



UNIVERSITÀ
POLITECNICA
DELLE MARCHE

FACOLTÀ DI INGEGNERIA

CORSO DI LAUREA IN INGEGNERIA INFORMATICA E DELL'AUTOMAZIONE

Algoritmi di Machine Learning e Sensori RGB-D per il Rilevamento del Segnale Respiratorio

MACHINE LEARNING ALGORITHMS AND RGB-D SENSORS FOR
RESPIRATORY SIGNAL DETECTION

Laureando:

Anass Chebbaki

Relatore:

Prof. Ennio Gambi

Correlatore:

Dr.ssa Linda Senigagliesi

Anno Accademico 2023-2024

UNIVERSITÀ POLITECNICA DELLE MARCHE
FACOLTÀ DI INGEGNERIA
CORSO DI LAUREA IN INGEGNERIA INFORMATICA E DELL'AUTOMAZIONE
Via Brezze Bianche – 60131 Ancona (AN), Italy

*“A Stack Overflow, per avermi insegnato che non esistono
domande stupide ma solo risposte complicate.”*

Indice

1	Introduzione	1
1.1	Stato dell'arte	2
1.1.1	YOLOv8 Ultralytics	2
1.1.2	OpenCV	3
1.1.3	Approcci simili in letteratura	4
2	Descrizione del Setup Utilizzato	9
2.1	Intel RealSense™ Depth Camera D455	9
2.2	Raccolta dei Dati	13
3	Estrazione del Segnale Respiratorio	17
3.1	Metodi per l'Estrazione della ROI	19
3.2	Elaborazione del Segnale Respiratorio Grezzo	23
3.2.1	Interpolazione dei Dati	24
3.2.2	Detrending e Filtraggio	25
4	Risultati e Discussione	27
4.1	Sensibilità a Distanza e Indumento	28
4.2	Sensibilità Luce	33
4.3	Stima di un Indice di Qualità	35
4.4	Prospettive Future	37
5	Conclusioni	39
	Bibliografia	41
	Ringraziamenti	45

Elenco delle figure

1.1	Schema dell'algoritmo YOLO.	2
1.2	YOLOv8 Pose	3
1.3	OpenCV	3
2.1	Intel RealSense D455.	9
2.2	Schema illustrativo.	10
2.3	Invalid Depth Band.	11
2.4	Setup Sperimentale.	14
2.5	Gerarchia delle directory.	15
3.1	Disposizione dei keypoints.	17
3.2	Diagramma di Flusso.	18
3.3	Keypoints selezionati.	19
3.4	1°Modello.	20
3.5	2°Modello.	21
3.6	3°Modello.	21
3.7	4°Modello.	22
3.8	Falso positivo.	23
3.9	Esempio di Interpolazione Lineare.	24
3.10	Azione del Filtro Mediano.	25
3.11	Esempio di Incremento Interpolazioni.	26
4.1	Disposizione dei grafici.	27
4.2	Proiettore Infrarossi (IR).	33
4.3	Esempio Maschera $M(f)$	35

Capitolo 1

Introduzione

Un monitoraggio accurato della frequenza respiratoria è cruciale in numerose applicazioni mediche. Tuttavia, i metodi tradizionali spesso richiedono l'uso di dispositivi che, sebbene non siano invasivi, possono essere scomodi o limitati. In risposta a questa sfida, la ricerca si è orientata verso l'utilizzo di tecnologie avanzate per il rilevamento del movimento toracico senza contatto diretto. Il presente studio si concentra sullo sviluppo di un sistema di monitoraggio respiratorio basato sull'utilizzo del sensore RGB-D **Intel RealSense™D455**. Questo dispositivo consente di acquisire dati di profondità ad alta risoluzione, consentendo una valutazione dettagliata dei movimenti toracici associati alla respirazione. Utilizzando algoritmi implementati in Python e sfruttando la potenza di **YOLO (You Only Look Once)** nella versione **v8 Pose**, sono stati estratti i segnali di respirazione dai dati di profondità forniti dalla telecamera. Lo scopo principale di questo studio è stato quello di valutare l'accuratezza e l'affidabilità del sistema proposto nel monitoraggio della frequenza respiratoria in diverse condizioni. Questo ha incluso la valutazione delle prestazioni del sistema a diverse distanze dal soggetto, nonché l'impatto di variazioni nelle condizioni di illuminazione e nell'abbigliamento del soggetto stesso. Inoltre, è stata presa in considerazione la precisione dei keypoints estratti da YOLO nel determinare i movimenti del torace. Queste valutazioni hanno richiesto un'approfondita analisi dei pattern respiratori e un'ottimizzazione degli algoritmi per garantire la massima precisione e affidabilità del sistema proposto. L'obiettivo ultimo è stato quello di fornire un metodo efficace e affidabile per il monitoraggio continuo della frequenza respiratoria, con possibili applicazioni in campo medico, assistenziale e di monitoraggio del benessere.

1.1 Stato dell'arte

In questo lavoro di tesi, sono state impiegate diverse tecnologie per l'elaborazione delle immagini RGB. Per l'analisi in tempo reale, YOLO v8 Pose, un potente framework di rilevamento fornito da **Ultralytics** è stato integrato con la libreria **OpenCV** in Python. Questa combinazione ha consentito di individuare pose umane e relativi keypoints nelle immagini catturate in tempo reale. Per l'analisi offline, è stato utilizzato l'SDK fornito direttamente da Intel per acquisire le immagini iniziali e successivamente elaborarle. Questo approccio ha consentito di applicare algoritmi più complessi e computazionalmente intensivi su un set di dati statico, senza le limitazioni temporali dell'analisi in tempo reale.

1.1.1 YOLOv8 Ultralytics

YOLO (*You Only Look Once*) di Ultralytics è un modello di Computer Vision basato su reti neurali convoluzionali (CNN), noto per la sua particolare velocità e accuratezza nel rilevamento in tempo reale di oggetti all'interno di immagini e video. L'architettura di YOLOv8, come citato nell'articolo [1], può essere schematizzata come nella seguente figura [1.1].

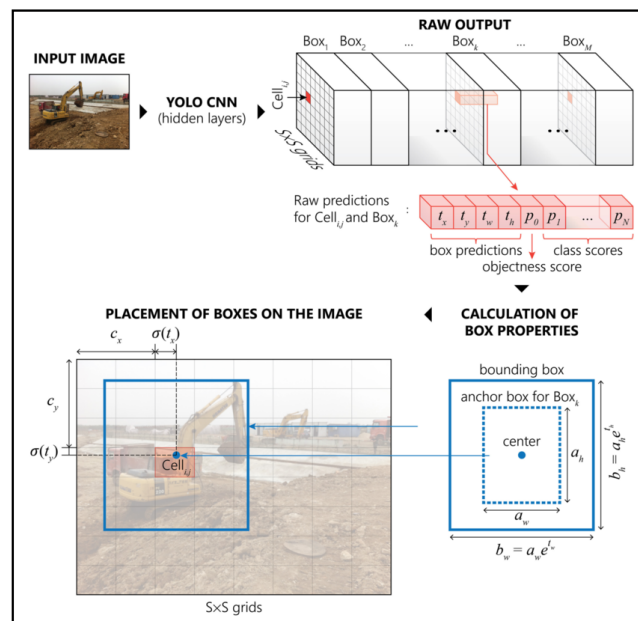


Figura 1.1: Schema dell'algoritmo YOLO.

YOLOv8 Pose [2] è invece una variante del popolare framework di object detection, progettata specificamente per il rilevamento e l'analisi delle pose umane nelle immagini. Utilizza un modello di deep learning per individuare la posizione di punti chiave anatomici sul corpo umano, come articolazioni e caratteristiche distintive che svolgeranno un ruolo fondamentale nell'obiettivo della mia ricerca. Questi punti vengono identificati nelle immagini insieme alla relativa certezza di accuratezza. YOLO Pose è efficace per l'analisi delle pose in tempo reale e può essere utilizzato in una vasta gamma di applicazioni, come il monitoraggio del movimento, la sorveglianza e l'analisi del gesto umano



Figura 1.2: YOLOv8 Pose

1.1.2 OpenCV

OpenCV è una libreria open source per la visione artificiale e l'elaborazione delle immagini. Offre strumenti per analizzare, manipolare e comprendere le immagini attraverso algoritmi e tecniche di visione artificiale. È ampiamente utilizzata per sviluppare applicazioni di riconoscimento facciale, tracciamento di oggetti, correzione dell'immagine e altro ancora.



Figura 1.3: OpenCV

1.1.3 Approcci simili in letteratura

L'analisi del respiro con sensori RGB-D richiede una comprensione dei precedenti approcci in letteratura. La tabella [1.1] esamina i principali studi, evidenziando differenze in metodi, algoritmi, strumenti e dataset.

Tabella 1.1: Stato dell'arte.

Ref.	Sensori Utilizzati	Dataset	Sensore e segnale di riferimento	Algoritmo	Performance
[3], 2022	Intel RealSense D455	3 sogg. maschili (20 anni); 2 sogg. maschili (40 anni); Ambiente sperimentale: Laboratorio	Spirometro (Spiro Sift SP-370COPD), segnale respiratorio	Skeleton Tracking SDK by CUBEMOS per definire la ROI	Discrete, la tenuta dello spirometro nelle mani ha ridotto l'accuratezza del rilevamento dei keypoints
[4], 2022	Microsoft Azure Kinect e 3 Radar IR-UWB	5 sogg. maschili; 5 sogg. femminili; fino a due soggetti simultaneamente; Ambiente sperimentale: Lab. di ricerca	Go Direct, Fascia respiratoria	Azure Kinect Body Tracking SDK per l'estrazione dei punti anatomici; Segnale radar dal sensore radar meglio posizionato per l'estrazione del respiro; Random Forest, SVM, KNN, RNN per classificazione dei pattern respiratori	errore assoluto medio da 0.2 a 1.92 respiri/min singolo soggetto, da 0.3 a 1.15 respiri/min per soggetti multipli
[5], 2023	Intel RealSense D415	7 sogg. sani; Ambiente clinico in diverse condizioni di illuminazione e posture	Un capnografo (Capnostream 35, Medtronic, Boulder, CO) ed uno spirometro (SpiroFlo, SpiroSonic, Uscom Kft, Budapest) utilizzati per la validazione dei dati	La ROI è stata definita manualmente mediante il posizionamento del paziente.	Per avere una buona validità della frequenza respiratoria debba essere compresa tra 4 e 40 respiri/min
[6], 2020	2 dispositivi Microsoft Kinect v2	1 sogg. maschio; 1 sogg. femmina; Manichino per eseguire esperimenti; Ambiente: ospedaliero (terapia intensiva pediatrica)	Respiratore meccanico (Servo i, Maquet Inc, Sweden) il confronto dei dati;	Allineamento dei due sensori per la rilevazione della ROI, e in seguito tramite SDK della Kinect avviene l'estrazione della nuvola dei punti (k-NN), della quale viene misurata la variazione tramite metodo di ricostruzione di Poisson.	Paziente 1: gli errori medi sono stati di 5.4 mL per Vt, 0.75 respiri al minuto per RR e 0.22 L/min per MV; Paziente 2: sono stati di 6.4 mL per Vt, 0.68 respiri al minuto per RR e 0.24 L/min per MV. Questi risultati confermano la capacità del sistema di fornire misurazioni precise e affidabili anche in condizioni reali.
[7], 2020	Microsoft Kinect	84 volontari; Ambiente sperimentale interno ad una temperatura di 26°C.	"Biopac MP150" per la monitoraggio del respiro e validazione dei dati	Skeletal Tracking di Kinect per la determinazione dei keypoints e la ROI mentre per la classificazione delle caratteristiche sono stati utilizzati SVM e Random Forest	Accuratezza media del 90%
[8], 2020	Microsoft Kinect	39 volontari con indumenti differenti.	Spirometro collegato ad un computer secondario per poter sincronizzare Kinect e il sensore di riferimento tramite un segnale di trigger.	"Cascade Classifier (OpenCv)" utilizzato per il riconoscimento della parte posteriore del torace	Il risultato sperimentale ha mostrato che la stima RR del loro algoritmo aveva un tasso di errore massimo dell'1,5% (cinque errori su un totale di 344 respiri)
[9], 2022	Microsoft Kinect v2	1 soggetto (58 anni, 75kg); Ambiente sperimentale: Lab. di ricerca.	Il volontario respirava con un ventilatore (Puritan Bennett™ 980, Medtronic, Carlsbad, CA)	Per l'estrazione del respiro è stata creata manualmente la ROI (svantaggioso) sulla zona toracica del paziente e con il supporto di un algoritmo scritto in linguaggio C++ è stato possibile estrarre la frequenza respiratoria.	Alto grado di accordo tra la frequenza respiratoria determinata dalla Kinect e il segnale di riferimento, dove è stato raggiunto un RMSD (Root Mean Square Deviation) di 0.51 respiri/min corrispondeva a stretti limiti di accordo compresi tra -1.02 e 0.98 respiri/min
[10], 2020	Microsoft Kinect v2	24 sogg. (di cui 7 donne, con un'età compresa tra i 22 e i 57 anni); Ambiente sperimentale ben illuminato al chiuso	Cintura respiratoria (Verrier GDX-RB)	Tramite "Kinect SDK 2.0" vengono estratti i dati di profondità e tramite delle "maschere di occlusione" viene definita la ROI	I risultati mostrano che la frequenza respiratoria può essere rilevata con il nostro metodo con una precisione compresa tra il 92 e il 97% da una distanza di due metri
[11], 2020	Microsoft Kinect v2	14 soggetti in posizione leggermente reclinata; Ambiente sperimentale clinico.	Capnografo (modello non specificato)	Non specificato l'algoritmo di estrazione della ROI	I risultati hanno mostrato una forte concordanza tra il tasso respiratorio calcolato dall'algoritmo (RRdepth) e quello misurato dal capnografo di riferimento (RRcap). La differenza media tra i due era di soli 0.04 bpm, con una deviazione media quadratica di 0.66 bpm
[12], 2018	Microsoft Kinect v2, RGB-D Asus Xtion Pro	7 soggetti seduti a varie distanze; Ambiente sperimentale: Lab. di ricerca	Non viene specificato nessun sensore di riferimento per la validazione dei dati	In una prima fase viene processata l'immagine per togliere eventuali rumori mentre "Kinect SDK 2.0" fornisce i punti di giunzione dai quali viene estratta la regione di interesse che è definita come un rettangolo generato a seconda delle proporzioni del soggetto.	A distanze ravvicinate si ha un elevato grado di precisione ma all'aumentare della distanza (circa 4m) si notano degradi nel rilevamento del segnale respiratorio acquisito.

