

# POLITECNICO DI TORINO

Tesi di Laurea Magistrale



Laurea Magistrale in Ingegneria Biomedica

## **Rilevamento e caratterizzazione automatica del movimento nel sonno tramite analisi di sequenze video IR/depth**

Relatore:  
Prof.ssa Gabriella Olmo

Candidata:  
Arianna Pellegrino

Correlatore:  
Dr. Roberto Nerino, CNR-IEIIT

Anno 2020/2021

## **Abstract**

*Lo studio del sonno ha una notevole importanza per la valutazione sia delle condizioni psicofisiche che dello stato di salute degli esseri umani. L'analisi dei segnali fisiologici e dei movimenti durante il sonno consente di valutarne la qualità, nei soggetti sani, e l'entità del disturbo, in quelli affetti da patologie.*

*Per quanto riguarda analisi del movimento, gli attuali approcci di valutazione quantitativa dei movimenti si basano sull'elettromiografia e/o sull'accelerometria. Oltre ad essere approcci invasivi e poco confortevoli per il paziente, essi non forniscono una visione globale del movimento del soggetto, ma solo valutazioni locali che riguardano gli specifici muscoli o arti su cui sono stati posizionati i sensori.*

*Lo scopo di questo lavoro è sviluppare un approccio non invasivo all'analisi e alla caratterizzazione dei movimenti nel sonno, che consenta di catturare lo spettro completo dei movimenti del soggetto.*

*Questo è reso possibile tramite sensori ottici, e in particolare camere infrarosse e di profondità, quali Microsoft Kinect Azure. Queste camere acquisiscono, senza contatto e al buio, le sequenze video dei movimenti, consentendone la successiva analisi.*

*In questo lavoro è stato realizzato un setup sperimentale in ambiente domestico, ed è stata impiegata Azure Kinect, che è in grado di fornire contemporaneamente sequenze video sia nell'infrarosso che di profondità.*

*È stato così possibile esplorare due approcci specifici, uno più convenzionale, che utilizza le sequenze video nell'infrarosso, e uno che utilizza quelle di profondità e l'informazione tridimensionale del soggetto e della scena analizzata.*

*Il software è stato sviluppato in ambiente MATLAB™ e comprende il software di acquisizione e filtraggio delle sequenze video, nonché gli algoritmi di rilevamento del movimento (motion detection), di inseguimento (tracking), e di valutazione della quantità di movimento.*

*Per la motion detection sono stati valutati diversi algoritmi di rimozione dello sfondo (background subtraction), tra cui quello delle misture di gaussiane e quello delle differenze assolute. L'analisi è stata condotta sia per le sequenze infrarosse che per quelle di profondità.*

*Il sistema costituito dall'hardware e dal software sviluppato consente un'analisi actigrafica globale dei movimenti del soggetto durante il sonno, permettendo anche di rilevarne la presenza nel letto. Queste informazioni forniscono la stima di alcuni parametri di qualità del sonno, tra cui dell'efficienza, evitando l'utilizzo di sensori indossabili che potrebbero influire sulla qualità del sonno stesso. Il sistema fornisce anche un istogramma bidimensionale di movimento a varie scale temporali, che consente di stimare quali parti del corpo si sono mosse maggiormente, e quali pose sono predominanti.*

*Nel complesso, il sistema realizzato costituisce un laboratorio di base per l'implementazione di algoritmi di analisi di movimenti specifici di corpo ed arti, e*

*in particolare di movimenti correlati a disturbi patologici del sonno, quali ad esempio la sindrome delle gambe senza riposo.*

*Sarà comunque necessario un ulteriore lavoro per quantificare le relazioni tra disturbi del sonno e tipologia e quantità di movimento di determinate parti del corpo. Questo consentirà di ampliare i risultati di questo lavoro, permettendo un monitoraggio meno invasivo e più efficace di soggetti sia sani che patologici.*

## ***Ringraziamenti***

Ringrazio la professoressa Olmo per la sua disponibilità e per avermi concesso l'opportunità di svolgere questo lavoro di tesi.

Un grande ringraziamento al Dr. Roberto Nerino del CNR-IEIIT per avermi supportata durante tutto il lavoro, per avermi fornito il materiale di cui avevo bisogno durante questi mesi di tesi e per gli utili suggerimenti e consigli.

Desidero ringraziare la mia famiglia per il sostegno datomi durante questi anni universitari e per aver creduto in me.

Un grazie ai miei amici e ai miei compagni universitari per questi anni trascorsi insieme e per tutti i momenti condivisi.

# Sommario

1	Introduzione .....	1
2	Fisiologia del sonno - Sonno normale e patologico .....	5
2.1	Segnali fisiologici nel sonno di un soggetto sano .....	5
2.1.1	Patologie del sonno generico .....	7
2.1.2	Il sonno nelle malattie neurodegenerative.....	8
2.2	Stadiazione del sonno.....	9
2.2.1	Segnali fisiologici acquisiti con polisonnografia e dispositivi utilizzati ...	10
2.2.2	Polisonnografia - Valutazione clinica standard.....	11
2.2.3	Polisonnografia – valutazione clinica automatica e Machine Learning....	12
2.3	Ulteriori sistemi e tecniche di valutazione del sonno.....	14
2.3.1	Principi fisici delle tipologie indossabili presentate.....	14
2.3.2	Tecniche basate su approcci video 2D-3D-miste.....	15
2.3.3	Qualità e affidabilità della stadiazione automatica per le tipologie proposte .....	18
3	Segmentazione e caratterizzazione dinamica dei movimenti nel sonno .....	20
3.1	Segmentazione: fase preliminare per la stadiazione automatica e per il riconoscimento di posture, movimenti normali e patologici .....	20
3.2	Stato dell’arte degli algoritmi di segmentazione della dinamica dei movimenti nel sonno .....	22
3.2.1	Approcci IR video-based (2D).....	22
3.2.2	Approcci Depth-based (3D).....	25
3.2.2.1	Motion detection e blob/object tracking.....	25
3.2.2.2	Motion detection, object tracking e individuazione skeleton con tecniche di ML e CNN.....	26
3.2.2.3	Approcci misti.....	28
4	Setup sperimentale e sviluppo di algoritmi specifici per la segmentazione 2D-3D dei movimenti .....	30
4.1	Descrizione HW del sistema di acquisizione utilizzato .....	30
4.2	Descrizione SW del sistema di acquisizione utilizzato .....	33
4.3	Setup sperimentale .....	34
4.4	Descrizione del SW sviluppato per la segmentazione della dinamica dei movimenti nel sonno nell’IR e in Depth .....	37
4.4.1	Pre-elaborazione dei file video acquisiti con Azure Kinect.....	38
4.4.2	Algoritmi di Motion-Body Tracking e Blob Analysis .....	40
4.4.3	Tecniche per il rilevamento del primo piano.....	41
4.4.3.1	Gaussian Mixture modeling .....	41
4.4.3.2	Metodo delle differenze assolute con sfondo dinamico .....	46
4.4.3.3	Tecnica di Background Subtraction scelta e commenti .....	48

4.4.4	Algoritmo di tracking con filtro di Kalman .....	49
5	Risultati sperimentali .....	52
5.1	Risultati Motion-Body Tracking nell'infrarosso .....	52
5.1.2	Check sull'occupazione del letto e analisi del movimento .....	54
5.1.3	Valutazione del rumore nelle immagini e istogramma di movimento .....	57
5.2	Risultati Motion-Body Tracking in profondità .....	60
6	Discussione finale e sviluppi futuri .....	63
7	Bibliografia .....	66

## **Elenco delle tabelle**

**Tabella 2.1** - Dispositivi in PSG e funzionalità

**Tabella 2.2** – Indicatori di valutazione. Tabella tratta [16].

**Tabella 4.1** – Confronto tra le tre versioni di Kinect. Tabella tratta da [40].

**Tabella 4.2** – Modalità di supporto per la telecamera di profondità. Tabella tratta da [44].

**Tabella 4.3** – Modalità di supporto per la telecamera a colori. Tabella tratta da [44].

**Tabella 4.4** – Modalità di temporizzazione per la telecamera di profondità. Tabella tratta da [44].

**Tabella 4.5** – Tipi di formato di immagini. Tabella tratta da [45].

**Tabella 4.6** – Coppia Parametri-Valori utilizzati per vision.BlobAnalysis

**Tabella 4.7** – Valori dei parametri ottimali

**Tabella 4.8** – Coppia Parametri-Valori utilizzati per vision.ForegroundDetector

## Elenco delle figure

**Figura 2.1** – Ciclo circadiano. Immagine tratta da [1].

**Figura 2.2** – Onde tipiche ricavate tramite EEG. Immagine tratta da [35].

**Figura 2.3** – Configurazione Kinect v2. Immagine tratta da [22].

**Figura 3.1** – Schema dei componenti principali di SkelPose. Immagine tratta da [23].

**Figura 3.2** – Microsoft Kinect v.2. Immagine tratta da [24].

**Figura 4.1** – Da sinistra a destra: Kinect v1, Kinect v2 e Azure Kinect. Immagine tratta da [40].

**Figura 4.2** – Schema dispositivo Azure Kinect. Immagine tratta da [40].

**Figura 4.3** – Posizionamento della telecamera Kinect sopra il letto – (a) Vista destra – (b) Vista frontale – (c) Vista sinistra.

**Figura 4.4** – Interfaccia grafica. Immagine tratta da [41].

**Figura 4.5** – Distribuzione di probabilità del pixel con  $K = 3$ . Immagine tratta da [43].

**Figura 4.6** – Risposta all'impulso di un filtro gaussiano. Immagine tratta da [39].

**Figura 4.7** – (a) Istogramma sulla distribuzione di probabilità del pixel sul Foreground - (b) Maschera di Foreground

**Figura 4.8** – (a) Istogramma sulla distribuzione di probabilità del pixel sul Background - (b) Immagine di background originale: letto e ambiente.

**Figura 5.1** – Esempio di due frame estratti dalla sequenza video di input.

**Figura 5.2** – Nel riquadro in alto a sinistra si riporta il frame relativo al file originale IR; in alto a destra la maschera di Foreground. In basso, il grafico che indica la quantità di movimento frame per frame.

**Figura 5.3** – Nel riquadro in alto a sinistra è rappresentato il frame relativo al file originale IR; in alto a destra è riportata la maschera di Foreground. In basso, il grafico che indica la quantità di movimento frame per frame. La differenza rispetto alla Figura 5.1 consiste nella presenza di un unico blob. Il pallino in rosso si riferisce al picco relativo al movimento nel frame considerato.

**Figura 5.4** – Situazione in cui il letto è vuoto. La parte destra dell'immagine è tutta nera perché non è presente il soggetto e, quindi, il primo piano.

**Figura 5.5** – Sovrapposizione tra il grafico relativo all'occupazione del letto (in nero, valore binario 0-1) e il grafico relativo alla quantità di movimento normalizzata (in rosso), frame per frame.

**Figura 5.6** – (a) Intensità media e (b) varianza di una sequenza di immagini IR dello sfondo (unità arbitraria).

**Figura 5.7** – Rilevamento del movimento dalla differenza (differenza tra frame normalizzata tra -1/+1).

**Figura 5.8** – Istogramma della quantità di movimento complessiva della sequenza. In colori più caldi, le zone a maggior movimento.

**Figura 5.9** – Istogramma della quantità complessiva di movimento a sei diverse scale temporali (partendo dall'alto a sinistra) di durata 0, 1, 2, 4, 8, 16 secondi.

**Figura 5.10** – A sinistra si riporta il frame originale estratto dalla sequenza depth. A destra il risultato del metodo delle differenze assolute statico.

**Figura 5.11** – Sovrapposizione tra il grafico relativo all'occupazione del letto (in nero) e il grafico relativo alla quantità di movimento (in rosso) per le sequenze di profondità.

## **Elenco degli acronimi**

**REM** – Rapid Eye Movement

**NREM** – Non Rapid Eye Movement

**OSAS** – Obstructive Sleep Apnea Syndrome

**RLS** – Restless Legs Syndrome

**PLMD** – Periodic Limb Movements Disorder

**RBD** – Rapid eye movement sleep Behavior Disorder

**AASM** – American Accademy of Sleep Medicine

**ECG** – Elettrocardiogramma

**EEG** – Elettroencefalogramma

**EMG** – Elettromiogramma

**EOG** – Elettrooculogramma

**HRV** - Heart Rate Variability

**PPG** – Photoplethysmography

**IR** – Infrared

**NIR** – Near Infrared

**SVM** – Support Vector Machines

**kNN** – K-nearest neighbors

**CNN** – Convolutional Neural Networks

**ML** – Machine Learning

**GMM** – Gaussian Mixture Model

# 1 Introduzione

Il monitoraggio dell'attività motoria di un soggetto durante il sonno è un modo utile per rilevare anomalie riconducibili, per esempio, a malattie neurologiche tramite le pose che il soggetto assume e tramite i suoi movimenti. Un soggetto sano spende buona parte del suo tempo nello stato fisiologico del sonno. Solo ultimamente, però, si sta dando importanza allo studio e all'analisi del sonno come aspetto importante della qualità della vita e, in particolare, alla valutazione sia della qualità del sonno in soggetti sani che della presenza di disturbi in soggetti affetti da patologie.

I disturbi del sonno infatti sono spesso sottovalutati. Alcuni disturbi del sonno sono correlati a malattie e patologie neurologiche, quali ad esempio Alzheimer, Parkinson e sindrome delle gambe senza riposo. Per questo motivo, focalizzarsi sui risultati ottenuti dall'analisi del sonno può essere importante per scoprire in maniera tempestiva un disturbo premonitore di malattia o, se preesistente, un suo aggravamento.

Gli approcci di valutazione quantitativa dei movimenti durante il sonno oggi esistenti si basano sull'elettromiografia e/o sull'accelerometria. L'analisi del sonno spesso avviene all'interno ospedali o ambulatori, fatto che comporta molti svantaggi. Il soggetto, infatti, non si trova in un ambiente a sé familiare e ciò influisce sulla valutazione della qualità stessa del sonno.

Un altro problema può essere caratterizzato dal fatto che i sistemi di analisi del sonno oggi conosciuti, come il polissonigrafo, possono risultare invasivi e, quindi, non confortevoli per il paziente analizzato. Con questi approcci non si riesce ad ottenere una visione globale dei movimenti durante il sonno. Si riesce solo a stimare la quantità del movimento degli arti o dei muscoli su cui sono stati posizionati i sensori.

Per questi motivi, oggi è di grande interesse effettuare l'analisi del sonno in ambiente domestico, direttamente presso la stessa abitazione del paziente. L'obiettivo è quello di utilizzare un sistema che sia minimamente invasivo, tale da non condizionare i risultati della valutazione. Si vuole quindi ridurre al minimo l'uso di strumenti indossati o attaccati al corpo del paziente, che potrebbero rappresentare fonte di disturbo.

Per raggiungere questo obiettivo riguardo all'analisi dei movimenti durante il sonno, in questo lavoro è stato utilizzato un approccio ottico. L'approccio risulta non invasivo perché i sensori non si trovano a contatto con il paziente e, inoltre, consente di ottenere una visione globale dei movimenti nel sonno.

Per l'analisi del movimento sono state utilizzate camere RGB-Depth e, in particolare, Azure Kinect [40][44]. Le telecamere appartenenti alla famiglia Kinect sono state spesso utilizzate per l'analisi del movimento di un soggetto durante il sonno [24][25][27]. Queste camere permettono di rilevare l'attività motoria e le pose assunte senza utilizzare altri dispositivi. Queste, infatti, non forniscono solamente informazioni relativamente all'attività motoria, ma sono in grado di estrarre altre informazioni (es. segnali acustici del respiro).

Per lo svolgimento di questa tesi, le sequenze video sono state acquisite con l'ultima versione del dispositivo Kinect rilasciata da Microsoft, ovvero Azure Kinect. Essa è in grado, analogamente alle sue precedenti versioni, di fornire contemporaneamente sequenze video sia nell'infrarosso che in profondità, ma con risoluzione e accuratezza migliori.

Lo scopo principale della tesi è lo sviluppo di un sistema automatico in grado di analizzare e caratterizzare i movimenti che il soggetto compie durante il sonno. Questo rappresenta un vantaggio rispetto all'utilizzo delle tecniche standard che richiedono un'analisi visiva delle sequenze video da parte di un operatore. È anche un vantaggio per il soggetto su cui si effettua l'analisi perché non si utilizzano metodi invasivi e non confortevoli.

Il seguente lavoro è diviso in questa Introduzione e in cinque successivi capitoli.

Nel secondo capitolo si analizza lo stato fisiologico del sonno. In particolare, vengono presentate le caratteristiche del sonno di un soggetto sano, quali il ritmo circadiano e le varie fasi del sonno. Vengono descritti anche disturbi e patologie correlate al sonno, evidenziandone principalmente gli aspetti motori. Nel suddetto capitolo, inoltre, è presentata la tecnica standard di analisi del sonno, la polisonnografia, e alcune metodologie di analisi dei movimenti nel sonno basate su elettromiografia e/o accelerometria.

Nella parte iniziale del terzo capitolo si analizza la correlazione tra qualità sonno e i movimenti e le pose del paziente. Nella seconda parte, invece, si descrive lo stato dell'arte degli algoritmi di analisi del movimento che utilizzano sequenze video e, in particolare, gli algoritmi di segmentazione e tracking; sono presentati sia approcci basati su telecamere nell'infrarosso che approcci basati su telecamere di profondità.

Nel quarto capitolo è descritta la parte sperimentale della tesi. Nella prima parte si descrive il sistema di acquisizione video utilizzato sia dal punto di vista dell'hardware che dal punto di vista del software. La camera utilizzata è Microsoft Azure Kinect, la quale permette di acquisire contemporaneamente sequenze video nell'infrarosso e di profondità. Questo consente di effettuare l'analisi del movimento confrontando entrambe le tipologie di informazione. Viene inoltre descritto il setup sperimentale utilizzato per l'acquisizione dei video in ambiente domestico.

Nella seconda parte del quarto capitolo si descrive il software sviluppato e gli algoritmi di motion detection e di tracking implementati in ambiente MATLAB™. Per la motion detection sono stati confrontati diversi algoritmi per effettuare la sottrazione dello sfondo (background subtraction). In particolare, sono stati valutati il metodo delle misture gaussiane e il metodo delle differenze assolute.

Nel quinto capitolo si riportano i risultati ottenuti. Il software sviluppato consente di realizzare un "actigrafo globale", in quanto permette di analizzare i movimenti dell'interno corpo durante il sonno; inoltre, consente di rilevare la presenza del soggetto nel letto. In aggiunta, il sistema fornisce la stima di alcuni parametri del sonno, quali l'efficienza e una descrizione temporale compatta del movimento tramite istogramma bidimensionale a varie scale temporali. Questo permette di capire quali parti del corpo si sono mosse maggiormente e quali pose predominano rispetto ad altre.

Il sesto e ultimo capitolo si focalizza sulla conclusione di questo lavoro e sui risultati ottenuti. Il sistema hardware e software (HW/SW) realizzato ha consentito l'analisi preliminare dei movimenti di parti del corpo durante il sonno tramite l'implementazione di una serie di algoritmi in ambiente MATLAB™. L'ambiente HW/SW implementato costituisce un laboratorio che consentirà di

ampliare i risultati qui ottenuti, permettendo lo sviluppo e la sperimentazione di algoritmi di analisi di movimenti specifici correlati a disturbi patologici del sonno, come ad esempio la sindrome delle gambe senza riposo.

## **2 Fisiologia del sonno - Sonno normale e patologico**

Il sonno è uno stato fisiologico che ricopre circa un terzo della vita umana. Fin da quando si è neonati dormire è una necessità. Un soggetto sano, in media, dorme circa otto ore al giorno. È stato riscontrato che la stragrande maggioranza della popolazione è affetta da problematiche relative al sonno, come insonnia, apnea notturna, sonnambulismo. I disturbi del sonno, tuttavia, sono spesso sottovalutati. Soltanto negli ultimi decenni una particolare attenzione è stata rivolta allo studio e all'analisi dei disturbi del sonno. Si è trovata, infatti, una correlazione tra le problematiche relative al sonno e determinate patologie, come la malattia di Parkinson o le malattie neurologiche in generale. Altri fattori esercitano una notevole influenza sulla qualità del sonno, come obesità, stress, stato emotivo e psicologico. Analizzare i disturbi del sonno, per un soggetto patologico, spesso può aiutare a diagnosticare precocemente una malattia e a trattarla in modo tempestivo. Per un soggetto sano, invece, monitorare il sonno aiuta a valutarne la qualità.

### **2.1 Segnali fisiologici nel sonno di un soggetto sano**

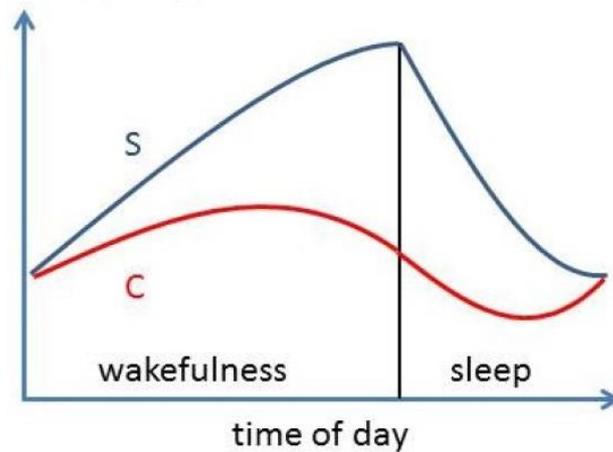
Comprendere i meccanismi del sonno è importante in quanto il sonno svolge un ruolo cruciale in molti processi fisiologici come la formazione della memoria, la funzione immunitaria, la funzione endocrina, il funzionamento cardiovascolare, l'umore [10]. Durante il sonno, infatti, si verifica un'alterazione dell'attività cerebrale, muscolare, oculare, cardiaca e respiratoria.

