



**UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA**



**DIPARTIMENTO  
DI INGEGNERIA  
DELL'INFORMAZIONE**

**DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE**

**CORSO DI LAUREA IN INGEGNERIA BIOMEDICA**

**“Elaborazione di immagini per training di una rete neurale per markerless motion capture”**

**“Image processing to train a neural network for markerless motion capture”**

**Relatrice: Prof.ssa Zimi Sawacha, PhD**

**Laureando: Aymen Louachama**

**Correlatore: Ing. Rigoni Giulio, PhD student**

**ANNO ACCADEMICO 2022 – 2023**

**Data di laurea 16/11/2023**

# INDICE

<b>ABSTRACT</b>	pag.03
<b>1. INTRODUZIONE</b>	pag.04
<b>2. ANALISI DEL MOVIMENTO</b>	pag.05
2.1. Introduzione	
2.2. Stereofotogrammetria	
2.2.1. Calibrazione e triangolazione	
2.2.2. Tracking	
2.2.3. Errori in stereofotogrammetria	
<b>3. MARKERLESS MOTION CAPTURE</b>	pag.10
3.1. Vantaggi	
3.2. Applicazioni	
3.2.1. Theia3D	
3.2.2. OpenCap	
3.3. Limiti attuali delle soluzioni markerless	
<b>4. MATERIALI E METODI</b>	pag.17
4.1. Soggetti	
4.2. Labelling	
<b>5. RISULTATI</b>	pag.22
<b>6. CONCLUSIONI</b>	pag.24
<b>BIBLIOGRAFIA</b>	pag.25

## Abstract

La tecnologia attualmente più utilizzata per l'analisi del movimento è la stereofotogrammetria, capace di fornire, in ambito clinico per esempio, importanti informazioni per la riabilitazione dei pazienti, la diagnosi di patologie neuromuscolari, il supporto alle decisioni del medico.

Il suo utilizzo però è limitato da necessità della tecnologia stessa, come il bisogno di laboratori controllati e lunghe e specifiche procedure di preparazione dei soggetti e calibrazione del sistema.

Per superare alcuni di questi limiti si è pensato di passare a tecniche *markerless*, basate su principi di intelligenza artificiale e col vantaggio di non aver bisogno di ambienti specifici e marcatori sul corpo dei soggetti per studiarne il movimento. I sistemi *markerless* attualmente disponibili non raggiungono però valori sufficientemente bassi di errore per essere ampiamente impiegati come alternativa alla stereofotogrammetria.

Un progetto del *BioMovLab* dell'Università di Padova, in collaborazione con l'azienda BBSof, al fine di sviluppare un sistema *markerless* adeguato all'utilizzo in ambito clinico è in corso. La mia attività di tirocinio presso il *BioMovLab* ha avuto come scopo il *training* della rete neurale su cui si basa tale progetto, tramite la creazione di un insieme di immagini elaborate appositamente per permettere al sistema di ricostruire i parametri del movimento con più precisione dell'attuale stato dell'arte del *markerless motion capture*.

## 1. INTRODUZIONE

L'analisi del movimento umano è uno strumento essenziale utilizzato in un ampio spettro di attività diverse, dall'industria dell'intrattenimento alle applicazioni in ambito clinico e sportivo. Ai fini dell'intrattenimento i dati ottenuti da tecnologie di *motion capture* sono ad esempio utilizzati per animare personaggi 3D nei film e videogiochi, applicazioni cliniche includono diagnosi e trattamento di condizioni associate a patologie del movimento, nello sport invece l'analisi del movimento si concentra sulla prevenzione di infortuni e sul miglioramento delle prestazioni.

Un utilizzo più ampio e ordinario delle tecniche di analisi del movimento è ostacolato dai requisiti delle tecnologie disponibili per un'accurata misurazione dei parametri legati al moto, come la necessità di un ambiente di laboratorio con attrezzatura specifica e la necessità di applicare marker e sensori sul corpo del soggetto. Questi requisiti richiedono tempo, risorse, introducono artefatti dovuti allo scorrimento della pelle sulle ossa, fattori che prevengono il loro utilizzo in ambienti non controllati, come ad esempio all'aperto nei campi sportivi.

Un sistema *markerless* adeguato, che possa abbattere le limitazioni dei metodi sopra citati, potrebbe espandere in maniera sostanziale il campo di applicazione del *motion capture*. La mia attività di tirocinio presso il laboratorio di bioingegneria del movimento (BioMovLab) dell'Università degli Studi di Padova mi ha permesso di entrare in contatto con un progetto il cui obiettivo è lo sviluppo di un sistema di analisi del movimento *markerless* (*markerless motion capture* (MMC)), basato su intelligenza artificiale, che possa essere considerato una valida alternativa alla stereofotogrammetria, attuale *gold standard* dell'analisi del movimento, e che superi alcuni svantaggi impliciti nei sistemi *markerless* già diffusi.

## 2. ANALISI DEL MOVIMENTO

### 2.1. INTRODUZIONE

Una definizione dell'analisi del movimento è data dal professor David A. Winter:

*“la disciplina che descrive, analizza, definisce il movimento dell'uomo, dal cammino, al sollevamento di un oggetto, alla performance di un'atleta. Tutto è regolato dagli stessi principi biologici, ciò che cambia è lo specifico task motorio”* [5].

Lo scopo di questa disciplina è ottenere informazioni, durante atti motori selezionati, su:

- il movimento di punti selezionati,
- il movimento di parti o dell'intero corpo,
- i movimenti articolari,
- la funzione muscolare,
- i carichi agenti sui tessuti.

La raccolta ed elaborazione di tali informazioni avviene tramite osservazioni visive dirette, registrazioni effettuate con videocamere e fotocamere, e misurazioni eseguite con sensori di vario tipo.

Le misurazioni in questo ambito sono mirate alla determinazione quantitativa di variabili cinematiche, quali la posizione, velocità ed accelerazione, e di variabili dinamiche.

Per misure di natura cinematica vengono impiegati sistemi di *motion capture*, che possono essere *marker-based* se prevedono l'utilizzo di marcatori fisici, attivi o passivi, oppure *markerless* se non li prevedono. Le variabili dinamiche possono riguardare la stima delle forze interne o esterne, nel caso di forze esterne agenti sul soggetto in esame, vengono ottenute tramite piattaforme di forza e/o solette di pressione [9].

## 2.2. STEREOFOTOGRAMMETRIA

La stereofotogrammetria, una delle tecniche più utilizzate per l'analisi cinematica del movimento, consiste nel rilevamento di marcatori (o *marker*) tramite sistemi optoelettronici con lo scopo di ricostruire la traiettoria all'interno di un volume di registrazione.