Notiamo che mentre alcuni studi hanno impiegato sensori come le “Depth Camera”, concentrandosi sull’analisi dei movimenti del torace o dell’addome, altri potrebbero aver utilizzato Microsoft KinectTM per ragioni economiche. È importante anche considerare le performance degli algoritmi utilizzati, indicando le metriche di valutazione utilizzate per confrontarli in termini di accuratezza, velocità e robustezza. Vediamo infatti che lo studio che include più soggetti, dunque con un dataset molto ampio, è quello di *Yuhao Shan, Shigang Li, and Tong Chen* [7], che prende in riferimento 84 volontari in un ambiente sperimentale interno ottenendo un’accuratezza del 90%. Questo articolo presenta un framework per la rilevazione e la classificazione dello stress umano utilizzando il sensore Kinect. Utilizzando *Skeletal Tracking* per la determinazione dei keypoints e la definizione della *Region of Interest* (ROI), l’articolo ha applicato algoritmi di machine learning come *Support Vector Machine* (SVM) e *Random Forest* per la classificazione delle caratteristiche. Rispetto agli altri articoli, questo si concentra sull’analisi delle caratteristiche della respirazione in relazione allo stress psicologico, ottenendo buoni risultati nella classificazione dei diversi stati emotivi.

L’articolo in [3] propone un sistema di misurazione della funzione respiratoria senza contatto utilizzando la telecamera a profondità RealSense D455 di Intel. Attraverso il tracciamento scheletrico che avviene tramite *Skeleton Tracking SDK di Cubemos*, l’articolo dimostra la capacità di misurare lo sforzo respiratorio in modo non invasivo. Tuttavia, si nota una ridotta accuratezza quando lo spirometro è tenuto nelle mani rendendo difficile il rilevamento dei keypoints. D’altra parte, l’articolo [4] affronta il problema della stima senza contatto del tasso respiratorio e la classificazione dei pattern respiratori per uno o due soggetti in uno spazio confinato. Utilizzando tre radar IR-UWB e una Azure Kinect, l’articolo propone algoritmi avanzati per la localizzazione dei soggetti e la selezione del radar, ottenendo un’accuratezza accettabile nella stima del tasso respiratorio e nella classificazione dei pattern respiratori riportando un errore assoluto medio che varia da 0.2 a 1.92 respiri al minuto per i singoli soggetti e da 0.3 a 1.15 respiri al minuto per soggetti multipli. Mentre invece l’articolo [5] si concentra sull’accuratezza del monitoraggio continuo del tasso respiratorio mediante la telecamera RealSense D415, riportando un’alta precisione nella misurazione del tasso respiratorio in condizioni cliniche varie. Usando come sensore di riferimento un capnografo (Capnostream 35, Medtronic, Boulder,

CO) ed uno spirometro (SpiroFlo, SpiroSonic, Uscom Kft, Budapest) utilizzati per la validazione dei dati. In questo caso sette soggetti sani hanno eseguito una serie di respiri da 4 a 40 respiri al minuto in condizioni diverse, mentre la telecamera acquisiva informazioni sulla profondità della regione del torace. L'algoritmo quindi dopo il posizionamento del paziente nella zona di interesse elaborava questi dati per calcolare il tasso respiratorio (RRdepth) confrontandolo con quello di riferimento. L'accuratezza complessiva è stata elevata, con una radice dell'errore quadratico medio (RMSD) di 0,69 respiri al minuto e un intervallo di concordanza di -1,42 a 1,36 respiri al minuto. Inoltre, il sistema è risultato accurato anche in sotto-range di bassi, normali e alti tassi respiratori, ritenuti clinicamente rilevanti.

Una prospettiva diversa emerge dall'articolo [6], che descrive un nuovo approccio per valutare quantitativamente la respirazione in unità di terapia intensiva pediatrica, utilizzando due telecamere Kinect per registrare le sequenze video dei movimenti del torace e dei fianchi dei pazienti. L'articolo dimostra un accordo significativo tra il nuovo sistema di imaging 3D proposto e il metodo standard per l'analisi delle forme d'onda respiratorie. In un contesto più domestico, l'articolo [8] propone un algoritmo per stimare il volume corrente e il tasso respiratorio utilizzando la telecamera Kinect per misurare la respirazione a casa per i pazienti cardiopatici. Sebbene l'algoritmo mostri una buona precisione nella stima del tasso respiratorio, si notano grandi errori nella stima del volume corrente, soprattutto durante la respirazione profonda (Gli errori relativi assoluti medi della stima del *Volume Corrente* per i maschi era del 14,0% senza vestiti e del 10,7% con indosso la maglietta mentre, gli errori relativi ai soggetti di sesso femminile erano rispettivamente del 14,7% e del 15,5%). Continuando con l'esplorazione delle applicazioni cliniche, l'articolo [9] valuta le prestazioni di un sistema di monitoraggio della frequenza respiratoria basato sulla telecamera Kinect, confrontandolo con un ventilatore di riferimento (Puritan Bennett™ 980, Medtronic, Carlsbad, CA). L'articolo dimostra un alto grado di concordanza tra la frequenza respiratoria determinata dalla Kinect e il segnale di riferimento del ventilatore, suggerendo che la Kinect sia promettente per il monitoraggio preciso della frequenza respiratoria in ambito clinico. Un articolo che invece riporta un'accuratezza del 92-97% nella rilevazione del tasso respiratorio da una distanza di due metri è [10] che presenta un metodo innovativo per il monitoraggio della respirazione utilizzando la telecamera

Kinect a distanze fino a due metri e come sensore di riferimento una cintura respiratoria (Vernier GDX-RB). Infine, l'articolo [11] propone un algoritmo che mostra un'eccellente concordanza con il tasso respiratorio di riferimento determinato da un capnografo, suggerendo che la tecnologia della telecamera a profondità potrebbe essere altamente affidabile per il monitoraggio continuo della respirazione in ambienti clinici. Uno studio che invece utilizza un depth camera diversa da quelle precedentemente menzionate è l'articolo [12], come riportato nella tabella, esplora l'utilizzo delle telecamere Kinect v2 e RGB-D Asus Xtion Pro per monitorare la respirazione umana. Lo studio coinvolge sette soggetti seduti a varie distanze, con l'obiettivo di valutare la capacità di queste telecamere nel rilevare il tasso respiratorio. I risultati evidenziano che a distanze ravvicinate si ottiene un elevato grado di precisione nella rilevazione del segnale respiratorio, tuttavia si notano degni nella qualità del segnale all'aumentare della distanza, soprattutto oltre i 4 metri. In particolare, l'ampiezza del segnale respiratorio risulta influenzata dalla distanza del soggetto dalla telecamera e dalla frequenza di campionamento della stessa. L'algoritmo utilizzato consisteva in una prima fase dove l'immagine viene processata per togliere eventuali rumori, mentre in una seconda fase grazie "Kinect SDK 2.0" che fornisce i punti di giunzione del soggetto, si riusciva a creare la regione d'interesse. Quest'ultima è definita come un rettangolo generato a seconda delle proporzioni del soggetto. In breve, questi articoli mostrano diverse tecniche e dispositivi per il monitoraggio non invasivo della respirazione umana, come telecamere a profondità, radar IR-UWB e sensori di movimento. Pur avendo punti di forza e limitazioni specifiche, insieme forniscono una base robusta per questa tesi e per lo sviluppo futuro di sistemi di monitoraggio respiratorio più precisi e accessibili.

Capitolo 2

Descrizione del Setup Utilizzato

2.1 Intel RealSense™ Depth Camera D455

La Intel RealSense D455 è una telecamera di profondità stereo avanzata che offre una serie di caratteristiche impressionanti. Questo dispositivo è dotato di un sensore RGB a otturatore globale, un'unità di misura inerziale (IMU) e una lunga portata di 4 metri con basso errore. In termini di prestazioni, la D455 è in grado di fornire una risoluzione di output di profondità fino a 1280×720 e una frequenza dei fotogrammi di profondità fino a 90 fps. Il sensore RGB offre una risoluzione di 1 MP con una frequenza dei fotogrammi di 30 fps. Queste specifiche tecniche permettono di catturare immagini di alta qualità e di profondità con dettagli nitidi. Uno dei principali punti di forza di questo dispositivo è la sua versatilità. Può essere utilizzata sia in ambienti interni che esterni e supporta una vasta gamma di applicazioni, tra cui l'evitamento delle collisioni, la scansione 3D e la misurazione volumetrica. Inoltre, quest'ultima utilizza lo stesso software (*SDK Intel RealSense 2.0* [13]) open source di tutte le altre telecamere di profondità di generazione attuale.



Figura 2.1: Intel RealSense D455.

Capitolo 2 Descrizione del Setup Utilizzato

La tecnologia di profondità stereo della Intel RealSense D455 come mostrato in Figura [2.2] utilizza due telecamere per calcolare una stima della profondità di un determinato oggetto rispetto al sensore. Ecco il principio di funzionamento:

1. **Cattura delle immagini:** Le due telecamere catturano due immagini dello stesso oggetto da due punti di vista leggermente diversi, proprio come fanno i nostri occhi.
2. **Calcolo della disparità:** Il sistema calcola la differenza o “disparità” tra le due immagini.
3. **Calcolo della profondità:** Utilizzando la disparità e la geometria delle telecamere (la distanza tra le telecamere e l’angolo di visione), il sistema può calcolare il campo visivo di profondità (Depth FOV) a qualsiasi distanza (Z) con la seguente equazione [2.1]

$$\text{Depth FOV} = \frac{\text{HFOV}}{2} + \tan^{-1} \left(\tan \left(\frac{\text{HFOV}}{2} \right) - \frac{B}{Z} \right) \quad (2.1)$$

Dove:

Depth FOV = Campo Visivo in Profondità

HFOV = Campo Visivo Orizzontale dell’immagine sinistra sul modulo di profondità

B = Baseline

Z = Distanza della scena dal modulo di profondità

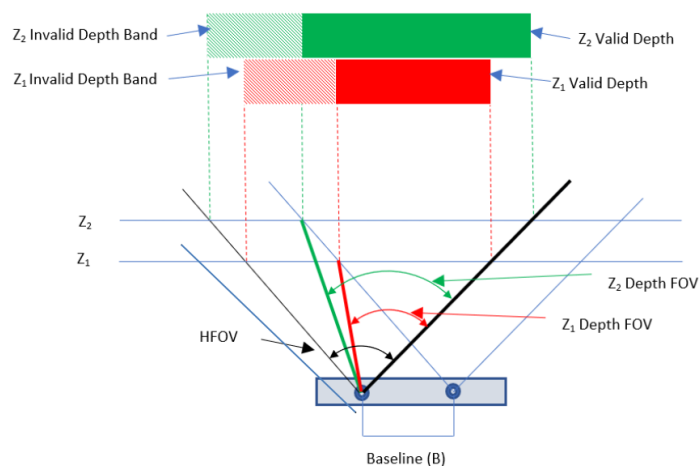


Figura 2.2: Schema illustrativo.

2.1 Intel RealSense™ Depth Camera D455

Come già accennato la telecamera è dotata di sensori di profondità distanziati di 95 millimetri, il che contribuisce a ridurre l'errore di misurazione della profondità a meno del 2% quando si misurano distanze di 4 metri. Inoltre, l'unità di misura inerziale consente all'applicazione di ottimizzare la percezione della profondità in qualsiasi situazione di movimento della telecamera.

Tuttavia, come con qualsiasi dispositivo, la D455 ha anche alcuni punti di debolezza. Ad esempio, la distanza di profondità minima (**Min-Z**) alla massima risoluzione è di circa 52 cm, il che potrebbe limitare la sua utilità in alcune applicazioni che richiedono una maggiore precisione a distanze più brevi. Inoltre, nonostante la sua lunga portata di 4 metri, l'accuratezza della profondità è inferiore al 2% a 4 metri. Un'altra considerazione da sottolineare è la cosiddetta "**Invalid Depth Band**" [2.3] che si riferisce a una porzione dell'immagine acquisita in cui la profondità dei pixel non è stata rilevata correttamente o non è affidabile. Questo può accadere per varie ragioni, ad esempio a causa di condizioni di illuminazione sfavorevoli, oggetti trasparenti o riflettenti nell'ambiente, o a causa di problemi tecnici nel funzionamento del sensore stesso. La larghezza di questa banda può variare a seconda della distanza della telecamera dall'oggetto che sta osservando, perciò è di fondamentale importanza effettuare una severa calibrazione del dispositivo prima dell'utilizzo.