Il sonno rappresenta uno stato in cui la mobilità e la coscienza nei confronti del mondo esterno sono ridotti. Durante il sonno la mente consolida la memoria, la tensione muscolare diminuisce e si può osservare un decremento della pressione sanguigna, della frequenza cardiaca e della temperatura corporea. Solitamente, un soggetto sano presenta un'attività fisica ridotta durante il sonno, nonostante l'attività cerebrale in alcune fasi del sonno [7][10][14] come la REM (Rapid Eye Movement) sia molto presente [6].

La registrazione dei segnali fisiologici durante il sonno e la sua conseguente analisi è importante per poter effettuare eventuale diagnosi e trattamenti. I segnali fisiologici, infatti, contengono informazioni relative alle fasi del sonno [7]. Il segnale elettroencefalografico (EEG) è quello maggiormente utilizzato per estrarre

informazioni relativamente agli stadi del sonno. In particolare, si analizzano e si valutano caratteristiche del dominio nel tempo, caratteristiche spettrali, caratteristiche tempo-frequenza [14]. Attraverso l'elettrocardiogramma (ECG), invece, si monitora l'attività elettrica del cuore. Nel caso di un soggetto sano i segnali ECG sono strutturati e le singole componenti del segnale possono essere identificate attraverso l'ispezione visiva. La frequenza respiratoria e la frequenza cardiaca variano a seconda delle fasi del sonno. L'elettrooculogramma (EOG) permette di monitorare i movimenti oculari, fornisce quindi informazioni importanti per il rilevamento dello stadio REM in cui gli occhi si muovono rapidamente [6].

La regolazione dell'alternanza di sonno e veglia si basa sul *ritmo circadiano*. Questo rappresenta per l'uomo una sorta di orologio biologico. In **Figura 2.1** si riporta il modello a due processi per la regolazione del sonno, in particolare si raffigura il processo S e il processo C. Il primo consiste nella fase omeostatica che favorisce il sonno; l'omeostasi è il processo che garantisce e regola l'equilibrio tra sonno e veglia. Il processo C invece si occupa di mantenere la veglia, si tratta di un orologio circadiano ovvero un complesso sistema neurochimico che utilizza segnali dall'ambiente in modo tale da ricreare un ritmo interno giorno-notte. Il processo C contrasta la spinta omeostatica del sonno durante il giorno e la aumenta di notte. Questo processo è molto influenzato dai cambiamenti di luce ed è caratterizzato da una temperatura corporea bassa; durante questa fase avviene un grande rilascio di melatonina. L'esposizione anche a piccole quantità di luce può sopprimere la secrezione di melatonina e favorire la veglia e aumentare la temperatura corporea. Il processo S consiste nella privazione del sonno. La privazione del sonno causa onde cerebrali più lente nella corteccia frontale, capacità di attenzione ridotta, maggiore ansia, disturbi della memoria e umore scontroso. Contrariamente, un organismo ben rilassato è più predisposto ad avere una memoria più robusta e a mantenere un buon umore. Nel punto in cui il livello della curva S raggiunge il suo apice e contemporaneamente il livello di C è basso, si passa alla fase del sonno. Si osserva, inoltre, come il bisogno di dormire aumenti durante la fase diurna e diminuisca durante la notte, infatti più a lungo un organismo è sveglio più sente il bisogno di dormire, in questo senso si parla di "debito del sonno" [1].



**Figura 2.1** – Ciclo circadiano. Immagine tratta da [1].

### 2.1.1 Patologie del sonno generico

I disturbi del sonno generici sono molto comuni e a volte possono essere fonte di allarme di una malattia neurologica. L'insonnia rappresenta [2] uno dei disturbi più comuni al mondo ed è la difficoltà che un soggetto ha nel mantenere il sonno o nel riuscire ad addormentarsi. Recenti studi hanno dimostrato che l'insonnia può essere un sintomo molto comune in malattie neurodegenerative, come la malattia di Parkinson e la malattia di Alzheimer [2].

Un altro disturbo molto comune è rappresentato dalla sindrome di apnea notturna ostruttiva (OSAS). I soggetti affetti da questo disturbo presentano spesso una sonnolenza diurna, il sonno è intermittente e caratterizzato da numerosi episodi di veglia. L'apnea notturna è spesso associata a problemi respiratori che si possono ripercuotere in problemi neurologici più gravi e in problemi cardiaci. A causa dell'apnea notturna, il paziente si sveglia regolarmente durante la notte per riprendere la respirazione. Il risveglio frequente si traduce in una qualità del sonno molto scarsa e in un eccessivo affaticamento diurno. Di solito, l'apnea notturna può essere accompagnata da un forte russamento che potrebbe essere monitorato da un microfono [35].

La narcolessia è una patologia neurologica associata ad un'eccessiva sonnolenza diurna, caratterizzata da un sonno intermittente e frammentato e da continui attacchi di sonno incontrollati.

I disturbi del sonno, inoltre, possono essere associati ai disturbi motori. Tra questi si annoverano la sindrome delle gambe senza riposo (RLS) e il disturbo periodico del movimento degli arti (PLMD) che è una conseguenza di RLS. PLMD è caratterizzato dal movimento ripetuto degli arti (soprattutto inferiori) durante il sonno [3]. L' RLS rappresenta una condizione che comporta un impulso incontrollabile a muovere le gambe, di solito causato da una sensazione di disagio.

Alcuni disturbi possono essere classificati come disturbi del sonno NREM (non REM) e altri come disturbi del sonno REM. I primi consistono in stati di confusione, sonnambulismo, disturbi alimentari legati al sonno. I disturbi del sonno REM, invece, consistono nella comparsa di un'attività motoria associata, per esempio, alla situazione che si riproduce durante un sogno; in queste fasi il soggetto potrebbe anche emettere dei suoni. [35]

### **2.1.2 Il sonno nelle malattie neurodegenerative**

I disturbi del sonno possono essere associati a malattie neurodegenerative molto gravi, tra cui Alzheimer e Parkinson [4]. In particolare, nei soggetti parkinsoniani il sonno risulta essere molto frammentato e poco stabile. Recenti studi hanno dimostrato una correlazione tra la difficoltà di movimento che il soggetto riscontra durante la veglia e la difficoltà dello stesso nell'addormentarsi o nel mantenere il sonno. Durante il sonno i soggetti parkinsoniani tendono a muoversi circa il doppio rispetto ad un soggetto sano, questo è dovuto alla disfunzione del sistema dopaminergico che si verifica durante il Parkinson [4]. I disturbi prevalenti di soggetti parkinsoniani consistono per esempio in apnee ostruttive del sonno (OSA), disturbi comportamentali (RBD) e sindrome delle gambe senza riposo (RLS). Un paziente parkinsoniano tende ad avere molti episodi di sonnolenza diurna, questo comporta numerosi episodi di risveglio durante la notte con conseguente fasi di sospensione del respiro della durata di molti secondi. A differenza di un soggetto sano, il paziente affetto da Parkinson perde il ruolo ristorativo del sonno. Nei soggetti sani, inoltre, il corpo durante la fase REM è rilassato, cosa che non si verifica in un soggetto parkinsoniano, il quale a causa di disfunzioni di particolari strutture del Sistema Nervoso Centrale presenta un tono muscolare non inibito. Il paziente, infatti, può parlare, urlare e compiere dei bruschi movimenti. Questi

episodi si verificano soprattutto nella fase REM, in cui può comparire anche la sindrome RLS in cui il paziente sente il bisogno incessante di muovere le gambe. Studi dimostrano come i soggetti Parkinsoniani mostrano un'attività EMG maggiore durante il sonno REM e NREM nel muscolo mentale e nel muscolo tibiale rispetto ai soggetti sani [11].

Anche nella malattia di Alzheimer il sonno ha un ruolo cruciale. Questa è correlata a riduzioni di caratteristiche specifiche dello stadio di sonno leggero N2, come i fusi del sonno e i complessi K, che diventano poco formati, con ampiezze e frequenze più basse [5].

## **2.2 Stadiazione del sonno**

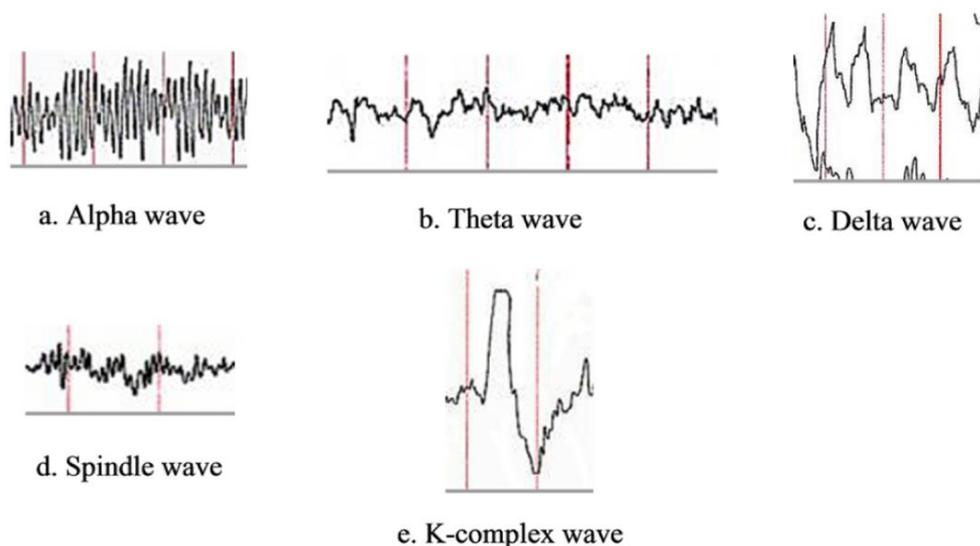
Il sonno di un soggetto sano è caratterizzato da un ciclo che si ripete più volte durante tutto il periodo, ciclo che cambia in base all'individuo. Ogni ciclo ha la durata di circa 90 minuti ed è caratterizzato dall'alternanza tra la fase REM e la fase NREM. La fase REM è caratterizzata da movimenti rapidi degli occhi e da un'attività dell'elettroencefalogramma (EEG) corticale con frequenza più rapida simile alla fase di veglia, accoppiata con atonia motoria scheletrica. Alcuni autori, infatti, denominano la fase REM come “sonno attivo” o “sonno paradossale” [6]. Rispetto alla fase NREM, questa rappresenta circa il 20-25% del sonno complessivo.

L'American Academy of Sleep Medicine (AASM) suddivide la fase NREM in ulteriori stadi: N1, N2 e N3 [10][14]. In totale, quindi, si ottengono cinque stadi di classificazione del periodo del sonno: veglia, REM, N1, N2 e N3. I vari stadi sono caratterizzati dall'alternanza delle onde cerebrali, ovvero dall'attività elettrica ritmica e ripetitiva del tessuto nervoso del sistema nervoso centrale.

Le onde alfa sono oscillazioni neurali della frequenza di circa 8-12 Hz. Hanno origine dal lobo occipitale durante la fase di veglia, mentre il soggetto si sta rilassando e tiene gli occhi chiusi. Le onde alfa si riducono nel momento in cui si aprono gli occhi, ci si addormenta oppure si è nella fase di sonnolenza. Le onde theta, invece, presentano componenti di frequenza nella gamma 4-7 Hz. Sono più frequenti nei bambini piccoli; negli adulti queste onde appaiono durante gli stati meditativi, sonnolenti, ipnotici, ma non durante le fasi più profonde del sonno.

Diversi tipi di patologie cerebrali possono dare origine a onde theta corticali anormalmente forti o persistenti. Le onde delta presentano frequenze intorno allo 0.5-4 Hz. Le onde delta caratterizzano il sonno a onde lente e aiutano a caratterizzare la profondità dello stato di sonno (**Figura 2.2**).

Lo stadio N1 occupa dal 2 al 5% del tempo di sonno totale, corrisponde allo stato di transizione dalla veglia al sonno, ovvero si verifica il passaggio dall'attività caratterizzata da onde alfa a una con frequenza mista con una maggiore predominanza di onde theta. Si verifica, inoltre, una diminuzione dell'attività motoria rilevabile tramite il segnale elettromiografico (EMG). Durante la fase N2 si riscontra un aumento dell'ampiezza e un aumento dell'attività EEG a bassa frequenza, sono visibili inoltre i complessi K. Infine, nella fase N3, l'EEG è caratterizzato da una predominanza delle onde delta sincronizzate, bassa attività muscolare e una risposta limitata agli stimoli esterni. Man mano che il sonno progredisce, gli intervalli N3 diminuiscono. Quest'ultimi corrispondono a circa il 15-20 % del tempo di sonno totale e più si progredisce con l'età, più la durata degli intervalli N3 si riduce.



**Figura 2.2** – Onde tipiche ricavate tramite EEG. Immagine tratta da [35].

### 2.2.1 Segnali fisiologici acquisiti con polisonnografia e dispositivi utilizzati

Analizzare il sonno di un soggetto richiede il monitoraggio di alcuni segnali vitali, come la frequenza cardiaca, la pressione sanguigna, la temperatura corporea, la frequenza respiratoria. La polisonnografia (PSG) è il metodo standard utilizzato per

la valutazione del sonno. Si tratta di un sistema di analisi del sonno molto invasivo che permette la diagnosi di numerosi disturbi associati. La valutazione polisomnografica avviene comunemente in laboratorio tramite un polisomnografo, un sistema di acquisizione e analisi dei diversi segnali fisiologici caratterizzanti il sonno. I dispositivi integrati nel polisomnografo sono generalmente l'elettroencefalogramma (EEG), l'elettrooculografo (EOG), un elettromiografo (EMG) e un elettrocardiografo (ECG). Nella **Tabella 2.1** si riportano le funzionalità di ogni dispositivo.

<i><b>Dispositivi in PSG e funzionalità</b></i>	
<i>ECG</i>	Monitoraggio dell'attività cardiaca
<i>EEG</i>	Monitoraggio dell'attività elettrica del cervello
<i>EMG</i>	Monitoraggio dell'attività muscolare
<i>EOG</i>	Monitoraggio del movimento degli occhi

**Tabella 2.1** - Dispositivi in PSG e funzionalità

Il paziente tipicamente indossa molti elettrodi per consentire l'acquisizione dei vari segnali di interesse sopracitati e potrebbe essere necessario aggiungere ulteriori sensori, quali pulsossimetri, utilizzati per valutare la saturazione di ossigeno nel sangue. Normalmente l'acquisizione dei segnali fisiologici dura per l'intera notte di sonno del soggetto e, in seguito, si passa alla fase di valutazione. Si evince che effettuare l'esame di polisomnografia non è un'operazione immediata e richiede del tempo.

### **2.2.2 Polisomnografia - Valutazione clinica standard**

La valutazione clinica del sonno viene generalmente effettuata da un neurologo esperto in sonnologia e si basa sull'analisi dei segnali fisiologici in intervalli tipicamente di 30 secondi (epoche), e la loro classificazione nei singoli stadi REM, N1, N2, N3. Durante l'analisi di una registrazione, viene valutata la quantità e la qualità del sonno. Si deve tenere conto di quante volte il soggetto si sveglia e per quanto tempo, il tempo di sonno totale e come si distribuisce il sonno nelle varie fasi. Tramite la PSG si può notare come il sonno REM sia caratterizzato da una bassa ampiezza del segnale EEG, accompagnato da rapidi movimenti degli occhi e da ipotensione nella muscolatura posturale. Nel sonno NREM, invece, con l'EEG si può osservare un aumento dell'ampiezza. Tramite l'utilizzo di opportuni sensori (es. pulsiossimetri), è possibile riconoscere episodi di apnea, molto frequenti del

sonno. La tecnica del PSG permette una valutazione delle singole fasi del sonno e il riconoscimento di eventuali disturbi [7].

L'analisi della qualità del sonno tramite PSG si basa sulla valutazione di determinate metriche. Le metriche possono essere suddivise per durata, intensità, continuità, stabilità, frequenza ed episodi di sonno [16]. Nella **Tabella 2.2** sono descritte le metriche di valutazione.

<i><b>Metriche PSG</b></i>	
<i>Durata</i>	Metrica basata sulla durata temporale degli stadi del sonno
<i>Intensità</i>	Percentuale del tempo speso in uno specifico stadio del sonno
<i>Continuità</i>	Grado di frammentazione del sonno
<i>Stabilità</i>	Eventi di disturbo del sonno
<i>Frequenza</i>	Ripetizione di una stessa fase del sonno
<i>Episodi di sonno</i>	Descrizione dei periodi di sonno

**Tabella 2.2** – Indicatori di valutazione. Tabella tratta [16].

La polisonnografia risulta però una tecnica molto scomoda per il paziente, il quale è costretto a passare l'intera notte in un laboratorio indossando molti elettrodi collegati con fili ai vari dispositivi e sensori. La valutazione manuale clinica, inoltre, richiede molto tempo, per cui se qualcosa non è stato rilevato, ripetere tutta la procedura è sicuramente molto complicato. Per questo motivo, negli ultimi anni si privilegiano tecniche automatiche di Machine Learning [8][13][14].

L'utilizzo di tecniche automatiche tramite Machine Learning permette di effettuare una valutazione più efficiente, si risparmia in termini di tempo e si è più veloci nell'effettuare l'analisi di una registrazione. Grazie a queste tecniche si è in grado di ottenere buone performance per quanto riguarda il riconoscimento e l'analisi delle fasi del sonno. Un problema di questi metodi è caratterizzato dall'affidabilità. Spesso gli esperti sono costretti a rifare il punteggio delle registrazioni per ottenere risultati affidabili. Bisogna anche ricordare che l'accordo tra esperti nella valutazione degli stadi del sonno è parziale, intorno all'80% [8].

### **2.2.3 Polisonnografia – valutazione clinica automatica e Machine Learning**

Negli ultimi decenni le tecniche di Machine Learning hanno acquisito un ruolo fondamentale nella valutazione del sonno in quanto permettono un'analisi

automatica e più veloce. I classificatori, infatti, vengono spesso addestrati in modo da automatizzare il processo di classificazione. In base, quindi, ad un set di addestramento si trovano i valori per cui il test risulti accurato. Grazie al continuo progresso computazionale e tecnologico gli algoritmi adesso sono più accessibili anche dal punto di vista economico e permettono di effettuare le analisi in tempo reale [9].

Alcuni dei classificatori più utilizzati includono: SVM (Support Vector Machines), kNN (k-neighbors), reti neurali artificiali. Le reti SVM, per esempio sono dei classificatori di tipo binario utilizzate per la classificazione delle immagini, per il riconoscimento del testo, per il riconoscimento del movimento [9]. Potrebbero quindi essere utili nel momento in cui si vuole analizzare l'attività motoria di un soggetto durante il sonno. Con le reti neurali, invece, si riesce a raggiungere una precisione maggiore [8][17].

Alcune tecniche di Machine Learning, inoltre, sfruttano le relazioni tra HRV (Heart Rate Variability) e fasi del sonno. Si è dimostrato come diversi algoritmi sono in grado di valutare automaticamente le fasi del sonno in base all'HRV, tipicamente misurato con l'ECG, talvolta in combinazione con lo sforzo respiratorio [12]. Inoltre, molti metodi di apprendimento automatico si incentrano sulla rilevazione e sulla previsione degli stati sonno-veglia a partire dai risultati ricavati dall'actigrafia; si tratta di metodi generalizzati addestrati utilizzando i dati della popolazione. I metodi generalizzati, però, non lavorano bene sul singolo individuo in quanto non considerano la differenza delle caratteristiche in termini di età, fattori ambientali, stili di vita. Per questo si sono sviluppati metodi di apprendimento automatico in cui i parametri possono variare in base alla tipologia di individuo. In questo modo si ottiene una suddivisione delle fasi sonno-veglia più veritiera e affidabile [13].

Le tecnologie di apprendimento automatico affidano il lavoro alle macchine consentendo al personale medico di svolgere il loro lavoro in modo più veloce. Le tecniche di Machine Learning, inoltre, permettono di ottenere un monitoraggio del sonno più efficiente [9] e un calcolo delle metriche di valutazione del sonno automatico, veloce e facilmente generalizzabile [16].

## **2.3 Ulteriori sistemi e tecniche di valutazione del sonno**

Esistono altri approcci per l'analisi del sonno, nonostante il PSG rappresenti lo standard principale. Nell'ultimo decennio, infatti, sono stati sviluppati nuovi metodi e tecniche in grado di registrare l'attività cardiaca e respiratoria durante il sonno di un soggetto. Il sonno può essere monitorato tramite tecniche di actigrafia, balistocardiogramma (BCG), elettrocardiografia (ECG), elettromiografia (EMG), fotopletismografia (PPG) e analisi del respiro (fasce toraciche respiratorie). Le tecniche elencate hanno come fattore comune la presenza di sensori indossabili e dispositivi attaccati al corpo umano.

Negli ultimi decenni l'interesse è rivolto verso sistemi di valutazione del sonno che permettono il monitoraggio attraverso dispositivi non indossabili. Per la parte che riguarda il movimento, questo è reso possibile grazie al progresso della visione computerizzata tramite l'utilizzo di videocamere 2D-3D.