I *marker* possono essere di due tipi: *attivi* se dotati di diodi ad emissione di luce nelle frequenze degli infrarossi, o *passivi*, generalmente di forma sferica, realizzati in materiale plastico ricoperto di una pellicola catarifrangente. Questi ultimi richiedono una strumentazione capace di emettere luce, in genere un segnale infrarosso con lunghezza d'onda tra i 780-820 nm [7].

I *marker* vengono posizionati sul corpo del soggetto seguendo protocolli ben definiti e, durante l'acquisizione, il sistema stereofotogrammetrico ricostruisce la posizione tridimensionale di ogni punto rilevato in relazione al sistema di riferimento definito sul volume di registrazione. Ogni videocamera del sistema acquisisce una proiezione bidimensionale della scena; perciò, per poter calcolare la posizione nello spazio di un punto, è necessario che venga osservato da almeno due obiettivi distinti e sincronizzati, requisito essenziale per il processo di triangolazione [4].

### 2.2.1. CALIBRAZIONE E TRIANGOLAZIONE

Passaggio fondamentale per un utilizzo corretto della stereofotogrammetria è la calibrazione, il cui scopo è determinare i parametri delle telecamere impiegate nel sistema [4].

Questi parametri possono essere divisi in due tipi:

- i parametri interni, o intrinseci, come la lunghezza focale, le coordinate del piano principale e i coefficienti di distorsione;
- i parametri esterni, o estrinseci, che consistono nella posizione del sistema di riferimento della telecamera rispetto a quello assoluto.

Prima di procedere con la calibrazione bisogna associare alle telecamere un modello matematico adeguato; solitamente viene adoperato un modello *pin-hole*, detto anche a foro stenopeico, in cui si interpreta l'ottica come un foro di dimensioni infinitesime posto al centro di prospettiva.

La calibrazione si articola solitamente in due fasi:

- calibrazione dinamica, in cui una bacchetta di *marker* posti a distanze note viene mossa all'interno del volume di registrazione al fine di individuare i confini dello spazio di lavoro e calcolare i parametri di calibrazione delle telecamere;

- calibrazione statica, in cui una terna di aste perpendicolari e dotate di marker a distanze note viene utilizzata per definire il sistema di riferimento assoluto.

Dato ora un punto  $P$  nello spazio, grazie ai parametri di sistema è possibile risalire alla sua proiezione sul piano immagine di una telecamera tramite i seguenti passaggi:

- trasformazione rigida delle coordinate dal sistema assoluto a quello della telecamera;
- trasformazione prospettica dal sistema di riferimento della telecamera a quello del piano immagine;
- trasformazione 2D dal piano immagine al sistema di riferimento del sensore.

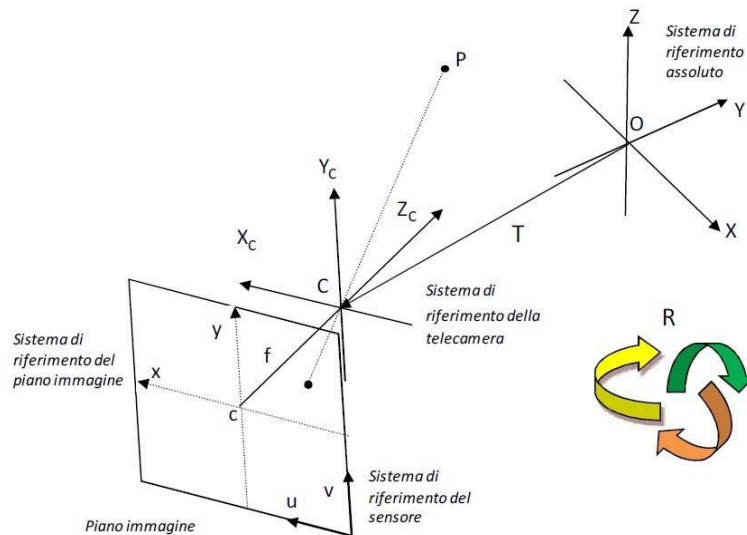


Figura 2.1: Proiezione sul piano immagine di un punto  $P$ . [9]

### 2.2.2. TRACKING

Il sistema calibrato può quindi determinare la posizione di qualsiasi *marker* visualizzato nel volume di lavoro tramite la procedura di triangolazione: il marcatore visto contemporaneamente da due telecamere, conoscendo la posizione dei punti nodali e la posizione ed orientamento dei piani principali delle stesse, avrà la sua posizione ricostruita, al netto di approssimazioni, calcolando l'intersezione delle rette che originano dalle coordinate 2D sul piano immagine di ciascuna telecamera e passanti dal centro ottico del relativo obiettivo.

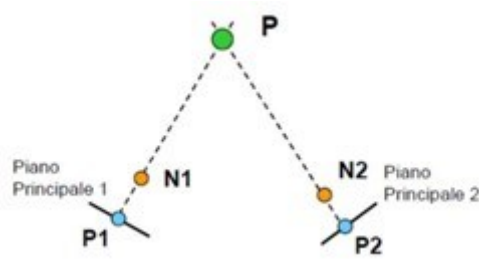


Figura 2.2: Ricostruzione del punto P dalle sue proiezioni. [9]

Dalle posizioni istantanee di ogni *marker* si procede con la ricostruzione delle traiettorie (*tracking*), operazione la cui complessità è legata al numero di marcatori utilizzati e al tipo di movimento compiuto dal soggetto.

Ad ogni marker viene associata inoltre un'etichetta che lo identifica, operazione che prende il nome di *labelling* e può essere effettuata sia in automatico, con i sistemi di acquisizione più recenti, che manualmente.

### 2.2.3. ERRORI IN STEREOFOTOGRAMMETRIA

I risultati dell'acquisizione permettono la stima della cinematica dei vari segmenti corporei esaminati, individuando per ciascuno posizione e orientamento a partire dalle coordinate spaziali e dalle traiettorie dei *marker*. La bontà della stima è però affetta da errori e artefatti, che possono provenire da fonti diverse:

- errori strumentali: errori intrinseci al sistema di misura stesso che introducono una variazione delle grandezze misurate in istanti diversi, anche in condizioni statiche. Gli errori strumentali si suddividono a loro volta in errori sistematici, solitamente dovuti a una calibrazione scorretta del sistema, ed errori casuali, segnali ad alta frequenza dovuti principalmente al rumore elettronico (*flickering*) e all'imprecisione nella conversione digitale in punti immagine della posizione dei marcatori. Gli errori strumentali possono essere ridotti tramite calibrazioni più attente e frequenti, mentre gli errori casuali possono essere contenuti utilizzando opportune tecniche di filtraggio nei domini del tempo e della frequenza [4];
- errori nell'individuazione dei punti di repere anatomico: errori dovuti al posizionamento inconsistente dei marcatori sui punti di repere, a causa della difficoltà nel definire un punto ad alta precisione sul corpo umano e a causa della presenza di uno strato di tessuto più o meno spesso sopra al punto stesso;
- errori da artefatto da tessuto molle: errori causati dal movimento relativo tra i *marker*, posizionati sulla cute, e il segmento osseo che definiscono. Questo movimento è dovuto allo scorrimento della pelle sui tessuti sottostanti, portando a



una variazione delle distanze tra i marcatori e inducendo errori nella ricostruzione del movimento rigido dell'articolazione. Gli artefatti da tessuto molle hanno frequenza simile alla cinematica in osservazione, perciò, non sono eliminabili tramite tecniche di filtraggio [6].