In conclusione, la Intel RealSense D455 è un dispositivo potente e versatile che offre una serie di caratteristiche impressionanti. Nonostante alcune limitazioni, la sua flessibilità e le sue capacità avanzate la rendono una scelta eccellente per una vasta gamma di applicazioni.

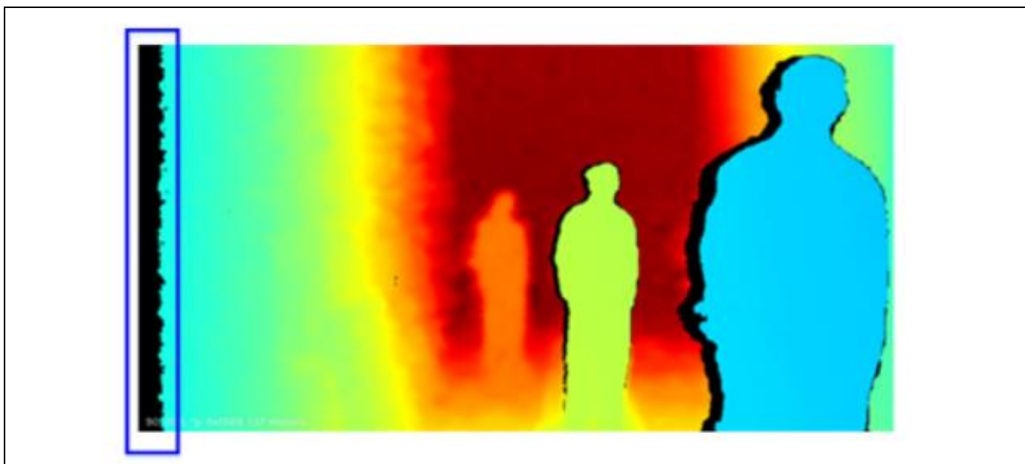
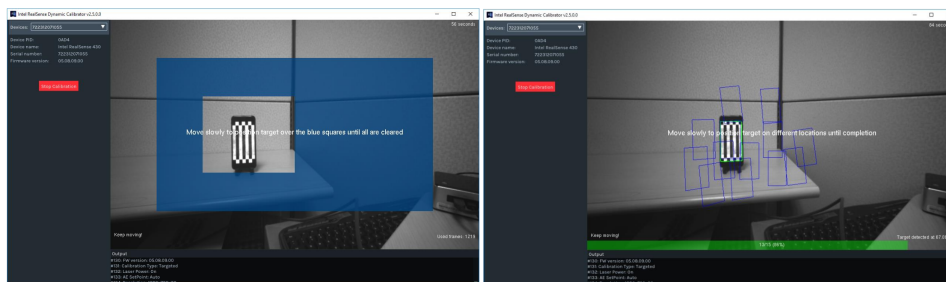


Figura 2.3: Invalid Depth Band.

Capitolo 2 Descrizione del Setup Utilizzato

Per ottenere stime precise sulla profondità, è consigliabile eseguire una procedura di **calibrazione** della telecamera. Una di queste è la calibrazione della telecamera tramite il software "**Dynamic Calibrator**" messo a disposizione da Intel stessa nella seguente guida [14]. La procedura si è suddivisa nei seguenti passaggi:

1. **Download del Target:** Prima della fase di calibrazione è utile scaricare e stampare il *Target* delle dimensioni adeguate che viene fornito nella guida già menzionata
2. **Regolazione dell'esposizione:** In questa fase l'esposizione viene regolata automaticamente per riconoscere il target di calibrazione in diverse condizioni di illuminazione. Se il target non è rilevato a causa di riflessi o illuminazione estrema, lo strumento regolerà l'esposizione finché non sarà individuato.
3. **Fase di rettifica:** Come mostrato in figura [2.4a] occorrerà puntare la fotocamera verso il target di calibrazione e muoverla lentamente per posizionare il target sui quadrati blu fino a quando tutti sono stati cancellati. Una volta eliminati tutti i quadrati blu, la fase di rettifica è completata.
4. **Fase di calibrazione della scala:** Dopo la rettifica, si avvia la calibrazione della scala (figura [2.4b]). Si deve puntare la telecamera verso il target e muoversi lentamente mentre una barra verde mostra il progresso. Verranno analizzate 15 immagini positive del target. Alla fine, la calibrazione del sensore sarà completata.



(a) Rettifica.

(b) Calibrazione della scala.

2.2 Raccolta dei Dati

In questa sezione, ci addentreremo nel processo di raccolta dei dati, che si avvale dell'uso di un file `.bag` precedentemente registrato. La registrazione è stata realizzata attraverso uno script che avvia automaticamente una registrazione di **trenta secondi** tramite l'SDK Intel RealSense Viewer, fornendo un segnale acustico per notificare all'utente l'inizio della registrazione. Al completamento della registrazione, il file viene conservato nella directory di destinazione per l'elaborazione successiva. Come precedentemente menzionato nel capitolo [1], l'algoritmo inizialmente raccoglieva i dati in tempo reale, analizzando i frame direttamente dalla telecamera. Tuttavia, abbiamo constatato che l'approccio offline garantisce una maggiore precisione, una latenza ridotta durante la fase di compilazione del codice, inoltre consente di verificare le performance dei quattro modelli nella singola acquisizione.

Le librerie che hanno reso possibile questo processo sono le seguenti:

- **Libreria `csv`:** Utilizzata per leggere e scrivere file `.csv` (*Comma Separated Values*), quest'ultima è ampiamente utilizzata per memorizzare dati tabulari in formato testuale. In questo contesto, la libreria `csv` serve per registrare i dati di profondità insieme ai relativi **timestamps** in due colonne distinte. Questo metodo di registrazione consente di strutturare i dati in un formato facilmente analizzabile e di garantire il plot di essi in un momento successivo.
- **Libreria `time`:** La libreria `time` di Python è uno strumento per la gestione del tempo nelle applicazioni. In particolare la funzione `time.perf_counter()` permette di misurare con precisione i timestamp relativi a ciascun frame acquisito, facilitando una registrazione accurata dei dati nel file `.csv`. Questo approccio assicura che ogni dato di profondità sia associato al corretto momento temporale, migliorando la qualità e l'utilità dei dati raccolti per l'analisi successiva.

Al termine dell'esecuzione del codice, il file viene automaticamente salvato nella cartella corrente, rendendolo facilmente accessibile per la fase di plot in un algoritmo secondario.

Setup Sperimentale

La seguente figura [2.4] rappresenta una ricostruzione accurata dell'ambiente in cui sono state effettuate le acquisizioni, seguendo le procedure descritte precedentemente. In questa fase, l'obiettivo principale è stato quello di raccogliere un dataset di acquisizioni il più variegato possibile, utilizzando indumenti diversi, varie distanze e differenti condizioni di illuminazione.

Per garantire una maggiore precisione, sono stati applicati dei marker adesivi sul pavimento, che indicano la distanza esatta dal sensore. Inoltre, per garantire una copertura visiva completa dell'individuo e permettere al modello YOLOv8 di determinare con precisione le pose e i keypoints, la telecamera è stata posizionata ad un'altezza di 0.90 m . Infine, la disposizione dei marker a distanze di 1.0 m , 2.0 m , 3.0 m e 2.5 m dal sensore ha consentito di verificare la qualità delle delle acquisizioni al variare della distanza dalla sorgente luminosa principale della stanza.

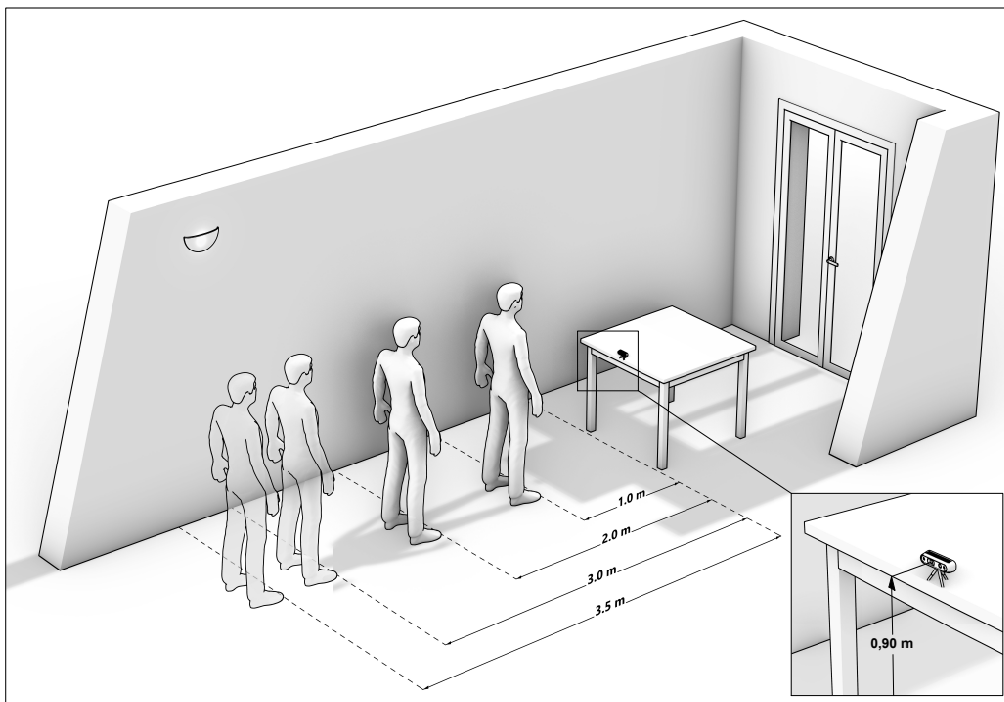


Figura 2.4: Setup Sperimentale.

L'iter per effettuare le registrazioni è stato il seguente:

1. **Avvio dello script di acquisizione:** Dopo l'avvio dello script Python per l'acquisizione, viene fornito un intervallo di 10 secondi per assicurare un posizionamento adeguato alla distanza desiderata. Successivamente, un segnale acustico indica l'inizio della registrazione, segnando il passaggio alla fase successiva.
2. **Respirazione:** Questa fase prevede l'esecuzione di circa 7-8 respiri durante l'intera durata dell'acquisizione video. L'obiettivo è eseguire il gesto respiratorio nel modo più naturale possibile.
3. **Arresto della registrazione:** Al termine dell'acquisizione, ovvero dopo trenta secondi lo script fornisce un secondo feedback sonoro che indica la fine. Il file viene quindi salvato automaticamente in una directory specifica per l'elaborazione successiva. Un esempio di come è strutturata la gerarchica delle cartelle è in figura [2.5].

Questo processo è stato applicato a tutte le distanze indossando le categorie di indumenti selezionate per i test, che includono una maglia termica aderente, t-shirt e una felpa. Inoltre, abbiamo effettuato una serie di acquisizioni in condizioni di scarsa illuminazione (circa 2/5 Lux), misurate con un luxmetro, per valutare le prestazioni anche in ambienti poco illuminati.

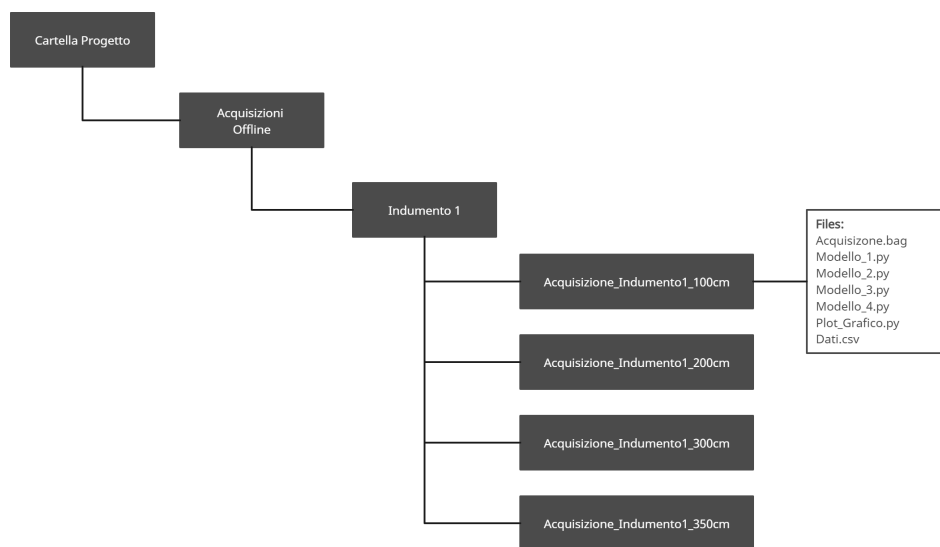


Figura 2.5: Gerarchia delle directory.

Capitolo 3

Estrazione del Segnale Respiratorio

Durante il corso di questo lavoro, sono stati impiegati il modello pre-addestrato YOLO v8 Pose, un algoritmo per l'identificazione delle articolazioni del corpo umano. Dopo l'estrazione dei keypoints dalla finestra RGB e la loro classificazione, tali dati sono stati trasferiti alla finestra Depth. Particolarmente utile in questo processo è stato l'articolo [15], il quale illustra un approccio per l'estrazione e la classificazione dei principali punti di giunzione riconosciuti da YOLO. Poiché YOLO non fornisce direttamente l'etichetta dei singoli keypoints ma soltanto le rispettive coordinate (x, y) , nella Figura [3.1] è possibile osservare come sono realmente disposti i vari keypoints. Questo passaggio ha permesso di acquisire informazioni di profondità relative ai suddetti punti, fondamentali per comprendere la disposizione delle articolazioni nello spazio. Quest'ultima è un aspetto cruciale per determinare la **ROI** (*Region Of Interest*), che in questo caso è rappresentata dalla zona toracica.

