerate “rumore” se l’obiettivo è stimare una traiettoria media o predire la posizione a breve termine [27]. Pertanto, l’applicazione di metodi di correzione e filtraggio diventa un passaggio cruciale per migliorare la qualità, l’affidabilità e l’interpretabilità dei dati cinematici estratti, rendendoli più idonei per successive analisi quantitative e applicazioni cliniche.

Una delle problematiche più comuni nei dati di movimento grezzi è la presenza di

rumore ad alta frequenza (jitter), che si manifesta come rapide e piccole oscillazioni nelle traiettorie dei keypoint o negli angoli articolari calcolati. Questo rumore può compromettere la stima di derivate del movimento, come velocità e accelerazioni, e rendere più difficile l'identificazione di pattern di movimento significativi. Per affrontare questo problema, vengono comunemente impiegati filtri passa-basso, il cui scopo è attenuare o rimuovere le componenti di frequenza più elevate del segnale, considerate prevalentemente rumore, preservando al contempo le componenti a bassa frequenza associate al movimento effettivo. Tra i filtri passa-basso più utilizzati in biomeccanica vi è il filtro di Butterworth, spesso applicato con un ordine e una frequenza di taglio scelti in base alle caratteristiche del movimento analizzato e del rumore presente [27, 15]. La scelta della frequenza di taglio è critica: un valore troppo alto potrebbe non eliminare sufficientemente il rumore, mentre un valore troppo basso potrebbe attenuare componenti significative del movimento, distorcendo il segnale (oversmoothing) [27].

Un'alternativa o un complemento ai filtri in frequenza sono i filtri basati su regressioni polinomiali locali, tra i quali spiccano i filtri di Savitzky-Golay [27]. Questi filtri operano adattando un polinomio di un certo grado a una finestra mobile di campioni di dati e utilizzando il valore del polinomio al punto centrale della finestra come stima filtrata. Il vantaggio principale dei filtri di Savitzky-Golay risiede nella loro capacità di preservare meglio le caratteristiche del segnale originale, come l'altezza e la larghezza dei picchi, rispetto ai filtri media mobile o ad altri filtri passa-basso lineari, specialmente quando il segnale presenta variazioni rapide ma significative [27, 28]. Essi agiscono come un filtro passa-basso ottimale per segnali che possono essere approssimati localmente da polinomi e sono noti per la loro capacità di conservare la forma e i momenti del segnale [27]. La loro implementazione richiede la scelta del grado del polinomio e della dimensione della finestra (numero di punti). Versioni più recenti di questi filtri affrontano il problema del ritardo introdotto dai filtri simmetrici e la gestione dei dati agli estremi del segnale proponendo l'uso di finestre asimmetriche [27]. L'applicazione di filtri come Savitzky-Golay può essere particolarmente utile in contesti, come l'analisi di specifici pattern motori in patologie neurologiche o psichiatriche, dove è cruciale non "appiattire" eccessivamente il segnale, per non perdere informazioni sulle sottili variazioni che potrebbero essere clinicamente rilevanti.

È importante notare che, se l’obiettivo è filtrare fluttuazioni a frequenze molto basse (corrispondenti a periodi lunghi, come quelle intrinseche dell’alternanza del passo), i filtri convenzionali potrebbero introdurre ritardi o distorsioni significative se non attentamente parametrizzati [27].

Oltre al filtraggio del rumore ad alta frequenza, un’altra sfida importante è la gestione delle occlusioni locali o della perdita temporanea di tracciamento di alcuni keypoint. In questi casi, i dati possono mancare o essere palesemente errati per brevi intervalli. Tecniche di interpolazione (lineare, spline) possono essere utilizzate per stimare i dati mancanti, ma la loro efficacia dipende dalla durata dell’interruzione e dalla natura del movimento. In contesti di analisi video in tempo reale o quasi, il filtro di Kalman si è dimostrato uno strumento potente per la stima e la predizione dello stato di un sistema dinamico (come la posizione e la velocità di un keypoint) in presenza di rumore di misura e di processo [26]. Il filtro di Kalman opera in due fasi ricorsive: una fase di predizione, in cui lo stato corrente viene stimato sulla base dello stato precedente e di un modello dinamico del sistema, e una fase di aggiornamento (o correzione), in cui la stima predetta viene corretta utilizzando le nuove misurazioni osservate [26]. Questo approccio permette di ottenere una stima ottimale (nel senso dei minimi quadrati) dello stato del sistema, riducendo l’impatto del rumore e fornendo stime anche in caso di misurazioni mancanti temporanee, basandosi sul modello predittivo [26]. L’efficacia del filtro di Kalman dipende dalla corretta definizione del modello di stato e delle matrici di covarianza del rumore di processo e di misurazione. Hu et al. (2022) hanno introdotto il filtro di Kalman nella stima della posa per ridurre il rumore nelle coordinate dei keypoint e per gestire occlusioni locali, dimostrando un miglioramento nell’accuratezza della stima dei keypoint e una maggiore stabilità dei dati, eliminando oscillazioni e variazioni repentine [26].

In alcuni contesti specifici, come la predizione a breve termine della posizione umana durante il cammino, le fluttuazioni naturali del passo possono essere considerate un disturbo per i modelli cinematici predittivi. González et al. (2021) hanno investigato come rimuovere tali fluttuazioni per migliorare le capacità predittive. Hanno confrontato filtri in frequenza (Butterworth band-stop per eliminare la frequenza del passo) e filtri di Savitzky-Golay con un approccio basato sulla biomec-

canica del cammino. Quest’ultimo metodo proponeva di sfruttare l’orientamento del bacino, che durante il cammino normale ruota in controfase rispetto al tronco e alla direzione di spostamento, per correggere le oscillazioni dell’orientamento generale del soggetto [27]. Sebbene i filtri di Savitzky-Golay mostrassero un errore minore nella stima della posizione iniziale rispetto al segnale ideale, introducevano un errore maggiore nell’orientamento iniziale. Al contrario, il metodo basato sulla ricostruzione del segnale tramite l’orientamento del bacino, pur avendo un errore maggiore sulla posizione iniziale a causa di possibili derive, forniva un errore minore sull’orientamento iniziale, risultando in una riduzione significativa dell’errore di predizione a orizzonti temporali più lunghi (es. 1 secondo), dove l’errore sull’orientamento iniziale ha un impatto maggiore [27].

La scelta del metodo di correzione e filtraggio più appropriato dipende quindi strettamente dalla natura dei dati grezzi, dal tipo di rumore o artefatto presente, dagli obiettivi specifici dell’analisi e dalle caratteristiche del movimento che si intende preservare o enfatizzare. Nel contesto dell’analisi del cammino in pazienti schizofrenici, dove si potrebbero cercare alterazioni sottili nei pattern di movimento o nella loro variabilità, è fondamentale adottare tecniche di filtraggio che non eliminino queste caratteristiche potenzialmente informative. L’uso di filtri come quelli di Savitzky-Golay, con una parametrizzazione attenta, o di approcci adattivi che modulino l’intervento del filtro in base alla qualità del segnale, come proposto nella presente tesi, potrebbe rappresentare una strategia valida per bilanciare la necessità di ridurre il rumore con quella di preservare i dettagli significativi del movimento.

## 2.5 Il Contesto di Riferimento: il Progetto SPECTRA

Il presente lavoro di tesi si inserisce e costituisce una componente fondamentale del più ampio progetto di ricerca **SPECTRA** (*Supporting Schizophrenia Patients Care with Artificial Intelligence*). Finanziato nell’ambito del Piano Nazionale di Ripresa e Resilienza (PNRR) e dell’iniziativa europea NextGenerationEU, SPECTRA si pone l’ambizioso obiettivo di fornire agli psichiatri strumenti di analisi diagnostica avanzati per supportare la diagnosi precoce di una specifica e complessa forma di schizofrenia, la Schizofrenia Resistente al Trattamento (TRS) [29]. La diagnosi di

TRS è particolarmente sfidante e i pazienti che ne sono affetti incontrano notevoli difficoltà nel loro percorso clinico; una diagnosi tempestiva è quindi cruciale per migliorare la loro qualità della vita [29].

Il nucleo tecnologico del progetto si basa sull'impiego di tecniche all'avanguardia di Intelligenza Artificiale (AI) e, in particolare, di Intelligenza Artificiale Spiegabile (*Explainable Artificial Intelligence - XAI*). Quest'ultimo aspetto è di fondamentale importanza, poiché mira a superare la natura di "scatola nera" (black-box) di molti modelli di AI, fornendo ai clinici spiegazioni interpretabili per gli output del sistema e favorendo così la fiducia e l'adozione di tali strumenti nel processo decisionale clinico [29].

Per raggiungere i suoi obiettivi, SPECTRA adotta un approccio multimodale, integrando valutazioni standard con analisi basate su tecnologie informatiche per investigare le principali dimensioni della disorganizzazione sintomatologica nella schizofrenia. Il progetto è strutturato in quattro moduli di analisi principali, ciascuno dedicato a un dominio specifico [29]:

1. **Analisi del Linguaggio:** Si focalizza sull'elaborazione di dati linguistici, trascritti da sessioni cliniche, per identificare marcatori oggettivi associati ai disturbi del pensiero e del linguaggio (*Formal Thought Disorders*), utilizzando la scala TALD come riferimento e una pipeline che include modelli come Whisper per la trascrizione.
2. **Analisi del Movimento:** Indaga la presenza di anomalie motorie attraverso l'analisi dei dati di movimento. Questo modulo, in cui si colloca il presente lavoro di tesi, esamina la cinematica di anca, ginocchio, caviglia, collo e le metriche generali del cammino (lunghezza e larghezza del passo, velocità) per rilevare alterazioni e variabilità nella cadenza e nel ciclo del passo.
3. **Analisi della Struttura Cerebrale:** Impiega tecniche di neuroimaging, come la risonanza magnetica strutturale e funzionale (MRI/fMRI), per addestrare modelli di Deep Learning (es. CNN, Autoencoder, Transformer) capaci di identificare anomalie strutturali o funzionali del cervello associate alla TRS.

4. **Analisi della Risposta Emotiva:** Si occupa della valutazione delle risposte emotive dei pazienti attraverso l'analisi di segnali fisiologici.

Il modulo di **Analisi del Cammino**, in particolare, prevede una metodologia che parte dall'acquisizione del movimento tramite sistemi multi-camera, procede con il tracciamento dello scheletro (menzionando esplicitamente l'uso di tecnologie come MediaPipe), e si conclude con l'estrazione di pattern di movimento significativi tramite modelli avanzati come i Transformer. Per garantire l'interpretabilità dei risultati, il modulo prevede l'uso di tecniche XAI come SHAP (*SHapley Additive exPlanations*) per visualizzare come specifiche caratteristiche del cammino (es. postura, variabilità del passo) influenzino la decisione del modello di classificazione [29].

Il software sviluppato in questa tesi rappresenta quindi il "sotto-modulo" di estrazione, ricostruzione e correzione del dato cinematico, un passaggio fondamentale e abilitante per tutte le successive fasi di analisi dei pattern motori previste dal progetto SPECTRA.

# CAPITOLO 3

---

## Metodologia

---

### 3.1 Architettura e Setup Sperimentale

Il sistema sviluppato per questo lavoro di tesi è stato concepito per permettere un'analisi quantitativa, accurata e accessibile del cammino umano. L'architettura complessiva integra una componente hardware per l'acquisizione dei dati in un ambiente controllato e una componente software dedicata all'elaborazione offline, alla correzione e all'analisi dei dati di movimento. La scelta di un'elaborazione offline è stata dettata dalla necessità di applicare algoritmi computazionalmente intensivi per il filtraggio e la correzione, garantendo la massima qualità del dato finale, un requisito fondamentale per l'analisi clinica.

#### 3.1.1 Architettura Hardware e Setup di Acquisizione

L'infrastruttura hardware e il relativo setup di acquisizione sono stati definiti con l'obiettivo di creare un ambiente di misurazione standardizzato, a basso costo e facilmente replicabile, pur basandosi su una revisione critica delle alternative tecnologiche presenti in letteratura.

La scelta di un sensore RGB-D come strumento primario di acquisizione si allinea con la crescente tendenza a utilizzare tecnologie accessibili per l’analisi del movimento [19, 20]. In letteratura sono state esplorate diverse soluzioni. I sensori Microsoft Kinect (sia V1 a luce strutturata, sia V2 a tecnologia Time-of-Flight) sono stati ampiamente utilizzati, dimostrando buone capacità ma anche alcune limitazioni, come una performance ottimale in viste prettamente frontali a causa di possibili occlusioni degli arti in angolazioni oblique [30]. Il successore, Azure Kinect, ha mostrato una maggiore robustezza a diversi angoli di visuale grazie a modelli di stima della posa più avanzati, suggerendo i benefici di un posizionamento obliquo per visualizzare simultaneamente i piani sagittale e frontale [30].

Per questo lavoro, la scelta è ricaduta sulla famiglia di telecamere **Intel RealSense**, in quanto rappresentano una tecnologia matura, in attivo sviluppo e con un robusto supporto software tramite il loro SDK. Sebbene il software sviluppato sia compatibile con diversi modelli della serie D400 (come la D435), per l’acquisizione è stato impiegato il modello **Intel RealSense D455f**. Questa telecamera, le cui specifiche principali sono riassunte in Tabella 3.1, si basa sulla tecnologia di **stereoscopia attiva**. Questo approccio utilizza due sensori di profondità (aumentando la baseline rispetto ai modelli precedenti per una maggiore accuratezza a distanza) e un proiettore di pattern a infrarossi (IR) per calcolare una mappa di profondità densa e affidabile, anche in scene con scarsa texture visiva [21, 22].