### 3. MARKERLESS MOTION CAPTURE

#### 3.1. VANTAGGI

La valutazione delle dinamiche del movimento umano è importante per comprendere e gestire patologie muscoloscheletriche e neuromuscolari, eppure queste metriche sono raramente presenti in ambito clinico. Seppure l'analisi quantitativa del movimento possa fornire informazioni più profonde e riproducibili rispetto a semplici valutazioni visive e test funzionali, il suo impiego in larga scala è ostacolato dai costi proibitivi e dalla mole di risorse necessarie al suo funzionamento. I metodi tradizionalmente più usati richiedono uno spazio di laboratorio equipaggiato di strumentazione specifica e costosa. I parametri cinematici sono, nella maggior parte dei casi, misurati con sistemi *marker-based* che richiedono anche più di otto telecamere apposite per catturare le traiettorie tridimensionali dei marcatori posti sul soggetto; e i parametri cinetici vengono stimati con un'ulteriore misura delle forze di reazione al suolo tramite piattaforme montate sotto al pavimento. Per completare i modelli muscoloscheletrici necessari all'investigazione del controllo motorio del soggetto vengono inoltre utilizzati anche i dati di attivazione muscolare ottenuti tramite sensori per elettromiografia. L'intero processo, dalla raccolta dei dati di un solo partecipante alla generazione di simulazioni sul modello muscoloscheletrico, può richiedere anche diversi giorni di lavoro di un esperto.

Risulta evidente il bisogno di un'alternativa economica, versatile e accurata per estendere l'applicabilità dell'analisi del movimento.

Strumenti portatili per la stima della cinematica sono già stati sviluppati, ma la maggior parte continuano ad essere troppo costosi e poco pratici per passare a un esteso utilizzo clinico, e nessuno permette un'analisi del movimento del corpo intero. Le unità di misura inerziale (IMU e MIMU), i più comunemente usati tra gli strumenti sopra citati, sono in grado di stimare con precisione i parametri cinematici, ma i sensori disponibili in commercio non sono economici e operano su algoritmi proprietari.

I recenti progressi in campo di computer vision e deep learning hanno portato a tecniche di MMC in grado di individuare e tracciare il moto umano in due e tre dimensioni dai video. Questi sistemi utilizzano reti neurali allenate con grandi *dataset* di immagini, etichettate manualmente, di persone impegnate sia in attività quotidiane che sportive. Per ogni frame gli algoritmi di stima della posa 2D riescono ad individuare *landmark* anatomici o centri articolari dell'individuo, derivandone poi la posa in quell'istante. Un approccio a più telecamere sincronizzate è possibile anche con i sistemi *markerless*, in

cui la posa 3D viene calcolata a partire da *keypoint* identificati nei video di ogni videocamera. I sistemi *markerless* multi-camera offrono una valida alternativa allo stato dell'arte della tecnologia *marker-based* per il *motion capture*. Vantaggio immediato del MMC è la velocità e facilità d'utilizzo, il posizionamento dei *marker*, infatti, richiede l'identificazione di punti di riferimento anatomici tramite il tatto, abilità che richiede conoscenze ed esperienza per non introdurre una grande fonte di errore, la mancanza di questo passaggio rende l'analisi cinematica accessibile anche ad utenti meno esperti, senza sacrificarne l'affidabilità. Inoltre, l'eliminazione del bisogno di marcatori fisici porta benefici nel tempo di preparazione prima di ogni acquisizione e nella natura degli errori commessi. Il tempo di preparazione del paziente è infatti notevolmente ridotto e non si corre il rischio che i *marker* influenzino la naturalezza del movimento da eseguire. Gli errori nell'individuazione dei punti di repere anatomico dovuti all'operatore umano non sono presenti, di conseguenza diventa possibile consolidare dataset e risultati diversi senza incorrere in discrepanze dovute ad errori sistematici legati all'esaminatore stesso. La ridotta complessità dei sistemi *markerless* in materia di accessori necessari, unita a tempi di impostazione della strumentazione e preparazione dei soggetti contenuti, permette una mobilità altrimenti impensabile per sistemi stereofotogrammetrici. Il vantaggio che ne deriva è un'espansione notevole dell'applicabilità dell'analisi cinematica: i soggetti possono essere ripresi in un ambiente a loro familiare, i loro movimenti risultano più naturali, è possibile coinvolgere un numero maggiore di persone eliminando il bisogno di far arrivare i partecipanti in un laboratorio e imponendo requisiti di abbigliamento aderente al corpo, e situazioni altamente dinamiche, come i movimenti in un campo sportivo, possono ora diventare più accessibili.

### **3.2. APPLICAZIONI**

I sistemi di MMC hanno dunque il potenziale per fornire un'alternativa economica e accessibile ai tradizionali sistemi a marcatori per l'analisi biomeccanica del movimento. Prima che possano essere impiegati, però, è di vitale importanza valutare rigorosamente la loro precisione nella stima delle cinematiche articolari, soprattutto per quanto concerne l'ambito clinico.

Di seguito vengono presentati due software per MMC, basati su algoritmi di intelligenza artificiale, impiegati in prove sperimentali per determinare quantitativamente la loro

accuratezza e confrontarla con sistemi *marker-based*, il *gold standard* dell'analisi del movimento.

### 3.2.1. Theia3D

*Theia3D* è un approccio al MMC basato su algoritmi di *deep learning* che utilizza dati di video sincronizzati per eseguire la stima in tre dimensioni della posa del soggetto. Il sistema usa reti neurali convoluzionali profonde, che sono state allenate con più di cinquecentomila immagini digitali etichettate manualmente, per rilevare le posizioni di 51 punti salienti sugli umani presenti in nuove immagini fornite al sistema.