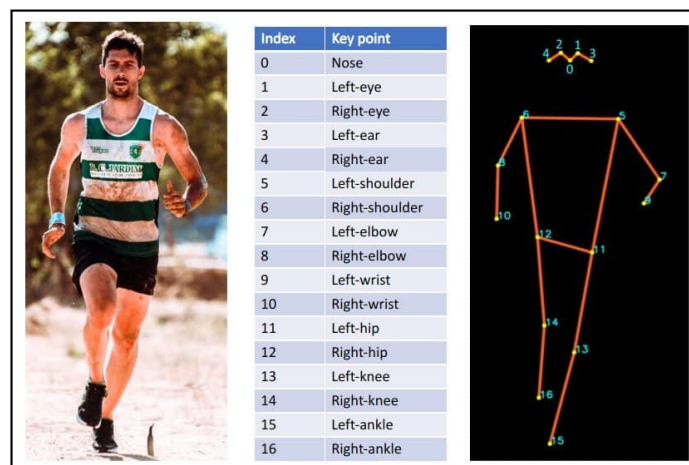


Figura 3.1: Disposizione dei keypoints.

Il seguente diagramma [3.2] fornisce una rappresentazione breve ma coincisa dei metodi che verranno impiegati per l'estrazione dei segnali dalle diverse Regioni di Interesse, fino alla loro successiva visualizzazione.

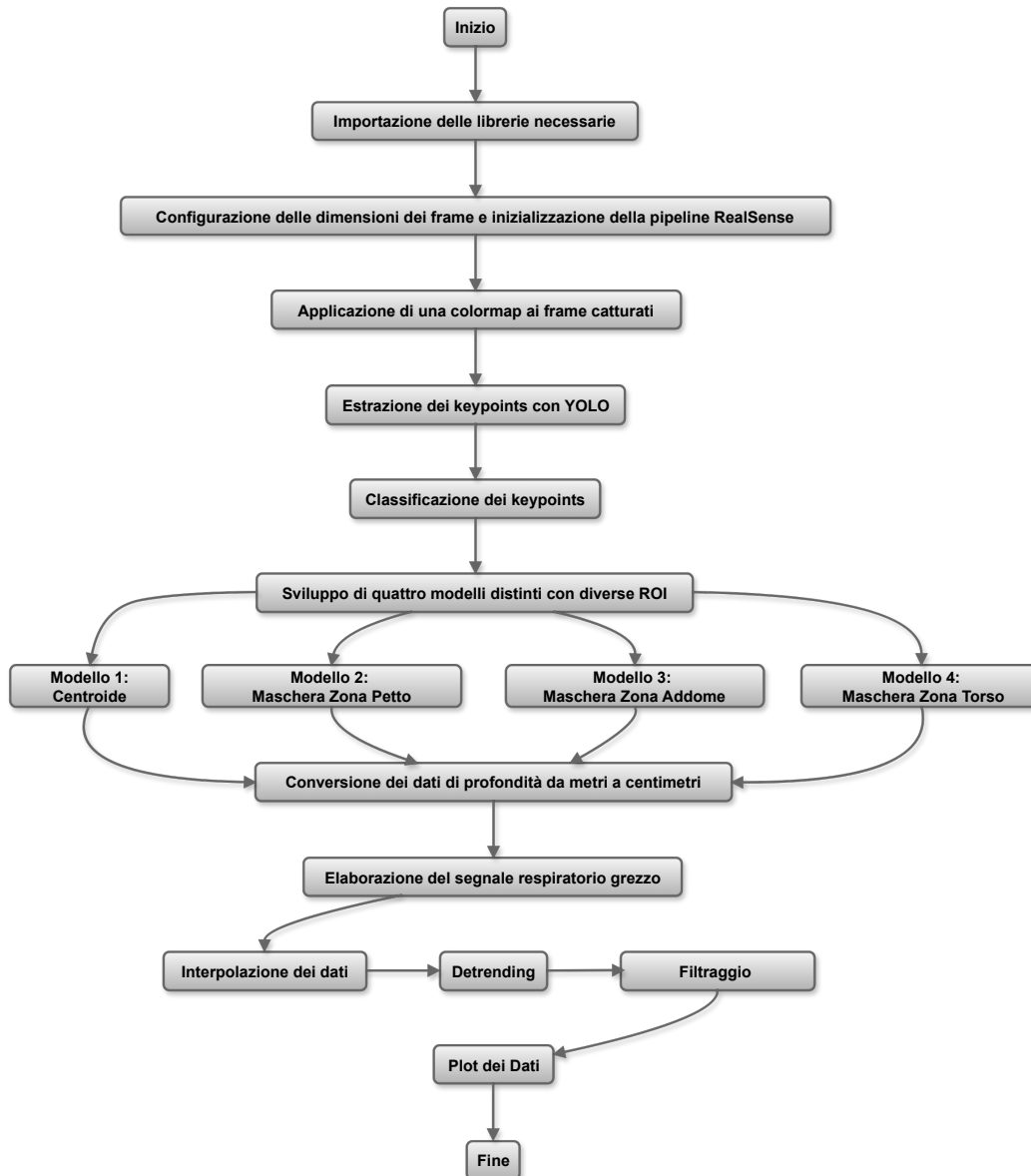


Figura 3.2: Diagramma di Flusso.

3.1 Metodi per l'Estrazione della ROI

Dopo aver importato le librerie necessarie, tra cui **cv2 (OpenCv)** per l'elaborazione delle immagini, **numpy** per la manipolazione degli array, **pyrealsense2** per l'interfacciamento alla telecamera e **ultralytics.YOLO** per implementare il modello **YOLOv8 Pose Nano**, si procede configurando le dimensioni dei frame e inizializzando la pipeline RealSense per catturare i frame RGB e Depth. Successivamente, viene applicata una *colormap* ai frame catturati, una funzione che assegna a ciascun valore di intensità in un'immagine un colore corrispondente consentendo una visualizzazione più intuitiva e comprensibile dei dati. L'algoritmo implementato è quindi in grado di estrarre i keypoints che vengono forniti in output da YOLO sotto forma di coordinate e classificarli con delle etichette come mostrato in figura [3.3], allo scopo di sviluppare quattro modelli distinti con diverse regioni di interesse. Il primo modello si basa sul calcolo del centroide del poligono formato dalla zona superiore del torace. Il secondo, terzo e quarto modello si fondano invece sull'analisi delle maschere, ossia sull'inclusione di tutti i punti presenti nelle aree delimitate dai poligoni. In sintesi, ogni fotogramma catturato dalla fotocamera fornisce una mappa di profondità, che assegna ad ogni pixel della scena una distanza approssimativa in centimetri tra il sensore e l'oggetto che occupa quel punto nella scena e di conseguenza, si ha una stima sulla variazione del torace nel tempo. Il vantaggio di questa scelta è dovuta al fatto che nonostante il soggetto sia in movimento questi modelli sono calcolati sulla base dei landmarks generati, perciò si otterranno misurazioni anche in posizioni non perfettamente frontali al sensore.

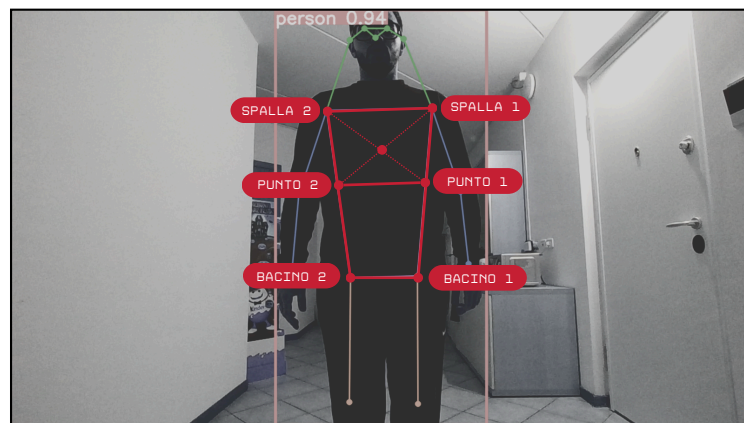


Figura 3.3: Keypoints selezionati.

1° Modello

Questo è stato il modello di partenza e si concentra su un singolo punto ottenuto come **centroide** del poligono formato dall'unione dei keypoints etichettati come: *spalla-2*, *spalla-1*, *punto-1* e *punto-2* come mostrato in figura [3.4]. Una volta individuato il centroide e ottenute le sue coordinate nei frame RGB, queste coordinate vengono trasferite in Depth, poiché i flussi video vengono inizializzati e allineati simultaneamente. In altre parole, le coordinate RGB del centroide vengono mappate sul frame Depth corrispondente per ottenere la profondità associata a quel punto nello spazio.

Successivamente, una volta che il punto è stato rappresentato nella schermata di interesse, vengono estratti i dati di profondità. È importante notare che i sensori RealSense forniscono questi dati in metri per default. Tuttavia, per una maggiore sensibilità ai movimenti del torace, è stata eseguita una conversione da metri a centimetri. Questa conversione è stata motivata dagli studi condotti nell'articolo [16], che indicano che il range di espansione del torace di un adulto sano è compreso tra 2 e 5 centimetri. La conversione in centimetri rende quindi più agevole l'interpretazione dei dati di profondità in relazione ai movimenti del torace.

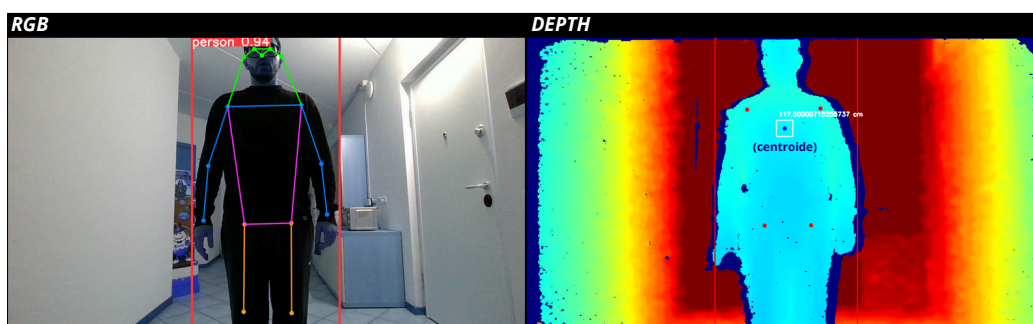


Figura 3.4: 1°Modello.

2° Modello

Il modello illustrato nella figura 3.5 si concentra sulla creazione della mappa dei punti nella zona toracica, ottenuta attraverso l'interconnessione dei keypoints: *spalla-2*, *spalla-1*, *punto-1* e *punto-2*. È essenziale notare che l'ordine di chiamata delle etichette avviene sempre in senso orario per garantire la corretta generazione della figura geometrica. Questo poligono delimita un'area definita, composta da una matrice di pixel in formato RGB, successivamente analizzati nei frame della profondità, come mostrato nell'illustrazione. Per ottenere una stima media del movimento toracico, è stato cruciale calcolare la media dei valori contenuti in questa matrice. Ciò ha consentito di ottenere un singolo valore rappresentativo della variazione di profondità nella regione di interesse.

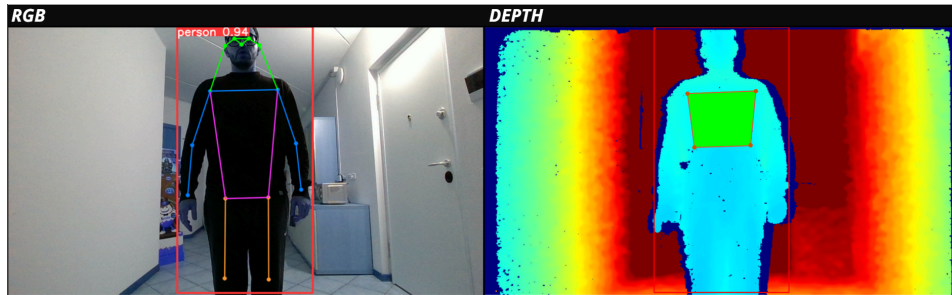


Figura 3.5: 2°Modello.

3°Modello

Il terzo modello [3.6] prende in esame invece la superficie costituita dall'unione dei keypoints: *punto-2*, *punto-1*, *bacino-1* e *bacino-2*. Si analizzano quindi le variazioni di profondità della zona addominale durante la fase respiratoria.

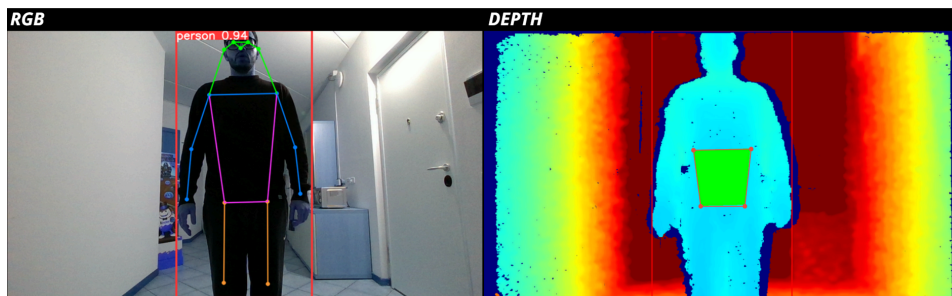


Figura 3.6: 3°Modello.