**Tabella 3.1:** Caratteristiche tecniche principali della telecamera Intel RealSense D455f.

Caratteristica	Specifiche
Tecnologia di Profondità	Stereoscopia Attiva con proiettore IR
Range di Profondità Ideale	0.4 m - 6 m
Risoluzione Profondità	Fino a 1280 x 720
Frame Rate Profondità	Fino a 90 FPS
Risoluzione RGB	Fino a 1920 x 1080
Frame Rate RGB	Fino a 30 FPS
Campo Visivo (FoV) Profondità	87° x 58°

Il setup sperimentale, progettato per un’analisi del cammino in un ambiente

controllato che simula un contesto clinico o domiciliare, è stato così definito:

- **Strumentazione:** Un **tappeto di camminata** (treadmill) lungo 6 metri è stato utilizzato come percorso standardizzato. La telecamera **RealSense D455f** è stata montata su un treppiedi stabile a un'altezza di circa 1 metro dal pavimento.
- **Posizionamento:** La telecamera è stata posizionata **frontalmente** rispetto al tappeto, con un angolo di 0 gradi rispetto alla direzione di marcia del soggetto. Questa configurazione è supportata da studi che la indicano come efficace per catturare molteplici cicli del passo in modo completo [31]. Per garantire la massima coerenza tra le acquisizioni, la telecamera è stata allineata con precisione al centro del bordo iniziale del tappeto tramite un marker di riferimento.
- **Condizioni Ambientali:** Per assicurare una qualità ottimale dell'immagine RGB, fondamentale per la successiva stima della posa, tutte le acquisizioni sono state condotte in un ambiente illuminato da luce naturale diurna.



**Figura 3.1:** Configurazione del sistema di acquisizione video impiegato per la registrazione delle sessioni sperimentali con i pazienti. Sono visibili le videocamere, i cavalletti e il tappeto di camminata.

Per la fase di acquisizione dati, ovvero la registrazione del materiale video che costituisce l'input per la successiva pipeline di analisi, è stato impiegato lo strumento software **Intel RealSense Viewer**. Si tratta dell'applicazione ufficiale, pronta all'uso e disponibile per Windows, fornita da Intel insieme all'SDK RealSense 2.0. Questo software ha permesso di visualizzare in tempo reale i flussi video provenienti dalla

telecamera D455f, consentendo un controllo preciso sul posizionamento del soggetto e sull'inquadratura durante la preparazione dell'esperimento. La sua funzionalità principale, utilizzata in questo lavoro, è stata la capacità di registrare simultaneamente e in modo sincronizzato tutti i flussi di dati del sensore (in particolare, il flusso a colori RGB e la mappa di profondità) e di immagazzinarli in un unico file contenitore con estensione .bag. Questo formato è fondamentale per la metodologia offline adottata, poiché preserva non solo i dati video, ma anche tutti i metadati essenziali, come i timestamp di ogni frame e i parametri intrinseci della telecamera, garantendo la totale riproducibilità e l'accuratezza dell'elaborazione successiva.

### 3.1.2 Architettura Software e Pipeline di Elaborazione

La componente software, sviluppata in Python, costituisce il nucleo metodologico di questa tesi. Essa implementa una pipeline di elaborazione che automatizza il processo di analisi a partire dai file .bag registrati. Il software si avvale di librerie open-source di riferimento nel campo della computer vision e del calcolo scientifico, tra cui **pyrealsense2** per l'interfacciamento con i dati RealSense, **OpenCV** per la manipolazione di immagini e video, e **NumPy/SciPy** per le operazioni numeriche e di filtraggio.

La fase centrale di estrazione dei keypoint scheletrici è affidata alla libreria **MediaPipe** di Google, e in particolare al modello **Holistic**. La scelta di questo modello è motivata dalla sua capacità di rilevare un ricco set di 33 landmark corporei. Per garantire la massima qualità del dato grezzo in input alle fasi successive di correzione, il modello è stato configurato con parametri specifici volti a massimizzare l'accuratezza e la sensibilità. In particolare, la complessità del modello è stata impostata al valore massimo (2), mentre le soglie di confidenza per il rilevamento e il tracciamento sono state abbassate a 0.4 per aumentare la probabilità di individuare i keypoint anche in condizioni non ottimali. È stato inoltre disabilitato il livellamento (smoothing) temporale interno di MediaPipe (`smooth_landmarks=False`) per poter applicare un filtraggio personalizzato e controllato a posteriori. La Tabella 3.2 riassume la configurazione completa utilizzata.

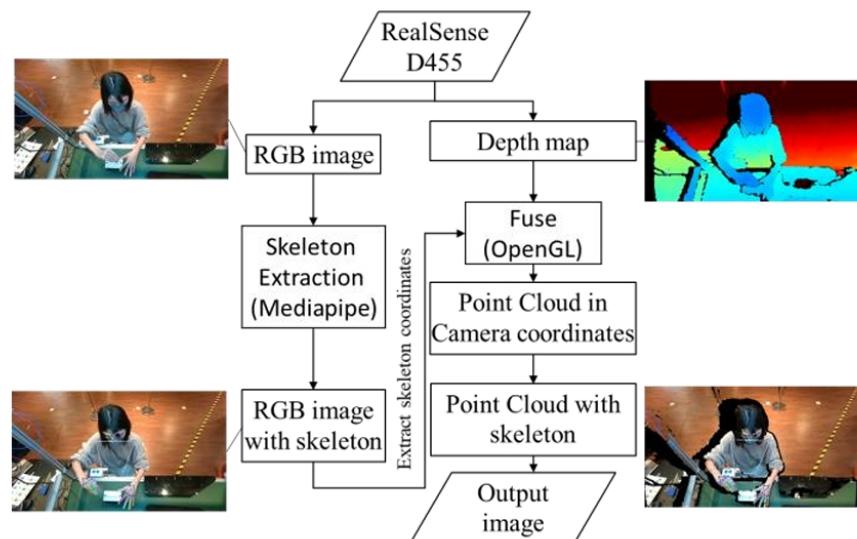
La configurazione adottata, come descritto in Tabella 3.2, è stata finalizzata a

**Tabella 3.2:** Configurazione dei parametri del modello MediaPipe Holistic.

Parametro	Valore Impostato
static_image_mode	True
model_complexity	2
smooth_landmarks	False
enable_segmentation	False
smooth_segmentation	False
refine_face_landmarks	True
min_detection_confidence	0.4
min_tracking_confidence	0.4

massimizzare la quantità e la qualità dei dati grezzi estratti, che costituiscono l'input per le successive fasi di ricostruzione 3D e di correzione.

La logica di elaborazione del software segue una pipeline strutturata, come schematizzato in Figura 3.2.


**Figura 3.2:** Riassunto della metodologia utilizzata.

Il processo può essere scomposto nei seguenti passaggi fondamentali:

- 1. Acquisizione dei Flussi Dati:** Il sistema legge il file .bag e separa i flussi video sincronizzati: l'immagine a colori (RGB) e la mappa di profondità (Depth Map).

2. **Estrazione dello Scheletro 2D:** L'immagine RGB di ogni frame viene processata da MediaPipe, che restituisce le coordinate 2D (in pixel) dei 33 landmark scheletrici. Questo risultato può essere visualizzato sovrapponendo lo scheletro all'immagine originale.
3. **Fusione Selettiva e Ricostruzione 3D:** Questo è il passaggio chiave della pipeline. Invece di processare l'intera e densa nuvola di punti generata dalla mappa di profondità, il sistema adotta un approccio di "fusione a singolo punto" [21]. Le coordinate 2D di ogni keypoint estratto da MediaPipe vengono utilizzate per interrogare puntualmente la mappa di profondità e recuperare il valore di distanza (asse Z) corrispondente. Questo metodo permette di ottenere le coordinate 3D di ogni keypoint in modo computazionalmente efficiente, scartando tutte le informazioni di profondità non pertinenti allo scheletro [21].
4. **Generazione dell'Output:** Le coordinate 3D dello scheletro così ottenute costituiscono il dato "grezzo" che verrà poi passato alle fasi successive di validazione e filtraggio. L'output finale del sistema può essere visualizzato come uno scheletro 3D all'interno della nuvola di punti o come scheletro 2D/3D proiettato sull'immagine originale, pronto per l'analisi cinematica.

La configurazione specifica di MediaPipe e questa pipeline di fusione selettiva sono state progettate per massimizzare la qualità del dato grezzo, che costituisce l'input per le fasi successive di correzione, la cui implementazione è descritta nel dettaglio più avanti.

### 3.1.3 Procedura di Acquisizione Dati

Per garantire la standardizzazione e la riproducibilità delle misurazioni, è stata definita una procedura di acquisizione rigorosa, applicata a ogni soggetto partecipante.

1. **Preparazione e Consenso:** In una fase preliminare, ogni paziente è stato informato dettagliatamente sugli obiettivi e le modalità dell'esperimento. È stato quindi raccolto il consenso informato scritto, nel pieno rispetto delle normative sulla privacy e sulla ricerca clinica.

**2. Esecuzione della Prova di Cammino:**

- (a) Il paziente veniva invitato a posizionarsi all'inizio del tappeto di camminata.
- (b) Veniva effettuata una breve fase di riscaldamento di circa 30 secondi, durante la quale il paziente poteva familiarizzare con il tappeto e l'ambiente.
- (c) Il paziente veniva istruito a camminare continuativamente per un totale di 4 minuti. La prova era suddivisa in due fasi consecutive: i primi **3 minuti** a un'andatura confortevole e auto-selezionata ("ritmo normale"), seguiti da **1 minuto** a un'andatura più sostenuta ("camminata accelerata"). Questa variazione di velocità ha lo scopo di sollecitare il sistema motorio e di osservare eventuali cambiamenti nei pattern di cammino al variare del ritmo.

**3. Conclusione e Verifica dei Dati:**

- (a) Al termine dei 4 minuti, la registrazione veniva interrotta e i dati acquisiti (file .bag) venivano immediatamente salvati e sottoposti a una verifica preliminare di integrità per assicurarsi che la registrazione fosse completa e priva di errori macroscopici.
- (b) Ai pazienti veniva fornito supporto e assistenza per concludere la sessione in modo confortevole.

### **3.1.4 Considerazioni Etiche**

L'intera procedura sperimentale, dalla raccolta del consenso alla gestione dei dati, è stata condotta nel rigoroso rispetto dei principi etici sanciti dalla Dichiarazione di Helsinki per la ricerca clinica. È stata garantita la massima tutela della privacy e dell'anonimato dei partecipanti e il pieno rispetto del loro benessere fisico e psicologico durante tutte le fasi dello studio.

# CAPITOLO 4

---

## Implementazione

---

### 4.1 Ambiente di Sviluppo e Tecnologie Software

Il sistema software, cuore di questo lavoro di tesi, è stato interamente sviluppato in linguaggio **Python**, scelto per la sua versatilità, la sintassi chiara e, soprattutto, per il suo vasto ecosistema di librerie open-source dedicate al calcolo scientifico e alla computer vision. Questa ricchezza di strumenti ha permesso di integrare soluzioni di state-of-the-art per ogni fase della pipeline di elaborazione, dalla lettura dei dati grezzi fino all'analisi e visualizzazione dei risultati, garantendo al contempo una notevole rapidità di sviluppo e prototipazione.

La realizzazione della pipeline si basa sull'interazione di diverse librerie fondamentali, ognuna con un ruolo specifico all'interno dell'architettura software:

- **OpenCV-Python (cv2):** È la libreria cardine per tutte le operazioni di manipolazione di immagini e video. In questo progetto, OpenCV è stata impiegata per leggere i singoli frame estratti dal file `.bag`, per eseguire trasformazioni geometriche come la rotazione delle immagini, per disegnare gli overlay grafici (come lo scheletro e le etichette testuali) e, infine, per assemblare e scrivere i file video di output in formato MP4.

- **NumPy:** Rappresenta le fondamenta per qualsiasi operazione numerica. Tutte le strutture dati relative a immagini (che vengono trattate come array multidimensionali), coordinate dei keypoint e matrici di trasformazione sono gestite tramite array NumPy. La sua efficienza nel calcolo vettoriale è essenziale per garantire prestazioni accettabili durante l'elaborazione di decine di migliaia di frame.
- **pyrealsense2:** È il wrapper Python ufficiale per l'**Intel RealSense SDK 2.0**. Questa libreria è lo strumento software fondamentale che funge da ponte tra l'hardware di acquisizione e la pipeline di elaborazione. È stata utilizzata per caricare e gestire i file di registrazione in formato `.bag`, permettendo di accedere in modo programmatico ai flussi di dati sincronizzati di colore (RGB) e di profondità. Inoltre, tramite `pyrealsense2`, è stato possibile estrarre i parametri intrinseci della telecamera (distanza focale, punto principale, baseline), dati indispensabili per la ricostruzione 3D metrica e per il calcolo dell'errore di profondità.
- **MediaPipe:** Il framework di Google per la percezione via machine learning è stato impiegato come motore per la stima della posa. Nello specifico, si è utilizzato il modello pre-addestrato **Holistic**, che fornisce una soluzione integrata per il rilevamento simultaneo di 33 landmark del corpo, del volto e delle mani. La sua ottimizzazione per l'efficienza ha permesso di ottenere stime della posa accurate per ogni frame in tempi contenuti, anche con la configurazione di massima complessità.
- **SciPy:** Questa libreria, complementare a NumPy, offre algoritmi più avanzati per il calcolo scientifico. In questo lavoro, è stata fondamentale per la fase di correzione del segnale, attraverso l'utilizzo della funzione che fornisce un'implementazione robusta ed efficiente del filtro di Savitzky-Golay.
- **Altre librerie di utilità:** Il software si avvale anche di librerie standard di Python come **json** per la serializzazione e il salvataggio dei dati strutturati dei keypoint, **argparse** per la gestione degli argomenti da riga di comando (rendendo lo script flessibile e configurabile), e **tqdm** per la visualizzazione

di barre di progresso, che migliorano l’esperienza utente durante le lunghe sessioni di elaborazione.

L’integrazione sinergica di queste tecnologie ha permesso di costruire una pipeline di elaborazione potente e modulare, in grado di trasformare i dati grezzi di un sensore RGB-D in un output quantitativo e corretto, pronto per l’analisi clinica.