### **2.3.1 Principi fisici delle tipologie indossabili presentate**

L'actigrafia è un metodo che si utilizza quando il PSG non può essere effettuato o quando si vogliono effettuare analisi semplificate in ambiente domiciliare. I dispositivi di actigrafia sono in grado di monitorare l'attività motoria per un periodo di tempo di 24 ore [17]. L'actigrafia consiste nell'indossare un accelerometro al polso in grado di rilevare i movimenti durante il sonno, indicativi di stati di veglia. L'actigrafo si basa principalmente sui dati dell'accelerometro, anche se potrebbero includere giroscopi e magnetometri. Per la valutazione ci si basa generalmente sull'osservazione del sonno del paziente di sette giorni consecutivi [17]. L'actigrafo può essere utilizzato anche a casa, ciò è un vantaggio in quanto il paziente assume un comportamento del tutto naturale e quindi si trova in una situazione di comfort [17]. Tuttavia, è una tecnica poco precisa in quanto è molto sensibile a tutti i movimenti, anche se poco rilevanti ad indicare lo stadio effettivo del sonno. Un problema è dato dal fatto che l'actigrafo raccoglie i dati di un solo polso; se a muoversi è un'altra parte del corpo il movimento non verrà rilevato in maniera chiara ed efficace.

Ulteriori tecniche basate sui segnali del PSG si concentrano sulla misurazione dell'attività elettrica del cuore tramite elettrodi posti nelle lenzuola [9]. Altre, come

il balistocardiografia (BCG), si occupano di analizzare l'attività meccanica di cuore e polmoni durante il sonno. In particolare, le accelerazioni impresse al corpo dal flusso sanguigno e dai movimenti respiratori, che consentono l'acquisizione di frequenza cardiaca e respiratoria e l'analisi dei loro valori e della loro variabilità nel tempo, permettono di ottenere informazioni indirette sugli stadi e sulla qualità del sonno [3].

Il sonno può essere analizzato anche a partire da piccoli sensori indossabili miniaturizzati; generalmente questi sono posti in una o più delle seguenti aree corporee: polso, torace, caviglia, anca [17]. Possono essere applicati più sensori contemporaneamente in modo da ottenere un'informazione più dettagliata degli stadi del sonno e delle posizioni che il corpo assume durante questi. Alcuni sensori possono includere anche dei termometri che contribuiscono a discriminare gli stati di veglia da quelli di sonno in base al valore della temperatura corporea [17]. Un'altra tecnica presentata in [17] e che può essere utilizzata per la valutazione del sonno è la PPG (Fotopletismografia): si tratta di una tecnica ottica semplice in grado di individuare i cambiamenti volumetrici nel sangue nella circolazione periferica.

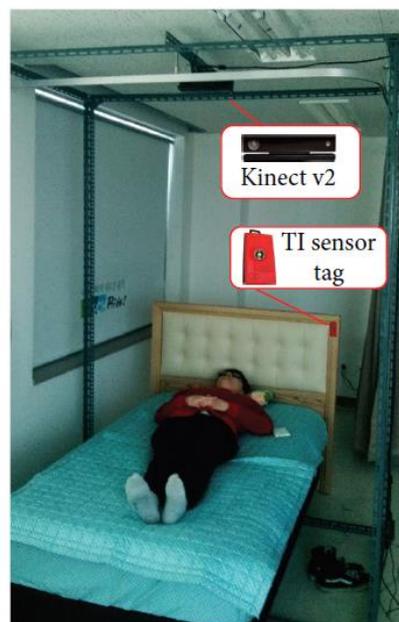
Un sistema utile per prevenire le ulcere da pressione è costituito da un tappetino dotato di sensori di pressione in grado di registrare le immagini e fornire una distribuzione della pressione [31]. Il sistema fornisce un monitoraggio continuo del paziente, ma è stato progettato per soggetti allettati. Tuttavia, potrebbe essere riadattato per monitorare i pazienti durante la notte in modo da mandare un avviso al caregiver nel caso in cui il soggetto resti nella stessa posizione per un lungo periodo. I sensori di pressione si occupano soltanto della fase di acquisizione, la parte di analisi e quindi la fase di addestramento è eseguita in modalità offline. Un problema dei sensori di pressione è rappresentato dal rumore: letture rumorose della pressione possono introdurre errori nelle mappe di pressione, per smorzare questo problema si utilizzano dei filtri [15].

### **2.3.2 Tecniche basate su approcci video 2D-3D-miste**

Grazie al progresso della visione computerizzata, oggi è possibile estrarre ricche informazioni dalle sequenze video di un soggetto che dorme. Generalmente si utilizzano camere sensibili all'infrarosso (IR), per evitare luci che disturberebbero

il sonno del soggetto analizzato. Nelle ricerche più recenti sull'analisi del sonno, e in particolare nell'analisi dei movimenti durante il sonno, oltre alle camere IR sono state utilizzate anche camere termiche e camere di profondità [9].

In particolare, da più di un decennio esistono sul mercato consumer camere RGB-Depth che sono in grado di acquisire sia video nello spettro visibile che nell'infrarosso. Tramite illuminatori IR queste camere sono in grado di illuminare la scena con luce IR senza disturbare il soggetto analizzato. Inoltre, tramite la particolare struttura e/o modulazione della luce proiettata, sono in grado anche di stabilire la distanza e la profondità dei punti nello spazio illuminati e acquisiti dalla camera IR, fornendo di fatto un video della mappa di profondità della scena. Ad esempio, nel caso della camera RGB-Depth Microsoft Kinect v1, la profondità viene calcolata tramite luce strutturata IR e triangolazione, mentre nella Microsoft Kinect v2 la profondità degli elementi della scena è ottenuta con la tecnologia di modulazione della luce IR proiettata e del suo tempo di volo [9]. Microsoft Kinect v2 fornisce il rilevamento delle articolazioni del corpo umano procurandando la stima del grado del movimento di un modello semplificato del corpo (skeleton) [20]. Kinect v2 richiede che ci sia una distanza minima per il rilevamento degli oggetti, la telecamera va quindi posizionata ad una distanza di circa 0.5-4.5 m dall'oggetto [20]. In **Figura 2.3** si riporta una possibile configurazione del sistema di monitoraggio di Kinect v2 [22].



**Figura 2.3** – Configurazione Kinect v2. Immagine tratta da [22].

Le camere IR possono quindi operare anche in una stanza completamente buia, col vantaggio per quelle RGB-Depth citate di poter acquisire non solo sequenze video bidimensionali nell'infrarosso, ma anche l'informazione di profondità associata alla scena acquisita. Inoltre, operando solo su informazioni di profondità, esse consentono di preservare meglio la privacy del paziente stesso [9].

Le tecniche analizzate in **2.3.1** sono state studiate anche con approcci basati su video 2D-3D. Tra le metodologie più semplici basate su video rientra la video-polisonnografia. La video-polisonnografia consiste nella registrazione di un'intera notte tramite una telecamera ad infrarossi e nell'analisi della stessa in un secondo momento [18]. Tramite questa tecnica si riesce a ridurre l'intera registrazione in una più breve in modo da avere solamente i frames in cui viene rilevato il movimento. Da un video iniziale di 7-8 ore se ne ricava uno di circa 30-60 minuti. Dall'analisi di questo breve video si potrebbe risalire ad eventuali disturbi del sonno [18]. Un'altra tecnica basata su video è la video-actigrafia, metodo che si occupa di stimare il movimento a partire dalle immagini infrarosse acquisite [19]. Questa tecnica è in grado di individuare varie posizioni di sonno sotto diverse condizioni di illuminazioni, angoli di visione, letti e coperte. Il sistema presentato in [19] lavora in real-time ed è in grado di fornire informazioni relative al sonno di un soggetto in modo dettagliato senza la presenza di sensori. Il sistema lavora nel vicino infrarosso (NIR) con una telecamera NIR invisibile all'occhio umano ed utilizza un algoritmo di analisi video in grado di estrarre i movimenti della persona che dorme [19]. Un'ulteriore tecnica automatizzata è la video-sonnografia (VSG), una tecnica che consiste nella registrazione e nella valutazione degli stadi del sonno e di veglia in adulti e bambini; in questo caso si tratta di un approccio 3D in quanto viene presa in considerazione anche la componente della profondità.

Gli approcci 2D, inoltre, possono essere combinati con altri tipi di approcci come quelli termici. La termografia è un metodo passivo senza contatto in grado di rilevare la radiazione emessa da particolari parti del corpo umano nella gamma infrarossa (IR) dello spettro elettromagnetico [21]. Tramite una termocamera la termografia permette di misurare il segnale fisiologico come i cambiamenti di temperatura o le differenze di temperatura dovute al flusso sanguigno. Tutti i metodi proposti in questa sezione non prevedono che i sensori siano in diretto contatto con il paziente, sono quindi dei metodi senza alcun contatto.

Una tecnica basata su un approccio 2D consiste nell'utilizzo di un sensore ottico che è in grado di misurare sia i movimenti del corpo sia di calcolare la forma d'onda respiratoria di un soggetto durante il sonno, superando così i limiti dell'actigrafia del polso [34]. Lo studio proposto in [34] si basa sull'utilizzo di una telecamera che a differenza di tutti gli altri sistemi non è posizionata in alto sulla parete o sul soffitto con vista sul letto. La telecamera può trovarsi su una superficie bassa, come sul comodino del paziente che deve essere monitorato. La telecamera in questione lavora nell'infrarosso ed è in grado di visualizzare le registrazioni in un ambiente buio. La novità proposta in [34] è che il paziente può essere monitorato anche se nel letto si trova un'altra persona. Il primo step, infatti, consiste nello sviluppo di un algoritmo di segmentazione in grado di discriminare l'area del paziente e l'area della persona accanto in modo poi da effettuare l'analisi solo sul paziente.

### **2.3.3 Qualità e affidabilità della stadiazione automatica per le tipologie proposte**

Le tecniche in grado di fornire una stadiazione automatica del sonno risultano più veloci e meno costose di quelle manuali. Spesso, però, i risultati che si ottengono non sono tanto affidabili quanto quelli ottenuti tramite tecniche manuali. Per esempio, la stadiazione automatica effettuata attraverso l'utilizzo delle telecamere Kinect può fornire informazioni meno accurate e affidabili se la telecamera si trova al confine degli oggetti o se gli oggetti sono molto lontani dal sensore. Tuttavia, si tratta di sistemi molto economici [15] che possono fornire informazioni meno accurate ma comunque utili.

Inoltre, gli approcci basati sul video sono più inclini a problemi di privacy. Tuttavia, con i progressi nell'imaging a infrarossi e nella tecnologia dei sensori di visione 3D, la gestione dei problemi di privacy è stata molto più efficace rispetto all'utilizzo dei sistemi convenzionali basati su telecamere 2D poiché non elaborano alcuna informazione cromatica che potrebbe comportare possibilità di rivelare l'identità dell'individuo [15].

I metodi di imaging 3D proteggono la privacy dell'individuo poiché in tali metodi viene elaborata solo l'informazione di profondità del soggetto [15].

La qualità dei sistemi automatici, inoltre, varia in base alla presenza o all'assenza di una coperta. La presenza di una trapunta o di una coperta, infatti, è problematica

nel monitoraggio del sonno. La precisione del rilevamento è limitata dallo spessore della trapunta [15]. Le telecamere NIR presentate in **3.3.2** hanno il vantaggio di essere a basso costo, ma presentano una bassa risoluzione.

La sfida consiste nello sviluppare sistemi robusti e accurati con una quantità adeguata di informazioni visive utilizzando tecnologie di rilevamento a basso costo.

### **3 Segmentazione e caratterizzazione dinamica dei movimenti nel sonno**

Sviluppare un robusto sistema di vision tracking è una sfida che molti studiosi stanno affrontando negli ultimi anni. Bisogna tenere conto di vari fattori, come la presenza di uno sfondo complesso, mancanza di tessitura e parziali o totali occlusioni del soggetto analizzato (es. coperte) [37]. In letteratura esistono molti approcci che si prestano a ricavare a partire da sequenze video informazioni riguardo lo stato di un paziente nel sonno.

Nella prima parte del capitolo si presenta la relazione tra il sonno e il movimento e le posture che un determinato soggetto, affetto per esempio da malattie neurologiche, può assumere.

Nella seconda parte del capitolo si presenta lo stato dell'arte degli algoritmi di segmentazione, sia attraverso approcci basati sull'infrarosso sia tramite approcci basati su telecamere 3D. Si riportano le tecniche e gli approcci già presenti in letteratura relativamente al tracking degli oggetti, alla sottrazione dello sfondo, al rilevamento del movimento.

#### **3.1 Segmentazione: fase preliminare per la stadiazione automatica e per il riconoscimento di posture, movimenti normali e patologici**

La postura che il paziente assume durante il sonno è una delle indicazioni più importanti per determinare la qualità del sonno e può essere utilizzata in diverse diagnosi mediche [31]. Per esempio, per un soggetto allettato in seguito ad un intervento chirurgico, il rischio di ulcera è elevato in quanto il soggetto è immobile e non riesce a muoversi. Le posizioni assunte da una persona durante il sonno e la loro durata possono essere indice di varie condizioni cliniche. Ad esempio, l'apnea notturna si presenta principalmente quando si è sdraiati supini, la posizione prona può essere associata ad episodi di epilessia, così come dormire per lungo tempo nella stessa posizione può provocare l'insorgere di ulcere da pressione [27].

Il lavoro di segmentazione è preliminare per la stadiazione automatica. Tra queste rientrano le tecniche di sottrazione dello sfondo, le quali costituiscono un passo importante nelle tecniche di visione artificiale. La sottrazione dello sfondo è una fase di pre-elaborazione comunemente utilizzata nella visione artificiale e consiste nel confrontare una serie di fotogrammi per verificare la presenza di movimento [36].

Il primo passaggio consiste nell'inizializzazione dello sfondo, seguito dalla sottrazione dello stesso in modo da mettere in rilievo l'oggetto in primo piano. Lo schema da seguire per la sottrazione dello sfondo e presente in [38] è riportato di seguito.

1. *Inizializzazione dello sfondo*: si calcola l'immagine di riferimento dello sfondo.
2. *Modellizzazione dello sfondo*: consiste nel descrivere il modello da usare per rappresentare lo sfondo.
3. *Mantenimento dello sfondo*: il modello viene aggiornato continuamente in modo da adattarsi ai cambiamenti
4. *Classificazione dei pixel*: i pixel nell'immagine possono essere classificati in due categorie, appartenenti allo sfondo o appartenenti all'oggetto in movimento.

I metodi di sottrazione dello sfondo sono classificati in cinque categorie: modelli di base (BM), modelli statistici (SM), modelli fuzzy (FZ), modelli basati su reti neurali (NN) e altri modelli (OTHR) [36]. Data la complessità e vastità degli approcci, nel lavoro di tesi si prenderanno in considerazione solamente le prime due categorie.

I modelli di base possono essere costruiti seguendo due possibili strade: Frame Difference e Filtro Mediano. Il primo metodo funziona bene in presenza di sfondi statici e non contenenti oggetti in movimento, è la tecnica più efficiente. Il metodo del filtro mediano invece è meno efficiente rispetto al primo [36].

I modelli statistici sono in grado di rilevare lo sfondo tramite descrizioni statistiche dell'intensità dei pixel dell'immagine. Si tratta di metodi che uniscono diversi classificatori statistici in modo da consentire al sistema di migliorare la precisione della maschera di primo piano rilevata [33]. La descrizione statistica può ricorrere all'utilizzo di una singola gaussiana o di una mistura di varie gaussiane [39].

## 3.2 Stato dell'arte degli algoritmi di segmentazione della dinamica dei movimenti nel sonno

In questa sezione sono esaminati gli algoritmi utilizzati per la segmentazione della dinamica dei movimenti che si verificano durante il sonno. Sono descritti metodi di *motion detection* e di *blob/object tracking* a partire da sequenze video IR e Depth con tecniche elementari. Le tecniche sono analizzate con o senza algoritmi di apprendimento automatico, Machine Learning (ML) o attraverso Convolutional Neural Networks (CNN). Inoltre, il soggetto può essere coperto con delle lenzuola oppure no. Altri algoritmi, invece, prevedono l'individuazione dello skeleton e, quindi il riconoscimento delle varie posizioni che il soggetto assume nel letto. Questi algoritmi funzionano con soggetto senza coperte. Infine, vengono presentati gli approcci misti, ovvero approcci 2D-3D combinati con altri tipi di approcci, come quelli con camere termiche. Notiamo che i sistemi basati su video migliorano il comfort del paziente, in particolare per il monitoraggio di lungo periodo.

### 3.2.1 Approcci IR video-based (2D)

Motion detection e tracking costituiscono alcune tra le principali tecniche di visione artificiale. Utilizzando delle telecamere 2D si possono analizzare sequenze video sulle quali è possibile applicare le tecniche sopracitate per ricavare informazioni relative al movimento. Un algoritmo proposto in letteratura per rilevare il movimento è presentato in [28]. L'algoritmo in questione come prima cosa rileva una ROI, ovvero una regione di interesse contenente un certo numero di pixel. Nel momento in cui la telecamera individua il movimento, il sistema avvia un algoritmo di tracciamento che dopo aver determinato la ROI esamina i pixel adiacenti alla stessa. Il rilevamento del movimento è condotto con una tecnica su cui si basano molti studi, quella della differenza tra pixel corrispondenti tra due o più immagini (**Equazione 3.1**) [28]. Il movimento è rilevato quando la differenza assoluta media tra le intensità dei pixel corrispondenti tra due frames supera una soglia prefissata.

$$Idif(x, y) = |I_{current}(x, y) - I_{previous}(x, y)| \quad (3.1)$$

Nell'**Equazione 3.1**  $(x, y)$  rappresentano le coordinate  $x, y$  del pixel. Se il valore risultante dalla differenza dei due pixel corrispondenti nei due frame è diverso da zero vuol dire che per quel pixel è stato rilevato movimento. Il metodo cattura anche

ogni singolo pixel che ha un valore diverso da zero indesiderato causato dall'illuminazione improvvisa o dalla presenza di ombre; questo problema viene risolto con l'applicazione di appositi filtri. Il filtro morfologico di erosione (erode), per esempio, si occupa di rimuovere i pixel rumorosi in maniera casuale in quanto la maggior parte delle fotocamere producono un'immagine rumorosa. L'algoritmo di tracking ottiene in modo iterativo il valore medio dei pixel nella ROI.

Un gruppo di pixel adiacenti nell'immagine (blob) ed etichettati di movimento con la definizione precedente, può essere associato a una regione in movimento nella scena, presumibilmente corrispondente al movimento di una o più parti del corpo. Per ottenere una stima globale e più robusta del movimento di questi blob, si può utilizzare il *filtro di Kalman*. Il filtro è uno strumento matematico in grado di stimare in maniera ottimale il movimento quando siano noti o assunti alcuni parametri iniziali del movimento del blob (es. accelerazione, velocità e posizione) e della quantità di rumore presente nei dati. Il filtro di Kalman in questo caso stima le posizioni successive e il modo del blob in base a quelle precedenti [28] [46].

L'estrazione della parte in movimento da un flusso video e la stima del suo movimento (tracking) è un tema centrale analizzato ampiamente in molti studi di computer vision. Lo sfondo della scena (background) è generalmente immobile e, per rendere più semplice l'analisi, lo si separa dagli oggetti in movimento (background subtraction) [29]. La maggior parte dei metodi di tracking richiedono l'uso di informazioni temporali calcolate da una successione di frame in modo da evitare falsi rilevamenti. L'oggetto in movimento nell'immagine viene spesso rappresentato tramite un'immagine binaria dove i pixel assumono due valori, a seconda che siano quelli appartenenti all'oggetto in movimento o allo sfondo [29].

La segmentazione dell'oggetto dallo sfondo si complica nel momento in cui si presentano interferenze di illuminazione nell'ambiente. In [29] sono riportati vari metodi utili per il rilevamento di un oggetto. La metodologia più utilizzata è la sottrazione dello sfondo. Si tratta di un processo che trova la differenza tra l'immagine corrente e l'immagine di sfondo per rilevare la regione di movimento. L'immagine di sfondo viene prima inizializzata e successivamente viene continuamente aggiornata. Questa rappresenta la tecnica delle differenze assolute con sfondo dinamico. Anche in questo caso la valutazione verte sul valore

dell'intensità del pixel: se è superiore ad una determinata soglia allora questo è classificato come primo piano. Un ulteriore metodo presentato in [29] è l'algoritmo del modello di misture di gaussiane (GMM), esso si basa sul fatto che la distribuzione statistica della luminanza dello sfondo (background) e dell'oggetto in movimento (foreground) siano abbastanza diverse. Inoltre, si assume che la varianza di luminanza dello sfondo sia minima rispetto a quella dell'oggetto.

Tramite approcci basati su video 2D è possibile sviluppare algoritmi in grado di rilevare la postura che i soggetti assumono nel sonno. Questo studio è affrontato in [30], dove si propone un sistema di tracking 2D. L'apprendimento è di tipo supervisionato e può essere utilizzato anche in modalità off-line e quindi essere generalizzato e applicato su altri pazienti che non fanno parte del dataset su cui, invece, l'approccio è stato addestrato. Ogni fotogramma registrato viene inviato al modulo di riconoscimento della postura, in questo modo le varie pose sono identificate e facilmente classificate in base al modello.

Oltre alla differenza tra due frame per la determinazione del movimento di un oggetto, sono stati proposti altri metodi [32]. Un ulteriore approccio consiste nel confrontare il fotogramma corrente con il primo fotogramma della sequenza video, impostato come fotogramma di sfondo. Quest'ultimo è il metodo delle differenze assolute con sfondo statico. Un'altra tecnica ancora si basa su una particolare tecnica di rilevamento dei blob. L'immagine dei pixel viene trasformata in una raccolta di super pixel. Se i pixel rientrano in una particolare gamma di colori allora sono raggruppati in un insieme e costituiscono delle regioni [32]. In questo modo si creano aree di interesse molto specifiche, come per esempio l'area di un arto, della testa, ecc. Il rilevamento dei blob fornisce in output informazioni utili, come il numero di oggetti, posizione e dimensione rappresentati in un'immagine binaria.