4°Modello

Infine, l'ultimo modello è costituito dall'intera superficie racchiusa nel poligono delimitato da: *spalla-2*, *spalla-1*, *bacino-1* e *bacino-2* come illustrato in figura [3.7].

Quest'ultimo prenderà quindi in considerazione non solo i movimenti del torace ma bensì anche quelli dell'addome, perciò durante la fase di inspirazione si rileveranno tramite il sensore valori di profondità maggiori nella zona addominale e viceversa nella zona superiore.

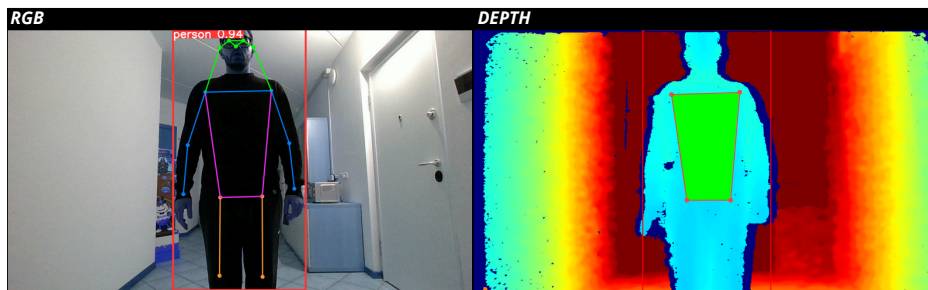


Figura 3.7: 4°Modello.

3.2 Elaborazione del Segnale Respiratorio Grezzo

Prima di poter visualizzare i dati, è stato fondamentale sottoporli a un processo di elaborazione e manipolazione per ottimizzare la loro risoluzione. Questa fase si è rivelata cruciale a causa della presenza di vari fattori di disturbo durante le rilevazioni. Tra questi, le condizioni di illuminazione sono state una causa significativa dei cosiddetti '*falsi positivi*', ovvero situazioni in cui vengono rilevati soggetti che non sono effettivamente presenti nella scena, con un basso punteggio di confidenza. Un esempio di questo fenomeno è illustrato nella Figura [3.8]. Inoltre, le interferenze causate dalla superficie irregolare degli indumenti hanno rappresentato un ulteriore fattore di disturbo. Di conseguenza, il post-processing dei dati, realizzato utilizzando le librerie specifiche, si è rivelato di vitale importanza per la corretta interpretazione dei risultati.



Figura 3.8: Falso positivo.

3.2.1 Interpolazione dei Dati

Una volta effettuato uno shift dei timestamps per farli partire da zero, sottraendo il valore minimo, si avvia la fase di interpolazione dei dati. Questa operazione viene eseguita utilizzando la libreria **SciPy**. Un esempio di questo processo è illustrato nella figura [3.9], come descritto nell'articolo [17].

Questa libreria è risultata particolarmente utile per gestire i dati mancanti, come ad esempio le distanze che risultavano essere pari a zero. Questo può accadere quando, durante la fase di acquisizione, si presentano frame con falsi positivi come accennato prima. Gli indici con valori zero vengono dunque sostituiti con **NaN** (*Not a Number*), un valore utilizzato per marcare i dati mancanti o non validi. Successivamente, i dati vengono interpolati linearmente utilizzando la funzione **interp1d** che viene messa a disposizione dalla libreria stessa. Questo processo ci assicura che i dati mancanti vengano stimati in modo appropriato, inoltre trattandosi di un set di dati di dimensioni relativamente piccole il metodo di interpolazione lineare è risultato il prescelto poiché è tra i più semplici e veloci. La velocità di questo metodo è dovuta al fatto che ogni coppia di punti adiacenti è semplicemente unita da un segmento, che può essere calcolato in maniera indipendente dagli altri.

Se denotiamo con (x_i, y_i) e (x_{i+1}, y_{i+1}) la coppia di punti adiacenti, la funzione interpolante è definita come:

$$f_i(x) = \frac{x_{i+1} - x}{x_{i+1} - x_i} \cdot y_i + \frac{x - x_i}{x_{i+1} - x_i} \cdot y_{i+1}$$

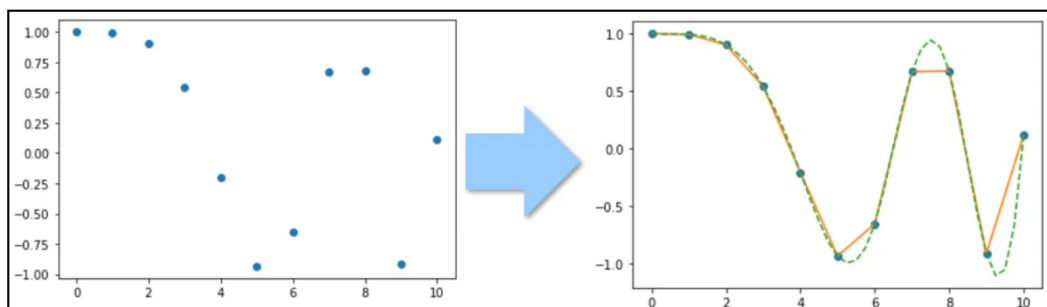


Figura 3.9: Esempio di Interpolazione Lineare.

3.2.2 Detrending e Filtraggio

Dopo aver eseguito l'operazione di interpolazione, abbiamo ritenuto utile rimuovere i trend attraverso un'operazione chiamata “**detrending**”. Questa operazione è stata eseguita utilizzando la funzione `'detrend'` della libreria che abbiamo menzionato. Il detrending ci ha permesso di ottenere un segnale centrato, eliminando gli effetti delle tendenze a lungo termine e mettendo in risalto i cambiamenti a breve termine. Successivamente, abbiamo applicato un **filtro mediano** utilizzando la funzione `'medfilt'`. Questa funzione consente di effettuare uno “smoothing” pesato dei campioni raccolti, eliminando così il rumore provocato da vari fattori, come le condizioni di luce e la tipologia di indumenti che presentano una superficie irregolare. Nella figura [3.10], è possibile infatti osservare nel dettaglio come il segnale viene processato dal filtro (segnale in blu) per eliminare il cosiddetto “*rumore sale e pepe*” e ottenere un segnale più smussato. Nel nostro caso, il “kernel size”, ovvero la dimensione della finestra utilizzata dal filtro, è impostata a cinque. Questo significa che il filtro mediano scorre questa finestra (o kernel) attraverso i dati e, per ogni posizione, sostituisce il valore centrale con il valore mediano dei valori all'interno della finestra. È importante notare che la scelta del “kernel size” può influenzare significativamente l'efficacia del filtro mediano. Un kernel size troppo piccolo potrebbe non essere in grado di rimuovere efficacemente il rumore, mentre un kernel size troppo grande potrebbe portare alla perdita di dettagli importanti nel segnale. Pertanto, la scelta del kernel size è un compromesso tra la rimozione del rumore e la conservazione dei dettagli del segnale.

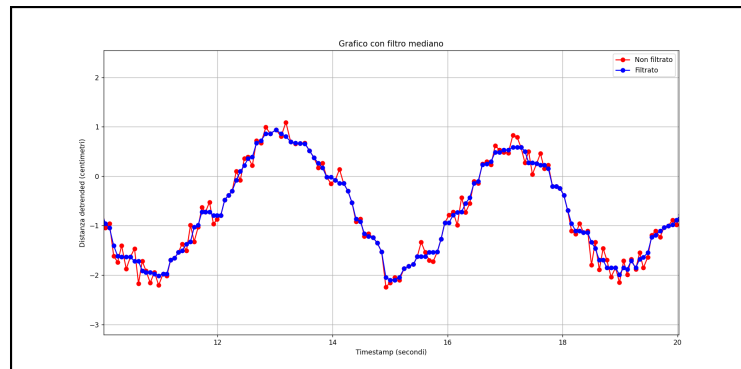


Figura 3.10: Azione del Filtro Mediano.

Dopo aver elaborato il segnale utilizzando le metodologie menzionate, abbiamo trovato utile incrementare le interpolazioni di un fattore **1000** tra un timestamp e l'altro. Questo ha permesso di ottenere un output meno discretizzato e con una maggiore definizione, come illustrato in figura [3.11].

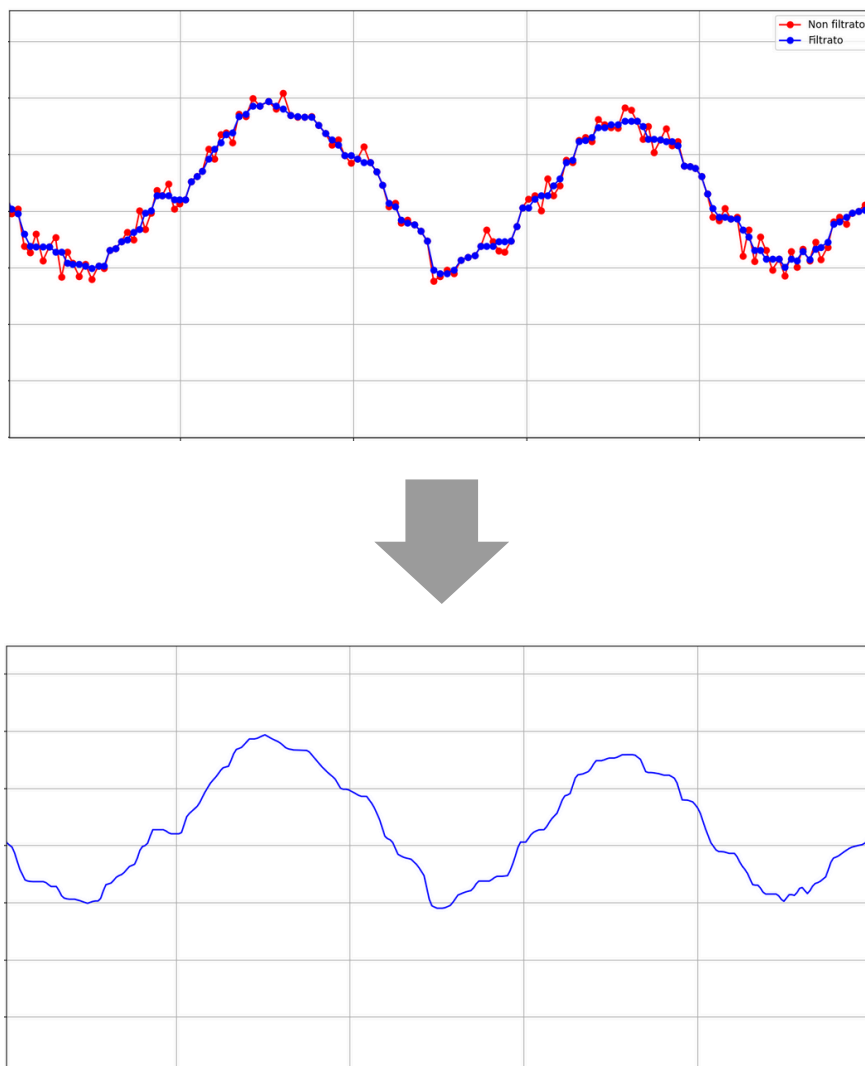


Figura 3.11: Esempio di Incremento Interpolazioni.

Capitolo 4

Risultati e Discussione

In questa fase, procederemo all'analisi e alla discussione dei risultati ottenuti attraverso i metodi precedentemente illustrati. I grafici svolgono un ruolo cruciale nella comprensione delle prestazioni dei quattro modelli sviluppati. In particolare, ci permettono di identificare i modelli che evidenziano più chiaramente i picchi respiratori rilevati durante l'intervallo di tempo di acquisizione. Per una visione più chiara e intuitiva della disposizione dei grafici che stiamo per esaminare, si può fare riferimento alla figura [4.1] dove ogni colonna rappresenta l'indumento indossato, mentre ogni riga indica la distanza alla quale è stata effettuata la rilevazione. Questo ci permette di avere un quadro visivo immediato dei risultati ottenuti.

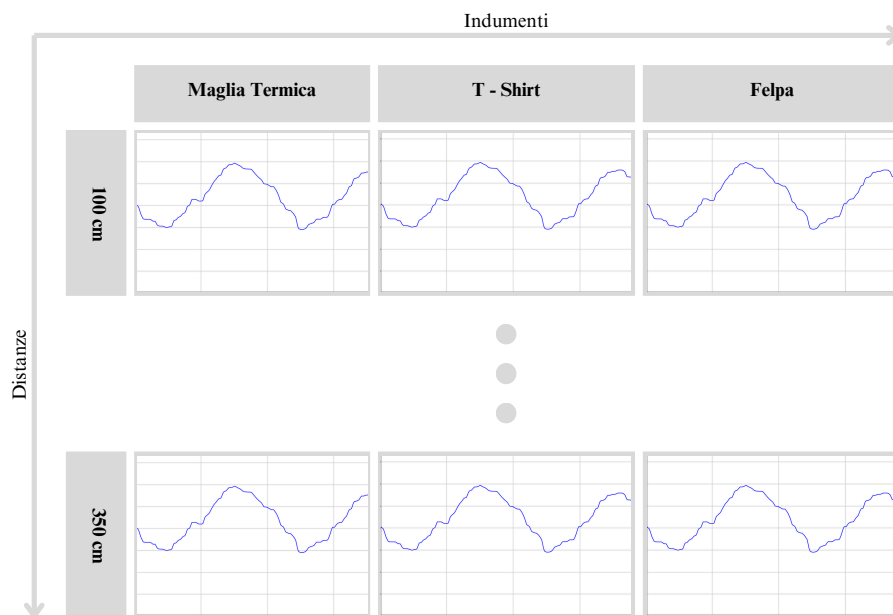
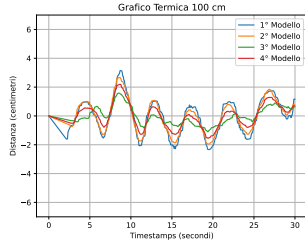
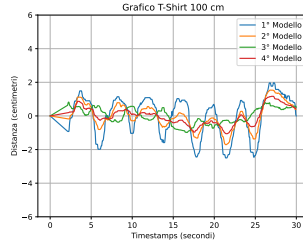


Figura 4.1: Disposizione dei grafici.