## 4.2 Panoramica della Pipeline di Elaborazione

Questo capitolo descrive in dettaglio l’implementazione pratica della metodologia di analisi del movimento delineata nel capitolo precedente. Il cuore del sistema è una pipeline di elaborazione software, progettata per l’analisi offline di registrazioni video acquisite tramite telecamere RGB-D della serie Intel RealSense. L’approccio offline è stato scelto per consentire l’applicazione di algoritmi di correzione e validazione computazionalmente complessi, privilegiando la massima accuratezza e affidabilità del dato finale rispetto a requisiti di elaborazione in tempo reale.

La pipeline automatizza l’intero processo, dal caricamento del dato grezzo fino alla generazione di output quantitativi e visivi pronti per l’analisi. Il flusso di lavoro logico, che verrà approfondito nelle sezioni successive di questo capitolo, può essere sintetizzato nei seguenti passaggi fondamentali:

1. **Inizializzazione e Caricamento Dati:** Il processo ha inizio con il caricamento di un file di registrazione in formato .bag. Lo script inizializza i parametri di elaborazione, inclusa la configurazione del modello di stima della posa MediaPipe Holistic e le strutture dati necessarie per la gestione della cronologia dei keypoint e per il filtraggio.
2. **Estrazione Sincronizzata dei Frame:** Il sistema itera attraverso il file .bag, estraendo per ogni istante temporale la coppia di frame sincronizzati: l’immagine a colori (RGB) e la corrispondente mappa di profondità (Depth Map). Viene inoltre eseguita una rotazione di 90 gradi per allineare l’orientamento dei frame alle convenzioni di visualizzazione.

3. **Stima della Posa 2D tramite MediaPipe:** L'immagine RGB di ogni frame viene fornita in input al modello MediaPipe Holistic, che restituisce le coordinate 2D, in pixel, e lo score di visibilità per ciascuno dei 33 landmark corporei previsti dal modello.
4. **Ricostruzione 3D e Quantificazione dell'Incertezza:** Per ogni keypoint 2D rilevato, viene eseguita la ricostruzione tridimensionale. Questo passaggio chiave include:
  - **Fusione Robusta della Profondità:** La coordinata Z (profondità) non viene letta da un singolo pixel, ma è il risultato di una stima robusta che calcola la media dei valori di profondità validi in un intorno spaziale del keypoint e, in caso di dati mancanti, utilizza una cronologia temporale per fornire una stima coerente.
  - **Calcolo dell'Errore di Profondità:** Viene calcolato l'errore RMS teorico associato a ogni misura di profondità, fornendo una metrica di incertezza quantitativa per ogni punto 3D.
5. **Valutazione della Qualità del Dato Grezzo:** I dati 3D di ogni frame vengono analizzati per calcolare uno *score di affidabilità* ('reliability\_score'). Questa metrica, sviluppata ad-hoc, valuta la coerenza temporale del movimento (jitter) e penalizza la presenza di keypoint con profondità assente o bassa visibilità.
6. **Filtraggio Adattivo della Sequenza:** Una volta elaborati tutti i frame, l'intera sequenza temporale di keypoint 3D grezzi viene processata dall'algoritmo di correzione. Un filtro di Savitzky-Golay **adattivo** analizza la qualità di ogni traiettoria e applica un livellamento di intensità variabile, intervenendo in modo mirato solo sulle sezioni di segnale giudicate rumorose o instabili.
7. **Generazione degli Output:** La pipeline si conclude con il salvataggio dei risultati. Vengono generati file in formato JSON contenenti le coordinate 3D, sia grezze che filtrate, arricchite con tutte le metriche calcolate (errore di profondità, score di affidabilità). Vengono inoltre prodotti output visivi, come video con la sovrapposizione dello scheletro e filmati di confronto, per una validazione qualitativa dell'efficacia del processo di correzione.

Le sezioni che seguono analizzeranno in dettaglio ciascuna di queste fasi, illustrando la logica sottostante tramite pseudocodice e discutendo le scelte implementative che caratterizzano l’originalità di questo lavoro.

## 4.3 Dall’Immagine ai Dati Grezzi: Estrazione e Ricostruzione 3D

Una volta inizializzato il sistema e caricato il file `.bag`, la pipeline entra nel suo ciclo di elaborazione principale, il cui scopo è trasformare ogni coppia di frame (RGB e profondità) in un insieme strutturato di coordinate scheletriche tridimensionali. Questa sezione analizza in dettaglio le funzioni e gli algoritmi implementati per realizzare questo passaggio cruciale.

### 4.3.1 Stima della Posa 2D tramite MediaPipe

Il primo passo all’interno del ciclo di elaborazione è l’identificazione dei landmark corporei nel dominio 2D dell’immagine a colori. Questo compito è affidato alla funzione `get_significant_keypoints()`, che funge da interfaccia con il modello MediaPipe Holistic.

Come illustrato nello Pseudocodice 1, la funzione riceve in input l’oggetto restituito da MediaPipe e l’immagine originale. Per prima cosa, verifica la presenza di landmark validi. Successivamente, itera su una lista predefinita di 33 indici, che rappresentano i keypoint corporei di interesse. Le coordinate di ciascun landmark, fornite da MediaPipe in formato normalizzato (tra 0.0 e 1.0), vengono convertite in coordinate pixel assolute, moltiplicandole per le dimensioni dell’immagine. Per allineare l’orientamento della posa a una visualizzazione più consona per l’analisi del cammino (con il soggetto che si muove orizzontalmente), le coordinate vengono ruotate di 90 gradi in senso orario. Infine, la funzione restituisce una lista di oggetti, ciascuno contenente le coordinate pixel ruotate e lo score di visibilità originale, pronti per la successiva fase di ricostruzione 3D.

---

**Algorithm 1** Estrazione dei Keypoint 2D Significativi

---

```

1: function GETSIGNIFICANTKEYPOINTS(mediaPipeResults, image)
2:   landmarkList  $\leftarrow []$ 
3:   if mediaPipeResults.pose_landmarks non è nullo then
4:     height, width  $\leftarrow$  Dimensioni di image
5:     for all indice i in INDICI_SIGNIFICATIVI do
6:       landmark  $\leftarrow$  mediaPipeResults.pose_landmarks.landmark[i]
7:       x_pixel  $\leftarrow$  landmark.x  $\times$  width
8:       y_pixel  $\leftarrow$  landmark.y  $\times$  height
9:       x_rot, y_rot  $\leftarrow$  ROTATEPOINT(x_pixel, y_pixel, width, height)  $\triangleright$  Ruota di
   90°
10:      Crea un nuovo oggetto lmk_obj con x_rot, y_rot, landmark.visibility
11:      Aggiungi lmk_obj a landmarkList
12:    end for
13:  end if
14:  return landmarkList
15: end function

```

---

### 4.3.2 Ricostruzione 3D e Gestione dell’Incertezza

Una volta ottenute le coordinate 2D dei keypoint, il passo successivo è la loro proiezione nello spazio tridimensionale attraverso la fusione con i dati della mappa di profondità. Questo processo è gestito da due funzioni principali, volte non solo a calcolare la coordinata Z, ma anche a garantirne la robustezza e a quantificarne l’incertezza.

#### Stima Robusta della Profondità

Per affrontare l’intrinseca rumorosità dei dati di profondità dei sensori low-cost, è stata implementata la funzione `get_depth_around_point()`. Invece di leggere il valore di un singolo pixel, che potrebbe essere errato o nullo, questa funzione adotta un approccio più resiliente, come descritto nello Pseudocodice 2.

La funzione implementa una logica a due livelli. Prima tenta una **media spaziale**: analizza una finestra di 5x5 pixel attorno al keypoint sulla mappa di profondità e calcola la media di tutti i valori validi (non nulli) trovati. Questo mitiga l’effetto di fluttuazioni isolate. Se questo approccio fallisce (ad esempio per occlusioni), scatta il **fallback temporale**: il sistema recupera gli ultimi 5 valori di profondità validi registrati per quello stesso keypoint nei frame precedenti e ne calcola la media. Questa cronologia temporale, mantenuta in una struttura dati globale, assicura la continuità della traiettoria 3D anche in caso di perdite di dati momentanee, restituendo un valore nullo solo come ultima risorsa.

#### Quantificazione dell’Incertezza della Misura

Oltre al valore di profondità, è fondamentale stimarne l’affidabilità. La funzione `calculate_depth_error()` si occupa di questo, fornendo una metrica di incertezza quantitativa per ogni keypoint.

Come mostrato nello Pseudocodice 3, la funzione implementa la formula teorica dell’errore RMS per la profondità in un sistema a stereoscopia attiva. Utilizza la distanza misurata (`depthValue`), la baseline della telecamera (distanza tra i sensori IR) e la sua lunghezza focale. Questi parametri intrinseci vengono letti direttamente dai metadati del file `.bag` all’avvio della pipeline, garantendo un calcolo specifico

---

**Algorithm 2** Stima Robusta della Profondità

---

```
1: function GETDEPTHAROUNDPOINT(depthFrame, x, y, keypointID)
2:   depth_values  $\leftarrow$  []
3:   Inverti la rotazione di (x, y) per ottenere le coordinate originali (x_orig, y_orig)
4:   for all pixel nell’intorno 5x5 di (x_orig, y_orig) do
5:     depth  $\leftarrow$  Valore di profondità del pixel da depthFrame
6:     if depth > 0 then
7:       Aggiungi depth a depth_values
8:     end if
9:   end for
10:  if depth_values non è vuoto then
11:    mean_depth  $\leftarrow$  Media di depth_values
12:    Aggiorna la cronologia keypoint_history[keypointID] con mean_depth
13:    return mean_depth
14:  else if keypoint_history[keypointID] esiste ed è non vuoto then
15:    mean_history_depth  $\leftarrow$  Media della cronologia per keypointID
16:    return mean_history_depth                                 $\triangleright$  Fallback temporale
17:  else
18:    return 0.0                                               $\triangleright$  Nessun dato valido disponibile
19:  end if
20: end function
```

---

---

**Algorithm 3** Calcolo dell’Errore di Profondità

---

```

1: function CALCULATEDEPTHERROR(depthValue, cameraInfo)
2:    $D_{mm} \leftarrow depthValue \times 1000$ 
3:    $B \leftarrow cameraInfo.baseline$                                  $\triangleright$  in mm
4:    $f \leftarrow cameraInfo.focal\_length$                              $\triangleright$  in pixel
5:    $e_{subpixel} \leftarrow 0.08$                                       $\triangleright$  Errore subpixel tipico
6:   if  $f \times B == 0$  then
7:     return 0.0
8:   else
9:      $error\_rms \leftarrow (D_{mm}^2 \times e_{subpixel}) / (f \times B)$ 
10:    return  $error\_rms$ 
11:   end if
12: end function

```

---

per il sensore utilizzato. Il risultato è un valore di errore in millimetri che viene associato a ogni keypoint, permettendo di pesare o scartare le stime meno affidabili nelle successive fasi di analisi, aggiungendo un livello di robustezza fondamentale all’intero processo.

L’implementazione della funzione `calculate_depth_error()` si basa sulla formula teorica standard per la stima dell’errore di profondità nei sistemi a visione stereoscopica, come le telecamere Intel RealSense. La formula, espressa matematicamente, è la seguente:

$$E_{RMS\_depth} = \frac{D^2 \times e_{subpixel}}{f_{px} \times B} \quad (4.3.1)$$

dove la lunghezza focale in pixel,  $f_{px}$ , è derivata dalla risoluzione orizzontale del sensore,  $W_{px}$ , e dal campo visivo orizzontale,  $\theta_{HFOV}$ :

$$f_{px} = \frac{W_{px}/2}{\tan(\theta_{HFOV}/2)} \quad (4.3.2)$$

L’Equazione 4.3.1 è di fondamentale importanza perché quantifica come l’incertezza sulla misura di profondità non sia costante, ma vari in funzione della geometria del sistema e della distanza dell’oggetto. Analizziamo i singoli termini:

- $D^2$  (**Distanza al quadrato**): È il termine più influente. L’errore di profondità aumenta con il quadrato della distanza dalla telecamera. Ciò significa che la precisione del sensore degrada rapidamente man mano che il soggetto si allontana. Ad esempio, un punto a 2 metri di distanza avrà un errore teorico quattro volte superiore rispetto a un punto a 1 metro, a parità di altre condizioni. Questa relazione quadratica sottolinea l’importanza di eseguire le acquisizioni all’interno del range operativo ottimale del sensore per garantire la minima incertezza possibile.
- $B$  (**Baseline**): Rappresenta la distanza fisica, in millimetri, tra i centri ottici dei due sensori a infrarossi della telecamera stereo. Trovandosi al denominatore, una baseline maggiore contribuisce a ridurre l’errore di profondità. Una maggiore separazione tra i “due occhi” del sistema permette infatti una migliore triangolazione, specialmente per oggetti più lontani. Modelli come il D455 sono dotati di una baseline più ampia rispetto a modelli precedenti, migliorando le loro prestazioni a distanze maggiori.
- $f_{px}$  (**Lunghezza Focale in pixel**): Questo valore, calcolato tramite l’Equazione 4.3.2, mette in relazione il campo visivo della telecamera con la risoluzione del suo sensore. Anch’esso si trova al denominatore, indicando che una maggiore lunghezza focale (equivalente a un campo visivo più stretto, o un effetto “zoom”) contribuisce a ridurre l’errore di profondità.
- $e_{subpixel}$  (**Errore Subpixel**): Questo termine rappresenta il limite di precisione intrinseco dell’algoritmo di corrispondenza stereo (stereo matching), ovvero la sua capacità di abbinare i punti corrispondenti tra l’immagine del sensore sinistro e quella del sensore destro. È un valore piccolo, tipicamente inferiore a 0.1 pixel, che quantifica l’incertezza minima del sistema. Nel nostro caso, è stato utilizzato un valore tipico di 0.08, come riportato nelle specifiche di sistemi analoghi.