Attualmente, se si guarda alla letteratura, si possono trovare sempre più articoli riguardanti la valutazione dell'affidabilità di questa applicazione tramite confronto con soluzioni di *marker-based motion capture* (MBMC), generalmente in un ambiente di laboratorio controllato su un piccolo campione di giovani adulti sani. Nelle mani degli sviluppatori il sistema *Theia3D* MMC ha prodotto cinematiche di cammino affidabili, con una varianza inter-sessione minore di quelle riportate da sistemi MBMC [8]. In uno studio pubblicato da Kanko et al. (2022) *Theia3D* ha dimostrato un'ottima correlazione dei propri parametri spazio-temporali sia con quelli ottenuti tramite tecnologie *marker-based* sia con parametri ottenuti tramite pedane di forza. Dati relativi alla cinematica 3D degli arti inferiori sono stati estratti in contemporanea con un sistema MBMC e *Theia3D*, il confronto ha evidenziato uno scarto quadratico medio inferiore ai 25 mm di distanza per centri articolari corrispondenti, tranne per l'anca che risulta di 36 mm. Gli angoli articolari ricostruiti, relativi al sistema di riferimento globale, hanno indicato una stima della posa dei segmenti molto simile fra entrambi i sistemi, con una differenza quadratica media inferiore ai 5.5 gradi per tutti gli angoli, eccetto quelli che rappresentano rotazioni attorno all'asse longitudinale del segmento [8]. Questi risultati, insieme ad altri studi concentrati sulla ripetibilità inter-sessione del MMC e sulla possibilità di valutare il *gait* ai livelli dei sistemi *marker-based*, indicano che il *motion capture* con *Theia3D* può essere un'alternativa adeguata nei casi in cui sono preferibili i benefici pratici della raccolta dati con un sistema *markerless*.

Le prestazioni di questa tecnologia sono dovute all'attenzione degli sviluppatori ad applicazioni nel campo della biomeccanica. Ogni aspetto di *Theia3D*, dalle *features* bidimensionali individuate, alla definizione dei segmenti e vincoli articolari, è stato selezionato per garantire precisione biomeccanica e risultati all'altezza delle pratiche e convenzioni standard. *Theia3D* si distingue inoltre da altri approcci al MMC tramite la

sua abilità nel fornire soluzioni accurate ma comunque generalizzate per il settore, altri sistemi markerless specializzati sono spesso limitati a particolari attività motorie, mentre Theia3D è in grado di riconoscere e tracciare un'ampia gamma di movimenti in quasi qualsiasi ambiente.

Risulta, ad oggi, molto raro, però, trovare studi che abbiano ottenuto risultati particolarmente soddisfacenti con soggetti patologici.

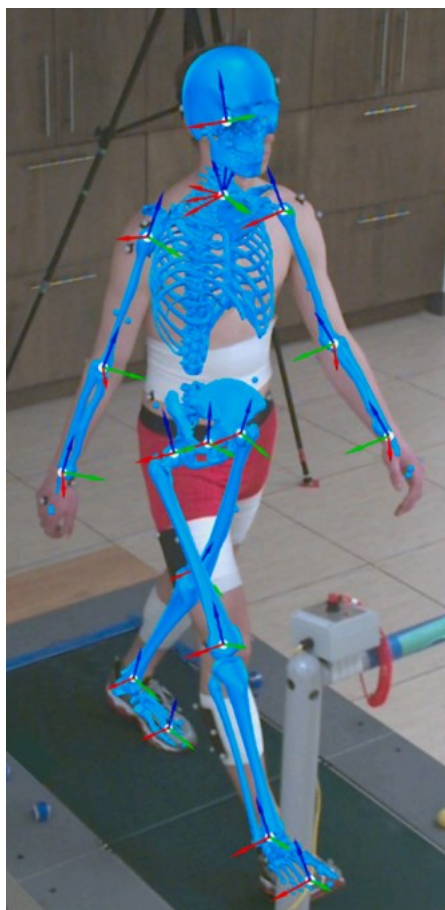


Figura 3.1: Modello scheletrico ottenuto con Theia3D. [8]

### 3.2.2. OpenCap

I sistemi di MMC in commercio sono in grado di stimare con buona precisione le variabili cinematiche, ma solitamente richiedono comunque un gran numero di telecamere cablate, software proprietari e specifiche risorse computazionali [13]. È per ovviare a quest'ultime limitazioni e compiere un ulteriore passo avanti nel campo del MMC che è stato introdotto *OpenCap*. *OpenCap* è un software *open-source*, *web-based*, disponibile a chiunque e capace di stimare variabili cinematiche e cinetiche del movimento in tre dimensioni a partire dai video girati con due o più smartphone [1].

OpenCap sfrutta i progressi in computer vision e simulazione muscoloscheletrica per rendere l'analisi del movimento disponibile senza il bisogno di hardware, software e competenze specializzati.

Le caratteristiche principali di quest'applicazione possono essere riassunte nei seguenti punti:

- basso costo e facilità d'uso: il sistema richiede solo due dispositivi con sistema operativo iOS (iPhone, iPad), dei tripodi per supportarli, un computer e una connessione a Internet. L'impostazione del sistema richiede pochi minuti e l'utente è guidato durante tutte le fasi d'utilizzo, dalla calibrazione alla visualizzazione delle cinematiche 3D;
- accuratezza di moti e forze: il movimento dello scheletro è ottenuto da un modello biomeccanico vincolato, le forze muscoloscheletriche sono stimate da simulazioni dinamiche;
- veloce e portatile: è possibile raccogliere dati ovunque sia presente una connessione a Internet, in cinque minuti il sistema è operativo e i risultati sono calcolati in automatico facendo affidamento su risorse nel *cloud*.