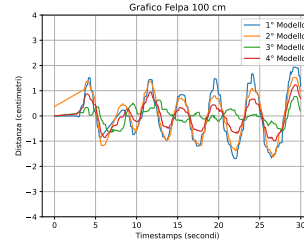
4.1 Sensibilità a Distanza e Indumento



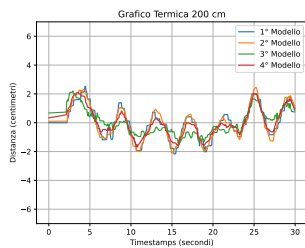
(a) Termica 100 cm.



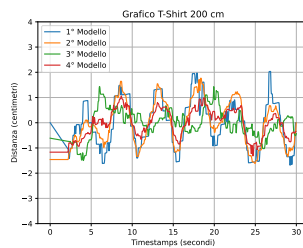
(b) T-Shirt 100 cm.



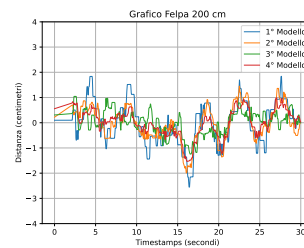
(c) Felpa 100 cm.



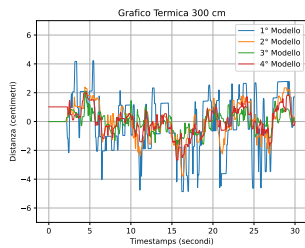
(d) Termica 200 cm.



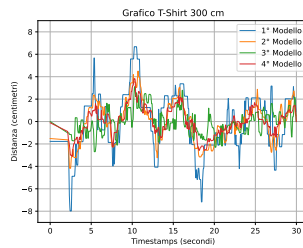
(e) T-Shirt 200 cm.



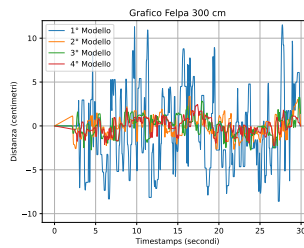
(f) Felpa 200 cm.



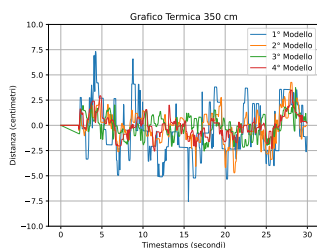
(g) Termica 300 cm.



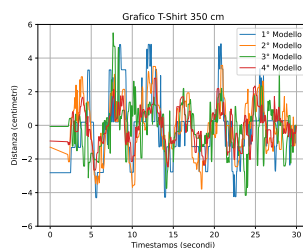
(h) T-Shirt 300 cm.



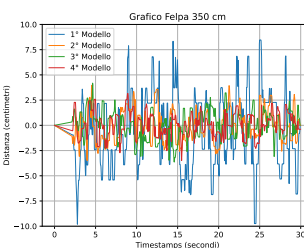
(i) Felpa 300 cm.



(j) Termica 350 cm.



(k) T-Shirt 350 cm.



(l) Felpa 350 cm.

4.1 Sensibilità a Distanza e Indumento

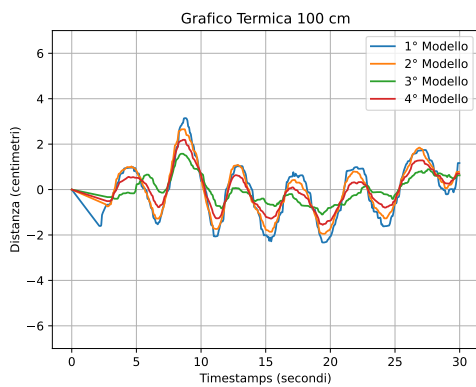
Dai grafici precedenti, si osserva che l'indumento che conserva i segnali respiratori più fedelmente è la maglia termica. Essendo attillata, questa garantisce una visualizzazione ottimale delle variazioni toraciche. Tuttavia, si registra un notevole degrado del segnale alla distanza di 300 cm [4.2e], sebbene sia ancora possibile rilevare, seppur con difficoltà, i picchi del segnale respiratorio. A 350 cm di distanza, il segnale diventa quasi indistinguibile per tutti i capi. La felpa risulta essere il capo peggiore, in quanto già a 200 cm [4.2f] il segnale inizia a mostrare segni di degrado. Questo è dovuto sia ai disturbi causati dalla superficie irregolare del capo, sia allo spessore che non facilita misurazione da parte del sensore. Per quanto riguarda le performance dei quattro modelli, il centroide (1°Modello) sembra offrire le migliori prestazioni in termini di risoluzione del segnale. Invece, la maschera di punti nella zona pettorale (2°Modello) risulta essere la migliore in termini di fedeltà rispetto alle mappe di punti prese in considerazione anche all'aumentare della distanza. Questo è probabilmente dovuto alla minore complessità del poligono considerato e al minor numero di punti presi in considerazione, che ha comportato una minore potenza di calcolo per la sua realizzazione e quindi una migliore risposta ai movimenti del torace. Un altro aspetto importante da considerare è l'impiego di YOLOv8 Pose nella versione '**Nano**'. Questo modello, essendo più piccolo e leggero rispetto agli altri disponibili da Ultralytics e come mostrato nella tabella [4.1], presenta una diminuzione drastica della confidence con l'aumento della distanza. Questo comporta una notevole instabilità dei keypoints durante la fase di acquisizione, costituendo un significativo fattore di disturbo, come evidenziato nel grafico [4.2l].

Model	Size (pixels)	mAP ^{pose} 50-95	mAP ^{pose} 50	Speed CPU-ONNX (ms)	Speed A100 TensorRT (ms)	Params (M)	FLOPs (B)
YOLOv8n-pose	640	50.4	80.1	131.8	1.18	3.3	9.2
YOLOv8s-pose	640	60.0	86.2	233.2	1.42	11.6	30.2
YOLOv8m-pose	640	65.0	88.8	456.3	2.00	26.4	81.0
YOLOv8l-pose	640	67.6	90.0	784.5	2.59	44.4	168.6
YOLOv8x-pose	640	69.2	90.2	1607.1	3.73	69.4	263.2
YOLOv8x-pose-p6	1280	71.6	91.2	4088.7	10.04	99.1	1066.4

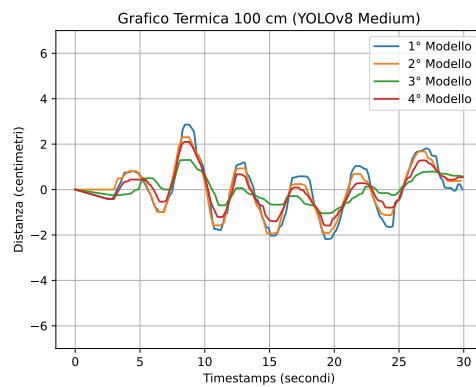
Tabella 4.1: Tabella dei Modelli YOLOv8 Pose.

Confronto con YOLOv8 Medium

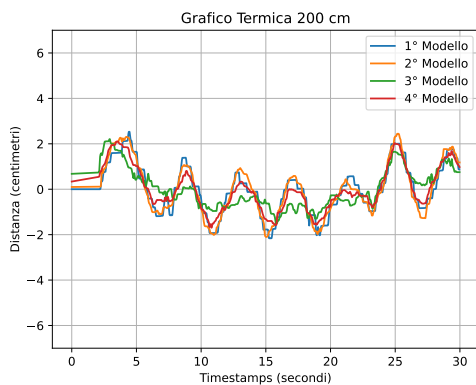
Per ottenere una maggiore dettagliatezza e definizione del segnale respiratorio, si è optato per l'utilizzo del modello **YOLOv8 Medium**. Come indicato nella tabella precedente, questo modello presenta un'architettura più ampia rispetto a quella precedentemente impiegata. Prendendo come riferimento l'indumento che ha mostrato le migliori prestazioni in termini di accuratezza, ovvero la 'maglia termica', si osserva che le acquisizioni realizzate con il modello di dimensioni maggiori tendono a essere meno influenzate da disturbi e a mostrare una precisione superiore.



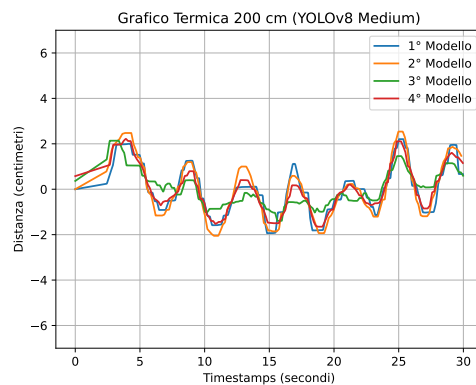
(a) Termica 100cm (YOLOv8 Nano).



(b) Termica 100cm (YOLOv8 Medium).



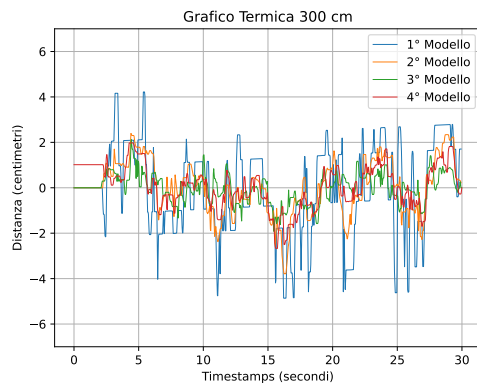
(c) Termica 200cm (YOLOv8 Nano).



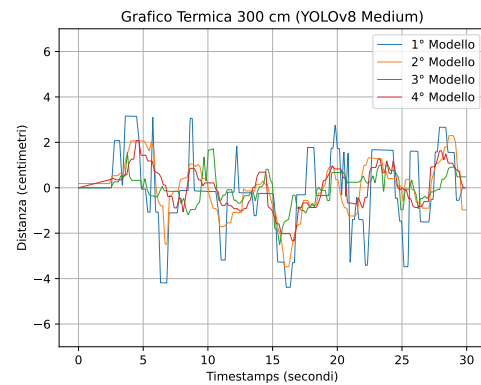
(d) Termica 200cm (YOLOv8 Medium).

4.1 Sensibilità a Distanza e Indumento

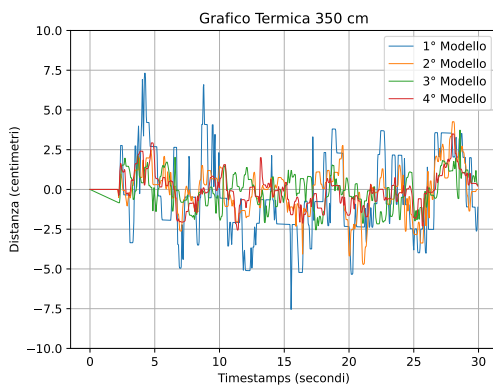
L'utilizzo della versione 'Medium' ha inoltre mostrato vantaggi per le rilevazioni effettuate alle distanze di 300cm e 350cm . Come si può osservare, queste distanze presentano segnali più degradati, soprattutto nel caso del 1° Modello, che analizza un singolo punto. Al contrario, il segnale che risulta più simile e fedele alla forma ideale del segnale respiratorio è quello del 2° Modello come mostrato in figura [4.2a]. In questo caso, i valori sono più vicini all'effettiva espansione toracica, rimanendo in un intervallo compreso tra $+2,0\text{cm}$ e $-2,5\text{cm}$.



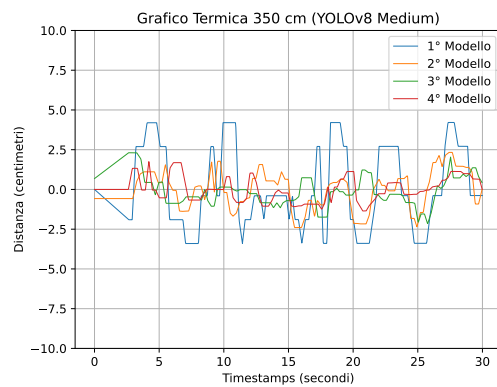
(a) Termica 300cm (YOLOv8 Nano).



(b) Termica 300cm (YOLOv8 Medium).



(c) Termica 350cm (YOLOv8 Nano).