Nell’implementazione software, tutti i parametri geometrici della telecamera ( $B$ ,  $W_{px}$ ,  $\theta_{HFOV}$ ) vengono letti direttamente dai metadati del file .bag, assicurando che il calcolo dell’errore sia specifico per la sessione di acquisizione. Associare questo

valore di errore calcolato a ogni singolo keypoint 3D arricchisce il dato grezzo di una metrica di affidabilità fondamentale, che può essere sfruttata nelle fasi successive per pesare le osservazioni o per scartare quelle con un'incertezza troppo elevata.

## 4.4 Valutazione della Qualità del Dato: Score di Affidabilità Temporale

Una volta ottenuta la sequenza di keypoint 3D grezzi, è fondamentale disporre di una metrica oggettiva per valutarne la qualità. A questo scopo, è stata sviluppata la funzione `calculate_jitter_score()`, il cui obiettivo è assegnare a ogni frame uno score di affidabilità quantitativo. Questo score non si limita a valutare la presenza dei keypoint, ma analizza la coerenza temporale del loro movimento, un fattore cruciale per l'analisi cinematica. Uno score elevato indica un movimento fluido e verosimile, mentre uno score basso segnala la presenza di jitter, instabilità o dati mancanti.

Lo Pseudocodice 4 illustra la logica della funzione. Per ogni keypoint presente nel frame corrente, l'algoritmo calcola la distanza euclidea 3D rispetto alla sua posizione nel frame precedente. Questa distanza, che rappresenta una stima della sua velocità istantanea, viene confrontata con la media delle distanze registrate in una finestra temporale mobile (di default, 15 frame). Il rapporto tra la distanza attuale e la media storica (*movement\_ratio*) è l'indicatore chiave:

- Un rapporto vicino a 1 indica che il keypoint si sta muovendo con una velocità costante e prevedibile, indice di un movimento fluido, a cui viene assegnato uno score massimo (1.0).
- Un rapporto molto maggiore di 1 suggerisce un "salto" improvviso e anomalo (jitter), che viene pesantemente penalizzato con uno score vicino a zero.
- Rapporti intermedi ricevono score proporzionali, premiando la coerenza del movimento.

Inoltre, lo score viene penalizzato da fattori che indicano una bassa qualità del dato grezzo. Una forte penalità viene applicata se il keypoint ha una stima di profondità

**Algorithm 4** Calcolo dello Score di Affidabilità del Frame (Jitter Score)

---

```

1: function CALCULATEJITTERSCORE(currentKeypoints, frameNum)
2:   Init frameScores  $\leftarrow []$ , depthPenaltyApplied  $\leftarrow$  Falso, visibilityPenalty  $\leftarrow 0.0$ 
3:   if non ci sono previousKeypoints then
4:     Salva currentKeypoints come previousKeypoints
5:     return (0.5, score individuali a 0.5)
6:   end if
7:   for all keypoint in currentKeypoints do
8:     kpID  $\leftarrow$  keypoint.id
9:     kpScore  $\leftarrow 0.5$                                       $\triangleright$  Score di base
10:    Applica penalità iniziale se keypoint.z è 0.0 o keypoint.visibility è bassa
11:    if kpID è presente in previousKeypoints then
12:      dist  $\leftarrow$  Distanza 3D tra keypoint corrente e precedente
13:      Aggiorna la finestra temporale current_window[kpID] con dist
14:      if la finestra per kpID ha abbastanza dati then
15:        avg_dist  $\leftarrow$  Media delle distanze passate nella finestra
16:        if avg_dist  $> 0 then
17:          movement_ratio  $\leftarrow$  dist / avg_dist
18:          if movement_ratio è molto alto ( $> 3.0$ ) then
19:            kpScore  $\leftarrow 0.0$                                       $\triangleright$  Jitter elevato
20:            else if movement_ratio è vicino a 1.0 then
21:              kpScore  $\leftarrow 1.0$                                       $\triangleright$  Movimento fluido
22:            else
23:              Assegna score intermedi...
24:            end if
25:          end if
26:        end if
27:      end if
28:      Applica penalità finali a kpScore
29:      Aggiungi kpScore a frameScores
30:    end for...
31: end function$ 
```

---

Calcolo dello Score di Affidabilità del Frame (Jitter Score)

---

```

1: function CALCULATEJITTERSCORE(currentKeypoints, frameNum)
    ...
2:     overallScore ← Media di frameScores
3:     Applica penalità globali a overallScore basate su profondità e visibilità
4:     Aggiorna previousKeypoints con currentKeypoints
5:     return (overallScore, score individuali)
6: end function
```

---

nulla ( $z=0.0$ ) o se il suo score di visibilità, fornito da MediaPipe, è basso. Lo score complessivo del frame è calcolato come media degli score dei singoli keypoint, a sua volta ulteriormente penalizzata se molti keypoint nel frame risultano di bassa qualità. Questa metrica fornisce quindi uno strumento quantitativo fondamentale per identificare i segmenti del video che necessitano di maggiore attenzione e correzione.

## 4.5 Correzione del Segnale: Filtraggio Adattivo Savitzky-Golay

La fase finale dell'elaborazione è la correzione delle traiettorie dei keypoint 3D per rimuovere il jitter e gli artefatti di movimento. Un semplice filtraggio passa-basso, se applicato uniformemente, rischierebbe di smussare eccessivamente il segnale (*oversmoothing*), eliminando non solo il rumore ma anche le variazioni fisiologiche del movimento, che sono di particolare interesse nell'analisi di andature patologiche. Per ovviare a questo problema, è stato implementato un approccio di filtraggio più sofisticato e intelligente: un **filtro di Savitzky-Golay adattivo**, gestito dalla funzione `apply_adaptive_savgol_filter()`.

Il filtro di Savitzky-Golay opera effettuando una regressione polinomiale locale su una finestra di dati, preservando la forma del segnale meglio di una semplice media mobile. La caratteristica innovativa di questa implementazione risiede nella sua natura "adattiva" e "condizionale".

Come descritto nello Pseudocodice 5, l'algoritmo non applica un filtro uniforme a tutti i dati. Invece:

---

**Algorithm 5** Filtraggio Adattivo Savitzky-Golay

---

```

1: function APPLYADAPTIVESAVGOLFILTER(keypointsSequence, baseParams)
2:   filteredSequence  $\leftarrow$  Copia di keypointsSequence
3:   for all ID keypoint kpID in keypointsSequence do
4:     Estrai la traiettoria temporale per kpID (coordinate e visibilità)
5:     Calcola le metriche di qualità della traiettoria: mean_visibility e avg_variance
6:     adaptive_window  $\leftarrow$  baseParams.window
7:     adaptive_poly  $\leftarrow$  baseParams.poly_order
8:     adaptive_threshold  $\leftarrow$  baseParams.threshold
9:     if la traiettoria è molto problematica (avg_variance alta E mean_visibility
bassa) then
10:      Aumenta adaptive_window (es. 21)            $\triangleright$  Filtro più forte
11:      Diminuisci adaptive_poly (es. 1)            $\triangleright$  Filtro più rigido
12:    else if la traiettoria è problematica then
13:      Aumenta adaptive_window (es. 17)
14:    end if
15:    for all coordinata coord in (x, y, z) do
16:      trajectory  $\leftarrow$  Traiettoria della coordinata coord per kpID
17:      Calcola le variazioni inter-frame (diffs) di trajectory
18:      if  $\max(\text{diffs}) > \text{adaptive\_threshold}$  then            $\triangleright$  Applica filtro solo se
necessario
19:        filtered_trajectory  $\leftarrow$  SAVITZKYGOLAYFILTER(trajectory, adaptive_window, adaptive_poly)
20:        Aggiorna la coordinata coord in filteredSequence con filtered_trajectory
21:      end if
22:    end for
23:  end for
24:  return filteredSequence
25: end function

```

---

1. **Analizza ogni traiettoria individuale:** Per ogni keypoint, il sistema analizza la sua intera traiettoria temporale, calcolando metriche di qualità come la visibilità media fornita da MediaPipe e la varianza del movimento (una misura del jitter).
2. **Adatta i parametri del filtro:** In base a queste metriche, il sistema "diagnostica" la qualità della traiettoria. Se una traiettoria è giudicata "problematica" (ad esempio, ha una visibilità media bassa e un'alta varianza), i parametri del filtro vengono resi più aggressivi, aumentando la dimensione della finestra di livellamento e riducendo l'ordine del polinomio per una smussatura più intensa. Traiettorie di buona qualità mantengono parametri di filtraggio più leggeri.
3. **Applica il filtro in modo condizionale:** Il filtraggio non viene applicato indiscriminatamente. Per ciascuna coordinata ( $x, y, z$ ) di ogni keypoint, il filtro viene attivato solo se le variazioni inter-frame superano una soglia predefinita. Questo passaggio è fondamentale: se un movimento è già fluido e al di sotto della soglia di jitter, viene lasciato inalterato, preservando così la naturalezza del gesto. Vengono corrette solo le componenti del segnale che mostrano un'instabilità eccessiva.

Questo approccio intelligente assicura che il rumore venga rimosso efficacemente, minimizzando al contempo il rischio di alterare le caratteristiche fisiologiche del movimento, un aspetto di massima importanza per un'analisi clinica affidabile.

## 4.6 Gestione dell'Output e Visualizzazione dei Risultati

La fase finale della pipeline di elaborazione è dedicata alla persistenza e alla presentazione dei risultati. Una volta che i dati grezzi sono stati estratti e le traiettorie corrette tramite il filtraggio adattivo, il sistema genera un insieme completo di output, sia quantitativi che qualitativi. Questi output sono progettati per facilitare non solo la successiva analisi cinematica dei parametri del cammino, ma anche per permettere una validazione visiva immediata dell'efficacia degli algoritmi di correzione implementati.

### 4.6.1 Salvataggio dei Dati Quantitativi in Formato JSON

Il principale risultato quantitativo del software è una serie di file strutturati in formato JSON (JavaScript Object Notation), scelto per la sua leggibilità e la facile interoperabilità con la maggior parte dei linguaggi di programmazione e software di analisi dati. Il sistema produce due tipologie di output JSON per garantire sia un'analisi dettagliata per singolo frame, sia un caricamento agevole dell'intera sequenza temporale.

#### Output per Singolo Frame

Per ogni frame processato del video, viene generato un file JSON dedicato, salvato in sottocartelle distinte a seconda che si tratti di dati grezzi (`keypoints/`) o filtrati (`keypoints_filtered/`). Questa granularità permette un'ispezione puntuale e dettagliata. La struttura di ogni file JSON per singolo frame è un oggetto che contiene le seguenti chiavi principali:

- **"frame"**: Un valore intero che indica l'indice progressivo del frame all'interno della sequenza (es. 0).
- **"reliability\_score"**: Un valore decimale che rappresenta lo score di affidabilità temporale globale calcolato per il frame, come descritto nella Sezione 4.4.
- **"keypoint\_scores"**: Un oggetto contenente gli score di affidabilità individuali per ogni keypoint rilevato, dove la chiave è l'ID del keypoint (es. "13") e il valore è il suo score (es. 0 . 8).
- **"body"**: Un array di oggetti, dove ogni oggetto rappresenta un singolo keypoint scheletrico e contiene i seguenti campi:
  - **"x"**, **"y"**: Le coordinate 2D del keypoint in pixel, già ruotate.
  - **"z"**: La coordinata di profondità in metri, ottenuta tramite la fusione robusta. Può avere valore 0 . 0 se la stima non è riuscita.
  - **"depth\_error"**: L'errore RMS di profondità stimato in millimetri.

- **"visibility"**: Lo score di visibilità (da 0.0 a 1.0) fornito direttamente da MediaPipe.
- **"id"**: L'indice numerico (da 0 a 32) del landmark secondo la convenzione di MediaPipe.
- **"filtered"**: Un valore booleano ('true' o 'false') che indica se i dati contenuti nel file sono il risultato del processo di filtraggio.

### Output Aggregato per l'Intera Sequenza

Oltre ai file per singolo frame, il sistema produce due file JSON globali:

- `video_keypoints.json`
- `video_keypoints_filtered.json`

Questi file sono progettati per facilitare l'importazione e l'analisi dell'intera sequenza di movimento in strumenti come Pandas o MATLAB. La loro struttura consiste in un unico array JSON di primo livello. Ogni elemento di questo array rappresenta un frame ed è, a sua volta, un array contenente gli oggetti keypoint per quel frame (con la stessa struttura "x", "y", "z", ... descritta sopra). È importante notare che, per mantenere un formato snello e ottimizzato per l'analisi delle traiettorie, questi file aggregati contengono esclusivamente la sequenza degli array di keypoint e omettono i metadati per singolo frame come `"reliability_score"`.

### 4.6.2 Generazione degli Output Visivi per la Validazione

Per consentire una valutazione qualitativa e intuitiva dell'effetto del processo di correzione, la pipeline genera diversi output video a 60 FPS. Questa fase si avvale di funzioni dedicate per il disegno dello scheletro e la composizione dei video finali, utilizzando la libreria OpenCV.

I principali video generati sono:

- **Video con Dati Grezzi (`skeleton_animation.mp4`)**: Mostra lo scheletro ricostruito direttamente dai dati non filtrati. Serve come baseline per valutare la qualità del dato originale.