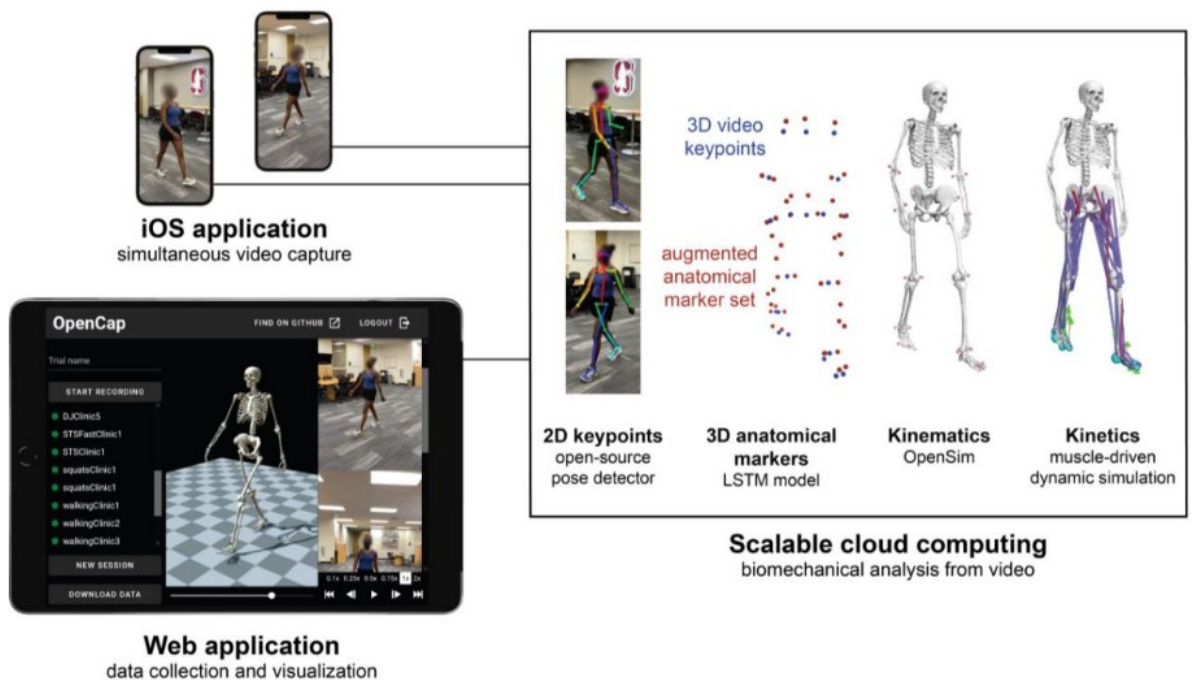


Figura 3.2: Ecosistema Opencap. [10]

Uno studio è stato condotto dalla *Stanford University* (Uhlrich et al., 2022) per validare l'accuratezza di OpenCap, confrontando la stima di parametri cinematici e cinetici con quelli ottenuti tramite MMC e MBMC.

I risultati di OpenCap sono stati ricavati utilizzando due iPhone per valutare diversi tipi di movimento (camminata, squat, leva da seduto e atterraggi dopo un salto) su un campione di dieci individui sani. Rispetto al MBMC, gli angoli articolari stimati dalla tecnologia *markerless* presentavano un errore assoluto medio (MAE) di 4.5 gradi, le forze di reazione al suolo un MAE pari al 6.2% del peso corporeo, e i momenti articolari un MAE pari al 1.2% del peso corporeo moltiplicato per l'altezza del soggetto [10].

La precisione delle stime di OpenCap è simile anche allo stato dell'arte del MMC, e gli errori osservati sono simili a quelli riportati per soluzioni a base di unità di misura inerziale [10]. Tali risultati dimostrano il potenziale per un utilizzo in campo clinico, come strumento di supporto alle decisioni del medico, o per la valutazione dell'efficacia di interventi e percorsi di riabilitazione.

### **3.3. Limiti attuali delle soluzioni Markerless**

I risultati ottenuti dalle tecnologie *markerless* non sono ancora sufficientemente accurati e affidabili per un impiego efficace in ambito clinico.

Un sistema ideale dovrebbe essere in grado di acquisire cinematiche articolari con un errore massimo accettabile inferiore ai cinque gradi [1]. Gli errori osservati negli studi presentati sono, nella maggior parte dei casi, superiori a questo limite e variano troppo da un caso all'altro. Tuttavia, gli algoritmi di *pose estimation* sono attualmente in una fase di rapida evoluzione, così come l'intero ambito dell'intelligenza artificiale, e di conseguenza possiamo prevedere che le prestazioni dei sistemi di MMC miglioreranno. Un fattore limitante per l'accuratezza è la diversa definizione dei sistemi di riferimento dei segmenti corporei tra MMC e stereofotogrammetria.

Il progetto a cui ho partecipato durante il periodo di tirocinio mira a sviluppare un sistema *markerless*, in grado di misurare con una precisione superiore ai sistemi già diffusi, affrontando questo limite con una rete neurale allenata con immagini etichettate con gli stessi protocolli (Figura 3.3) utilizzati per il sistema stereofotogrammetrico, colmando quel *gap* che si viene a verificare con tale *gold standard* rispetto alla definizione dei sistemi di riferimento per il calcolo degli angoli di rotazione articolare.

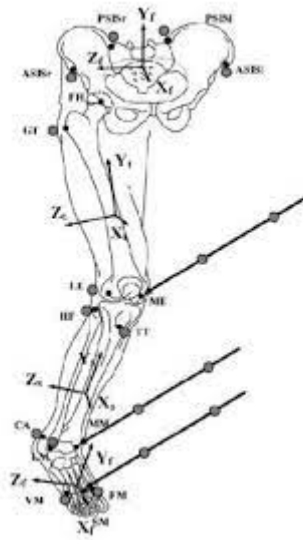


Figura 3.3: Protocollo IOR Gait. [9]



## 4. MATERIALI E METODI

L'obiettivo dell'attività di tirocinio è stato contribuire alla predisposizione di un *dataset* per il *training* della rete neurale su cui si basa un sistema di MMC.

Il progetto è frutto di una collaborazione tra il BioMovLab e l'azienda BBSof – Biomechanics Beyond Sports on Field, spin-off dell'Università degli Studi di Padova: esso consiste nello sviluppo di un sistema per MMC basato su una rete neurale di tipo convoluzionale allenata con immagini etichettate seguendo un protocollo ben preciso, al fine di ridurre l'errore nella stima di centri articolari e angoli ottenuti con MMC rispetto alla stereofotogrammetria.

### 4.1. SOGGETTI

Le immagini utilizzate sono state ricavate dalle registrazioni video di 12 soggetti impegnati in diversi *task* motori.

Per la validazione dei dati di cammino (*gait*) sono state utilizzate le acquisizioni di 5 soggetti, tutte ottenute in laboratori adibiti al *motion capture*. Per ogni soggetto sono state fornite le registrazioni di più di una camminata, catturate tramite quattro diverse videocamere poste in modo da evitare occlusioni di punti di interesse sul corpo del soggetto.

Per quanto riguarda l'applicazione all'esterno in ambiente non controllato, sono stati acquisiti 6 soggetti, giocatori di rugby, durante un movimento di rapido cambio di direzione con palla ovale in mano, simulando quindi un gesto tipico dello sport. Le riprese sono state fatte all'aperto, direttamente sul campo da gioco, con un setup delle telecamere simile a quello precedente.