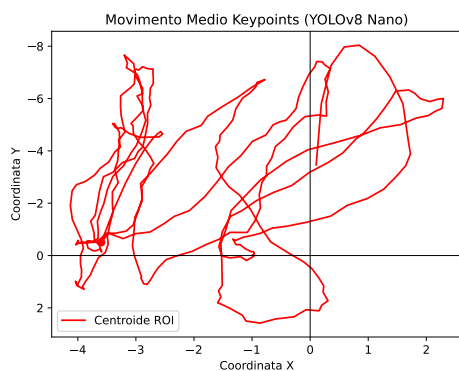


(d) Termica 350cm (YOLOv8 Medium).

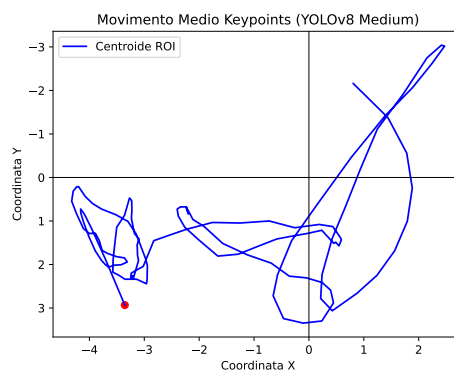
Capitolo 4 Risultati e Discussione

Come menzionato in precedenza, sono stati osservati movimenti indesiderati dei keypoints considerati per la generazione dei quattro modelli. Questi movimenti possono generare errori di misurazione, poiché i keypoints costituiscono il sistema di riferimento per la creazione delle Regioni di Interesse (ROI). Abbiamo riscontrato che un modello di dimensioni maggiori, con un dataset più ampio a disposizione, comporta una recall superiore. Di conseguenza, la stima dei keypoints risulta più precisa.

Per stimare il movimento dei keypoints nell'immagine RGB, abbiamo considerato il punto centrale dell'intera regione del torso rilevata dalla 'pose estimation'. Analizzando la variazione delle coordinate (x,y) di questo punto, possiamo ottenere un'idea degli spostamenti che avvengono durante l'acquisizione e di quanto questi possano influire sulla stima della profondità. Ricordiamo, infatti, che il punto selezionato nell'immagine RGB viene poi utilizzato parallelamente nell'immagine depth per estrapolare i dati di profondità. Nella figura [4.2a], è infatti possibile osservare come il movimento medio dei keypoints con il modello YOLO Pose Nano sia maggiore. Al contrario, il modello Medium, mostrato nella figura [4.2b], garantisce una maggiore stabilità con una conseguente riduzione dell'errore durante la stima della profondità e, quindi, del rumore sovrapposto al segnale.



(a)



(b)

4.2 Sensibilità Luce

Le telecamere di profondità, come l'Intel RealSense D455, offrono prestazioni eccezionali, anche in condizioni di scarsa illuminazione. Nonostante si basino sull'uso di immagini per la stima della profondità, queste telecamere, secondo quanto dichiarato dal produttore, sono in grado di effettuare rilevazioni ottimali a distanze che variano dai 0,4 ai 6 metri, a seconda delle condizioni di luminosità. L'integrazione di un *proiettore a infrarossi (IR)* garantisce che, anche in condizioni di scarsa illuminazione, la telecamera sia in grado di percepire i dettagli della profondità. Questo è stato dimostrato sia in uno studio [18] che in un paper [19]. Nella documentazione del produttore, la figura [4.2] fornisce una dimostrazione dell'efficacia del proiettore a infrarossi in questione.

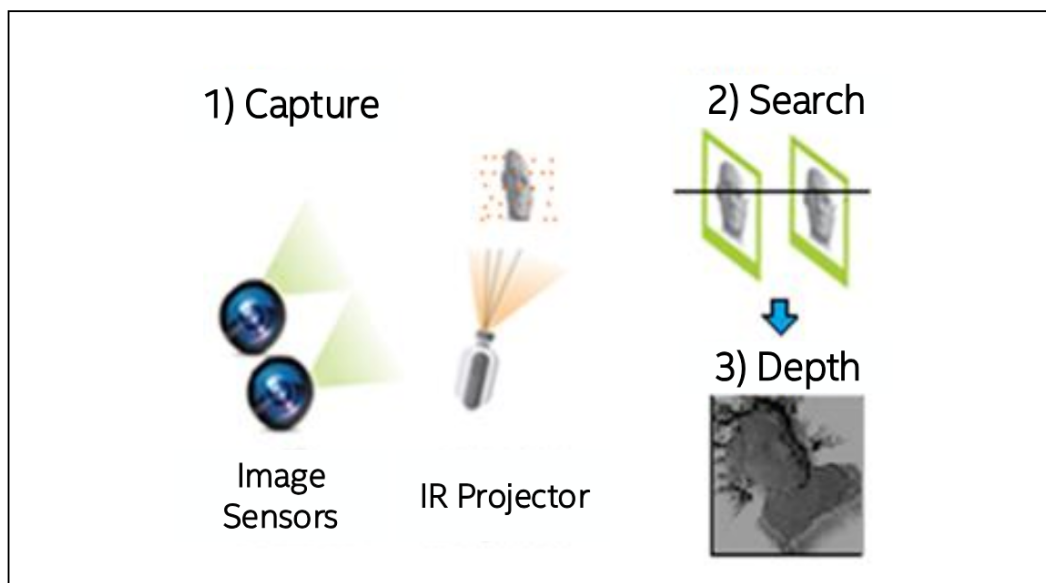
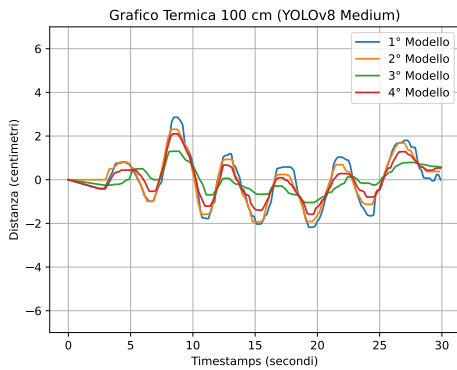


Figura 4.2: Proiettore Infrarossi (IR).

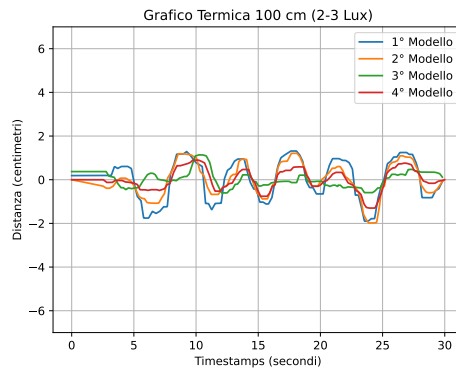
In questo scenario, abbiamo optato per l'impiego del modello di dimensioni medie per la raccolta dati. La scelta è stata dettata dal fatto che la versione 'nano' avrebbe mostrato una precisione ridotta in condizioni di scarsa illuminazione durante la fase di pose recognition comportando anche la generazione di falsi positivi che avrebbero aumentato ulteriormente il degrado del segnale. Le misurazioni, come illustrato nelle figure [4.3b] e [4.3d], sono state condotte in

Capitolo 4 Risultati e Discussione

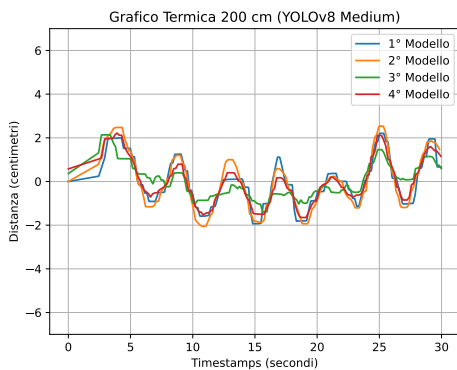
un intervallo di luminosità molto basso compreso tra 2 e 5 *Lux*, ovvero in condizioni di quasi totale oscurità. Questo contrasta con le rilevazioni precedenti, che variavano tra un minimo di 17 e un massimo di 38 *Lux*. Nonostante le condizioni di luce sfavorevoli, l'ausilio del proiettore infrarosso della telecamera ha permesso di rilevare chiaramente i picchi del segnale respiratorio sia a 100 che a 200 *cm*. Tuttavia, è presente del rumore sovrapposto, probabilmente dovuto all'instabilità dei keypoints perciò con l'aumento della distanza, il segnale diventa quasi completamente irricognoscibile. È interessante osservare come il terzo modello, anche a distanze ravvicinate, risulti il meno performante e leggermente fuori fase rispetto agli altri. Questo perché considera la mappa dei punti della zona addominale, che si muove in maniera alternata rispetto a quella del torace, risultando quindi sfasato rispetto agli altri segnali dei rispettivi modelli.



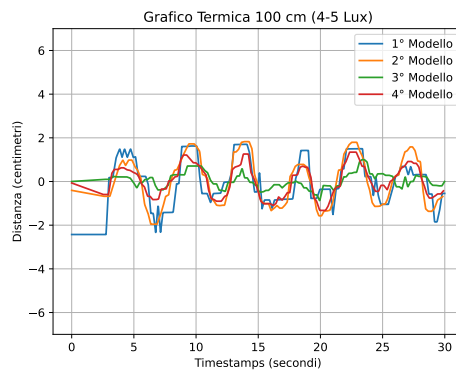
(a) Termica 100 cm (25 Lux).



(b) Termica 100 cm (2-3 Lux).



(c) Termica 200 cm (38 Lux).



(d) Termica 200 cm (4-5 Lux).

4.3 Stima di un Indice di Qualità

Per valutare il segnale del volume respiratorio, non è stato utilizzato un sensore di riferimento (Ground Truth). Pertanto, introduciamo un indice di qualità ispirato al Rapporto Segnale/Rumore (SNR), già utilizzato nella letteratura scientifica relativa alla registrazione senza contatto dei segni vitali [20], [21]. L'indice di qualità, che verrà chiamato **Rapporto Respirazione/Rumore (RNR)**, può essere calcolato come segue:

$$RNR = 10 \log_{10} \left(\frac{P_r}{P_n} \right), \quad (4.1)$$

Dove $P_r = \frac{\sum_{0.1\text{Hz}}^{F_{max}} (PS(f)M(f))}{\sum_{0.1\text{Hz}}^{F_{max}} (M(f))}$ e $P_n = \frac{\sum_{0.1\text{Hz}}^{F_{max}} (PS(f)(1-M(f)))}{\sum_{0.1\text{Hz}}^{F_{max}} (1-M(f))}$ rappresentano rispettivamente la potenza del volume respiratorio e la potenza del rumore. Mentre $M(f)$ è una maschera definita come:

$$M(f) = \begin{cases} 1 & \text{se } F_r - 0.05 \leq f \leq F_r + 0.05, \\ 0 & \text{altrimenti.} \end{cases} \quad (4.2)$$

Per ogni segnale registrato e pre-elaborato, viene calcolato lo spettro di potenza $PS(f)$ e la frequenza dominante, che viene considerata come la frequenza respiratoria F_r . La potenza di tutte le frequenze a una distanza di 0.05 Hz da quella dominante viene mediata per ottenere una stima di P_r . Tutte le frequenze al di fuori della banda selezionata contribuiscono a P_n . Un esempio visivo della maschera nel caso del 1° Modello a una distanza di 100 cm è mostrato in figura [4.3].

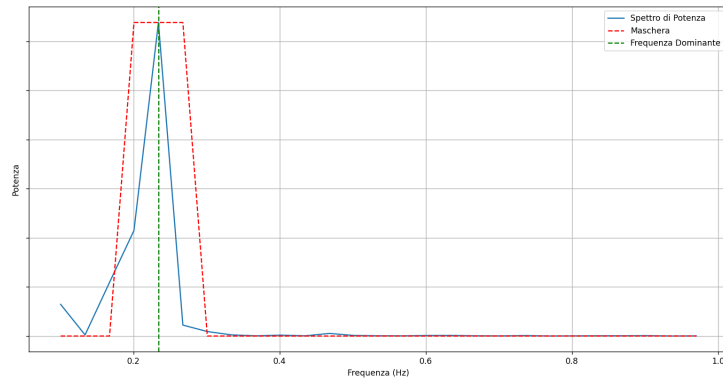


Figura 4.3: Esempio Maschera $M(f)$.

Capitolo 4 Risultati e Discussione

Nella tabella [4.2] sono riportati i valori medi di RNR per ciascuna distanza testata nei quattro modelli considerati. Si osserva un notevole aumento del rumore all'aumentare della distanza. Il miglior RNR si ottiene per distanze inferiori o uguali a 200 cm, con cali considerevoli fino a 5 dB quando la distanza supera i 300 cm.

Tabella 4.2: RNR Medio per Ciascuna Distanza e Indumento Testato.

<i>Distanza</i>	<i>RNR 1° Modello (dB)</i>	<i>RNR 2° Modello (dB)</i>	<i>RNR 3° Modello (dB)</i>	<i>RNR 4° Modello (dB)</i>
100 cm	14,18	12,56	9,41	11,50
200 cm	14,39	15,71	8,72	12,51
300 cm	9,05	10,89	5,67	10,81
350 cm	8,07	9,79	5,52	5,94
Media Totale	11,42	12,24	7,33	10,19

Nella tabella [4.3] sono invece riportati i RNR per ciascuna distanza e indumenti testati nei quattro modelli considerati. Si nota un RNR simile tra la maglia termica e la T-shirt, ma un calo di 2-3 dB quando si indossa la felpa.

Tabella 4.3: RNR per Ciascun Indumento, Stratificato per Modello.