**Algorithm 6** Disegno dello Scheletro su Immagine

---

```
1: function DRAWSKELETON(image, landmarks)
2:   Definisci CONNECTIONS come lista di coppie di indici (es. (11, 13) per spalla-gomito)
3:   for all landmark lm in landmarks do
4:     Disegna un cerchio su image alle coordinate (lm.x, lm.y)
5:   end for
6:   for all connessione c in CONNECTIONS do
7:     start_point  $\leftarrow$  landmarks[c[0]]
8:     end_point  $\leftarrow$  landmarks[c[1]]
9:     Disegna una linea su image da start_point a end_point
10:  end for
11:  return image
12: end function
```

---

- **Video con Dati Filtrati (`skeleton_animation_filtered.mp4`):** Mostra lo scheletro dopo l'applicazione del filtro adattivo Savitzky-Golay. Il confronto visivo con il video originale permette di apprezzare la riduzione del jitter.
- **Video di Confronto Affiancato (`skeleton_comparison.mp4`):** Crea un video a larghezza doppia, mostrando fianco a fianco il video originale e quello filtrato, per una comparazione diretta ed efficace dell'impatto della correzione.
- **Video con Scheletri Sovrapposti (`skeleton_overlay.mp4`):** Generato optionalmente, disegna entrambi gli scheletri (grezzo e filtrato) sulla stessa immagine con colori diversi, evidenziando come la traiettoria filtrata stabilizzi quella originale.

Questo insieme strutturato di output, sia numerici che visivi, fornisce un quadro completo e trasparente del lavoro svolto, abilitando sia un'analisi quantitativa rigorosa dei dati cinematici, sia una validazione qualitativa immediata dei metodi di correzione sviluppati.

# CAPITOLO 5

---

## Risultati Sperimentali

---

### 5.1 Presentazione del Caso di Studio e Protocollo Applicato

Questo capitolo presenta i risultati ottenuti dall'applicazione della pipeline software, descritta in dettaglio nei capitoli precedenti, a un'acquisizione sperimentale. L'obiettivo è validare l'efficacia del sistema di elaborazione e correzione attraverso un'analisi sia qualitativa che quantitativa, dimostrando la sua capacità di produrre dati cinematici affidabili a partire da un sensore RGB-D a basso costo.

Il software sviluppato si inserisce nel contesto del progetto SPECTRA, che ha previsto l'acquisizione di dati da numerosi partecipanti. Tuttavia, per le finalità di questa tesi, si è scelto di adottare un approccio basato sull'analisi approfondita di un **singolo caso di studio rappresentativo**. Questa scelta metodologica permette di illustrare in modo granulare e sequenziale l'intero flusso di lavoro, dalla gestione del dato grezzo fino al risultato filtrato, evidenziando in dettaglio l'impatto di ogni fase della pipeline.

Il caso selezionato per questa analisi è quello di una paziente identificata con l'identificativo anonimizzato **ID\_1\_INIT\_E\_A\_GENDER\_F AGE\_27 CLASS\_TRS\_01**.

Sulla base della struttura dell’identificativo, il partecipante è una paziente di 27 anni con diagnosi di Schizofrenia Resistente al Trattamento (TRS).

I dati di questa paziente sono stati acquisiti seguendo il protocollo standardizzato descritto nella Sezione 3.1.3. Per ricapitolare, la sessione ha previsto una prova di deambulazione su tappeto della durata complessiva di 4 minuti (3 minuti di camminata a ritmo confortevole, seguiti da 1 minuto di camminata a ritmo sostenuto). L’acquisizione è stata effettuata con una telecamera Intel RealSense D455f in posizione frontale, e i flussi di dati sono stati salvati in un singolo file in formato .bag. Questo file costituisce l’input unico per l’intera analisi presentata nelle sezioni seguenti, che mostreranno i risultati dell’applicazione della pipeline di correzione a segmenti rappresentativi di questa registrazione.

## 5.2 Validazione Qualitativa della Correzione del Segnale

La prima fase della validazione dei risultati consiste in un’analisi qualitativa, il cui scopo è dimostrare visivamente l’impatto positivo della pipeline di elaborazione sulla qualità dei dati di movimento. Questa analisi si concentra nel dimostrare non solo l’efficacia del filtro nel ridurre il rumore, ma soprattutto la sua capacità **adattiva**: intervenire in modo significativo solo quando necessario, preservando i dati di buona qualità. A tale scopo, vengono analizzati e confrontati due frame rappresentativi estratti dalla registrazione del caso di studio: un frame a basso jitter (frame 142) e un frame ad alto jitter (frame 185).

### 5.2.1 Caso 1: Analisi di un Frame a Basso Jitter (Frame 142)

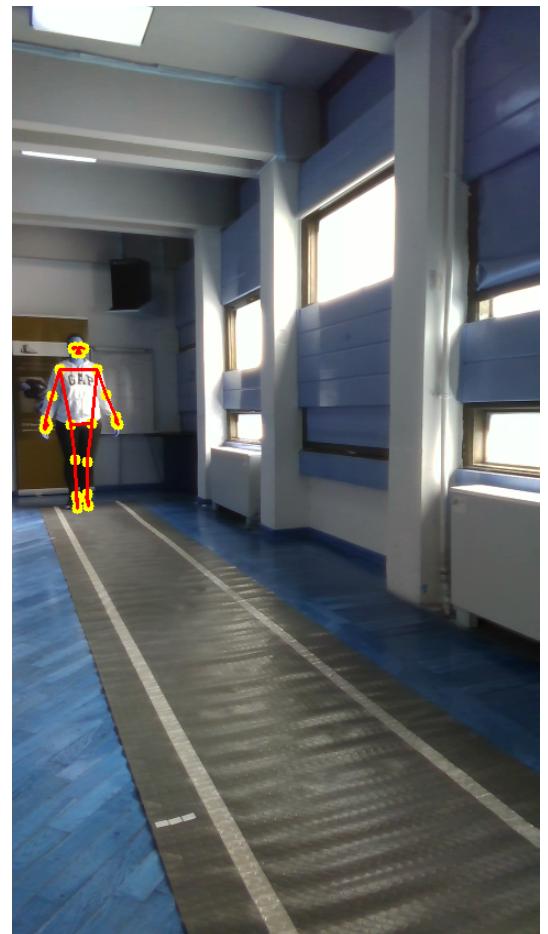
Il primo esempio riguarda il frame 142, un istante in cui il dato grezzo catturato dal sistema presentava già una buona qualità intrinseca. Come si può osservare dal confronto visivo in Figura 5.1, lo scheletro ricostruito dai dati grezzi (a) e quello ottenuto dopo il filtraggio (b) sono quasi indistinguibili.

Questa osservazione qualitativa è confermata dai dati numerici.

Il file `keypoints_0142.json` riporta uno "reliability\_score" complessivo



(a) Scheletro con dati grezzi.



(b) Scheletro con dati filtrati.

**Figura 5.1:** Confronto visivo per il frame 142 (basso jitter). Non si notano differenze significative tra lo scheletro grezzo (a) e quello filtrato (b), indicando che il filtro ha correttamente preservato il dato originale.

per il frame pari a **0.62**, un valore relativamente alto che segnala una buona coerenza temporale del movimento. Di conseguenza, l'algoritmo di filtraggio adattivo è intervenuto in modo molto conservativo. Analizzando, ad esempio, le coordinate del fianco sinistro (ID 23), si nota come i valori rimangano pressoché invariati:

- **Dato Grezzo:** "x": 72, "y": 379, "z": 8.141...
- **Dato Filtrato:** "x": 72.06..., "y": 379, "z": 8.141...

Questo dimostra la capacità del sistema di riconoscere un segnale di buona qualità e di non applicare un livellamento eccessivo (*oversmoothing*), preservando così l'integrità del dato originale.

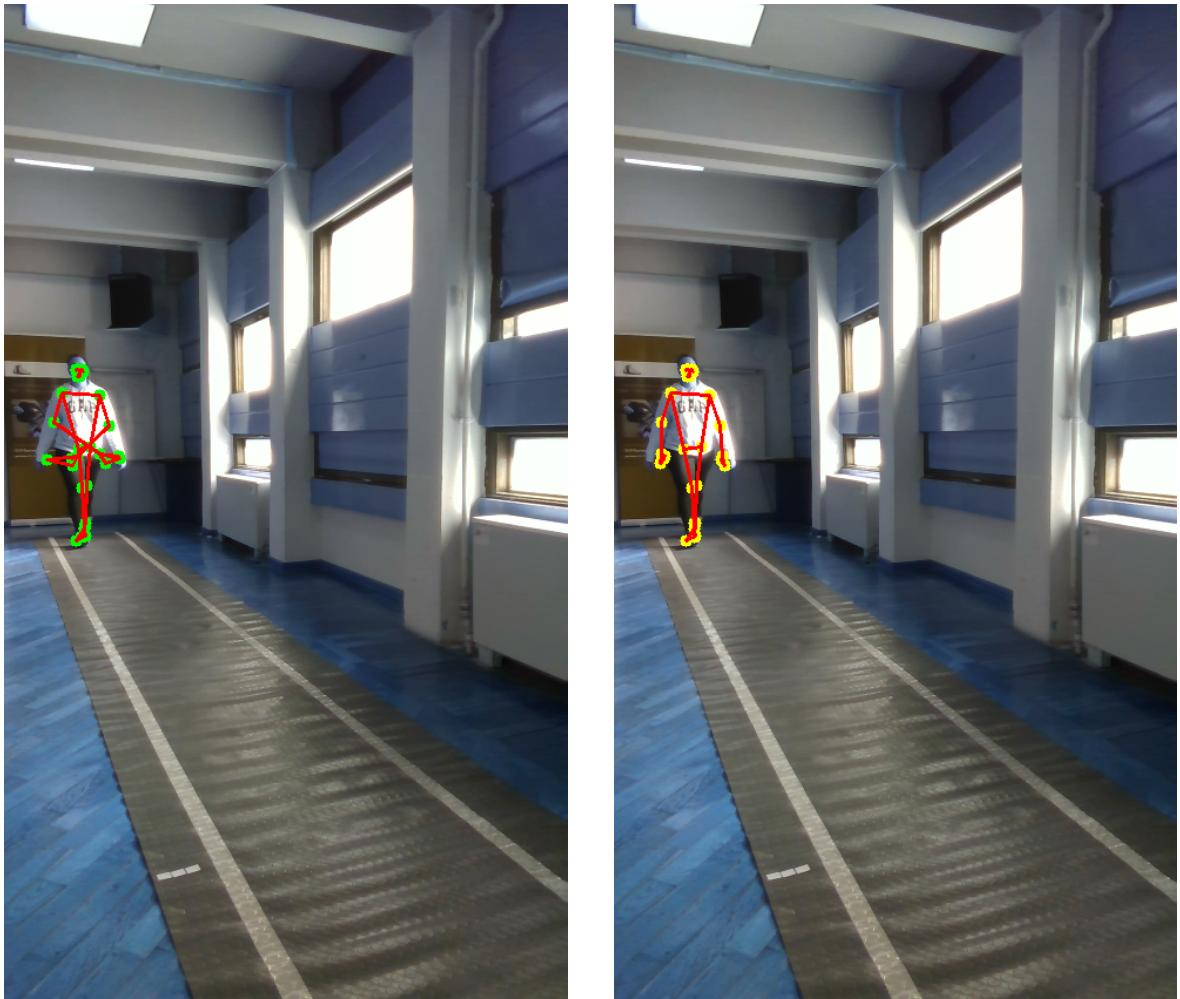
### 5.2.2 Caso 2: Analisi di un Frame ad Alto Jitter (Frame 185)

Il secondo esempio, relativo al frame 185, illustra il comportamento del sistema in presenza di dati rumorosi. In questo caso, il "reliability\_score" calcolato per il frame grezzo era significativamente più basso, pari a **0.51**, segnalando un'elevata instabilità. L'impatto del filtro è visivamente evidente in Figura 5.2.

Lo scheletro grezzo (a) mostra una posa palesemente innaturale, con numerosi keypoint posizionati in modo errato. Il filtro adattivo, attivato dalla bassa qualità del segnale, ha corretto in modo decisivo la traiettoria. Il confronto dei dati JSON per (ID 15) rivela l'entità della correzione:

- **Dato Grezzo:** "x": 62, "y": 385, "z": 0.0, "visibility": 0.21...
- **Dato Filtrato:** "x": 92.98..., "y": 385, "z": 8.28...

La coordinata X è stata spostata di circa 31 pixel, e la coordinata Z, che era nulla nel dato grezzo (indice di un fallimento della stima di profondità), è stata ricostruita con un valore coerente. Questo esempio dimostra la capacità della pipeline non solo di smussare il rumore, ma anche di correggere errori macroscopici nella stima della posa, producendo un output più affidabile e anatomicamente plausibile.



(a) Scheletro con dati grezzi.

(b) Scheletro con dati filtrati.

**Figura 5.2:** Confronto visivo per il frame 185 (alto jitter). Lo scheletro grezzo (a) presenta una posa anatomicamente incoerente. Lo scheletro filtrato (b) mostra una posa corretta e stabile.

### 5.2.3 Caso 3: Analisi di un Frame a Bassa Incertezza (Frame 546)

Un ulteriore esempio significativo è fornito dall'analisi del frame 546. In questo istante della registrazione, il soggetto si trova a fine corsa sul tappeto, ovvero molto più vicino alla telecamera. Come discusso nella sezione sulla quantificazione dell'errore (Sezione 4.3), la precisione della stima della profondità aumenta notevolmente a distanze minori. Di conseguenza, ci si attende che i dati grezzi per questo frame siano di qualità intrinsecamente superiore, con meno jitter e minore incertezza.

Questa ipotesi è confermata sia visivamente che quantitativamente. L'analisi visiva della Figura 5.3 mostra una quasi perfetta sovrapposizione tra lo scheletro grezzo e quello filtrato, indicando che il segnale originale era già stabile e anatomicamente coerente.