Infine, per l'attività di pattinaggio sul ghiaccio, sono state elaborate le immagini di un'atleta che esegue una piroetta (*axel*), riprese sempre da quattro telecamere in una pista da pattinaggio al chiuso.

Per ogni soggetto i dati grezzi sono stati forniti all'interno di cartelle con una struttura predefinita necessaria al funzionamento del programma utilizzato per l'elaborazione delle immagini. All'interno della cartella di ogni soggetto si trova un primo livello di sotto-cartelle, ognuna delle quali contiene i dati di un solo *task* acquisito, ad esempio i dati solo della prima camminata del soggetto. Per ogni *task* l'elaborazione genera nuovi *file* organizzati nelle seguenti cartelle:

- '2Dannotations': contiene le posizioni sul piano bidimensionale dei punti etichettati;

- ‘*annotations*’: contiene le posizioni in tre dimensioni stimate dall’intelligenza artificiale;
- ‘*blurred frames*’: contiene i *frame* in cui viene oscurato il viso del soggetto per proteggerne la privacy;
- ‘*frames*’: contiene le immagini, estratte dalle riprese video, su cui avviene l’elaborazione.

## 4.2. LABELLING

L’elaborazione delle immagini consiste principalmente in un’attività di *labelling*, cioè l’individuazione di punti di interesse a cui viene associata una etichetta, o *label*. L’algoritmo di intelligenza artificiale ‘impara’ così a riconoscere e seguire i punti etichettati sulle immagini che gli vengono fornite.

In laboratorio è stato utilizzato un programma, scritto in linguaggio ‘*python*’, creato e curato da un membro del progetto stesso.

```
--> Commands:
--> 'r' - reassign visible labels to new coordinates
--> 'v' - reassign hidden labels to new coordinates
--> 'u' - reassign hidden labels automatically
--> 'h' - convert a visible point to hidden
--> 'i' - convert a hidden point to visible
--> 'x' - delete a label
--> 'z' - save image with automatically-detected blurred faces
--> 'b' - blur faces manually
--> 'c' - delete blurred image
--> 'y' - project points from 3D
--> 'e' - project points with Perspective-n-Point
--> 'g' - display skeleton graph
--> 'w' - define bounding box
--> 'a' - go to previous frame
--> 'd' - go to next frame
--> 'n' - go to first frame of previous camera
--> 'm' - go to first frame of next camera
--> 'j' - go to current frame of previous camera
--> 'k' - go to current frame of next camera
--> 't' - refresh image
--> 'q' - quit
```

Figura 4.1: Lista dei comandi forniti dal programma.

Una volta avviato il programma, indirizzato sulla cartella di lavoro desiderata, si apre una finestra con il primo frame della prima telecamera. Tramite dei comandi accessibili da tastiera (figura 4.1) è possibile navigare tra i *frame*, cambiare telecamera, *blurrare* porzioni dell’immagine, modificare la posizione delle etichette, visualizzare le

traiettorie ricostruite dalla rete neurale e creare tabelle utili a controllare la regolarità dell'elaborazione svolta.

Attivando il comando per la modifica della posizione delle *label* si apre una seconda finestra più piccola, dalla quale possiamo selezionare l'etichetta del punto di reperi anatomico che desideriamo individuare sull'immagine.

All'avvio, i primi *frame* di ogni telecamera contengono tutti i *marker* della lista, pertanto il primo passaggio è l'eliminazione di quei punti che sono occultati dal punto di vista della telecamera, sempre tramite un comando attivabile da tastiera. Una volta ripulito il primo *frame* di una telecamera si può procedere con la modifica delle posizioni delle *label* rimaste, individuando i punti di interesse per via visiva, con l'ausilio dei marcatori fisici riconoscibili sui corpi dei soggetti. Di vitale importanza si è rivelata la funzione di ingrandimento sull'immagine, dato che ha permesso una più facile e precisa determinazione delle posizioni dei *marker*, altrimenti troppo piccoli nella risoluzione originale dei *frame*.



Figura 4.2: Esempio di finestra durante il *labelling* di un *frame*.

Sui *frame* successivi al primo vengono generate, dall'algoritmo di intelligenza artificiale, di volta in volta, le *label* stimandone posizione e numero in base alle immagini elaborate precedentemente. Questo permette di seguire in tempo reale le prestazioni della rete neurale, riducendo il tempo necessario al *labelling* e aiutando l'operatore a rilevare eventuali errori commessi nell'etichettatura di *frame* precedenti, dato che tali errori andranno a propagarsi nelle immagini successive.

Passaggio fondamentale per ogni frame è il *blurring* del volto del soggetto, se visibile che, come anticipato, è necessario per garantire la privacy dei soggetti acquisiti. L'apposito comando implementato nel programma permette di selezionare un'area dell'immagine da oscurare.

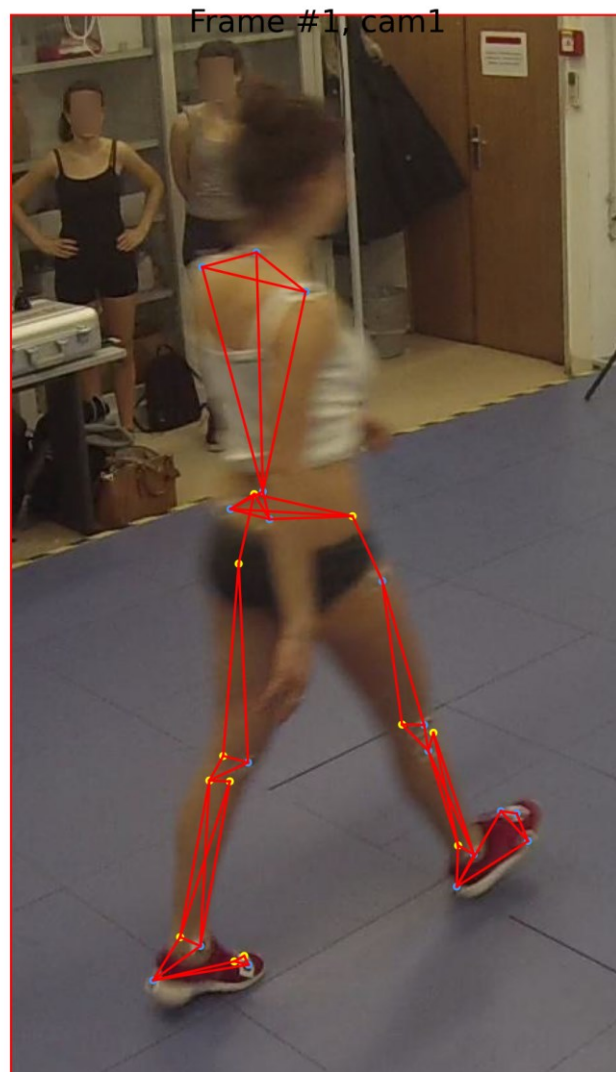


Figura 4.3: Scheletro del modello stimato dalla rete neurale (in rosso).