<i>Indumenti</i>	<i>RNR 1° Modello (dB)</i>	<i>RNR 2° Modello (dB)</i>	<i>RNR 3° Modello (dB)</i>	<i>RNR 4° Modello (dB)</i>
Maglia Termica	11,76	13,32	8,38	11,01
T-shirt	12,6775	13,2475	7,6825	11,195
Felpa	9,83	10,15	5,9275	8,37

Il modello migliore tra tutte le condizioni testate è il secondo, con un RNR medio complessivo di 12,24 dB. Seguono il centroide del petto (1° Modello) con una media di 11,42 dB, la media sull'intero torso (4° Modello) con 10,19 dB e la media sull'addome (3° Modello) con 7,33 dB. I risultati confermano che la media su più pixel dell'area del petto è preferibile rispetto al centroide in termini di riduzione del rumore. Per l'intera regione del torso, ci si aspetta di osservare un movimento periodico sincrono che segue il comportamento naturale della respirazione del soggetto. Pertanto, l'aggiunta coerente del contributo di tutte le piccole sottoregioni del petto migliora il segnale del volume respiratorio rispetto al rumore. Inoltre, i risultati meno soddisfacenti in termini di RNR per la regione dell'addome sono stati già osservati in letteratura utilizzando altri tipi di telecamere di profondità. Questo potrebbe essere dovuto al modo in cui il soggetto respira e ai vestiti che tendono a essere più larghi nella regione dell'addome [22]. In conclusione, la media sull'intera regione del torso presenta un RNR intermedio tra il petto e l'addome.

4.4 Prospettive Future

I risultati ottenuti aprono diverse possibilità per ulteriori ricerche e sviluppi. Alcuni aspetti che potrebbero essere esplorati sono:

- **Ampliamento del Dataset:** L'arricchimento del dataset con una maggiore varietà di indumenti e condizioni ambientali potrebbe incrementare la robustezza dei modelli. Un'ulteriore strategia potrebbe essere quella di posizionare i soggetti contro una parete per minimizzare le oscillazioni involontarie e garantire la qualità dei dati raccolti.
- **Ottimizzazione dei Modelli:** L'adozione di versioni più avanzate dei modelli YOLO potrebbe ridurre il rumore nelle rilevazioni a distanze maggiori, aumentando la versatilità del sistema. Ad esempio, l'impiego di *YOLO v8 Pose Large* potrebbe migliorare la precisione nella rilevazione dei keypoints e, di conseguenza, nell'identificazione delle ROI. Tuttavia, sarebbe necessario perfezionare gli algoritmi e utilizzare un hardware più performante per prevenire latenze durante la fase di elaborazione e raccolta dei dati.
- **Metodi Alternativi per l'Identificazione della ROI:** In alternativa all'utilizzo della superficie di punti, è possibile ricorrere alla nuvola di punti situata nella regione del torace. Questa può essere utilizzata per creare un volume tridimensionale e calcolare le variazioni di volume di tale solido, permettendo così una stima dell'atto respiratorio ancora più precisa. Inoltre, si potrebbero impiegare algoritmi di computer vision diversi da YOLO, come ad esempio '*Azure Kinect Body Tracking Joints*'. Quest'ultimo rappresenta una soluzione più economica rispetto ai sensori Intel Realsense.
- **Sensore di Riferimento (Ground Truth):** L'uso di un sensore per confrontare i segnali respiratori potrebbe fornire una stima precisa dell'efficacia del sistema sviluppato. Ad esempio, si potrebbe considerare l'uso di una fascia toracica o, preferibilmente, di uno spirometro.
- **Applicazioni in Altri Contesti:** Valutare l'applicabilità di questi modelli in contesti non clinici, come l'industria sportiva o la sicurezza, potrebbe aprire nuove opportunità di utilizzo.

Capitolo 5

Conclusioni

Questo studio ha analizzato l'accuratezza di quattro modelli sviluppati utilizzando il rilevamento dei keypoints forniti dal modello di riconoscimento delle pose 'YOLO v8 Pose'. Questi modelli hanno considerato diverse Regioni di Interesse (ROI) per monitorare i movimenti toracici durante la respirazione. L'analisi ha preso in esame vari fattori, come la distanza tra il soggetto e il sensore, il tipo di abbigliamento indossato e le condizioni di illuminazione dell'ambiente sperimentale. Per la rilevazione e l'acquisizione dei dati, è stato utilizzato il sensore RGB-D Intel RealSense D455, che ha permesso di ottenere informazioni dettagliate sulla profondità dei modelli analizzati.

Dall'analisi dei dati raccolti, è emerso che la performance dei modelli varia significativamente in base alla distanza di rilevazione. I risultati mostrano che le prestazioni migliori sono state ottenute a distanze ravvicinate, con il primo modello che si è dimostrato il più efficace in queste condizioni. A distanze maggiori, invece, il secondo modello ha prevalso. Tra i modelli testati per la mappatura dei keypoints, YOLOv8 Medium si è rivelato superiore al modello Nano, garantendo una maggiore stabilità della ROI e riducendo il rumore sovrapposto al segnale.

Il tipo di indumento indossato dai soggetti ha avuto un impatto significativo sulle rilevazioni. La maglia termica ha fornito le migliori prestazioni, consentendo di identificare chiaramente i picchi respiratori anche in condizioni di scarsa illuminazione. Al contrario, la T-shirt e la felpa hanno mostrato una maggiore variabilità e un aumento del rumore nei dati raccolti.

Capitolo 5 Conclusioni

Le condizioni di illuminazione si sono rivelate un fattore cruciale per la qualità delle rilevazioni. In scenari di quasi totale oscurità (2-5 Lux), l'uso del proiettore a infrarossi della telecamera ha permesso di mantenere una buona qualità del segnale, sebbene con un aumento del rumore per distanze superiori ai 200 cm. Anche in queste condizioni estreme, il modello YOLOv8 Medium ha dimostrato una precisione maggiore rispetto alla versione Nano. Inoltre, per valutare la qualità del segnale respiratorio senza l'uso di un sensore di riferimento (Ground Truth), è stato introdotto un indice di qualità ispirato al Rapporto Segnale/Rumore (SNR), denominato Rapporto Respirazione/Rumore (RNR). Questo indice ha permesso di misurare la qualità del segnale confrontando la potenza del volume respiratorio con quella del rumore. I risultati hanno mostrato un aumento significativo del rumore con l'aumentare della distanza, con il miglior RNR ottenuto per distanze inferiori o uguali a 200 cm e un calo considerevole fino a 5 dB oltre i 300 cm.

In conclusione, questo studio ha contribuito in modo significativo alla comprensione della rilevazione dei movimenti respiratori mediante modelli di machine learning. L'affidabilità degli algoritmi realizzati a distanza ravvicinata e l'efficacia del secondo modello suggeriscono potenziali applicazioni in ambito clinico per il monitoraggio respiratorio non invasivo, specialmente in contesti dove l'uso di dispositivi tradizionali potrebbe risultare scomodo per un possibile paziente.

Bibliografia

- [1] Nipun Nath and Amir Behzadan. Deep convolutional networks for construction object detection under different visual conditions. *Frontiers in Built Environment*, 6, 08 2020.
- [2] Kaiming Gu and Boyu Su. A study of human pose estimation in low-light environments using yolov8 model. *Applied and Computational Engineering*, 32:136–142, 01 2024.
- [3] Hirooki Aoki and Noriaki Fujita. Non-contact respiration measurement by using rgb-d camera. In *2022 Joint 12th International Conference on Soft Computing and Intelligent Systems and 23rd International Symposium on Advanced Intelligent Systems (SCIS&ISIS)*, pages 1–4, 2022.
- [4] Shan He, Zixiong Han, Cristóvão Iglesias, Varun Mehta, and Miodrag Bolic. A real-time respiration monitoring and classification system using a depth camera and radars. *Frontiers in Physiology*, 13, 2022.
- [5] Paul S Addison, André Antunes, Dean Montgomery, Philip Smit, and Ulf R Borg. Robust non-contact monitoring of respiratory rate using a depth camera. *Journal of Clinical Monitoring and Computing*, 37(4):1003–1010, 2023.
- [6] Haythem Rehouma, Rita Noumeir, Sandrine Essouri, and Philippe Jouvét. Quantitative assessment of spontaneous breathing in children: Evaluation of a depth camera system. *IEEE Transactions on Instrumentation and Measurement*, 69(7):4955–4967, 2020.
- [7] Yuhao Shan, Shigang Li, and Tong Chen. Respiratory signal and human stress: non-contact detection of stress with a low-cost depth sensing camera. *International Journal of Machine Learning and Cybernetics*, 11(8):1825–1837, Aug 2020.

Bibliografia

- [8] Wakana Imano, Kenichi Kameyama, Malene Hollingdal, Jens Refsgaard, Knud Larsen, Cecilie Topp, Sissel Højsted Kronborg, Josefine Dam Gade, and Birthe Dinesen. Non-contact respiratory measurement using a depth camera for elderly people. *Sensors*, 20(23), 2020.
- [9] Paul S. Addison, Philip Smit, Dominique Jacquel, Anthony P. Addison, Cyndy Miller, and Gardner Kimm. Continuous non-contact respiratory rate and tidal volume monitoring using a depth sensing camera. *Journal of Clinical Monitoring and Computing*, 36(3):657–665, Jun 2022.
- [10] Jochen Kempfle and Kristof Van Laerhoven. Towards breathing as a sensing modality in depth-based activity recognition. *Sensors*, 20(14), 2020.
- [11] Paul S. Addison, Philip Smit, Dominique Jacquel, and Ulf R. Borg. Continuous respiratory rate monitoring during an acute hypoxic challenge using a depth sensing camera. *Journal of Clinical Monitoring and Computing*, 34(5):1025–1033, Oct 2020.
- [12] Jochen Kempfle and Kristof Van Laerhoven. Respiration rate estimation with depth cameras: An evaluation of parameters. In *Proceedings of the 5th International Workshop on Sensor-Based Activity Recognition and Interaction*, iWOAR '18, New York, NY, USA, 2018. Association for Computing Machinery.
- [13] Intel Corporation. Intel realsense sdk. <https://github.com/IntelRealSense/librealsense.git>.
- [14] Intel Corporation. Intel realsense calibration. <https://dev.intelrealsense.com/docs/calibration>.
- [15] Ali Mustofa. Yolov8 pose estimation and pose keypoint classification using neural net pytorch. <https://github.com/Alimustooofaa/YoloV8-Pose-Keypoint-Classification>.
- [16] Muslim Talib Jabbar Ahmed Hasan Reheem, Hussam Riyadh Dakhil. Reference range of chest expansion in healthy adult living in al-muthanna governorate. *Medico Legal Update*, page 1566–1570, Nov. 2020.

- [17] Tirthajyoti Sarkar. How to interpolate data with scipy. <https://medium.com/productive-data-science/how-to-interpolate-data-with-scipy-d314143285bc>.
- [18] Shuntao Liu, Dedong Gao, Wang Peng, Xifeng Guo, Jing Xu, and Du-Xin Liu. A depth-based weighted point cloud registration for indoor scene. *Sensors*, 18:3608, 10 2018.
- [19] Anders Grunnet-Jepsen, John N. Sweetser, Paul Winer, Akihiro Takagi, and John Woodfill. Projectors for d400 series depth cameras. <https://dev.intelrealsense.com/docs/projectors>.
- [20] Juan Cheng, Runqing Liu, Jiajie Li, Rencheng Song, Yu Liu, and Xun Chen. Motion-robust respiratory rate estimation from camera videos via fusing pixel movement and pixel intensity information. *IEEE Transactions on Instrumentation and Measurement*, 2023.
- [21] Gerard De Haan and Vincent Jeanne. Robust pulse rate from chrominance-based rppg. *IEEE transactions on biomedical engineering*, 60(10):2878–2886, 2013.
- [22] Jochen Kempfle and Kristof Van Laerhoven. Breathing in-depth: A parametrization study on rgb-d respiration extraction methods. *Frontiers in Computer Science*, 3:757277, 2021.

Ringraziamenti

Desidero ringraziare il Prof. *Ennio Gambi* per avermi dato l'opportunità di completare questa tesi e il Dr. *Antonio Nocera* per il suo supporto paziente e dedicato, che ha reso il mio percorso significativo e gratificante. Un ringraziamento speciale alla Dr.ssa *Linda Senigagliaesi* per il suo competente sostegno durante la stesura della tesi.

Un ringraziamento speciale va ai miei genitori, *Mohamed e Meriem*, i miei primi sostenitori. Grazie alla vostra costante fiducia e supporto, sono riuscito a raggiungere questo grande traguardo. Desidero inoltre ringraziare anche i miei fratelli, *Ilyass e Yasser*, che mi hanno sempre guidato, protetto e indicato la strada migliore da seguire. La vostra presenza e il vostro affetto sono stati preziosi.

Il ringraziamento più sentito va a *Sara*, che da anni mi sostiene in ogni fase della mia vita. In questo percorso, la tua presenza è stata fondamentale. Grazie per il tuo amore, la tua pazienza e il tuo costante supporto. Sei stato il mio rifugio sicuro e la mia forza nei momenti difficili. La tua fiducia in me e il tuo sostegno mi hanno permesso di superare gli ostacoli e di raggiungere questo importante traguardo. Senza di te, questo cammino sarebbe stato molto più arduo, se non impossibile. Grazie di cuore per tutto ciò che fai e per essere sempre al mio fianco.

Infine, un sentito ringraziamento va ai miei amici di corso, con i quali ho condiviso feste, ansie e lunghe giornate di studio, rendendo ogni momento speciale.

Ancona, Luglio 2024

Anass Chebbaki