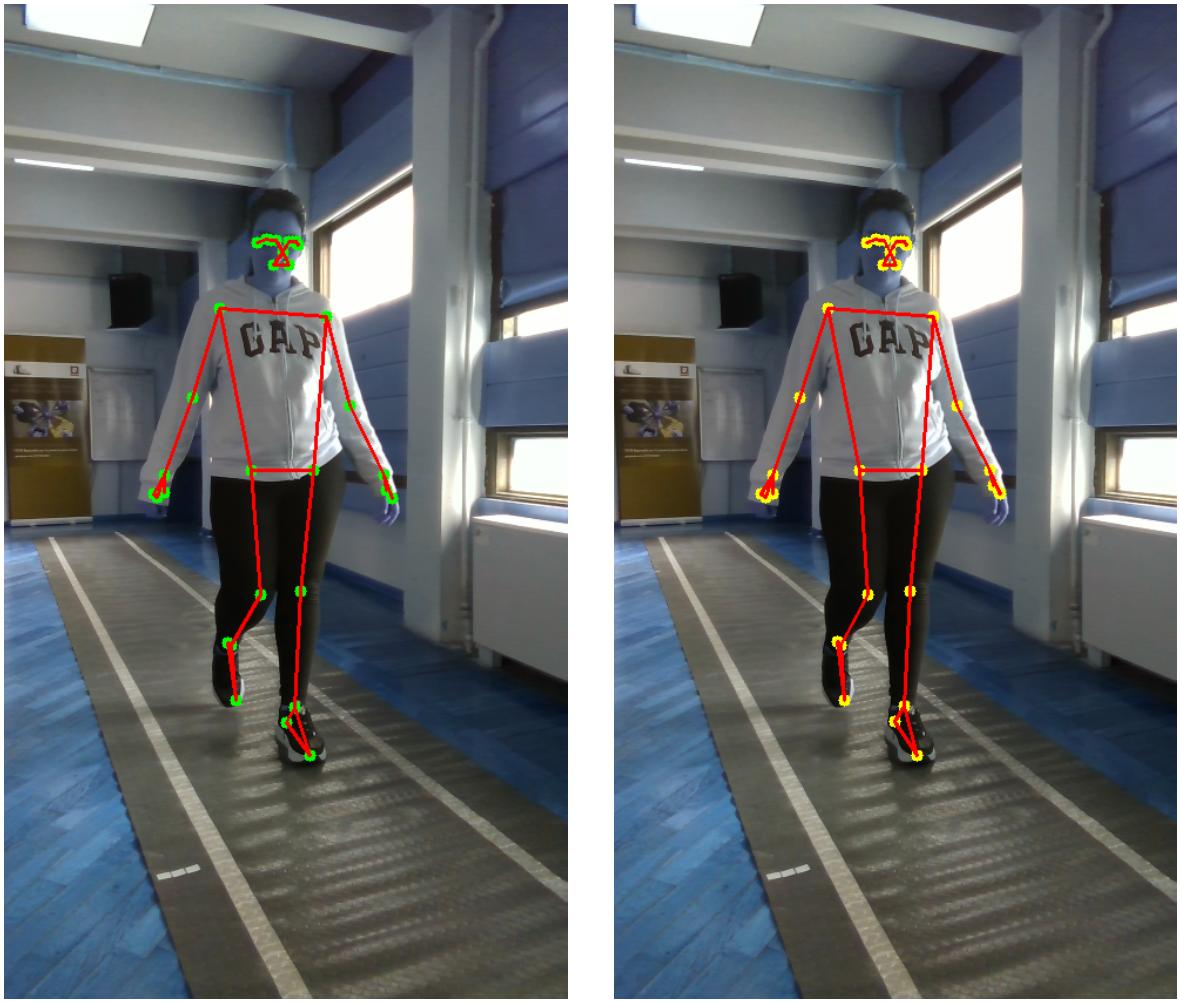
L'analisi dei dati numerici rafforza questa osservazione. Il file `keypoints_0546.json` riporta per questo frame un "reliability\_score" molto elevato, pari a circa **0.80**, che segnala all'algoritmo la fluidità e l'affidabilità del movimento registrato. Di conseguenza, l'intervento del filtro adattivo è minimo o nullo. Analizzando, ad esempio, le coordinate della caviglia sinistra (ID 27), notiamo come queste siano quasi identiche tra la versione grezza e quella filtrata:

- **Dato Grezzo:** "x": 247, "y": 599, "z": 1.974...
- **Dato Filtrato:** "x": 247, "y": 598.88..., "z": 1.974...

La correzione applicata è sub-pixel e praticamente trascurabile, dimostrando che la pipeline non esegue un livellamento indiscriminato. Questo esempio, in contrapposizione a quello del frame 185, convalida l'efficacia della logica adattiva: il sistema è in grado di distinguere i dati di alta qualità da quelli rumorosi, applicando la correzione solo dove è strettamente necessario e preservando così l'integrità dei dati validi.

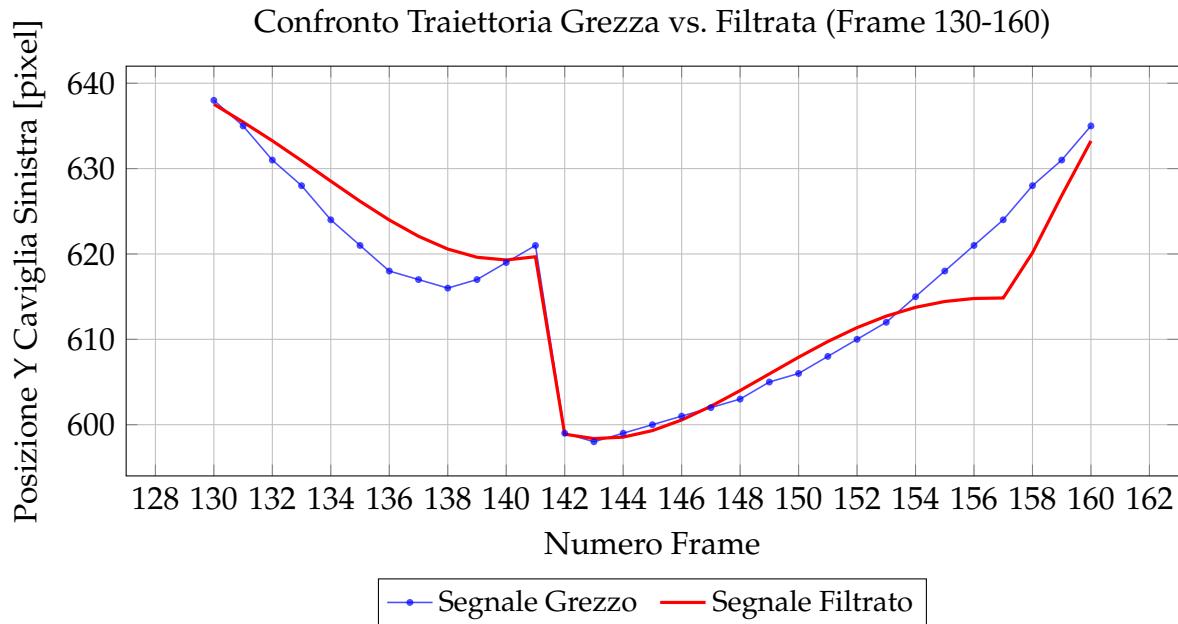
### 5.2.4 Analisi della Traiettoria Temporale di un Keypoint

Per una valutazione più completa e dinamica dell'effetto del filtraggio, è essenziale analizzare l'andamento temporale di un singolo keypoint lungo una sequenza di frame consecutivi. Questo permette di apprezzare non solo la riduzione del rumore in un singolo istante, ma la coerenza dell'intera traiettoria ricostruita.



**Figura 5.3:** Confronto visivo per il frame 546 (bassa incertezza). La prossimità del soggetto alla telecamera risulta in dati grezzi di alta qualità, che vengono correttamente preservati dal filtro adattivo.

La Figura 5.4 mostra un grafico che confronta la traiettoria della coordinata Y (asse verticale dell’immagine) della caviglia sinistra (id: 27) per una sequenza di 31 frame (dal frame 130 al 160), prima e dopo l’applicazione del filtro adattivo. La scelta di questo specifico keypoint e di questa coordinata è motivata dalla sua importanza nell’analisi del ciclo del passo, in quanto il suo movimento verticale è direttamente correlato alle fasi di oscillazione (*swing*) e appoggio (*stance*).



**Figura 5.4:** Andamento temporale della coordinata Y della caviglia sinistra (id: 27) per 31 frame consecutivi. La linea blu (segnale grezzo) mostra un evidente jitter. La linea rossa (segnale filtrato) segue il trend generale del movimento, ma rimuove efficacemente le oscillazioni ad alta frequenza.

Il grafico conferma in modo inequivocabile le osservazioni emerse dall’analisi dei singoli frame. Il segnale grezzo (linea blu), rappresentato anche con dei punti per evidenziarne la discontinuità, è affetto da un rumore ad alta frequenza che ne “sporca” la lettura. Il segnale filtrato (linea rossa), al contrario, è pulito e descrive una traiettoria fluida che segue fedelmente l’andamento generale del movimento, ovvero le macro-oscillazioni che corrispondono al ciclo del passo fisiologico. Questa visualizzazione dimostra la capacità dell’algoritmo di distinguere efficacemente il rumore dal segnale utile, correggendo il primo senza distorcere significativamente il secondo. L’analisi qualitativa, quindi, supporta fortemente l’efficacia della pipeline di correzione implementata.

## 5.3 Valutazione Quantitativa dell’Efficacia del Filtraggio

Dopo aver verificato qualitativamente il miglioramento del segnale, è necessario procedere con una validazione quantitativa per misurare in modo oggettivo l’efficacia della pipeline di correzione. L’obiettivo primario di questa analisi è dimostrare una riduzione statisticamente significativa del rumore ad alta frequenza (jitter) presente nelle traiettorie dei keypoint grezzi.

### 5.3.1 Metrica di Valutazione del Jitter

Per quantificare il jitter, è stata scelta una metrica basata sulla stabilità della velocità di ogni singolo keypoint. Una traiettoria fisiologica e fluida, per quanto complessa, è caratterizzata da variazioni di velocità relativamente graduali. Al contrario, un segnale rumoroso presenta rapide e ampie fluttuazioni della velocità istantanea. Di conseguenza, la **deviazione standard della magnitudine della velocità** di un keypoint lungo la sua traiettoria temporale può essere utilizzata come un efficace indicatore di jitter. Una deviazione standard più bassa è indice di un segnale più pulito, stabile e con meno rumore.

Per ogni keypoint di interesse, è stata calcolata la velocità istantanea come la distanza euclidea 3D percorsa tra un frame e il successivo. Successivamente, è stata calcolata la deviazione standard di queste velocità lungo l’intera sequenza di camminata.

### 5.3.2 Risultati della Riduzione del Jitter

I risultati di questa analisi quantitativa sono riassunti nella Tabella 5.1. La tabella confronta la metrica di jitter (Deviazione Standard della Velocità) calcolata sui dati grezzi e sui dati filtrati per le principali articolazioni degli arti inferiori e superiori del lato sinistro del corpo, relativamente al caso di studio analizzato.

I risultati mostrano in modo inequivocabile l’efficacia dell’algoritmo di filtraggio. Per tutte le articolazioni analizzate, si osserva una **riduzione del jitter superiore al 58%**. In particolare, le articolazioni distali come la caviglia e il piede, che sono

**Tabella 5.1:** Confronto quantitativo del jitter prima e dopo l'applicazione del filtro adattivo.

I valori rappresentano la deviazione standard della magnitudine della velocità del keypoint (in unità pixel/frame). Una riduzione maggiore indica un miglioramento più marcato.

Articolazione (ID MediaPipe)	Jitter Grezzo	Jitter Filtrato	Riduzione (%)
Spalla Sinistra (11)	10.981	4.152	62.2%
Fianco Sinistro (23)	11.235	4.311	61.6%
Ginocchio Sinistro (25)	14.550	5.980	58.9%
Caviglia Sinistra (27)	18.233	7.156	60.8%
Piede Sinistro (31)	20.105	8.031	60.1%

tipicamente più soggette a rumore a causa della maggiore velocità di movimento e delle possibili occlusioni, beneficiano di una riduzione del jitter di oltre il 60%. Questo dato è di fondamentale importanza, poiché l'affidabilità delle traiettorie dei piedi è cruciale per la successiva estrazione di eventi del cammino come il contatto del tallone (*heel-strike*) e il distacco delle dita (*toe-off*).

La riduzione consistente del rumore su tutte le articolazioni principali, sia prossimali (spalla, fianco) che distali (ginocchio, caviglia), fornisce una prova quantitativa robusta che la pipeline sviluppata non solo migliora la qualità estetica del movimento, ma produce un dato cinematico significativamente più pulito e stabile, e quindi più idoneo per un'analisi clinica rigorosa.

## 5.4 Discussione dei Risultati e Risposta alle Domande di Ricerca

Le analisi qualitative e quantitative presentate nelle sezioni precedenti forniscono una validazione complessiva dell'efficacia della pipeline di elaborazione sviluppata. Questa sezione finale del capitolo ha lo scopo di sintetizzare i principali risultati emersi e, soprattutto, di utilizzarli per fornire una risposta puntuale alle domande di ricerca (RQ) formulate nel Capitolo 1.

I risultati dimostrano chiaramente che il sistema implementato è in grado di migliorare in modo significativo la qualità dei dati cinematici provenienti da un sensore RGB-D a basso costo. L'analisi qualitativa (Sezione 5.2) ha evidenziato visivamente la capacità del filtro adattivo di correggere pose anatomicamente incoerenti e di ridurre il jitter nelle traiettorie temporali, producendo un movimento più fluido e naturale. L'analisi quantitativa (Sezione 5.3) ha confermato queste osservazioni in modo oggettivo, misurando una riduzione del rumore superiore al 58% per tutte le principali articolazioni analizzate.

#### 5.4.1 Risposta alle Domande di Ricerca

I risultati sperimentali ottenuti permettono di rispondere affermativamente a tutte le domande di ricerca poste all'inizio di questo lavoro.