Il contatore di *Blob* è una funzionalità molto pratica e può essere utilizzato in molte diverse applicazioni [32]. Riesce, infatti, a contare oggetti su un'immagine binaria, estrarli da essa ed isolarli. L'estrazione del blob è una procedura anch'essa interessante poiché una volta identificati i super-pixel, è possibile eseguire un semplice calcolo per vedere se il blob soddisfa determinate condizioni (in termini di altezza, larghezza o forma).

### 3.2.2 Approcci Depth-based (3D)

Recentemente è stato dimostrato che le telecamere di profondità riescono in modo più esaustivo a stimare le posizioni del corpo umano durante il sonno e la sua attività motoria, soprattutto per soggetti che dormono senza coperte [23]. Le telecamere di profondità forniscono un monitoraggio più completo rispetto alle telecamere 2D, in quanto sono invarianti all'illuminazione e riescono a localizzare gli oggetti in una scena tridimensionale. L'utilizzo di una telecamera 3D permette di ottenere uno spettro di movimenti più ampio, semplifica l'impostazione della registrazione e preserva il comportamento naturale del sonno [25]. Le telecamere 3D, inoltre, permettono di rilevare il movimento anche quando il soggetto è occluso dalle coperte. In quest'ultima condizione le telecamere a infrarosso rilevano una quantità minima di movimento, nelle immagini di profondità invece i cambiamenti sono facilmente osservabili [27].

#### 3.2.2.1 Motion detection e blob/object tracking

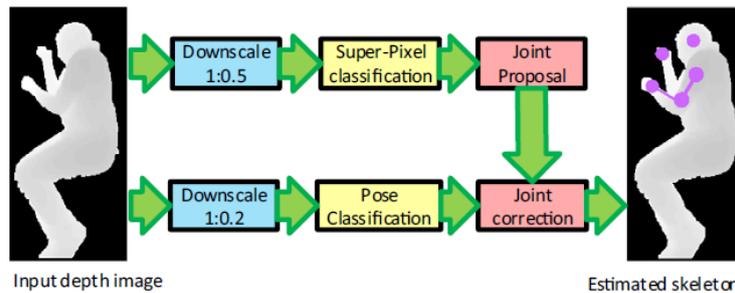
Le telecamere di profondità permettono di ricavare sequenze video da cui è possibile estrarre informazioni più dettagliate relativamente al movimento che il paziente compie nel sonno, anche quando i pazienti sono coperti da delle lenzuola [25]. In questo studio è stato utilizzato il dispositivo Microsoft Kinect v1 per la registrazione di video 3D.

In letteratura esistono molti software in grado di classificare le posizioni che un soggetto assume durante la notte. Tra questi studi si riporta l'analisi condotta in [27]. Il sistema presentato in questione utilizza il dispositivo Microsoft Kinect v2 e utilizza sia l'informazione infrarossa che di profondità. Il sistema presentato è in grado di effettuare il monitoraggio del movimento respiratorio nello stesso modo di un hardware commerciale. Le immagini di profondità rendono possibile la registrazione in un ambiente completamente buio e permettono di rilevare il movimento quando il soggetto è occluso dalle coperte. Il sistema è robusto ad occlusioni di coperte e a variazioni di illuminazione ambientale, e il tempo di analisi per una singola sessione notturna è inferiore a cinque minuti. Questo rappresenta un grande passo avanti rispetto alle tecniche convenzionali presentate nel **Capitolo 2**. Il software sviluppato in [27] (SoftwareX) può essere considerato

un punto di partenza per l'analisi del movimento nel sonno con camere di profondità. È caratterizzato da due parti, la parte di monitoraggio e la parte di analisi. La prima registra tramite la videocamera le posizioni del corpo, la seconda categorizza le immagini acquisite per determinare il tempo trascorso dal soggetto in ciascuna posizione. La componente di analisi, quindi, suddivide la notte in segmenti e ogni segmento ha la durata pari a quella che il soggetto ha impiegato in una determinata posizione. In output, quindi, si produce un report dal quale è possibile effettuare una valutazione sulla qualità del sonno.

### *3.2.2.2 Motion detection, object tracking e individuazione skeleton con tecniche di ML e CNN*

È possibile risalire direttamente ad un modello relativo alla posizione del corpo (skeleton) tramite metodi basati su CNN. Ad esempio, tramite le informazioni ricavate dal modello di scheletro e l'utilizzo di una telecamera di profondità, CNN-SkelPose [23] sviluppa un sistema di monitoraggio del sonno. CNN-SkelPose utilizza una rete neurale convoluzionale addestrata per estrarre le informazioni locali e globali dall'immagine di profondità. Un modello skeleton è utile per monitorare l'attività dei pazienti durante il sonno e le posizioni che questi assumono. Il modello è una rappresentazione compatta dei principali punti del corpo come testa, spalle, gomiti, mani, fianchi, ginocchia. Questi punti principali sono indicati come "giunti". Il modello analizzato nello studio in questione riguarda solo la parte superiore del corpo, la parte inferiore non è oggetto di analisi perché coperta dalle lenzuola. Il modello, inoltre, fornisce risultati affidabili in tutte le posizioni tipiche del soggetto. Anche in questo caso si utilizza la telecamera Kinect v2 per la registrazione dei video. CNN-SkelPose non richiede la segmentazione del soggetto dallo sfondo. Tutti i fotogrammi estrapolati dalla sequenza video acquisita vengono elaborati per segmentare il soggetto dallo sfondo. Questo viene fatto mediante il metodo di sottrazione dello sfondo che utilizza l'impostazione di una soglia per identificare i pixel appartenenti al soggetto in foreground. CNN-SkelPose valuta tutti i super-pixel candidati nell'immagine attraverso una finestra scorrevole. I componenti principali di CNN-SkelPose sono: classificazione dei super pixel, raggruppamento dei giunti identificati in seguito alla classificazione, rilevamento della posa e correzione dello scheletro in base alla posa identificata. Lo schema è rappresentato nella **Figura 3.1**.



**Figura 3.1** – Schema dei componenti principali di SkelPose. Immagine tratta da [23].

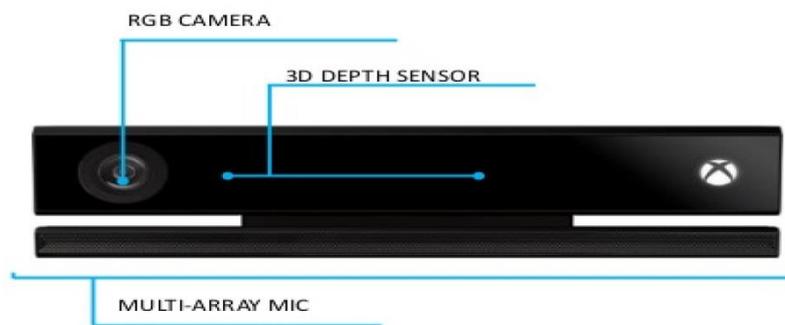
Tra il software disponibile per il tracking, LRSLibrary fornisce una raccolta di algoritmi in MATLAB™ [42]. Gli algoritmi presenti nella LRSLibrary prevedono tecniche di sottrazione dello sfondo e tecniche di segmentazione del movimento nei video. Questi metodi possono essere utilizzati anche per problemi di visione artificiale generici. La libreria contiene un totale di 103 algoritmi, e include moltissime funzioni di facile utilizzo, che permettono all'utente di effettuare l'elaborazione dei video.

Alcune tecniche di analisi video basate su Machine Learning sono in grado di riconoscere episodi di *apnea* (sospensione temporanea della respirazione) e di *ipopnea* (bassa frequenza respiratoria) [26]. Il monitoraggio di questi episodi può avvenire tramite telecamera 3D. Il sistema di monitoraggio presentato in questo lavoro consiste in una parte di compressione dei video di profondità e in una parte di analisi. Per l'acquisizione dei video si utilizza una camera Microsoft Kinect e per l'analisi si utilizza una rete SVM addestrata. Il sistema permette, inoltre, il riconoscimento di movimenti irregolari delle gambe, giramenti nel letto e altre tipologie di movimento.

Un metodo efficace per effettuare il tracking di oggetti in movimento è utilizzare le telecamere RGB-D [33]. Anche l'analisi condotta in questo lavoro si preoccupa di individuare lo sfondo confrontando il frame in questione con i frame precedenti che fungono come frame di riferimento. Gli algoritmi di sottrazione dello sfondo implementati si basano sull'ipotesi di uno sfondo statico. Combinando i dati di profondità con quelli a colore si riesce ad ottenere dei risultati accurati e consistenti. I metodi non parametrici modellizzano lo sfondo basandosi direttamente su dati osservati; lo svantaggio è che richiedono molto tempo computazionale e molta memoria.

### 3.2.2.3 Approcci misti

La telecamera Microsoft Kinect v2 è utilizzata per l'acquisizione di sequenze video in molti approcci video depth-based. In [24] la qualità del sonno del soggetto è determinata dal numero di movimenti effettuati durante il sonno. Tramite la distanza Euclidea si calcola la velocità dei nodi dello skeleton, e quindi delle articolazioni. Nello studio condotto la qualità del sonno è classificata in tre categorie: *buono*, *normale*, *pessimo*. Il sistema di monitoraggio proposto è misto: include una telecamera stereo che fornisce un'immagine di profondità 3D, una telecamera RGB e un microfono a schiera (**Figura 3.2**).



**Figura 3.2** – Microsoft Kinect v.2. Immagine tratta da [24].

I sensori di profondità delle camere RGB-D producono informazioni geometriche riguardanti l'area della scena osservata dal sensore [33]. Il valore di ogni pixel nel frame di profondità corrisponde alla distanza approssimativa da un determinato punto nel mondo reale al dispositivo. I dati di profondità, tuttavia, possono apparire abbastanza rumorosi e per questo motivo si è deciso di combinare i dati di profondità con quelli provenienti dalla camera a colori del dispositivo. Dalla combinazione delle informazioni di colore e di profondità dello sfondo, si è cercato di migliorare l'accuratezza dei risultati della sottrazione dello sfondo [33]

Questa tipologia di metodi si suddividono in due tipi. Nel primo tipo inizialmente si eseguono due segmentazioni indipendenti, una sull'immagine a colori e un'altra sui dati di profondità. Solo in un secondo momento, le due segmentazioni sono unite. Nel secondo tipo invece si esegue un'unica segmentazione sui dati RGB-D, ovvero sui dati di colore e profondità contemporaneamente [33].

Con i sensori Kinect v2 non solo si estraggono informazioni relative al movimento durante il sonno, ma si riesce ad ottenere un quadro completo del sonno generale [22].

## 4 Setup sperimentale e sviluppo di algoritmi specifici per la segmentazione 2D-3D dei movimenti

Questo capitolo si suddivide in due parti. Nella prima parte si descrive il sistema di acquisizione utilizzato per l'analisi del movimento durante il sonno. Il dispositivo utilizzato è la Kinect Azure [40] [44] [45] con la quale sono state registrate le sequenze video che riprendono il sonno di un soggetto. Le sequenze video sono state fornite dal CNR-IEIIT di Torino. Tramite l'SDK Azure Kinect [41] [45] è stato possibile ottenere gli streams video su cui effettuare l'analisi.

La seconda parte del capitolo consiste nella descrizione del setup sperimentale utilizzato per poter fare l'acquisizione e del software sviluppato per la segmentazione, l'analisi dei movimenti e per la valutazione della qualità del sonno. Si riportano i vari algoritmi e i parametri utilizzati e la loro applicazione alle sequenze video acquisite ed analizzate.

### 4.1 Descrizione HW del sistema di acquisizione utilizzato

Il sistema di acquisizione utilizzato è Azure Kinect, il successore di Kinect v1 e Kinect v2. I primi due dispositivi, dei quali si è ampiamente parlato nel **Capitolo 3**, sono stati ormai dismessi da Microsoft; esse sono state utilizzate per il rilevamento e il riconoscimento di oggetti, per il riconoscimento di gesti, per l'interazione uomo-macchina, ma gli algoritmi ivi sviluppati sono ancora utilizzabili nella nuova Kinect. Nel 2019, infatti, Microsoft ha rilasciato Azure Kinect, ovvero un kit per sviluppatori con sensori di intelligenza artificiale avanzati e in grado di creare modelli di visione artificiale e vocale. Azure Kinect contiene una camera RGB, una camera di profondità, un array di microfoni spaziali e un sensore di orientamento spaziale della camera. In **Figura 4.1** sono raffigurate le tre telecamere Kinect. Entrambe le versioni precedenti di Kinect sono costituite da una fotocamera a colori e da una fotocamera di profondità. In particolare, Kinect v1 sfrutta il principio di proiezione di pattern per misurare la profondità. Il pattern a infrarossi viene proiettato sulla scena e, in seguito, viene calcolata la profondità per triangolazione tra proiettore-sensore di immagine [40]. La telecamera Kinect v2 utilizza, invece, l'approccio di modulazione dell'intensità a onda continua (CW).



**Figura 4.1** – Da sinistra a destra: Kinect v1, Kinect v2 e Azure Kinect. Immagine tratta da [40]

Nella **Tabella 4.1** sono riportate le principali differenze tra Kinect v1, Kinect v2 e Azure Kinect.

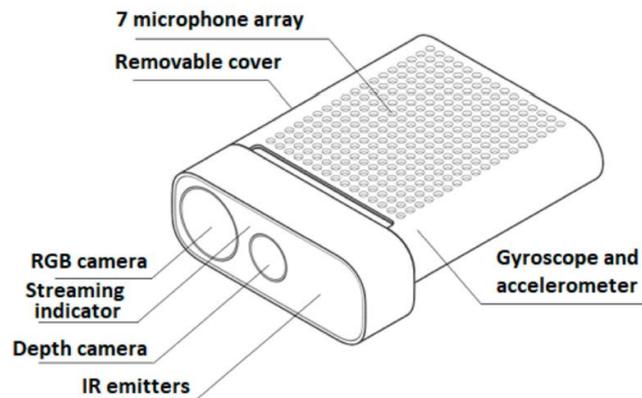
	Kinect v1 [17]	Kinect v2 [26]	Azure Kinect
Color camera resolution	1280 × 720 px @ 12 fps 640 × 480 px @ 30 fps	1920 × 1080 px @ 30 fps	3840 × 2160 px @ 30 fps
Depth camera resolution	320 × 240 px @ 30 fps	512 × 424 px @ 30 fps	NFOV unbinned—640 × 576 @ 30 fps NFOV binned—320 × 288 @ 30 fps WFOV unbinned—1024 × 1024 @ 15 fps WFOV binned—512 × 512 @ 30 fps
Depth sensing technology	Structured light—pattern projection	ToF (Time-of-Flight)	ToF (Time-of-Flight)
Field of view (depth image)	57° H, 43° V alt. 58.5° H, 46.6°	70° H, 60° V alt. 70.6° H, 60°	NFOV unbinned—75° × 65° NFOV binned—75° × 65° WFOV unbinned—120° × 120° WFOV binned—120° × 120°
Specified measuring distance	0.4–4 m	0.5–4.5 m	NFOV unbinned—0.5–3.86 m NFOV binned—0.5–5.46 m WFOV unbinned—0.25–2.21 m WFOV binned—0.25–2.88 m
Weight	430 g (without cables and power supply); 750 g (with cables and power supply)	610 g (without cables and power supply); 1390 g (with cables and power supply)	440 g (without cables); 520 g (with cables, power supply is not necessary)

**Tabella 4.1** – Confronto tra le tre versioni di Kinect. Tabella tratta da [40].

Azure Kinect, come Kinect v2, si basa su una telecamera a tempo di volo (ToF) a onda continua (CW). La telecamera a tempo di volo, infatti, permette di misurare in tempo reale la distanza che intercorre tra punti della scena e la telecamera, valutando il tempo necessario ad un impulso luminoso per percorrere questa distanza.

A differenza di Kinect v1 e Kinect v2, Azure Kinect supporta più modalità di rilevamento della profondità. La fotocamera a colori, inoltre, supporta una

risoluzione fino a 3840 x 2160 pixel [40]. Uno schema del dispositivo è rappresentato in **Figura 4.2**.



**Figura 4.2** – Schema dispositivo Azure Kinect. Immagine tratta da [40].

La telecamera di profondità di Azure Kinect supporta le modalità indicate in **Tabella 4.2**. Può funzionare in quattro modalità: *NFOV* (modalità di profondità del campo visivo stretto) *unbinned*, *WFOV* (modalità di profondità del campo visivo ampio) *unbinned*, *NFOV binned* e *WFOV binned* e *passive IR* [44].

Per le acquisizioni effettuate nella tesi la telecamera di profondità è stata impostata ad una modalità *NFOV unbinned*.

DEPTH CAMERA SUPPORTED OPERATING MODES					
MODE	Resolution	FoI	FPS	Operating range	Exposure time
<b>NFOV UNBINNED</b>	640x576	75°x65°	0, 5, 15, 30	0.5 - 3.86 m	12.8 ms
<b>NFOV 2X2 BINNED (SW)</b>	320x288	75°x65°	0, 5, 15, 30	0.5 - 5.46 m	12.8 ms
<b>WFOV 2X2 BINNED</b>	512x512	120°x120°	0, 5, 15, 30	0.25 - 2.88 m	12.8 ms
<b>WFOV UNBINNED</b>	1024x1024	120°x120°	0, 5, 15	0.25 - 2.21 m	20.3 ms
<b>PASSIVE IR</b>	1024x1024	N/A	0, 5, 15, 30	N/A	1.6 ms

**Tabella 4.2** – Modalità di supporto per la telecamera di profondità. Tabella tratta da [44].

La telecamera a colori, invece, supporta le modalità indicate in **Tabella 4.3**.

COLOR CAMERA SUPPORTED OPERATING MODES				
RGB CAMERA RESOLUTION (HXV)	Aspect Ratio	Format Options	Frame Rates (FPS)	Nominal FOV (HxV)(post-processed)
<b>3840X2160</b>	16:9	MJPEG	0, 5, 15, 30	90°x59°
<b>2560X1440</b>	16:9	MJPEG	0, 5, 15, 30	90°x59°
<b>1920X1080</b>	16:9	MJPEG	0, 5, 15, 30	90°x59°
<b>1280X720</b>	16:9	MJPEG/YUY2/NV12	0, 5, 15, 30	90°x59°
<b>4096X3072</b>	4:3	MJPEG	0, 5, 15	90°x74.3°
<b>2048X1536</b>	4:3	MJPEG	0, 5, 15, 30	90°x74.3°

**Tabella 4.3** – Modalità di supporto per la telecamera a colori. Tabella tratta da [44].

La fotocamera RGB è compatibile con la classe video USB 3 e può essere utilizzata senza Sensor SDK [44]. Nel caso della telecamera a colori la modalità utilizzata è quella che ha una risoluzione di  $1280 \times 720$  e che presenta tre diverse opzioni di formato immagine, ma il formato scelto è MJPEG.

Il sensore di profondità, inoltre, può presentare diverse modalità di temporizzazione, come indicato in **Tabella 4.4**. Per ottenerne le complete funzionalità, Kinect Azure richiede almeno un processore Intel® Core™ i5, scheda NVIDIA™ GEFORCE GTX 1070 o superiore, e una interfaccia USB3 dedicata.

DEPTH SENSOR RAW TIMING					
DEPTH MODE	IR Pulses	Pulse Width	Idle Periods	Idle Time	Exposure Time
NFOV UNBINNED	9	125 us	8	1450 us	12.8 ms
NFOV 2XX BINNED					
WFOV 2X2 BINNED					
WFOV UNBINNED	9	125 us	8	2390 us	20.3 ms

*Tabella 4.4 – Modalità di temporizzazione per la telecamera di profondità. Tabella tratta da [44]*

Rispetto alle sue versioni precedenti, il campo visivo di Azure Kinect è molto più ampio [40]. Azure Kinect comprende un array circolare di sette microfoni di alta qualità e un sensore di movimento (IMU) che include sia un giroscopio che un accelerometro.

Rispetto a Kinect v2, Azure Kinect ha un peso pari alla metà, quindi è più maneggevole. Non necessita di alimentazione e presenta una variabilità maggiore, in quanto può lavorare in quattro modalità differenti. È molto precisa e poco rumorosa e ha un'ottima risoluzione angolare.

## 4.2 Descrizione SW del sistema di acquisizione utilizzato

L'ambiente di sviluppo di Azure Kinect è caratterizzato dai seguenti SDK:

- *Sensor SDK* per accedere ai dispositivi e ai sensori di basso livello.
- *Body Tracking SDK* per tracciare i corpi e i loro movimenti in 3D
- *Speech Cognitive Services SDK* per permettere l'accesso al microfono e ai servizi vocali offerti da Azure

I due tipi di formati di immagine utilizzati sono riassunti in **Tabella 4.5**.

<b>K4A_IMAGE_FORMAT_DEPTH16</b>	Ogni pixel di dati DEPTH16 è costituito da due byte di dati di profondità senza segno. L'unità dei dati è espressa in millimetri dall'origine della telecamera.
<b>K4A_IMAGE_FORMAT_IR16</b>	Ciascun pixel di dati IR16 è costituito da due byte di dati di profondità senza segno. Il valore dei dati rappresenta la luminosità. Questo formato rappresenta la luce a infrarossi.

*Tabella 4.5 – Tipi di formato di immagini. Tabella tratta da [45].*

È possibile impostare diverse frequenze per l'acquisizione dei fotogrammi, in particolare 5 Hz, 15 Hz o 30 Hz [40]. Il frame-rate di Azure Kinect è abbastanza stabile, ma il problema è che non fornisce un valore di output stabile dopo l'accensione. È necessario che sia riscaldato per un po' di tempo, stimato a 60 minuti, in modo da ottenere un output stabilizzato. Vi è anche il problema dei pixel invalidati: negli angoli dell'immagine la luce IR del sensore viene spesso riflessa e alcuni pixel non risultano essere validi indicatori di profondità. Un altro svantaggio è la presenza dei pixel volanti: ai bordi degli oggetti i pixel possono contenere informazioni di distanza intermedie tra quella di primo piano e di sfondo [40].