Terminata l'elaborazione di un task motorio, cioè, terminati il *labelling* e il *blur* in tutti i *frame* di tutte e quattro le telecamere, si procede con un controllo dei requisiti per la ricostruzione tridimensionale del modello cinematico. Il programma fornisce un comando per la creazione di una tabella che riassume la presenza o meno di ogni *label*, in ogni *frame* di ogni telecamera. Questo permette di verificare il requisito fondamentale per la stima delle posizioni in un volume, ovvero che ogni *marker* sia visto in ogni istante da almeno due telecamere. Nella tabella vengono evidenziate in rosso le etichette che non rispettano questo vincolo, indicando anche i frame in cui manca una delle viste necessarie.

Il programma, infine, permette di visualizzare, in sovraimpressione ad ogni *frame*, il modello stimato dall'intelligenza artificiale usando le immagini elaborate. Ciò consente all'utente che sta effettuando il *labelling* di verificare la qualità e la correttezza del suo operato.

## 5. RISULTATI

I risultati dell'attività di tirocinio sono presentati sotto forma di comparative fra le predizioni del modello, ottenute con il *training* basato sulle immagini elaborate, e il *gold standard*, costituito, in questo caso dal corrispettivo *tracking* ottenuto tramite operazione manuale, frame per frame, eseguita su un software dedicato, sviluppato sempre in collaborazione con BBSof S.r.l, TrackOnField. Nella figura 5.1 si possono osservare due ricostruzioni tridimensionali: in rosso quella ottenuta tramite operazione manuale, in blu quella ottenuta tramite l'algoritmo allenato con le nostre *label*.

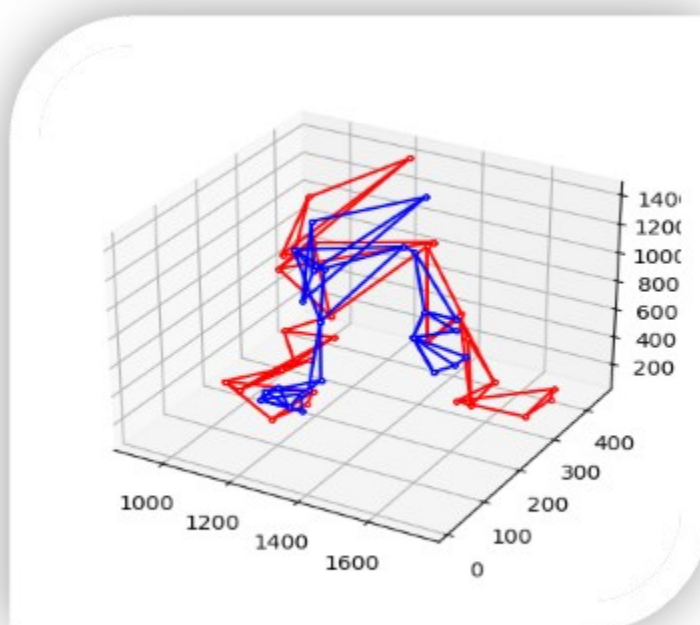


Figura 5.1: predizioni del *gold standard* (in rosso) e del modello AI (in blu).

Un'ulteriore metrica qualitativa è data dalle predizioni delle posizioni dei *marker* sui soggetti, presentata sotto forma di *heatmaps* in sovrapposizione alle immagini. Le *heatmaps* ottenute dalle predizioni del modello sono valutate in relazione a quelle ottenute sugli stessi *trial*, ma dal *gold standard*, che viene preso come riferimento.

Di seguito vengono riportate le immagini considerate, suddivise a coppie: ogni coppia di immagini riporta lo stesso *trial* di un soggetto, e in ogni riga a sinistra si trova il risultato ottenuto con l'algoritmo allenato e a destra il riferimento del *gold standard*.