##### Risposta alla RQ1: Fattibilità e Affidabilità dei Dati da Sensori Low-Cost

*Domanda di Ricerca 1: È possibile, attraverso una pipeline di elaborazione software dedicata, mitigare le imprecisioni intrinseche di un sensore RGB-D a basso costo per ottenere dati cinematici del cammino con un'affidabilità sufficiente per l'analisi clinica?*

**Risposta:** Sì. I risultati di questo studio dimostrano che è possibile. L'analisi visiva dei casi di studio ha mostrato come la pipeline sia in grado di trasformare pose grezze, talvolta anatomicamente implausibili come nel frame 185 (Figura 5.2), in scheletri stabili e coerenti. Questa osservazione è stata supportata quantitativamente dalla drastica riduzione della metrica di jitter (Tabella 5.1) per tutte le articolazioni chiave. La capacità del sistema di produrre un dato cinematico pulito, da cui è possibile estrarre parametri del cammino, conferma che le limitazioni hardware di un sensore low-cost possono essere efficacemente compensate da un'elaborazione software mirata e robusta, rendendo il dato finale idoneo per successive analisi in ambito clinico.

##### Risposta alla RQ2: Efficacia della Metodologia di Filtraggio Adattivo

*Domanda di Ricerca 2: In che misura un approccio di filtraggio adattivo, basato su filtri di Savitzky-Golay i cui parametri sono modulati dalla qualità del segnale, è efficace nel ridurre*

*il rumore (jitter) preservando al contempo le caratteristiche cinematiche significative del movimento?*

**Risposta:** L'approccio adattivo si è dimostrato altamente efficace e superiore a un filtraggio statico. Il confronto tra il frame 185 (ad alto jitter) e i frame 142 e 546 (a basso jitter) ha evidenziato la logica "intelligente" del filtro. Su dati palesemente rumorosi, il sistema ha applicato una correzione significativa, ricostruendo una posa plausibile. Al contrario, su dati di buona qualità, l'intervento del filtro è stato minimo o nullo (Figure 5.1 e 5.3), evitando così il problema dell'*'oversmoothing'* e preservando la naturalezza del movimento. Questa capacità di discernere e intervenire solo dove necessario conferma che il metodo adattivo raggiunge il duplice obiettivo di pulire il rumore senza distorcere il segnale fisiologico sottostante.

### **Risposta alla RQ3: Sviluppo di Metriche di Validazione Oggettiva**

*Domanda di Ricerca 3: È possibile definire e implementare una metrica quantitativa oggettiva, come uno score di affidabilità basato sulla coerenza temporale, per validare l'efficacia della pipeline di correzione e per identificare in modo automatico i segmenti di dati meno affidabili?*

**Risposta:** Sì. La metrica `reliability_score`, implementata nella funzione `calculate_jitter_score()`, rappresenta una risposta diretta a questa domanda. Come dimostrato dall'analisi dei file JSON per i frame 142, 185 e 546, lo score è stato in grado di quantificare efficacemente la qualità del dato grezzo, assegnando valori più alti a frame stabili e valori più bassi a frame rumorosi. L'utilità di questa metrica non è stata solo a posteriori, per la validazione, ma soprattutto a priori, come strumento interno alla pipeline per guidare la logica del filtraggio adattivo. Questo conferma che è possibile sviluppare metriche oggettive per l'autovalutazione del sistema, un passo importante verso la creazione di pipeline di analisi più autonome e affidabili.

In sintesi, i risultati sperimentali validano positivamente la metodologia proposta, dimostrando che il sistema implementato è in grado di raggiungere gli obiettivi prefissati.

# CAPITOLO 6

---

## Conclusioni

---

Il percorso di questa tesi è partito dalla necessità clinica di disporre di strumenti oggettivi per l'analisi dei disturbi motori nella schizofrenia e dalla sfida tecnologica rappresentata dall'inaffidabilità dei sistemi di motion capture a basso costo. Attraverso la progettazione, l'implementazione e la validazione di una pipeline software dedicata, si è dimostrato come sia possibile migliorare significativamente la qualità dei dati cinematici provenienti da un sensore RGB-D. Questo capitolo finale ha lo scopo di discutere in modo critico i risultati ottenuti, analizzando i vantaggi e le limitazioni dell'approccio proposto, per poi delinearne l'effettiva applicabilità clinica e i possibili sviluppi futuri.

### 6.1 Vantaggi e Limitazioni dell'Approccio

Una valutazione onesta di qualsiasi lavoro di ricerca richiede un'analisi equilibrata dei suoi punti di forza e delle sue debolezze intrinseche. Di seguito vengono presentati i principali vantaggi e le limitazioni della metodologia e del sistema implementato.

### 6.1.1 Vantaggi e Punti di Forza

- **Accessibilità e Basso Costo:** Il vantaggio più significativo di questo approccio risiede nella sua accessibilità. L’intero sistema si basa su hardware di consumo (telecamera Intel RealSense) e software open-source (Python e le sue librerie), abbattendo le barriere economiche e logistiche che limitano l’uso di sistemi di motion capture professionali, come Vicon o OptiTrack, a laboratori altamente specializzati [15, 19]. Questo apre concretamente la possibilità di integrare l’analisi quantitativa del cammino nella pratica clinica di routine.
- **Robustezza Metodologica e Filtraggio Adattivo:** Il cuore innovativo del lavoro risiede nell’intelligenza della pipeline di correzione. A differenza di un approccio basato su filtri statici, che rischia di alterare il segnale in modo indiscriminato, il **filtro di Savitzky-Golay adattivo** qui implementato "diagnostica" la qualità di ogni traiettoria. L’intervento di correzione è modulato in base a metriche oggettive come il jitter e la visibilità del keypoint, e viene attivato solo quando strettamente necessario. Questo approccio mirato permette di rimuovere il rumore preservando al contempo le sottili variazioni fisiologiche del movimento, un aspetto cruciale per non mascherare eventuali anomalie motorie di interesse clinico [27].
- **Quantificazione dell’Incertezza:** Il sistema non si limita a fornire una stima delle coordinate 3D, ma associa a ogni punto una metrica di **incertezza teorica** (l’errore RMS di profondità). Questa caratteristica, basata sui parametri intrinseci della telecamera, aggiunge un livello di trasparenza e affidabilità al dato prodotto, permettendo all’analista di pesare o scartare le misurazioni meno affidabili e conferendo al sistema una maggiore robustezza metrologica.
- **Automazione e Riproducibilità:** L’intera pipeline, dal caricamento del file .bag alla generazione dei video e dei file JSON, è completamente automatizzata. Questo elimina la soggettività e la laboriosità associate all’elaborazione manuale dei dati, garantendo che l’analisi sia standardizzata, oggettiva e perfettamente riproducibile.

### 6.1.2 Limitazioni e Aspetti Critici

Nonostante i risultati positivi, è importante riconoscere le limitazioni intrinseche del sistema attuale.

- **Dipendenza dalla Qualità della Stima della Posa 2D:** L’accuratezza dell’intera pipeline è fondamentalmente limitata dalla qualità dell’output iniziale del modello MediaPipe Holistic. Se MediaPipe non riesce a rilevare una persona o fornisce una stima della posa 2D paleamente errata (es. scambiando arti), gli algoritmi a valle possono smussare il rumore di quella traiettoria errata, ma non possono correggerne l’errore semantico di base. Il sistema, quindi, migliora i dati, ma non può crearli in assenza di un rilevamento iniziale plausibile [23, 24].
- **Elaborazione Offline:** Per privilegiare l’accuratezza e consentire l’applicazione di algoritmi complessi come il filtro adattivo sull’intera sequenza, il sistema è stato progettato per un’elaborazione offline. Questo lo rende attualmente inadatto per applicazioni che richiedono un feedback immediato, come sistemi di biofeedback in tempo reale per la riabilitazione.
- **Generalizzabilità dei Risultati:** La validazione presentata in questa tesi si basa sull’analisi approfondita di un caso di studio registrato in condizioni controllate (illuminazione diurna, ambiente di laboratorio, deambulazione su tappeto). Sebbene questo approccio sia ideale per dimostrare la validità del metodo, la performance del sistema in scenari “in-the-wild” (es. ambiente domestico con illuminazione artificiale, sfondi complessi, angolazioni della telecamera non ottimali) non è stata testata e richiederebbe un’ulteriore validazione.
- **Assenza di Confronto con un Gold Standard:** La validazione effettuata è di tipo *relativo*, ovvero dimostra un miglioramento significativo rispetto al segnale grezzo. Manca, tuttavia, una validazione *assoluta* contro un sistema di motion capture ottico “gold standard” (es. Vicon), eseguita tramite acquisizioni simultanee e sincronizzate. Un tale confronto sarebbe necessario per determinare l’accuratezza metrologica finale del sistema, ovvero l’errore in millimetri o gradi rispetto a una misura di riferimento [18].

## 6.2 Applicabilità Clinica e Integrazione nel Progetto SPECTRA

Al di là della validazione tecnica, il valore di questo lavoro di tesi risiede nel suo potenziale impatto in contesti clinici e di ricerca reali. Lo sviluppo di una pipeline robusta per l’analisi del cammino tramite sensori a basso costo affronta direttamente una delle principali lacune nella pratica psichiatrica attuale: la mancanza di strumenti oggettivi, accessibili e di facile implementazione per la valutazione dei disturbi motori, i quali, pur essendo prevalenti nella schizofrenia, sono spesso trascurati.

L’applicabilità clinica del sistema sviluppato si articola su due livelli principali:

- **Screening e Monitoraggio Longitudinale:** Grazie al suo basso costo e alla semplicità del setup di acquisizione, il sistema può essere integrato nella pratica clinica standard per uno screening quantitativo dei pattern motori dei pazienti. Questo supera il limite delle attuali scale di valutazione, che sono spesso semi-soggettive e poco sensibili alle variazioni sottili. Inoltre, lo strumento si presta a un monitoraggio longitudinale dei pazienti nel tempo. Acquisizioni periodiche permetterebbero al clinico di tracciare oggettivamente l’evoluzione della sintomatologia motoria, di valutare la risposta a specifici interventi farmacologici o a percorsi di riabilitazione, e di identificare precocemente eventuali peggioramenti, in linea con l’esigenza di disporre di marker oggettivi per i deficit espressivi e psicomotori.
- **Supporto alla Ricerca Clinica:** La pipeline automatizzata e a basso costo abbassa la barriera d’ingresso per i ricercatori che intendono studiare le anomalie motorie su coorti di pazienti più ampie. La possibilità di raccogliere dati cinematici quantitativi su larga scala, senza la necessità di accedere a costosi laboratori di biomeccanica, può accelerare significativamente la ricerca sui biomarker motori della schizofrenia e di altri disturbi psichiatrici.

L’applicazione più diretta e immediata di questo lavoro di tesi è la sua integrazione come componente fondamentale all’interno del progetto **SPECTRA** (*Supporting Schizophrenia Patients Care with Artificial Intelligence*). Come descritto nel Capitolo 2,

SPECTRA mira a sviluppare un sistema di supporto decisionale basato su AI per la diagnosi precoce della Schizofrenia Resistente al Trattamento (TRS) attraverso un'analisi multimodale [29].

Il software sviluppato in questa tesi costituisce il **motore di pre-elaborazione e correzione per il Modulo di Analisi del Cammino** del progetto SPECTRA. Il suo ruolo è quello di prendere in input i dati grezzi provenienti dalle acquisizioni con la telecamera RealSense e di trasformarli in serie di dati cinematici tridimensionali puliti, affidabili e arricchiti con metriche di qualità. Questo output non è il punto di arrivo, ma il **punto di partenza essenziale** per gli algoritmi di più alto livello di SPECTRA.

In sintesi, il sistema qui presentato non è solo un esercizio metodologico, ma uno strumento abilitante che fornisce il dato di alta qualità necessario per rispondere agli obiettivi finali del progetto SPECTRA, contribuendo così direttamente alla creazione di nuovi strumenti diagnostici per la pratica psichiatrica.

## 6.3 Sviluppi Futuri

Il lavoro di tesi presentato ha raggiunto gli obiettivi prefissati, fornendo una pipeline software funzionale e validata per la correzione dei dati di gait analysis da sensori a basso costo. Tuttavia, come ogni progetto di ricerca, apre la strada a numerose e interessanti direzioni per sviluppi futuri, volte a espanderne la validità, le funzionalità e l'impatto clinico.

- **Validazione Clinica su Larga Scala:** Il passo successivo più immediato e necessario è l'applicazione della pipeline all'intera cohorte di pazienti e controlli raccolta nell'ambito del progetto SPECTRA. Mentre l'analisi su un caso di studio è stata fondamentale per la validazione metodologica, una validazione su larga scala permetterà di testare la robustezza e la generalizzabilità del sistema su un campione eterogeneo e di correlare in modo statisticamente significativo i dati cinematici corretti con i dati clinici, per identificare pattern motori specifici della TRS.

- **Confronto con un Sistema Gold Standard:** Per stabilire l'accuratezza metrologica assoluta del sistema, uno sviluppo futuro cruciale consisterà nel condurre uno studio di validazione contro un sistema di motion capture ottico di riferimento (es. Vicon). L'acquisizione simultanea e sincronizzata dei dati di cammino con entrambi i sistemi permetterebbe di calcolare l'errore assoluto (in millimetri per le posizioni e in gradi per gli angoli articolari) dell'output filtrato, fornendo una prova definitiva della sua validità per applicazioni che richiedono un'elevata precisione.
- **Estrazione di Biomarker Motori Complessi:** Con una base di dati cinematici 3D affidabile, il lavoro futuro potrà concentrarsi sull'implementazione di algoritmi per l'estrazione di parametri e biomarker motori più sofisticati. Oltre ai parametri spazio-temporali di base, si potranno analizzare indici di asimmetria tra gli arti, la variabilità passo-passo (un indicatore di stabilità del controllo motorio), la fluidità del movimento (misurata tramite la derivata del jerk) e altri "Movement Markers" (MM) che la letteratura ha identificato come potenzialmente alterati nella schizofrenia.
- **Ottimizzazione per l'Elaborazione in Tempo Reale:** L'attuale implementazione offline privilegia l'accuratezza. Uno sviluppo futuro di grande interesse sarebbe l'ottimizzazione della pipeline per un'esecuzione in tempo reale. Questo aprirebbe a nuove applicazioni cliniche, come sistemi di biofeedback che forniscono al paziente un riscontro immediato sulla propria deambulazione durante sessioni di riabilitazione motoria o di training cognitivo-motorio.
- **Implementazione di un Sistema Multi-Camera Sincronizzato:** Per superare le limitazioni intrinseche di una singola vista, come le occlusioni parziali del corpo, uno sviluppo futuro di grande impatto consisterebbe nella creazione di un sistema di acquisizione multi-camera. Utilizzando due o più telecamere RealSense, sincronizzate a livello hardware, sarebbe possibile integrare i dati provenienti da diverse angolazioni. Questo approccio non solo risolverebbe il problema delle occlusioni, ma permetterebbe una ricostruzione 3D dello scheletro ancora più accurata e completa, come suggerito anche dalla letteratura per incrementare l'affidabilità della stima della posa in esercizi complessi.