A causa della tecnologia del tempo di volo, Azure Kinect non fornisce risultati affidabili se esposta alla luce diretta, quindi non è consigliata per applicazioni all'esterno. Da alcuni esperimenti eseguiti, tuttavia, in letteratura risulta che Azure Kinect presenta prestazioni migliori rispetto alle sue versioni precedenti in termini di precisione e ripetibilità [40]. Azure Kinect è un dispositivo dalle piccole dimensioni e versatile. È utilizzata per diversi usi come il riconoscimento o la ricostruzione di oggetti, per il tracciamento di oggetti e persone in movimento, per il rilevamento e l'identificazione di persone, per il riconoscimento dei gesti, per biometria e altro ancora.

### 4.3 Setup sperimentale

Le acquisizioni video sono state condotte in ambiente domestico. La telecamera è stata posizionata verticalmente in alto, a circa 180 cm sopra il letto e posta al centro del letto stesso. In **Figura 4.3** si rappresenta il setup completo all'interno della camera da letto. Il PC è posizionato sul ponte e collegato alla telecamera tramite la porta USB 3. La camera è alimentata con il suo alimentatore dalla rete. La telecamera riesce ad inquadrare completamente il letto, il quale rappresenta il focus

principale su cui si concentrerà l'analisi. Riesce comunque a riprendere anche l'ambiente circostante al letto in modo da avere una visione ampia della scena.

Azure Kinect permette di fornire in output contemporaneamente sequenze video sia nell'infrarosso che di profondità. L'approccio ottico utilizzato consente di eseguire l'acquisizione al buio e senza utilizzare ulteriori sensori, ovviamente senza contatto. La telecamera, quindi, non influisce in alcun modo sul sonno del soggetto. Si monitora l'attività del soggetto nel sonno senza che il soggetto sia disturbato.



**Figura 4.3** – Posizionamento della telecamera Kinect sopra il letto – (a) Vista destra – (b) Vista frontale – (c) Vista sinistra.

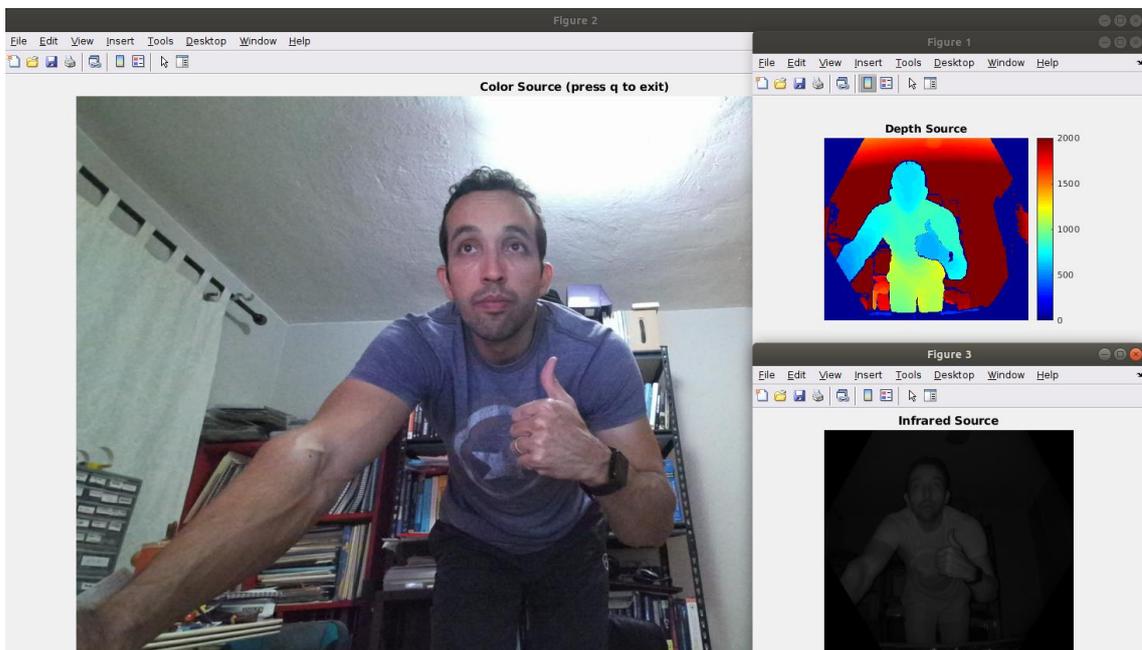
La telecamera deve interfacciarsi con l'SDK per avviare l'acquisizione e per estrarre i frame necessari alla fase di processing. Il software utilizzato è quello di Azure Kinect ed è modificabile nel linguaggio C++ [45].

In questa tesi è stata utilizzata una libreria in MATLAB™ che fa da interfaccia (wrapper) alle funzionalità dell'SDK Azure Kinect [41]. Questo software è stato modificato per consentire l'acquisizione delle sequenze video IR e Depth.

Successivamente all'installazione del software di interfaccia e alla modifica dei codici sorgente per il nostro caso di utilizzo, si installano le librerie grafico NVIDIA, si imposta in MATLAB™ il compilatore in C++ e si compilano i codici sorgenti in C, come indicato in [41]. Si riportano di seguito alcuni Demo presenti nel software descritto in [41]:

- *videoDemo.m*: visualizza i video Depth, a colori, IR e i dati del sensore
- *CalibrationDemo.m*: visualizza i parametri intrinseci della telecamera Depth e IR
- *bodyTrackingDemo.m*: visualizza la profondità, il colore e lo scheletro su entrambe le immagini.

In **Figura 4.4** si riporta l'interfaccia grafica ottenuta dal software. Si mostra sia l'immagine originale, sia la medesima in infrarosso e di profondità. L'immagine IR nella figura è nera perché i valori IR associati ai pixel del frame sono compresi tra 0 e  $2^{16}-1$  (16 bit), mentre i monitor standard hanno tipicamente una risoluzione di 8 bit.



**Figura 4.4** – Interfaccia grafica. Immagine tratta da [41].

Nel software originale è stato introdotto il salvataggio degli streaming video IR e Depth come sequenze di immagini codificate mPEG2000. In questo modo si è in grado di salvare i pixel delle immagini a uint16. Le sequenze Depth e IR sono

acquisite da Azure Kinect a 16 bit/pixel, ma solamente i primi 12 bit sono significativi e sono presi in considerazione per l'analisi.

Le sequenze video IR e Depth sono filtrate temporalmente con filtro gaussiano in modo da eliminare il rumore temporale sui frame oltre circa 2.5 Hz, frequenza massima entro cui considerare tutte le tipologie di movimento. Il filtro è mobile (a scorrimento) sulla sequenza. Si utilizza una gaussiana centrata sul frame  $n$  che si sposta da  $-3+n$  a  $3+n$ , fino ad arrivare a  $M-3$  con  $M$  che indica il numero dei frame presenti nella sequenza. Nel dettaglio, nella sequenza sono analizzati 7 frame contigui, pesati ognuno con i 7 valori della gaussiana. Successivamente all'elaborazione a 16 bit per la detection e il tracking, le sequenze video a 12 bit IR e Depth sono scalate nel range 0-255 ( $2^8-1$ ) in modo da renderli compatibili con le funzioni di visualizzazione utilizzate in MATLAB™.

#### **4.4 Descrizione del SW sviluppato per la segmentazione della dinamica dei movimenti nel sonno nell'IR e in Depth**

La stima e il monitoraggio del movimento rappresentano due punti chiave per l'analisi del sonno di un soggetto. Nella prima parte di questa sezione si descrivono le fasi preliminari all'analisi. Si spiega come i file video IR e Depth, ricavati con Azure Kinect, sono stati rielaborati in modo da rendere più semplice l'applicazione degli algoritmi di segmentazione.

Nel lavoro svolto, è stato possibile inoltre rilevare l'attività motoria durante il sonno tramite l'utilizzo di *Computer Vision Toolbox*<sup>TM</sup> offerto da MATLAB™. Il Toolbox in questione fornisce algoritmi utili per estrapolare delle informazioni a partire da video. Si tratta per esempio di algoritmi in grado di eseguire il tracking di un singolo oggetto o più oggetti in movimento, frame per frame.

Nel lavoro di tesi svolto sono stati esaminati due approcci specifici. Quello più tradizionale e ampiamente discusso in letteratura è l'approccio che utilizza le sequenze video nell'infrarosso. L'approccio più innovativo, invece, è quello che utilizza le sequenze video di profondità, che consente di ricavare informazioni tridimensionali sia riguardo al soggetto, sia riguardo alla scena analizzata.

Sono descritte le tecniche utilizzate per il rilevamento del primo piano e per il tracciamento dei blob, che si assume rappresentino le parti del soggetto in

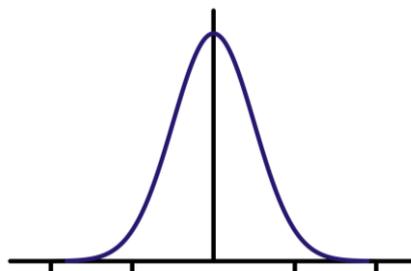
movimento. Si riportano anche le tecniche di *Foreground Detector* utilizzate per il rilevamento del primo piano. Sono stati implementati, inoltre, algoritmi base di valutazione della qualità del sonno. Lo scopo è rilevare quanto il soggetto si muove e quale parte del corpo muove di più, in modo da avere più informazioni per poter risalire al disturbo di cui potrebbe essere affetto. Quindi il software sviluppato, quindi, comprende algoritmi di rilevamento del movimento (motion detection), di inseguimento (tracking) e di valutazione della quantità di movimento.

Tutte le tecniche descritte sono state applicate sia per l'analisi nell'IR sia in Depth.

#### 4.4.1 Pre-elaborazione dei file video acquisiti con Azure Kinect

I file video IR e Depth di partenza sono quelli acquisiti direttamente dai video stream della Kinect Azure. Questi sono salvati con il formato *.mj2*. I file video sono codificati tramite MPEG 2000 a 320x288 pixel, ognuno a 16 bit. Le immagini in uscita in MATLAB™ avranno un formato di tipo *uint16* per IR e Depth e di tipo *uint8* per l'immagine a colore. Solamente i primi 12 bit sono significativi e contengono informazione utile per l'analisi [45]. Tutto ciò è possibile grazie all'utilizzo della classe *KinZ* e alle modifiche del software originale [41] che consentono di interfacciarsi con le funzionalità di Azure Kinect.

In seguito, si applica un filtraggio gaussiano passabasso ai video *.mj2* salvati. In questo modo si riduce il rumore e si filtrano i movimenti troppo veloci. Come accennato prima, il filtro gaussiano è un filtro uni-dimensionale che lavora come filtro temporale sulla sequenza di frame del video. La risposta all'impulso del filtro è una funzione gaussiana (in pratica approssimata), come mostrato in **Figura 4.5**.



**Figura 4.5** – Risposta all'impulso di un filtro gaussiano. Immagine tratta da [43].

Nell'**Equazione 4.1** si riporta la risposta all'impulso del filtro gaussiano descritto con variabile temporale  $x$ , o meglio  $x_i$ , per sequenze (discrete) di frame di  $M \times N$

pixel del video (dove  $x_i$  è di fatto l'immagine o frame  $i$ -esimo). Ciò in pratica equivale a realizzare  $M \times N$  filtri gaussiani temporali dove gli  $M \times N$  pixel corrispondenti nei frame  $i$  temporalmente successivi (variabile continua  $x$  o indice discreto  $x_i$ ) vengono filtrati ognuno indipendentemente dagli altri.

$$g(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{\frac{-x^2}{2 \cdot \sigma^2}} \quad (4.1)$$

Con  $\sigma$  si indica la deviazione standard. La deviazione standard nel dominio del tempo ( $\sigma_t$ ) è correlata con la deviazione standard nel dominio della frequenza ( $\sigma_f$ ) tramite la relazione dell'**Equazione 2**. [43]

$$\sigma_t \cdot \sigma_f = \frac{1}{2\pi} \quad (4.2)$$

La frequenza di taglio  $f_c$  del filtro (espresso in campione/s), la frequenza di campionamento del segnale FS (corrispondente a 30 sample/s per Kinect) e la larghezza della finestra temporale  $\sigma_t$  gaussiana di filtraggio passabasso sono correlate dalla relazione espressa nell'**Equazione 3**.

$$f_c = \frac{FS}{2\pi\sigma_t} \quad (4.3)$$

La finestra temporale  $\sigma_t$  è espressa in numero di campioni entro cui la gaussiana decade a 0,606 del suo valore a destra e a sinistra del massimo.

L'algoritmo di filtraggio gaussiano chiede in input i file da filtrare *.mj2*. Il valore di  $\sigma_t$  è di circa 4-9 campioni/s (9-19 frame di supporto per il filtro) per ottenere un filtraggio passabasso in frequenza a 0.2-0.5 Hz. In uscita, invece, si ottengono i file IR e Depth con la stessa estensione e formato *.mj2*, ma filtrati.

I file IR e Depth *.mj2* ottenuti sono convertiti in formato *.avi*, questo permette di gestire e visualizzare i file in modo più semplice. Le schede grafiche di solito non gestiscono i file *.mj2* con risoluzione di 16 bit per pixel; se si prova a visualizzarli questi appaiono neri. Di contro, i file dal formato *.avi* hanno una risoluzione di 8 bit per pixel, quindi non di 12 bit per pixel che sono quelli effettivamente codificati dal dispositivo e, quindi utili, e nemmeno di 16 bit per pixel che sono quelli del formato *.mj2*, comunque necessari per non perdere la risoluzione di 12 bit per pixel in fase di archiviazione delle sequenze originali ed elaborate.

#### 4.4.2 Algoritmi di Motion-Body Tracking e Blob Analysis

Per prima cosa, vengono creati gli oggetti Matlab (analoghi alle classi C++) che permettono la riproduzione dei frame del video. I file video IR e Depth con cui si lavora sono codificati in *.avi*, a soli 8 bit per pixel. L'algoritmo lascia all'utente la possibilità di salvare il file originale e/o il file contenente solamente il primo piano, entrambi con i blob sovrapposti.

Lo step successivo è stato quello di impostare lo strumento necessario per creare dei blob, ovvero regioni delimitate e raggruppanti pixel vicini. Questo è stato possibile grazie all'utilizzo dell'oggetto di `vision.BlobAnalysis(Name, Value)` MATLAB™. Questo strumento ha permesso di creare degli oggetti tramite i quali è possibile calcolare delle misure statistiche per le regioni connesse delle immagini (frame) IR e depth. Lo strumento richiede di trovare le combinazioni ottimali per i vari parametri (proprietà) dei metodi richiamati dall'oggetto.

I parametri utilizzati e i rispettivi valori sono analizzati di seguito.

- **BoundingBoxOutputPort:** può essere impostato come 'True' o 'False' e restituisce le coordinate dei riquadri di delimitazione. In questo caso è stato impostato su True.
- **AreaOutputPort:** può essere impostato come 'True' o 'False' e restituisce l'area dei riquadri di delimitazione. In questo caso è stato impostato su False, in quanto non è un'informazione rilevante.
- **CentroidOutputPort:** può essere impostato come 'True' o 'False' e restituisce le coordinate dei centroidi dei riquadri di delimitazione. Anche questo parametro è stato settato su False perché non rilevante per l'analisi.
- **MinimumBlobArea:** è un numero positivo intero scalare e corrisponde alla dimensione minima dei blob in pixels. In questo caso potrebbe rappresentare la dimensione minima di un arto, come per esempio il piede, in pixels. È stato impostato a 280.
- **MaximumCount:** è un numero positivo intero scalare e corrisponde al numero massimo di regioni presenti all'interno dell'immagine di input. In

questo caso rappresenta il numero  $n$  di parti in moto e indipendenti. Questo parametro è stato impostato al valore di 12 perché si considerano i quattro arti, la testa ed eventuali movimenti del busto.

Nella **Tabella 4.6** si riassumono i valori scelti.

'NAME'	VALUE
'BOUNDINGBOXOUTPUTPORT'	True
'AREAOUTPUTPORT'	False
'CENTROIDOUTPUTPORT'	False
'MINIMUMBLOBAREA'	280
'MAXIMUMCOUNT'	12

**Tabella 4.6** – Coppia Parametri-Valori utilizzati per vision.BlobAnalysis

Ogni blob rilevato, quindi, corrisponde ad una parte del corpo in movimento.

#### 4.4.3 Tecniche per il rilevamento del primo piano

In questa sezione si descrivono le tecniche applicate per il rilevamento del primo piano. Si tratta del *Gaussian Mixture modeling* e del *metodo delle differenze assolute*. Si riporta una descrizione di entrambe le tecniche, sottolineando come con il *metodo delle differenze assolute* sono stati ottenuti risultati migliori.

##### 4.4.3.1 Gaussian Mixture modeling

Una prima prova è stata eseguita applicando *Gaussian Mixture modeling*. Questo modello permette di confrontare un frame video a colori o in scala di grigi con un modello di sfondo per determinare se i pixel in questione fanno parte dello sfondo oppure del primo piano. Lo strumento utilizzato è impostato come di seguito: `vision.ForegroundDetector(Name,Value)`. Anche in questo caso è possibile, infatti, impostare una o più coppie di Nome-Valori. Il problema è stato trovare la combinazione ottimale dei vari valori per ottenere buoni risultati. In letteratura esistono pochi studi relativamente al `ForegroundDetector`, per cui trovare i parametri ottimali non è una procedura semplice [39].

I metodi di `ForegroundDetection` si basano sulla segmentazione del primo piano effettuando un'analisi pixel per pixel e non per intere regioni. La segmentazione degli oggetti in movimento si effettua calcolando per ogni pixel la probabilità di appartenenza a un pixel di foreground o background sulla base della sua intensità e sulla distribuzione statistica delle luminanze prevista sia per il background, o se il

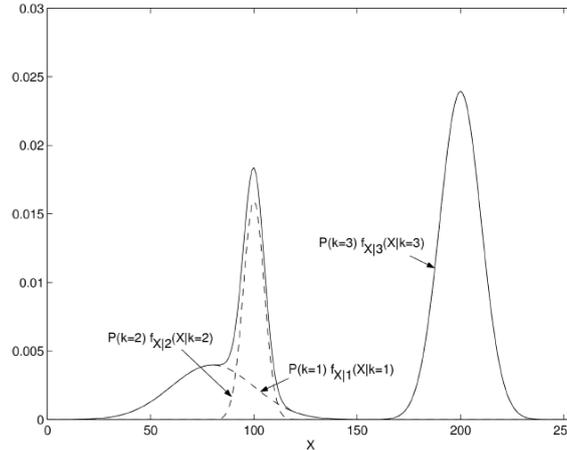
soggetto è nel letto, che per il foreground, o immagine di sfondo. Questi metodi, però, non funzionano bene se l'ambiente in cui si deve rilevare il movimento non è controllato come nel caso di un ambiente aperto. I cambiamenti di illuminazione e i movimenti dello sfondo possono essere fonte di disturbo [39]. In questa tesi, però, si analizza il movimento di soggetti durante il sonno in un ambiente interno, senza apprezzabili variazioni di luce IR.

In [39] si propone l'algoritmo di Stauffer-Grimson che consiste in un metodo adattivo in grado di modellare e segmentare una sequenza di immagini sfondo (background) e oggetto (foreground) tramite distribuzioni normali. L'algoritmo parte dal presupposto che lo sfondo sia più visibile rispetto al primo piano e che le mode delle gaussiane abbiano una varianza molto bassa. Si assume che la scena sia caratterizzata da  $K$  superfici, abbastanza uniformi in luminanza, costituenti la scena e l'oggetto. Ogni pixel dell'immagine è caratterizzato da uno stato  $k$  (appartiene ad una superficie, background o foreground) tra un insieme di stati  $k \in \{1, 2, \dots, K\}$ , dove  $K$  è il numero di superfici nella scena. Essendo la scena dinamica, lo stato  $k$  del pixel nell'immagine può variare dinamicamente. Alcuni stati  $k$  appartengono allo sfondo, altri al primo piano. Di solito, il parametro  $K$  è modellato con un valore costante compreso tra 3 e 7. L'obiettivo dell'algoritmo è modellare al meglio la mistura di gaussiane, delle quali alcune rappresentano la distribuzione di intensità del background, ed altre quella del foreground o oggetto in movimento. Il processo stocastico  $\mathbf{k}$  che genera lo stato  $k_i$  al tempo  $i$  per un dato pixel  $p(x,y)_i$  del frame  $F_{t=i}$ , cioè la probabilità che il pixel appartenga alla superficie  $k$ , è quindi modellato da un insieme di parametri di una mistura di gaussiane.

Il processo  $\mathbf{k}$  che genera  $k$  è nascosto (hidden) e si esplica tramite il valore del pixel  $X_i$  al tempo  $i$  tramite una certa distribuzione  $f(X|k)$  di cui è noto solo il valore di luminanza a posteriori  $X_i$  del pixel, e la forma della distribuzione, come mistura gaussiana. Nel dettaglio, il modello di Gaussians Mixture con  $K$  gaussiane è caratterizzato dai valori di media e varianza delle gaussiane (parametri  $\theta_k$ ) nella **Equazione 4.1**, una per ogni stato  $k$ . Considerando gli eventi  $k$  come disgiunti l'uno dall'altro, la distribuzione di luminanza  $X$  del generico pixel può essere modellata come la somma di  $K$  gaussiane  $f_{X|k}(X|k, \theta_k)$ , ognuna con probabilità a priori  $P(k)$  del pixel di appartenere alla superficie  $k$ , come descritto nell'**Equazione 4.1**.

$$f_x(X|\phi) = \sum_{k=1}^K P(k) f_{X|k}(X|k, \theta_k) \quad (4.1)$$

Una possibile distribuzione  $f_x(X|\phi)$  di probabilità di luminanza del pixel è rappresentata nella **Figura 4.6**, dove l'asse delle ascisse rappresenta i valori di luminanza dei pixel, tra 0 e 255.