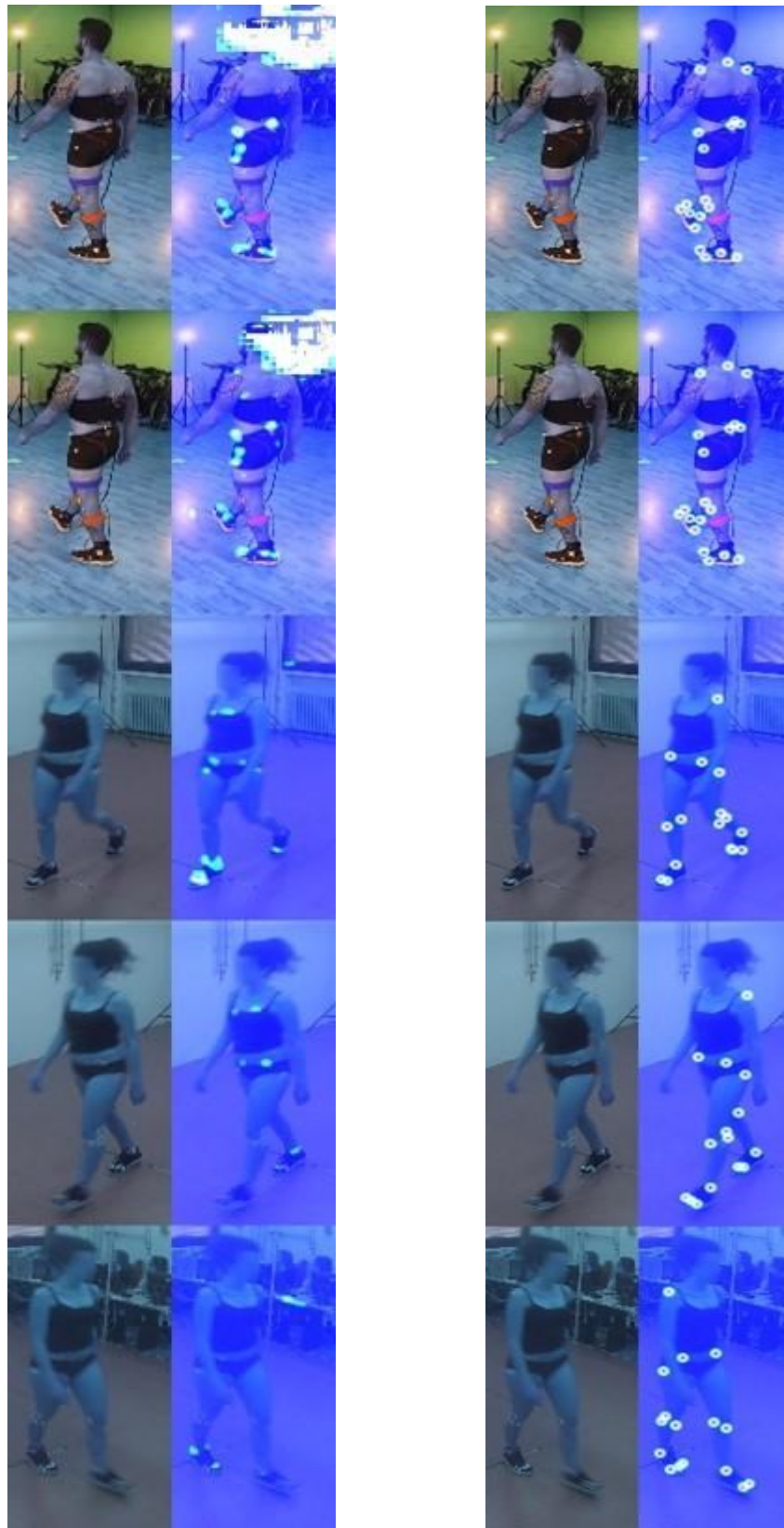


Figura 5.2: *heatmaps* delle predizioni del modello (a sinistra).

Figura 5.3: *heatmaps* di riferimento (a destra).

## 6. CONCLUSIONI

Per questo progetto di tesi sono state elaborate le immagini del movimento di 12 soggetti, 4 *trial* per ognuno di loro, e per ogni *trial* un numero che varia dagli 8 ai 34 *frame* acquisiti. Questa attività è stata condotta allo scopo di contribuire al *dataset* utilizzato per allenare la rete neurale di un modello adibito al MMC.

Nel complesso l'attività di elaborazione delle immagini si è svolta con successo, senza incontrare particolari ostacoli e riuscendo nel suo intento di creare immagini utili al progetto.

Una valutazione della precisione delle ricostruzioni rispetto al *gold standard* risulterebbe prematura nello stato attuale del sistema. I confronti con i dati di riferimento in questa fase del progetto sono utili per individuare eventuali problematiche di carattere tecnico, per evidenziare i progressi ottenuti nel tempo e seguire l'andamento delle prestazioni.

Tuttavia, da un'analisi qualitativa, si può comunque osservare come, nonostante un *dataset* di training ridotto, alcuni specifici punti di repere anatomico (si veda Figura 5.2), vengano individuati con particolare precisione da parte della rete neurale se confrontati con la *ground truth* (Figura 5.3). Inoltre, analizzando la posa descritta in Figura 5.1, e comparandola con il risultato ottenuto tramite tracking manuale a video, si nota come la forma restituita dalla rete abbia comunque un senso fisiologico, nonostante l'offset rispetto al riferimento in rosso, segnale che le funzionalità implementate all'interno della rete e le caratteristiche del *dataset* di training fornito sono conformi allo scopo prefissato.

La mancata individuazione di alcuni marcatori da parte del modello può essere attribuita in larga parte alla scarsa quantità di immagini attualmente utilizzate per allenare la rete neurale, si può predire che le prestazioni miglioreranno all'aumentare delle dimensioni del *dataset* usato per il *training*.



## BIBLIOGRAFIA

1. Brian Horsak, Anna Eichmann, Kerstin Lauer, Kerstin Prock, Philipp Krondorfer, Tarique Siragy, Bernhard Dumphart; *Concurrent validity of smartphone-based markerless motion capturing to quantify lower-limb joint kinematics in healthy and pathological gait*; 'Journal of Biomechanics', Volume 159, 2023.
2. Cappello A., Cappozzo A., Di Prampero P.E.; *Bioingegneria della postura e del movimento*; Patron Editore, 2003.
3. Cappozzo A., Della Croce U., Leardini A., Chiari L.; *Human movement analysis using stereophotogrammetry Part 1: theoretical background.*; Gait and Posture, 2005.
4. Chiari L., Della Croce U., Leardini A., Cappozzo A.; *Human movement analysis using stereophotogrammetry Part 2: Instrumental errors.*; Gait and Posture, 2005.
5. David A. Winter; *Biomechanics and Motor Control of Human Movement*, Fourth Edition, University of Waterloo, Ontario, Canada.
6. Leardini A., Chiari L., Della Croce U., Cappozzo A.; *Human movement analysis using stereophotogrammetry Part 3. Soft tissue artifact assessment and compensation.*; Gait and Posture, 2005.
7. Rigoni G.; *Sviluppo di un indice di rischio di infortunio al legamento crociato anteriore tramite analisi del movimento*, tesi di laurea magistrale, unipd, rel. Sawacha Z., 2023
8. Robert M. Kanko, Elise K. Laende, Elysia M. Davis, W. Scott Selbie, Kevin J. Deluzio; *Concurrent assessment of gait kinematics using marker-based and markerless motion capture*, 'Journal of Biomechanics', Volume 127, 2021.
9. Sawacha Zimi; *Dispense del corso di Bioingegneria del movimento e della riabilitazione*; Università degli studi di Padova, 2021.

10. Scott D. Uhlich, Antoine Falisse, Łukasz Kidziński, Julie Muccini, Michael Ko, Akshay S. Chaudhari, Jennifer L. Hicks, Scott L. Delp; *OpenCap: 3D human movement dynamics from smartphone videos*, 2022
11. Stefano Corazza, Lars Mündermann, Emiliano Gambaretto, Giancarlo Ferrigno, Thomas P. Andriacchi; *Markerless Motion Capture through Visual Hull, Articulated ICP and Subject Specific Model Generation*, 'Int J Comput Vis' 87, 2010.
12. Theresa E. McGuirk, Elliott S. Perry, Wandasun B. Sihanath, Sherveen Riazati and Carolyn Patten; *Feasibility of Markerless Motion Capture for Three-Dimensional Gait Assessment in Community Settings*, 'Frontiers in Human Neuroscience', 2022.
13. Uhlich SD, Falisse A., Kidziński Ł., Muccini J., Ko M., Chaudhari AS, et al.; *OpenCap: Human movement dynamics from smartphone videos.*; 'PLoS Comput Biol', 2023.
14. [www.opencap.ai](http://www.opencap.ai)
15. [www.theiamarkerless.ca/why-theia3d](http://www.theiamarkerless.ca/why-theia3d)
16. Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, Yaser Sheikh; *OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields*, 'IEEE transactions on pattern analysis and machine intelligence', vol. 43, no. 1, 2021.