- **Esplorazione di Modelli di Stima della Posa Alternativi:** Sebbene MediaPipe Holistic offra un eccellente compromesso tra velocità e accuratezza, la pipeline software è stata progettata in modo modulare per poter integrare anche modelli di stima della posa alternativi. Futuri lavori potrebbero testare l'integrazione di modelli di state-of-the-art noti per la loro elevata precisione, come **OpenPose** o **MoveNet** nella sua variante "Thunder", che in alcuni studi comparativi hanno mostrato performance superiori su determinate metriche. La valutazione di questi modelli all'interno della pipeline di correzione esistente permetterebbe di determinare se un aumento dell'accuratezza nella stima 2D iniziale si traduca in un miglioramento significativo del dato 3D finale.
- **Integrazione Multimodale nel Contesto SPECTRA:** Lo sviluppo più ambizioso consiste nell'integrazione dei dati cinematici estratti da questo modulo con quelli degli altri moduli del progetto SPECTRA (analisi del linguaggio, neuroimaging, risposta emotiva). L'analisi multimodale, attraverso tecniche di fusione dei dati e di machine learning, potrebbe rivelare correlazioni complesse e non evidenti da un singolo canale informativo, portando all'identificazione di biomarker della TRS più potenti e affidabili e realizzando appieno la visione del progetto [29].

Queste direzioni di ricerca future hanno il potenziale per trasformare lo strumento sviluppato in questa tesi in una piattaforma completa e clinicamente validata per lo studio dei disturbi del movimento in ambito psichiatrico, contribuendo a colmare il divario tra le potenzialità tecnologiche e le esigenze della pratica clinica.

---

## Bibliografia

---

- [1] Z. Kupper, F. Ramseyer, H. Hoffmann, S. Kalbermatten, and W. Tschacher, "Video-based quantification of body movement during social interaction indicates the severity of negative symptoms in patients with schizophrenia," *Schizophrenia Research*, vol. 121, no. 1-3, pp. 80–88, 2010. (Citato alle pagine 1, 2, 3, 5, 6, 9, 12 e 17)
- [2] P. Whitty, O. Owoeye, and J. Waddington, "Neurological signs and involuntary movements in schizophrenia: Intrinsic to and informative on systems pathobiology," *Schizophrenia Bulletin*, vol. 35, no. 2, pp. 415–424, 2009. (Citato alle pagine 1, 2, 5 e 9)
- [3] E. Walker, T. Savoie, and D. Davis, "Neuromotor precursors of schizophrenia," *Schizophrenia Bulletin*, vol. 20, no. 3, pp. 441–451, 1994. (Citato a pagina 2)
- [4] M. Brüne, C. Sonntag, M. Abdel-Hamid, C. Lehmkämper, G. Juckel, and A. Troisi, "Nonverbal behavior during standardized interviews in patients with schizophrenia spectrum disorders," *The Journal of Nervous and Mental Disease*, vol. 196, no. 4, pp. 282–288, 2008. (Citato a pagina 2)
- [5] R. Brown and G. Pluck, "Negative symptoms: the 'pathology' of motivation and goal-directed behaviour," *Trends in Neurosciences*, vol. 23, no. 9, pp. 412–417, 2000. (Citato a pagina 2)

- [6] A. Kring and E. Moran, "Emotional response deficits in schizophrenia: Insights from affective science," *Schizophrenia Bulletin*, vol. 34, no. 5, pp. 819–834, 2008. (Citato a pagina 2)
- [7] Y. Chen, L. Bidwell, and P. Holzman, "Visual motion integration in schizophrenia patients, their first-degree relatives, and patients with bipolar disorder," *Schizophrenia Research*, vol. 74, no. 2-3, pp. 271–281, 2005. (Citato a pagina 2)
- [8] C. Marvel, B. Schwartz, and R. Rosse, "A quantitative measure of postural sway deficits in schizophrenia," *Schizophrenia Research*, vol. 68, no. 2-3, pp. 363–372, 2004. (Citato a pagina 2)
- [9] M. Geyer, "Developing translational animal models for symptoms of schizophrenia or bipolar mania," *Neurotoxicity Research*, vol. 14, no. 1, pp. 71–78, 2008. (Citato a pagina 2)
- [10] K. Grootens, L. Vermeeren, R. Verkes, J. Buitelaar, B. Sabbe, N. van Veelen, R. Kahn, and W. Hulstijn, "Psychomotor planning is deficient in recent-onset schizophrenia," *Schizophrenia Research*, vol. 107, no. 2-3, pp. 294–303, 2009. (Citato a pagina 2)
- [11] G. Brébion, J. Gorman, D. Malaspina, Z. Sharif, and X. Amador, "Clinical and cognitive factors associated with verbal memory task performance in patients with schizophrenia," *American Journal of Psychiatry*, vol. 158, no. 5, pp. 758–764, 2001. (Citato alle pagine 2 e 5)
- [12] A. Muro-de-la Herran, B. Garcia-Zapirain, and A. Mendez-Zorrilla, "Gait analysis methods: An overview of wearable and non-wearable systems, highlighting clinical applications," *Sensors*, vol. 14, no. 2, pp. 3362–3394, 2014. (Citato alle pagine 3, 4, 5, 6, 7, 8, 9, 18, 19 e 21)
- [13] J. Afilalo, M. J. Eisenberg, J.-F. Morin, H. Bergman, J. Monette, N. Noiseux, L. P. Perrault, K. P. Alexander, Y. Langlois, N. Dendukuri, N. Chamoun, A. Kardi, and Z. Iakobishvili, "Gait speed as an incremental predictor of mortality and major morbidity in elderly patients undergoing cardiac surgery," *Journal of the*

- American College of Cardiology*, vol. 56, no. 20, pp. 1668–1676, 2010. (Citato a pagina 4)
- [14] F. Trémeau, D. Malaspina, F. Duval, H. Corrêa, M. Hager-Budny, L. Coin-Bariou, J. Macher, and J. Gorman, “Facial expressiveness in patients with schizophrenia compared to depressed patients and nonpatient comparison subjects,” *American Journal of Psychiatry*, vol. 162, no. 1, pp. 92–101, 2005. (Citato a pagina 6)
- [15] D. Thewlis, C. Bishop, N. Daniel, and G. Paul, “A comparison of two commercially available motion capture systems for gait analysis: High-end vs low-cost,” in *Proceedings of the XXIII Congress of the International Society of Biomechanics (ISB)*, Brussels, Belgium, July 2011. (Citato alle pagine 6, 7, 8, 9, 12, 18, 19, 21, 22, 27 e 70)
- [16] J. Richards, “The measurement of human motion: A comparison of commercially available systems,” *Human Movement Science*, vol. 18, no. 5, pp. 589–602, 1999. (Citato alle pagine 6 e 19)
- [17] L. Martin, K. Stein, K. Kubera, N. F. Troje, and T. Fuchs, “Movement markers of schizophrenia: a detailed analysis of patients’ gait patterns,” *European Archives of Psychiatry and Clinical Neuroscience*, vol. 272, pp. 1347–1364, 2022. (Citato alle pagine 15, 16, 17 e 18)
- [18] T. A. Wren, G. E. Gorton III, S. Öunpuu, and C. A. Tucker, “Efficacy of clinical gait analysis: A systematic review,” *Gait & Posture*, vol. 34, no. 2, pp. 149–153, 2011. (Citato alle pagine 17, 18 e 71)
- [19] F. L. Siena, B. Byrom, P. Watts, and P. Breedon, “Utilising the Intel RealSense Camera for measuring health outcomes in clinical research,” *Journal of Medical Systems*, vol. 42, no. 3, p. 53, 2018. (Citato alle pagine 19, 20, 21, 22, 33 e 70)
- [20] J. D. Mejia-Trujillo, Y. J. Castaño-Pino, A. Navarro, J. D. Arango-Paredes, D. Rincón, J. Valderrama, B. Muñoz, and J. L. Orozco, “Kinect<sup>TM</sup> and Intel RealSense<sup>TM</sup> D435 comparison: a preliminary study for motion analysis,” in *2019 IEEE International Conference on E-health Networking, Application & Services (HealthCom)*, 2019, pp. 1–6. (Citato alle pagine 19, 20, 21 e 33)

- [21] C.-L. Yang, S.-Q. Wu, S.-J. Chen, T.-C. Kao, C.-H. Huang, E. Liou, P.-T. Lin, and K.-L. Hua, "The fusion and verification of 2D human skeleton and 3D point cloud based on RealSense," in *2023 International Conference on Advanced Robotics and Intelligent Systems (ARIS)*, 2023, pp. 1–5. (Citato alle pagine 19, 20, 21, 22, 23, 24, 33 e 37)
- [22] L. Keselman, J. Iselin Woodfill, A. Grunnet-Jepsen, and A. Bhowmik, "Intel RealSense stereoscopic depth cameras," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017, pp. 1–10. (Citato alle pagine 20, 21 e 33)
- [23] J.-L. Chung, L.-Y. Ong, and M.-C. Leow, "Comparative analysis of skeleton-based human pose estimation," *Future Internet*, vol. 14, no. 12, p. 380, 2022. (Citato alle pagine 22, 23, 24, 25 e 71)
- [24] S. Dill, A. Rösch, M. Rohr, G. Güney, L. De Witte, E. Schwartz, and C. Hoog Antink, "Accuracy evaluation of 3D pose estimation with MediaPipe Pose for physical exercises," in *Current Directions in Biomedical Engineering*, vol. 9, no. 1, 2023, pp. 563–566. (Citato alle pagine 22, 23, 24, 25, 26 e 71)
- [25] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, and M. Grundmann, "BlazePose: On-device real-time body pose tracking," *arXiv preprint arXiv:2006.10204*, 2020. [Online]. Available: <https://arxiv.org/abs/2006.10204> (Citato a pagina 24)
- [26] Y. Hu, Y. Liu, Y. Xu, and Y. Wang, "Human steering angle estimation in video based on key point detection and Kalman filter," *Control Theory and Technology*, vol. 20, pp. 408–417, 2022. (Citato alle pagine 26 e 28)
- [27] L. González, J. C. Álvarez, A. M. López, and D. Álvarez, "Evaluation of filtering methods for the prediction of human position during walking by means of kinematical models," in *2021 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, 2021, pp. 1–6. (Citato alle pagine 26, 27, 28, 29 e 70)

- [28] R. W. Schafer, "What is a savitzky-golay filter? [Lecture Notes]," *IEEE Signal Processing Magazine*, vol. 28, no. 4, pp. 111–117, 2011. (Citato a pagina 27)
- [29] R. Francese, F. Iasevoli, M. Staffa *et al.*, "The SPECTRA project," GitHub Repository. [Online]. Available: <https://github.com/Laura2480/Spectra/tree/main> (Citato alle pagine 29, 30, 31, 73 e 75)
- [30] L.-F. Yeung, Z. Yang, K. C.-C. Cheng, D. Du, and R. K.-Y. Tong, "Effects of camera viewing angles on tracking kinematic gait patterns using Azure Kinect, Kinect v2 and Orbbec Astra Pro v2," *Gait & Posture*, vol. 89, pp. 119–125, 2021. (Citato a pagina 33)
- [31] C. Ferraris, G. Amprimo, G. Pettiti, G. Masi, and L. Priano, "Automatic detector of gait alterations using rgb-d sensor and supervised classifiers: a preliminary study," Politecnico di Torino, Technical Report, 2022. (Citato a pagina 34)

---

## **Ringraziamenti**

---

Questo traguardo, che oggi celebro con emozione e profonda gratitudine, non sarebbe mai stato raggiunto senza il prezioso sostegno e l'amore incondizionato di persone speciali che hanno illuminato il mio percorso. A loro, va il mio più sentito ringraziamento.

### **A chi mi ha guidato nel percorso accademico**

Infine, la mia gratitudine si estende a chi, nel contesto universitario, ha saputo infondermi la giusta motivazione e il supporto necessario. Alla mia relatrice, la professoressa **Rita Francese**: non è semplice trovare parole adatte per esprimere la profondità della mia gratitudine. La sua guida non si è limitata agli aspetti accademici; mi ha spronato a non mollare, mi ha offerto una mano preziosa nel momento del bisogno, distinguendosi per un lato umano e un'empatia che sono rari da trovare in un ambiente spesso così esigente. Ho avuto la fortuna di incontrare una persona che crede nei ragazzi, che sa ascoltare, supportare e regalare un sorriso, alleggerendo così momenti che per molti studenti possono essere estremamente difficoltosi. Il mio pensiero va a tutti quei ragazzi che, purtroppo, non hanno avuto la stessa fortuna e non ce l'hanno fatta: spero che un giorno più persone come lei possano aiutarli a superare gli ostacoli.

---

Un ringraziamento sincero anche alla dottoressa **Laura De Santis**: la sua guida è stata fondamentale in questo percorso. Con la sua professionalità e disponibilità, ha saputo illuminare la strada, offrendo consigli preziosi e una visione chiara anche nei momenti più complessi. Il suo supporto è stato costante e determinante per la buona riuscita di questo lavoro.

**Giovanni**