**Figura 4.6** – Distribuzione di probabilità del pixel con  $K = 3$ . Immagine tratta da [39].

Come è stato già detto, il numero pratico di superfici, e di conseguenza di gaussiane, varia tra 3 e 7. Un totale di  $K = 3$  gaussiane può essere considerato sufficiente per modellare due superfici di sfondo e una di primo piano. Con meno di due mode di sfondo sarebbe inutile applicare l'algoritmo e basterebbe utilizzare la sottrazione di un'immagine di sfondo mediata. Inoltre, è stato dimostrato che oltre a  $K = 5$  distribuzioni non si hanno miglioramenti significativi. Una volta stimato lo stato corrente  $k$ , è necessario determinare se rappresenta una superficie di primo piano o di sfondo.

La procedura di Stauffer-Grimson è un algoritmo che cerca, a partire da un insieme di osservazioni sulla luminanza del pixel  $X$ , di stabilire i parametri della distribuzione di gaussiane dell'**Equazione 4.1**, e tramite il Teorema di Bayes e la stima di massima verosimiglianza, la probabilità  $f(k_i|X_i)$  che il pixel appartenga ad uno dei valori  $k$  definiti di foreground o di background, scegliendo quel  $k$  che massimizza tale probabilità [39].

Nel dettaglio, la procedura di Stauffer-Grimson parte dalla classificazione dei  $K$  stati in base al rapporto  $\omega_k/\sigma_k$  (con  $P(k) = \omega(k)$  nell'**Equazione 4.1**), che risulta proporzionale all'ampiezza del picco della gaussiana  $k$ ,  $\omega_k f_{X|k}(X|k, \theta_k)$ . Una superficie ha la probabilità maggiore di essere uno sfondo se  $\omega_k$  è alto e  $\sigma_k$  è basso. Per individuare lo sfondo si fornisce una probabilità globale  $T$ , che indica la

probabilità a priori che nella scena un pixel appartenga allo sfondo. Potrebbe essere stimata come il rapporto tra area del soggetto in movimento rispetto all'area totale del frame. L'algoritmo prevede che nell'Equazione 4.2 il B più piccolo per cui la probabilità accumulata sia maggiore di T è considerato sfondo e il resto, invece, viene considerato di default primo piano (Equazione 4.2) [39].

$$B = \underset{b}{\operatorname{argmin}} (\sum_{k=1}^b \omega_k > T) \quad (4.2)$$

Considerando l'Equazione 4.1, e il teorema di Bayes, i parametri possono essere riscritti per  $k = 1, 2, \dots, K$  in modo da ottenere l'Equazione 4.3.

$$\omega_k = \frac{1}{N} \sum_{t=1}^N P(k|X_t, \phi) \quad (4.3)$$

Un modo semplice per convertire Equazione 4.3 in una media cumulativa che rende adattivo l'algoritmo ai cambiamenti limitati delle distribuzioni, causati da stime iniziali troppo approssimate o variazioni di luminanza dinamiche della scena (che non sono presenti però nel nostro caso) è definire un guadagno variabile nel tempo  $\alpha = 1/t$  e scrivere per  $k = 1, 2, \dots, K$  e  $t = 1, 2, \dots$  Equazione 4.4.

$$\omega_{k,t} = (1 - \alpha_t)\omega_{k,t} + \alpha_t P(k|X_t, \phi) \quad (4.4)$$

In questo modo ogni parametro dipende dal parametro di guadagno  $\alpha$ . Un ultimo parametro preso in considerazione in [39] è la *standard deviation*  $\lambda$ , varianza iniziale a priori per le gaussiane. Nella Tabella 4.7 si riportano i valori dei parametri del modello in una tipica applicazione in un ambiente chiuso. [39]

Symbol	Value	Assumption
$K$	3	
$\lambda$	2.5	
$\alpha$	0.005	5 frames/second
$T$	0.7	$K = 3$
$\omega_{\text{init}}$	0.05	$K = 3$
$\sigma_{\text{init}}$	30	$X \in [0, 255]$

Tabella 4.7 – Valori dei parametri ottimali

Gli algoritmi di Gaussians Mixture per la segmentazione di oggetti in movimento generalmente non sono utilizzabili nella loro forma pura. I loro parametri devono essere inizializzati all'inizio e re-inizializzati quando l'illuminazione cambia drasticamente.

Come già anticipato, lo strumento utilizzato in MATLAB™ che permette l'applicazione del modello analizzato precedentemente è `vision.ForegroundDetector(Name,Value)`. Di seguito si elencano le coppie di parametri-valori con cui sono state effettuate le prove.

- **NumTrainingFrames:** è il numero iniziale dei frame necessario per addestrare il modello. In questo caso è stato stabilito a 250.
- **InitialVariance:** è un parametro che dipende dall'intervallo dei valori dei pixel nel video. Nel modello presentato in 4.2.1.1 corrisponde a  $\sigma$ . Se il tipo dei dati dell'immagine è "double" o "single" con valori dei pixel compresi tra 0 e 1, allora il valore di InitialVariance si imposta a  $(\frac{30}{255})^2$ . Invece se i dati dell'immagine sono di tipo "uint8" con valori compresi tra 0 e 255, allora il valore si imposta a  $30^2$ . I frame con cui si lavora in questa analisi rientrano nella prima casistica, quindi il parametro di InitialVariance è stato impostato a  $(\frac{30}{255})^2$ .
- **NumGaussians:** è il numero di mode gaussiane presenti nel modello. È un numero intero, tipicamente compreso tra 3 e 5. Corrisponde al numero K di superfici. Questo parametro si imposta ad un valore più basso se lo sfondo è statico e ad un valore più alto se non lo è, ovvero se sono rappresentate delle scene all'aperto. In questo caso è stato scelto un NumGaussians pari a 3 perché l'analisi avviene in un ambiente interno e chiuso in cui si assume che lo sfondo sia statico.
- **AdaptLearningRate:** parametro impostato come "True" in modo da permettere all'algoritmo di adattare nel tempo la stima della gaussian mixture (learning rate).
- **LearningRate:** è il parametro che controlla la velocità con cui il modello adatta nel tempo la stima della gaussian mixture a condizioni di luminanza mutevoli. È un numero scalare e corrisponde al parametro  $\alpha$ . Se si imposta questo parametro ad un valore più elevato allora gli oggetti che si muovono lentamente potrebbero diventare parte dello sfondo. In caso contrario il

modello di sfondo potrebbe non adattarsi ai cambiamenti di illuminazione. Il parametro di LearningRate è stato impostato a 0.07.

- **MinimumBackgroundRatio:** è la soglia utile per determinare il modello di sfondo. Si tratta di un numero scalare e nel modello presentato in 4.2.1.1 corrisponde a T. Rappresenta il limite minimo per i pixel di essere considerati valori appartenenti allo sfondo. Gli sfondi multimodali non possono essere gestiti se questo valore è troppo piccolo. Il parametro di MinimumBackgroundRatio è stato impostato a 0.9.

Nella **Tabella 4.8** si riassumono i valori scelti.

'NAME'	VALUE
'NUMTRAININGFRAMES'	250
'INITIALVARIANCE'	$(\frac{30}{255})^2$
'NUMGAUSSIANS'	3
'ADAPTLEARNINGRATE'	True
'LEARNINGRATE'	0.07
'MINIMUMBACKGROUNDDRATIO'	0.9

**Tabella 4.8** – Coppia Parametri-Valori utilizzati per vision.ForegroundDetector

La maschera binaria associata ad ogni frame, costituita da valori 1 per i pixel stimati come foreground e da valori 0 per quelli di background necessita di una post-elaborazione che riempia buchi e spazi vuoti nella regione di foreground dovuti a rumore ed errori di stima. Questo è necessario per irrobustire l'analisi successiva. Per fare ciò sono state utilizzate in MATLAB™ le funzioni morfologiche standard. La prima operazione è stata la chiusura tramite il comando *imclose*, grazie a questa operazione, infatti, è stato possibile riempire gli spazi vuoti e quindi chiuderli. Per la chiusura l'elemento strutturale utilizzato è stato un quadrato dalla dimensione di dodici pixel. Successivamente è stata applicata l'operazione di erosione tramite il comando *imerode* in modo da ridefinire e snellire la maschera creata con l'operazione di chiusura. Ciò è stato possibile utilizzato come elemento strutturale un quadrato, ma di dimensione più piccola (tre pixel).

#### 4.4.3.2 Metodo delle differenze assolute con sfondo dinamico

Si tratta di un metodo molto semplice e veloce. I file video IR e Depth di input codificati possono avere una risoluzione di 8 bit (video *.avi*) o di 16 bit (*.mj2*) per

pixel, a seconda del compromesso tra accuratezza e velocità di elaborazione. Il metodo consiste nell'effettuare una differenza in valore assoluto tra il frame corrente (*currentFrame*) e il frame precedente (*previousFrame*). (**Equazione 4.3**).

$$\text{diff} = |\text{previousframe} - \text{currentFrame}| \quad (4.3)$$

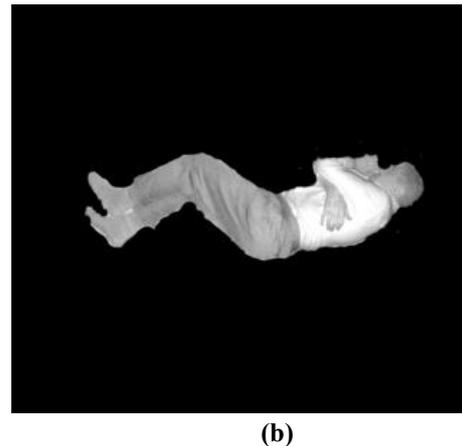
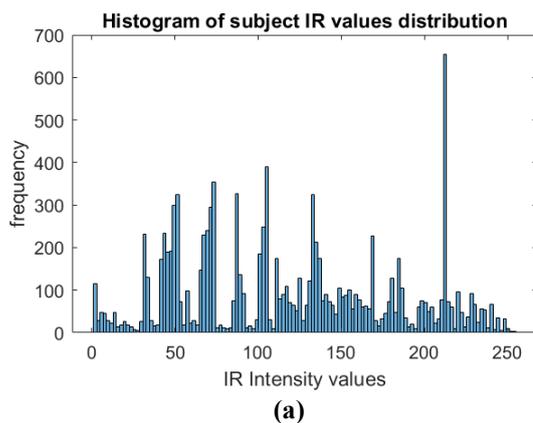
Si utilizza una codifica binaria, 0 e 1. Se la differenza ottenuta è minore di una determinata soglia  $Th$ , impostata sperimentalmente a 0.03, allora la codifica sarà impostata a 0 e, di conseguenza, l'immagine corrispondente alla differenza sarà nera e quindi rappresenterà lo sfondo. Altrimenti la codifica sarà impostata su 1 e l'immagine sarà bianca e, quindi, rappresenterà il primo piano. La soglia  $Th$  è stata determinata sperimentalmente come descritto nella **sezione 5.1.3** analizzando il rumore presente nelle immagini sia IR che di Depth.

Questo descritto è un metodo dinamico, a causa del continuo aggiornamento dell'immagine di sfondo. Anche in questo caso, come il metodo precedente, è stato necessario fare ricorso agli operatori morfologici in modo da ottenere una maschera migliore. È stato sufficiente applicare solo l'operazione di chiusura utilizzando come elemento strutturale un quadrato dalla dimensione di nove pixel. L'algoritmo consente di visualizzare l'input e l'uscita della blob detection ed eventualmente di salvare entrambi i video in un unico file di uscita. Inoltre, si ha anche la possibilità di salvare unicamente il file video relativo ai blob. L'algoritmo prevede che nel file di output queste immagini vengano rappresentate in un unico frame in cui si riporta, in basso, anche la quantità del movimento e un indicatore relativo all'occupazione del letto, Questi aspetti saranno meglio analizzato nel **Capitolo 5**.

Per quanto riguarda i video di profondità, l'approccio delle differenze assolute funziona meglio se i video sono codificati con una risoluzione di 12 bit per pixel, quindi si è preferito lavorare direttamente con i file dal formato *.mj2*. Il procedimento adottato è uguale a quello di prima, tranne per il fatto che il metodo è statico. Lo sfondo, infatti, non si aggiorna continuamente ma resta sempre fisso, ovvero corrispondente ad una media dei frame iniziali, quando il letto è vuoto. A differenza di prima, la soglia impostata per individuare e isolare il primo piano dallo sfondo è stata impostata a 40 mm (vedi **sezione 5.1.3**).

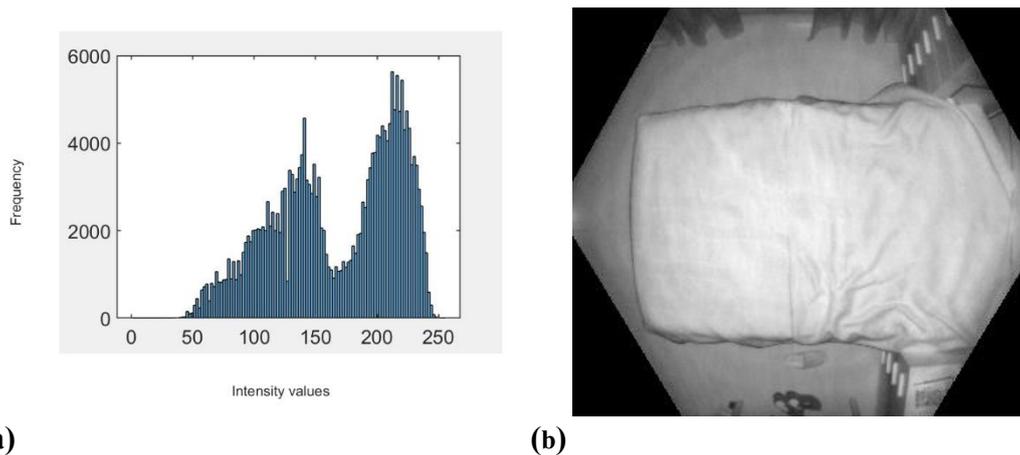
#### 4.4.3.3 Tecnica di Background Subtraction scelta e commenti

Con il metodo delle misture gaussiane, tuttavia, le performance raggiunte non sono buone come ci si aspettava. Tramite lo strumento MATLAB™ Histogram si sono ricavati due istogrammi, uno relativo al corpo del soggetto e quindi al primo piano e l'altro relativo allo sfondo, ovvero al letto e all'ambiente circostante. In **Figura 4.7** si riporta la distribuzione dei valori dei pixel di un frame estrapolato dal video della maschera. Si ricostruisce l'istogramma relativo al corpo del soggetto rilevato in primo piano. Si osserva che non si tratta di vere e proprie misture gaussiane. Dall'istogramma di **Figura 4.7** non è possibile individuare delle distribuzioni gaussiane; i pixel, infatti, hanno valori random. Il picco più alto è l'unico tra quelli adiacenti, per cui non si tratta sicuramente di una funzione gaussiana. I valori dei pixel sono molto diversi tra loro, la variabilità è molto alta e la distribuzione di background (**Figura 4.7**) si sovrappone in parte a quella di foreground (**Figura 4.8**).



**Figura 4.7** – (a) Iistogramma sulla distribuzione di probabilità del pixel sul Foreground.  
(b) Maschera di Foreground

Per quanto riguarda lo sfondo, invece, l'istogramma ricavato presenta due vere e proprie gaussiane in cui i due picchi sono ben visibili. I pixel non hanno valori random e i picchi non si discostano molto da quelli adiacenti. A differenza di prima la variabilità è molto ridotta. (**Figura 4.8**)



**Figura 4.8** – (a) Istogramma sulla distribuzione di probabilità del pixel sul Background.  
 (b) Immagine di background originale: letto e ambiente.

I due istogrammi presentano un andamento e dei valori di pixel molto diversi tra loro. Nel primo istogramma relativo al primo piano sono violate le condizioni sui cui si basano le misture gaussiane, non si tratta quindi di distribuzioni gaussiane. Nel secondo caso, relativo allo sfondo, i principi di misture gaussiane sono rispettate. Dato il comportamento differente per sfondo e primo piano, non è conveniente utilizzare il metodo delle misture gaussiane.

È preferibile utilizzare il metodo *delle differenze assolute* in quanto bisogna solamente effettuare una differenza e scegliere una soglia, a differenza del metodo delle misture gaussiane in cui scegliere le varie combinazioni parametri-valori non è semplice.

Ricordando che le sequenze video sono filtrate temporalmente con filtro gaussiano, la tecnica di background subtraction scelta richiede solo l'utilizzo di un operatore morfologico. L'unica operazione è la chiusura, a differenza del metodo delle misture gaussiane che richiedeva una doppia operazione (chiusura seguita da erosione). Questo perché le differenze tra due frame consecutivi dovute al movimento del soggetto, sono generalmente nettamente sopra il rumore delle immagini.

#### 4.4.4 Algoritmo di tracking con filtro di Kalman

È stato implementato, inoltre, un algoritmo di tracking del movimento in ambiente MATLAB™. Si tratta del filtro di Kalman, il quale permette di stimare il

movimento degli oggetti rilevati come in movimento (ad esempio gambe), di predirne la posizione dell'oggetto e la loro traccia temporale nella sequenza.

Per prima cosa è stato inizializzato il filtro di Kalman. Lo strumento utilizzato è stato:

```
configureKalmanFilter('MotionModel', 'InitialLocation', 'InitialEstimateError', 'MotionNoise', 'MeasurementNoise');
```

I parametri sono impostati come indicato di seguito.

- **MotionModel:** può essere impostato su due modelli differenti, 'ConstantVelocity' o 'ConstantAcceleration'. Nel modello 'ConstantVelocity' la velocità è mantenuta costante e si utilizza quando in una sequenza video tutte le parti si muovono alla stessa velocità. Nel modello 'ConstantAcceleration' è l'accelerazione ad essere costante. Si utilizza quando in una sequenza video varie parti si muovono ad una frequenza diversa.

Nelle sequenze video acquisite in questo lavoro i movimenti sono diversi e a diversa velocità, per cui il modello scelto per inizializzare il filtro di Kalman è stato quello di 'ConstantAcceleration'.

- **InitialLocation:** è un vettore numerico che restituisce due coordinate. In questo caso restituisce le coordinate del blob rilevato tramite lo strumento `vision.BlobAnalysis`.
- **InitialEstimateError:** può essere un vettore a due elementi (nel caso di 'ConstantVelocity') o a tre elementi (nel caso di 'ConstantAcceleration'). Valuta la stima dell'incertezza iniziale.
- **MotionNoise:** esprime quanto può deviare il modello in termini di tolleranza. Anche in questo caso può essere un vettore a due o a tre elementi in base al modello scelto.
- **MeasurementNoise:** è un numero scalare e indica quanto può essere inesatta la varianza di una determinata posizione. Un valore più alto di questo parametro comporta una diminuzione del rumore, ma si rischia di

adattarsi troppo al modello e di ottenere poche rilevazioni. In questo caso è stato impostato sperimentalmente a 25.

Se l'algoritmo rileva il movimento, allora consentirà anche di stimarne la traiettoria nel tempo.

Tuttavia, non sono emersi risultati soddisfacenti da questa analisi, in quanto non si riesce a ricavare un'informazione utile per capire il tipo di movimento svolto dal paziente. Per questo motivo, si è deciso di non riportare i risultati in quanto uno sviluppo futuro potrebbe essere la rielaborazione del filtro di Kalman in modo da ottenere un'analisi più veritiera circa la traiettoria del movimento del soggetto e quindi trovare una correlazione tra il tipo di movimento e un eventuale disturbo del sonno.

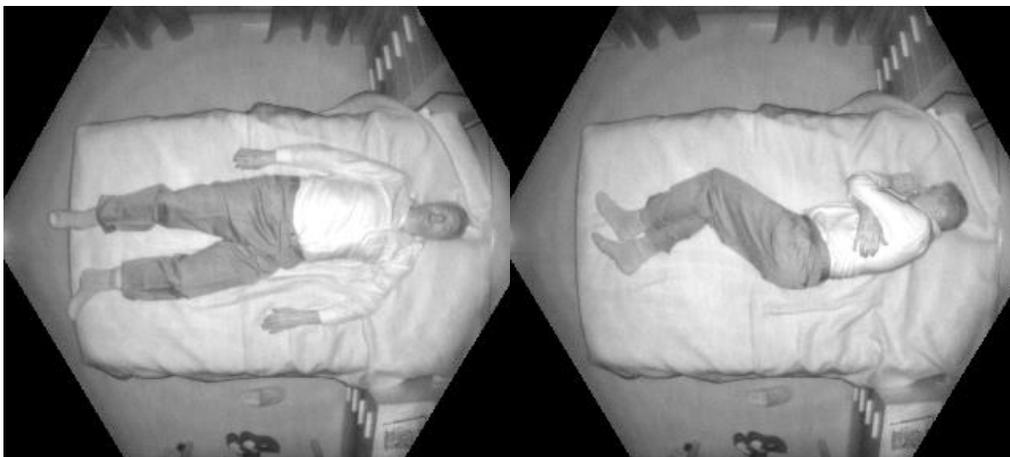
## 5 Risultati sperimentali

In questo capitolo verranno analizzati i risultati ottenuti in seguito all'applicazione degli algoritmi presentati nel **Capitolo 4**. Dalle informazioni ricavate si riesce a identificare quali parti del corpo del soggetto sono in movimento. La segmentazione di quest'ultime è messa in rilievo tramite lo strumento dei blob. Si ottiene in output anche un grafico che indica se il soggetto si trova nel letto oppure se questo è vuoto (bed occupancy) e un grafico che indica la quantità di movimento frame per frame. Si è effettuata anche un'analisi del movimento su più scale temporali per identificare le zone in cui si concentra maggiormente l'azione.

Si riportano i commenti relativi ai risultati, le immagini estrapolate dal video di output e i grafici ricavati con MATLAB™.

### 5.1 Risultati Motion-Body Tracking nell'infrarosso

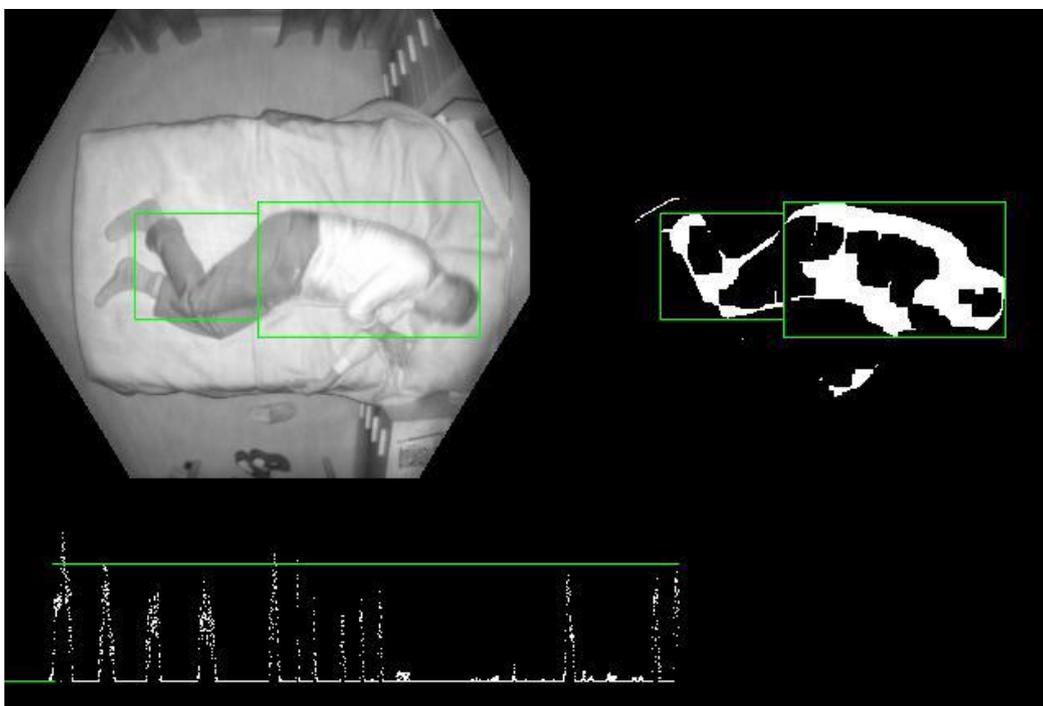
Le sequenze video di input acquisite con Azure Kinect mostrano un soggetto sdraiato nel letto che, durante la sessione analizzata, cambia pose e compie diversi movimenti. Nel video di input si osserva che durante la sessione di monitoraggio il soggetto può trovarsi in posizione prona, supina, sdraiato sul lato destro o sul lato sinistro. In **Figura 5.1** si riportano alcuni frame estratti dal video di input nell'infrarosso in cui si può osservare il cambiamento di posizione del soggetto. I cambiamenti di posizione sono molto frequenti durante tutta la sequenza acquisita; sono riportati due frame differenti in cui in quello a sinistra il soggetto assume la posizione supina, in quello a destra è sdraiato sul lato sinistro.



**Figura 5.1** – Esempio di due frame estratti dalla sequenza video di input.

Gli algoritmi implementati permettono di individuare il cambiamento di pose e di quantificare il movimento compiuto dal soggetto per spostarsi da una posizione all'altra.

Come già anticipato nel **Capitolo 4**, l'algoritmo consente di ottenere un video di output contenente il file video originale e la maschera di primo piano, entrambi con i blob sovrapposti e, in aggiunta, anche un grafico che indica la quantità di movimento rilevata durante la sessione di monitoraggio. In **Figura 5.2** si riporta un frame estrapolato dal video di output IR.



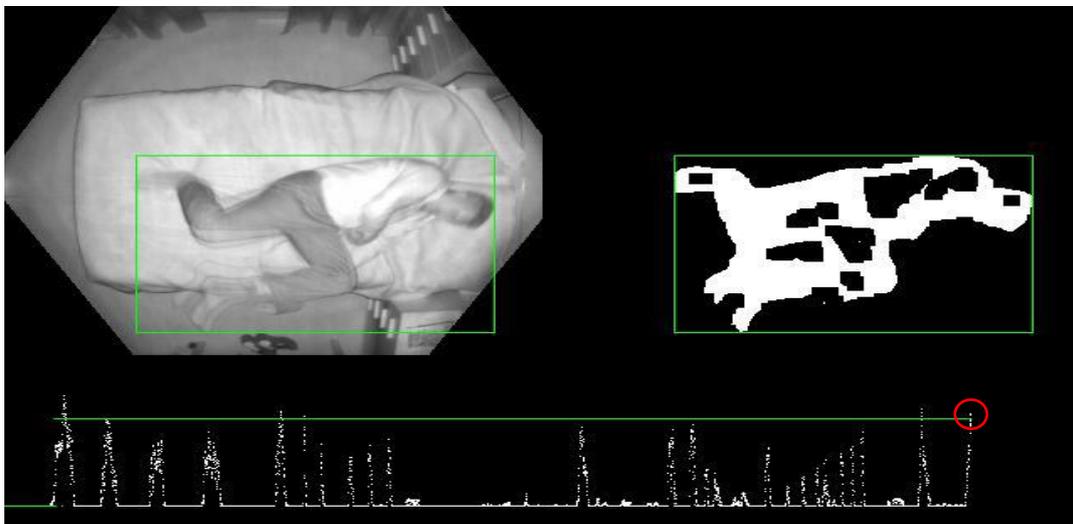
**Figura 5.2** – Nel riquadro in alto a sinistra si riporta il frame relativo al file originale IR; in alto a destra la maschera di Foreground. In basso, il grafico che indica la quantità di movimento frame per frame.

Nella figura in alto è riportato un frame estrapolato circa nel mezzo della sessione di monitoraggio. Si osserva come i blob rilevati siano due. Nel frame in questione è stato rilevato il movimento degli arti inferiori e della parte superiore del corpo, ovvero busto, testa e arti superiori. Lo strumento utilizzato, descritto in **4.2.1**, permette di individuare le varie parti del corpo che si muovono indipendentemente. Il riquadro relativo agli arti inferiori è più piccolo, perché i pixel coinvolti durante il movimento sono in quantità minore rispetto all'area superiore. Con la tecnica di

sottrazione dello sfondo si riesce meglio a focalizzare il movimento, in quanto è messo in primo piano il corpo del soggetto escludendo il letto e l'ambiente in cui si effettua il monitoraggio.

Come si può osservare nel riquadro inferiore della **Figura 5.2** si riporta la quantità di movimento nel tempo. La linea verde indica se il soggetto si trova nel letto oppure no, quando la linea coincide con l'asse allora significa che il letto è vuoto. Dal grafico si riesce a comprendere in quale punto della sessione di monitoraggio il paziente si trova in una determinata posa o compie un determinato movimento.

In **Figura 5.3** si riporta un altro frame, corrispondente quasi alla fine della sessione di monitoraggio come si evince dall'andamento del grafico del movimento. In questo caso, a differenza della **Figura 5.2** è rilevato un unico blob. Il blob racchiude infatti l'intero corpo del soggetto, in quanto si sta alzando dal letto. Si osserva come il picco che si sta creando, relativo al frame in questione, sarà mediamente più alto rispetto ai picchi che si osservano durante tutta la sessione (indicato in **Figura 5.3** in rosso). Questo è dovuto al fatto che il movimento coinvolge l'intero corpo.



**Figura 5.3** – Nel riquadro in alto a sinistra è rappresentato il frame relativo al file originale IR; in alto a destra è riportata la maschera di Foreground. In basso, il grafico che indica la quantità di movimento frame per frame. La differenza rispetto alla **Figura 5.1** consiste nella presenza di un unico blob. Il pallino in rosso si riferisce al picco relativo al movimento nel frame considerato.

### 5.1.2 Check sull'occupazione del letto e analisi del movimento

L'algoritmo sviluppato prevede che si faccia anche una verifica sull'occupazione del letto, ovvero verifica frame per frame se il soggetto si trova nel letto oppure se

questo è vuoto. Nel caso in cui il letto sia vuoto non è rilevato movimento e, di conseguenza, non sono rilevati blob. Ciò accade perché non si identifica una differenza tra un frame e quello successivo, l'immagine resta la stessa. In **Figura 5.4** si riporta la situazione in cui il letto è vuoto. Si osserva come nella parte destra dell'immagine non sia rilevato alcun primo piano, perché nella sequenza originale sono raffigurati solamente il letto e l'ambiente circostante. Quindi, il risultato del metodo di sottrazione dello sfondo in questo caso corrisponde ad un'immagine tutta nera.



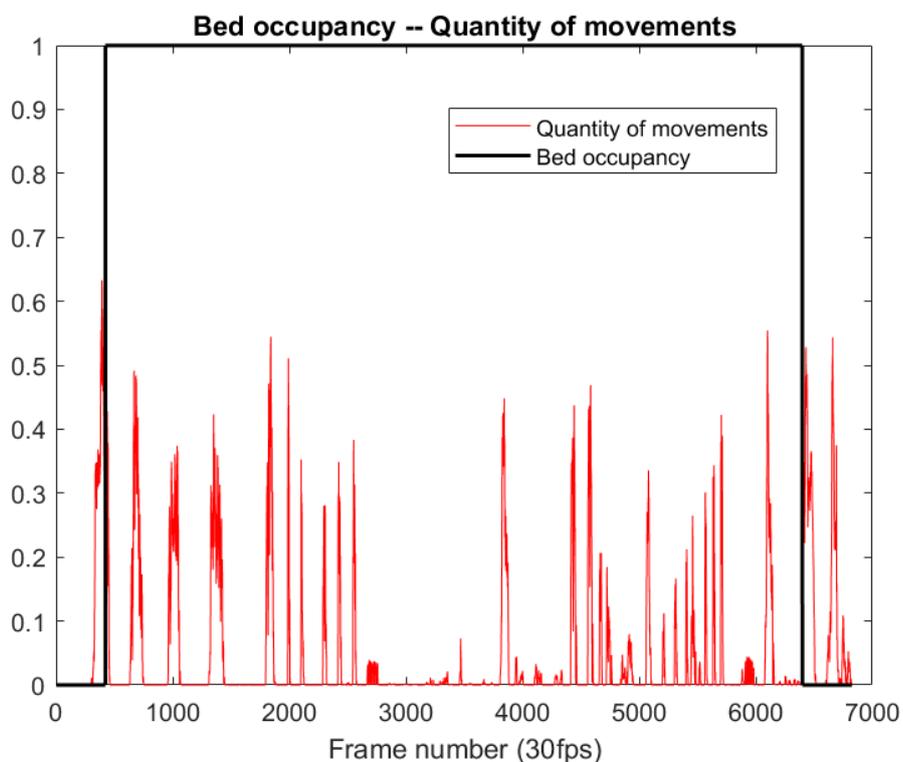
**Figura 5.4** –Situazione in cui il letto è vuoto. La parte destra dell'immagine è tutta nera perché non è presente il soggetto e, quindi, il primo piano.

La verifica sull'occupazione del letto si ricava calcolando la differenza tra un frame e il frame in cui il paziente non si trova sul letto, per esempio il frame iniziale. Se la differenza è superiore ad una data soglia, impostata allo 0,3 % allora significa che il paziente si trova sul letto e può essere rilevato il movimento. Il risultato che indica se il letto è occupato oppure no è dato in numero binario, ovvero 0 o 1. Se la differenza è inferiore alla soglia significa che il letto non è occupato e quindi il valore è impostato a 0; se la differenza è superiore alla soglia significa che il paziente si trova sul letto e quindi il valore è impostato a 1.

In **Figura 5.5** si riporta la sovrapposizione tra il grafico relativo all'occupazione del letto (in nero) e il grafico relativo all'attività motoria del soggetto (in rosso). Si osserva come il grafico che indica l'occupazione del letto può assumere dei valori corrispondenti solamente a 0 oppure a 1.

Analizzare la quantità del movimento è un altro step importante. La quantità di movimento è stata calcolata facendo un rapporto in termini di pixel tra i blob e l'area dei frame. In **Figura 5.5** si osserva l'andamento del movimento nei vari frame (in rosso). Più parti del corpo sono coinvolte nel movimento, più il picco sarà alto. Se, invece, a muoversi è solamente una piccola parte del corpo, come per esempio solo una mano o solo un piede, allora il picco diminuisce.

È interessante notare come all'inizio il paziente non si trova sul letto (valore = 0) e nello stesso intervallo non è rilevato il movimento, quindi si può osservare un'assenza di picchi. Le oscillazioni attorno al picco principale indicano che il paziente si sta muovendo nei vari frame. Meno oscillazioni indicano meno movimento nei frame vicini. Si potrebbe trattare di un movimento singolo, per esempio di uno scatto improvviso. Se invece nei frame vicini al picco ci sono delle oscillazioni, significa che si sta svolgendo un movimento più ampio. Il soggetto potrebbe trovarsi sul punto di cambiare posizione e per questo motivo si può osservare del movimento lungo i vari frame adiacenti. Quanto detto è riportato nella figura sottostante.

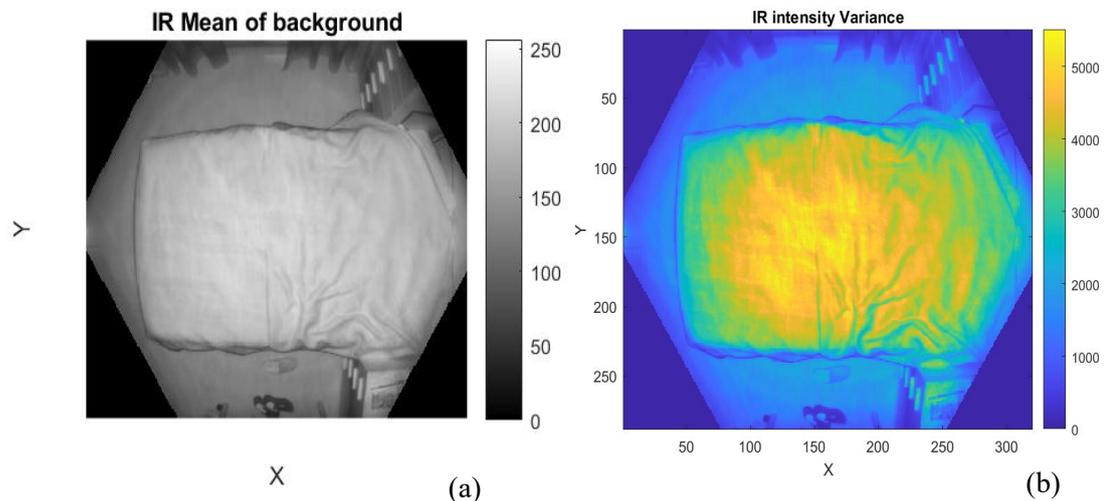


**Figura 5.5** – Sovrapposizione tra il grafico relativo all'occupazione del letto (in nero, valore binario 0-1) e il grafico relativo alla quantità di movimento normalizzata (in rosso), frame per frame.

### 5.1.3 Valutazione del rumore nelle immagini e istogramma di movimento

Per stabilire le soglie da utilizzare nella segmentazione del movimento utilizzate nei paragrafi precedenti è stato calcolato il rumore dei sensori utilizzati dalla Kinect contenuto nelle immagini IR e Depth, rumore rappresentato tramite la varianza e deviazione standard dei valori di luminosità dei pixel, mediate su tutta l'immagine.

Media e varianza sono calcolati su una sequenza di immagini che raffigurano la stessa scena statica. Sono stati utilizzati i semplici comandi MATLAB™ *mean* e *var*. In particolare, la varianza è stata calcolata considerando i valori di profondità e luminosità dei pixel. In **Figura 5.6** si riporta un esempio di output ottenuto da una sequenza video IR.

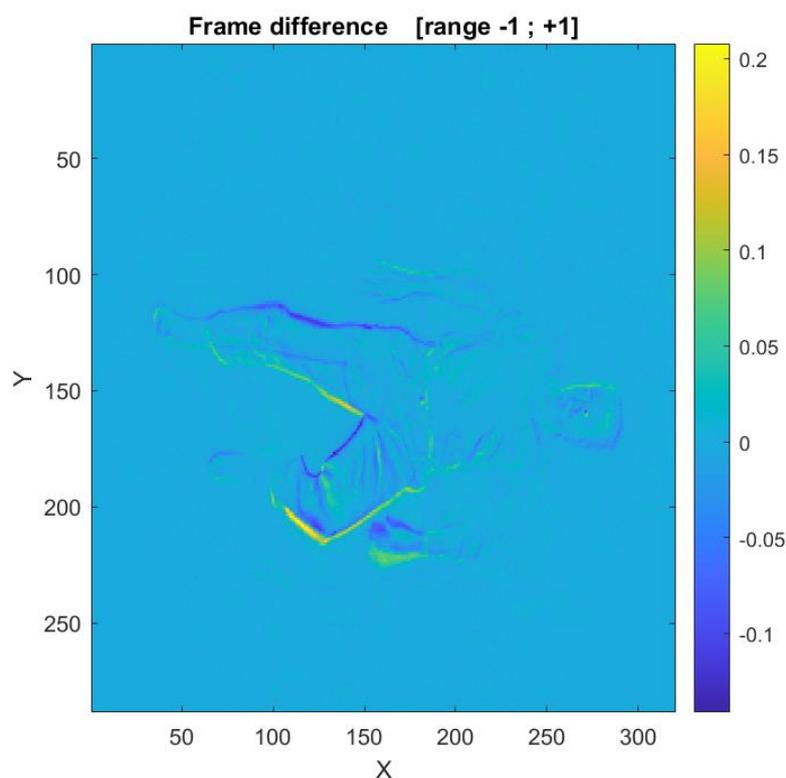


**Figura 5.6** – (a) Intensità media e (b) varianza di una sequenza di immagini IR dello sfondo (unità arbitraria).

Dall'immagine (a) di **Figura 5.6** si osserva la distribuzione della media ricavata. I colori più chiari si trovano al centro, quindi corrispondono al letto. I colori più scuri, invece, sono relativi al resto della stanza, che rappresenta lo sfondo. Dall'immagine (b) si osserva, invece, la distribuzione della varianza. Si osserva come i colori caldi, in questo caso il giallo, siano presenti nella parte centrale del letto, quella più luminosa, come previsto dalle leggi fisiche sulla intensità della luce.

È possibile visualizzare la differenza tra un frame e il successivo in una scala di colori compresi all'interno di un range tra -1 e +1, come mostrato in **Figura 5.7**. Il corpo del soggetto è ben visibile, ma si può osservare come le parti dal colore

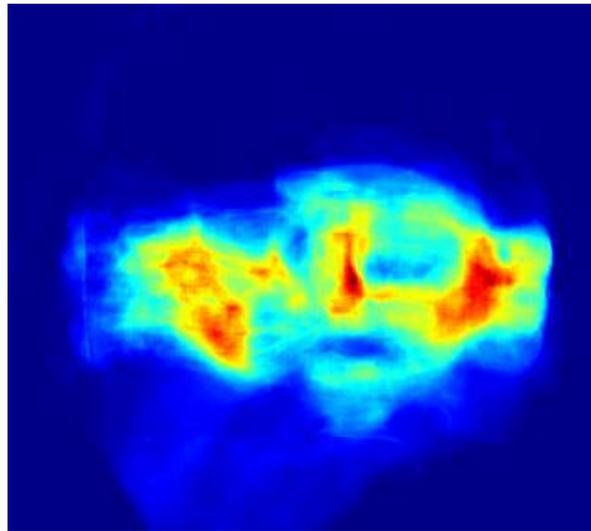
tendente al giallo corrispondano agli arti del soggetto. Questo a testimonianza del fatto che il soggetto tra un frame e l'altro si è mosso con le gambe, la mano e sicuramente ha fatto anche un lieve movimento della testa. La barra dei colori permette di avere un'idea riguardo all'intensità del movimento. La parte attorno al soggetto è omogenea e di un'intensità che è circa pari a 0. Questo dimostra che tra i due frame consecutivi non c'è stata nessuna variazione della scena circostante e, quindi, l'immagine di sfondo è rimasta uguale. Questo potrebbe aiutare a capire su quali parti del corpo focalizzarsi per risalire al possibile disturbo di cui il soggetto potrebbe essere affetto.



**Figura 5.7** – Rilevamento del movimento dalla differenza tra due frame (differenza tra frame normalizzata tra -1/+1).

Inoltre, il software sviluppato permette di ottenere in un file video di output l'evoluzione temporale del movimento valutata tramite l'istogramma della quantità di movimento complessiva calcolata in una finestra temporale  $\Delta T$  che scorre lungo la sequenza di input. Definendo opportunamente l'intervallo temporale  $\Delta T$  può fungere quindi da dinamogramma 2D, catturando movimenti a scale temporali diverse. Un esempio di frame estratto dal file di output in questione è rappresentato in **Figura 5.8**. È possibile effettuare una statistica dei movimenti del soggetto nel

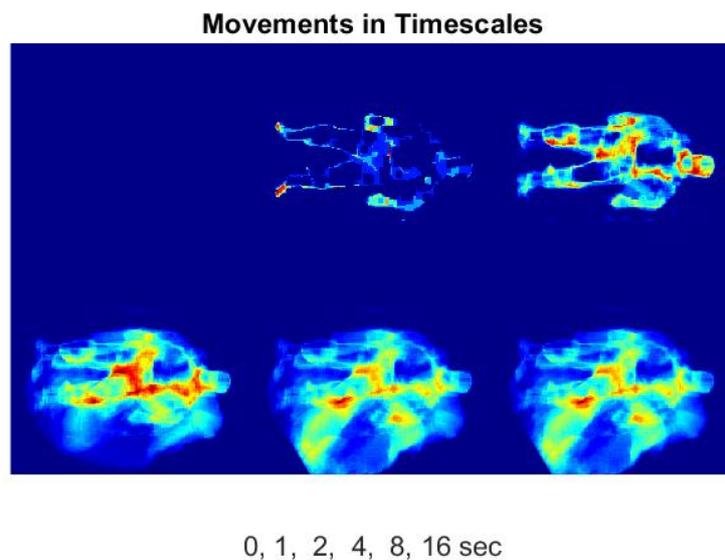
sonno. Si osserva come le parti coinvolte in movimenti più ampi sono quelle indicate in rosso o in giallo e corrispondono al corpo del soggetto che si muove. La parte del letto invece è caratterizzata da un blu più chiaro rispetto a tutto il resto, questo è dovuto al fatto che i movimenti del soggetto si concentrano nel letto e non nel resto della stanza.



**Figura 5.8** – Istogramma della quantità di movimento complessiva della sequenza. In colori più caldi, le zone a maggior movimento.

Il software sviluppato, inoltre, fornisce come file video di output l'evoluzione temporale dell'istogramma della quantità di movimento rilevata nella sequenza di input, ma a sei diverse scale temporali stabilite da diversi valori di  $\Delta T$  da 0.5 a 20 secondi. In questo modo si riesce a caratterizzare movimenti che avvengono su scale temporali diverse. La **Figura 5.9** riporta un esempio di frame in cui si analizza l'istogramma della quantità di movimento alle scale temporali di  $\Delta T$  uguali a 0.5 s, 1 s, 2 s, 4 s, 8 s, 16 s rispettivamente. Nell'immagine nel riquadro ad angolo in alto a sinistra si riporta la scala temporale più breve, invece nel riquadro in basso a destra la scala temporale più lunga, ovvero quella corrispondente a 16 secondi. Le scale temporali di 0.5 s, 1 s e 2 s sono le più brevi e le più adatte a catturare movimenti veloci, mentre le altre descrivono comportamenti motori via via più globali durante il sonno. I colori più caldi si ritrovano principalmente negli arti inferiori, negli arti superiori e in corrispondenza della testa. Le prime due sequenze temporali, in particolare, rilevano il movimento della testa. Nelle sequenze temporali più lunghe sembra che ci sia il movimento anche di altre parte oltre al corpo del soggetto. Attorno al letto, infatti, si crea una sorta di alone che non è di blu scuro, quindi

significa che è stato rilevato una sorta di movimento in questa area. Valutando le ultime tre sequenze temporali, si può dedurre come il soggetto durante il sonno tenda a muovere più la parte sinistra che la destra e tenda a dormire più sul lato sinistro. L'analisi di quale parte del corpo si muove maggiormente potrebbe rappresentare una base per un'ulteriore studio. Da qui, infatti, si potrebbe stimare quanto tempo il soggetto giace in una determinata posizione e quali pose sono predominanti rispetto ad altre. Ciò permetterebbe in modo più efficiente di risalire al tipo di disturbo del sonno di cui il paziente potrebbe essere affetto.

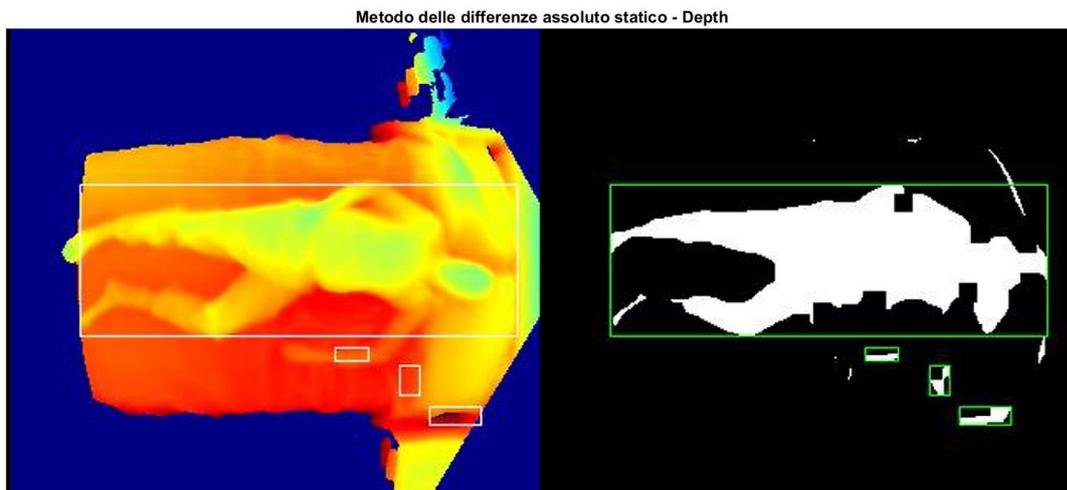


**Figura 5.9** – Istogramma della quantità complessiva di movimento a sei diverse scale temporali (partendo dall'alto a sinistra) di durata 0, 1, 2, 4, 8, 16 secondi.

## 5.2 Risultati Motion-Body Tracking in profondità

Come anticipato nel **Capitolo 4** il metodo delle differenze assolute con continuo aggiornamento dello sfondo, applicato alle sequenze in profondità, non permette di ottenere ottimi risultati come nel caso delle sequenze nell'infrarosso. Per questo motivo si è deciso di applicare il metodo delle differenze assolute con sfondo statico. Lo sfondo resta sempre lo stesso, corrispondente al frame iniziale in cui il letto è vuoto. Dal momento che l'immagine di sfondo non varia e quindi la profondità della scena si mantiene costante, ci si aspetterebbe una buona segmentazione del soggetto e dei suoi movimenti. Questo, invece, non accade perché la superficie del letto si deforma con il peso del soggetto (**Figura 5.10**).

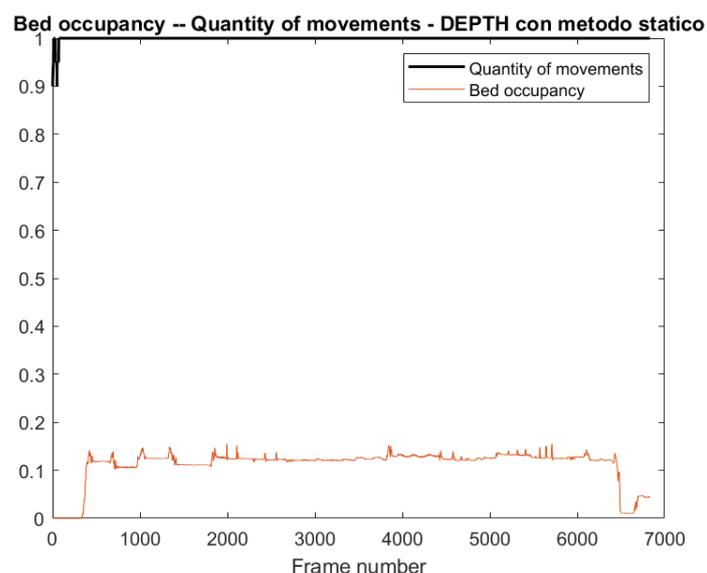
L'algoritmo riconosce la deformazione del letto come parte in movimento e, di conseguenza, sono segmentati come oggetti parti del materasso.



**Figura 5.10** – A sinistra si riporta il frame originale estratto dalla sequenza depth. A destra il risultato del metodo delle differenze assolute con sfondo statico.

Come si evince dalla **Figura 5.10** i risultati ottenuti mostrano come il metodo statico fornisce risultati peggiori del metodo dinamico. Nonostante le prestazioni ottenute con il metodo dinamico per le sequenze di profondità non siano delle migliori, non si verificano dei falsi rilevamenti.

Anche in questo caso si ottiene come output il grafico che riporta l'occupazione del letto sovrapposto al grafico che riporta la quantità di movimento. (**Figura 5.11**)



**Figura 5.11** – Sovrapposizione tra il grafico relativo all'occupazione del letto (in nero) e il grafico relativo alla quantità di movimento (in rosso) per le sequenze di profondità.

A differenza del grafico di movimento ottenuto con le sequenze IR, in questo caso non ci sono dei picchi ben visibili, ma sembrerebbe che il paziente si muove continuamente, quasi per tutta la durata della sessione. Osservando la sequenza originale, invece si osserva che non è così. Come detto prima, l'algoritmo con sfondo statico per le sequenze di profondità rileva come movimento la deformazione del materasso, quindi i picchi rappresentano false rilevazioni di movimento.

## 6 Discussione finale e sviluppi futuri

I risultati riportati nel **Capitolo 5** dimostrano come il sistema hardware e software sviluppato funzioni di base come un actigrafo 2D in grado di rilevare durante la sessione di monitoraggio notturna il movimento del soggetto e la sua intensità, fornendo così la possibilità di stimare l'efficienza del sonno. Inoltre, il sistema è in grado di rilevare la presenza del soggetto nel letto e fornisce in output un istogramma bidimensionale a varie scale temporali che indica quale parte del corpo si è mossa maggiormente. Attraverso il *color bar* si riesce a dire, per esempio, se il paziente muove di più la parte destra del corpo oppure la parte sinistra.

Si è riusciti, quindi, a sviluppare un sistema di monitoraggio del sonno che fornisce informazioni 2D sui movimenti compiuti dal soggetto. Il sistema in questione non è invasivo e rispetto ai sistemi tradizionali non richiede che nessun sensore e nessun dispositivo sia collegato al corpo del paziente. In questo modo il sonno del soggetto non è influenzato da fastidi esterni. Egli può dormire in un ambiente a sé familiare e senza l'utilizzo di nessun dispositivo invasivo.

È necessario, ovviamente, posizionare la telecamera in modo da riuscire a ben inquadrare il letto del soggetto. La telecamera funziona al buio. Per le analisi condotte in questa tesi si è utilizzata Azure Kinect, che a differenza delle telecamere tradizionali riesce ad acquisire sequenze video contemporaneamente nell'infrarosso e in profondità. Questo permette di ottenere una visione tridimensionale e completa del paziente e della scena analizzata.

Il sistema sviluppato comprende algoritmi che sono in grado di rilevare il movimento ed individuare quali parti del corpo si muovono. I video di output generati, quindi, mostrano se il soggetto ha mosso gli arti inferiori, superiori, il busto, la testa e si riesce a dire frame per frame se il soggetto si è mosso e quanto si è mosso.

Dall'analisi dei video di output si riesce a identificare quali pose sono predominanti rispetto ad altre. In questo modo, come anticipato nella **Sezione 4.2** si potrebbe risalire al tipo di disturbo di cui il soggetto potrebbe essere affetto.

Inoltre, con l'analisi condotta nell'infrarosso sono stati ottenuti risultati migliori rispetto all'analisi condotta in profondità. Come mostrato nel **Capitolo 5** nelle

sequenze di profondità le deformazioni del materasso sono considerate come parte del movimento e, di conseguenza, l'algoritmo segmenta anche delle sezioni di materasso. Uno degli sviluppi futuri potrebbe essere l'ulteriore sviluppo degli algoritmi per l'analisi delle sequenze video di profondità che consenta di segmentare correttamente le parti in movimento. Un altro possibile sviluppo è quello di utilizzare alcune meta-caratteristiche estratte dai movimenti per addestrare una rete neurale convoluzionale per l'analisi automatica di movimenti normali e patologici, estendendo l'approccio anche al caso di soggetti sotto le coperte.

Il software realizzato rappresenta un laboratorio di base che comprende algoritmi di motion detection, tracking e di valutazione del movimento. Si riesce a identificare il movimento di corpo e arti, in particolare di movimenti che possono essere correlati a disturbi patologici del sonno.

Sono tuttavia necessari ulteriori studi per determinare correttamente le pose e i movimenti rilevati. Questo consentirà di ampliare i risultati ottenuti dal lavoro svolto in questa tesi e di sviluppare un sistema efficiente, utile non solo per valutare la qualità del sonno di soggetti sani, ma anche per diagnosticare possibili disturbi del sonno in soggetti patologici.

## Lista dei simboli

Simbolo	Descrizione	Valore	Unità di misura
$\sigma$	Standard Deviation	-	-
$\sigma_t$	Standard Deviation nel dominio del tempo	-	-
$\sigma_f$	Standard Deviation nel dominio della frequenza	-	-
$f_c$	frequenza di taglio del filtro	-	$[\frac{campioni}{s}]$
FS	frequenza di campionamento del segnale	30	$[\frac{campioni}{s}]$
$\omega_k$	pulsazione	-	$[\frac{rad}{s}]$
$\alpha$	Guadagno	-	$[\frac{frame}{s}]$

## 7 Bibliografia

- [1] Nurmi, Sami. "Nocturnal sleep quality and quantity analysis with ballistocardiography." (2016).
- [2] Shamim, Sohaib A., et al. "Insomnia: risk factor for neurodegenerative diseases." *Cureus* 11.10 (2019).
- [3] Willemen, Tim. "Biomechanics based analysis of sleep." (2015).
- [4] Hwang, Shang-Rung, et al. "Association of periodic limb movements during sleep and Parkinson disease: A retrospective clinical study." *Medicine* 98.51 (2019).
- [5] Borges, Conrado Regis, et al. "Alzheimer's disease and sleep disturbances: a review." *Arquivos de neuro-psiquiatria* 77.11 (2019): 815-824.
- [6] Peever, John, and Patrick M. Fuller. "The biology of REM sleep." *Current Biology* 27.22 (2017): R1237-R1248.
- [7] Engstrøm, Morten, Eyvind Rugland, and Mona Skard Heier. "Polysomnography (PSG) for studying sleep disorders." *Tidsskrift for den Norske laegeforening: Tidsskrift for praktisk medicin, ny raekke* 133.1 (2013): 58-62.
- [8] Yildirim, Ozal, Ulas Baran Baloglu, and U. Rajendra Acharya. "A deep learning model for automated sleep stages classification using PSG signals." *International journal of environmental research and public health* 16.4 (2019): 599.
- [9] Kim, Taewoon. "Action Recognition in Bed for Sleep Monitoring from Depth Images."
- [10] Jawabri, Khalid H., and Avais Raja. "Physiology, Sleep Patterns." *StatPearls [Internet]* (2020).
- [11] Zoetmulder, Marielle, et al. "Increased motor activity during rem sleep is linked with dopamine function in idiopathic rem sleep behaviour disorder and Parkinson's disease." *Sleep Medicine* 14 (2013): e316.
- [12] Fonseca, P. "Home sleep monitoring." (2018).
- [13] Khademi, Aria, et al. "Personalized sleep parameters estimation from actigraphy: a machine learning approach." *Nature and Science of Sleep* 11 (2019): 387.

- [14]Faust, Oliver, et al. "A review of automated sleep stage scoring based on physiological signals for the new millennia." *Computer methods and programs in biomedicine* 176 (2019): 81-91.
- [15] Sathyanarayana, Supriya, et al. "Vision-based patient monitoring: a comprehensive review of algorithms and technologies." *Journal of Ambient Intelligence and Humanized Computing* 9.2 (2018): 225-251.
- [16]Mendonça, Fábio, et al. "A review of approaches for sleep quality analysis." *Ieee Access* 7 (2019): 24527-24546.
- [17]Fallmann, Sarah, and Liming Chen. "Computational sleep behavior analysis: A survey." *IEEE Access* 7 (2019): 142421-142440.
- [18]Islam, Md Zahidul, KM Talha Nahiyen, and Md Adnan Kiber. "A motion detection algorithm for video-polysomnography to diagnose sleep disorder." 2015 18th International Conference on Computer and Information Technology (ICCIT). IEEE, 2015.
- [19]Heinrich, Adrienne, Xavier Aubert, and Gerard de Haan. "Body movement analysis during sleep based on video motion estimation." 2013 IEEE 15th International Conference on e-Health Networking, Applications and Services (Healthcom 2013). IEEE, 2013.
- [20]Choe, Jeehyun, et al. "Sleep Analysis Using Motion and Head Detection." 2018 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI). IEEE, 2018.
- [21]Khanam, Fatema-Tuz-Zohra, Ali Al-Naji, and Javaan Chahl. "Remote monitoring of vital signs in diverse non-clinical and clinical scenarios using computer vision systems: A review." *Applied Sciences* 9.20 (2019): 4474.
- [22]Lee, Jaehoon, Min Hong, and Sungyong Ryu. "Sleep monitoring system using kinect sensor." *International Journal of Distributed Sensor Networks* 11.10 (2015): 875371.
- [23]Zavala-Mondragon, Luis A., et al. "CNN-SkelPose: a CNN-based skeleton estimation algorithm for clinical applications." *Journal of Ambient Intelligence and Humanized Computing* (2019): 1-12.
- [24]Febriana, Nityacas, Achmad Rizal, and Erwin Susanto. "Sleep monitoring system based on body posture movement using Microsoft Kinect sensor." *AIP Conference Proceedings*. Vol. 2092. No. 1. AIP Publishing LLC, 2019.

- [25] Gall, Markus, et al. "Automated detection of movements during sleep using a 3D time-of-flight camera: design and experimental evaluation." *IEEE Access* 8 (2020): 109144-109155.
- [26] Yang, Cheng, et al. "Sleep monitoring via depth video compression & analysis." 2014 IEEE International Conference on Multimedia and Expo Workshops (ICMEW). IEEE, 2014.
- [27] Masek, Martin, et al. "Sleep monitor: A tool for monitoring and categorical scoring of lying position using 3D camera data." *SoftwareX* 7 (2018): 341-346.
- [28] Sung, Kookjin. *A Real-time Motion Detection with Differential Images and Tracking with Mean-Shift and Kalman Filter*. Diss. 2017.
- [29] Mohanty, Aseema, and Sanjivani Shantaiya. "A survey on moving object detection using background subtraction methods in video." *International Journal of Computer Applications* 975 (2015): 8887.
- [30] Liu, Shuangjun, and Sarah Ostadabbas. "A vision-based system for in-bed posture tracking." *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 2017.
- [31] Pouyan, M. Baran, et al. "Continuous eight-posture classification for bed-bound patients." 2013 6th International Conference on Biomedical Engineering and Informatics. IEEE, 2013.
- [32] Yong, Ching Yee, Rubita Sudirman, and Kim Mey Chew. "Motion detection and analysis with four different detectors." 2011 Third International Conference on Computational Intelligence, Modelling & Simulation. IEEE, 2011.
- [33] Dorudian, Navid. *Nonparametric pixel-wise background modelling and segmentation to detect moving object with RGB-D camera*. Diss. Brunel University London, 2020.
- [34] Heinrich, Adrienne, et al. "Video based actigraphy and breathing monitoring from the bedside table of shared beds." *Journal of Ambient Intelligence and Humanized Computing* 6.1 (2015): 107-120.
- [35] Pan, Qiang, Damien Brulin, and Eric Campo. "JMIR Biomedical Engineering 5.1 (2020): e20921.
- [36] Mohammad Reza Keyvanpour<sup>1</sup>, Shokofeh Vahidian, Mahin Ramezani - HMR-vid: a comparative analytical survey on human motion recognition in video data

- [37] Dutta, Anjan, et al. "Vision tracking: A survey of the state-of-the-art." *SN Computer Science* 1.1 (2020): 1-19.
- [38] Bouwmans, Thierry, and Belmar Garcia-Garcia. "Background subtraction in real applications: Challenges, current models and future directions." *arXiv preprint arXiv:1901.03577* (2019).
- [39] Power, P. Wayne, and Johann A. Schoonees. "Understanding background mixture models for foreground segmentation." *Proceedings image and vision computing New Zealand*. Vol. 2002. 2002.
- [40] Tölgyessy, Michal, et al. "Evaluation of the Azure Kinect and Its Comparison to Kinect V1 and Kinect V2." *Sensors* 21.2 (2021): 413.
- [41] Juan R. Terven (2020). *KinZ-Matlab* (<https://github.com/jrterven/KinZ-Matlab/releases/tag/1.2>), GitHub. Retrieved November 30, 2020.
- [42] Andrews Sobral (2021). *LRSLibrary* (<https://github.com/andrewsobral/lrslibrary>), GitHub. Retrieved February 12, 2021.
- [43] [https://en.wikipedia.org/wiki/Gaussian\\_filter](https://en.wikipedia.org/wiki/Gaussian_filter)
- [44] <https://docs.microsoft.com/en-us/azure/kinect-dk/hardware-specification>
- [45] [https://microsoft.github.io/Azure-Kinect-Sensor-SDK/master/group\\_\\_\\_enumerations\\_gabd9688eb20d5cb878fd22d36de882dd.html#gabd9688eb20d5cb878fd22d36de882ddb](https://microsoft.github.io/Azure-Kinect-Sensor-SDK/master/group___enumerations_gabd9688eb20d5cb878fd22d36de882dd.html#gabd9688eb20d5cb878fd22d36de882ddb)
- [46] Chien, Shao-Yi, et al. "Video object segmentation and tracking framework with improved threshold decision and diffusion distance." *IEEE transactions on circuits and systems for video technology* 23.6 (2013): 921-934.