UNIONE EUROPEA

REGIONE PUGLIA
AREA POLITICHE PER LO SVILUPPO IL LAVORO E L'INNOVAZIONE

AIUTI A SOSTEGNO DEI CLUSTER TECNOLOGICI REGIONALI PER L'INNOVAZIONE

PERSON

**PERSON** - PERVASIVE GAME FOR PERSONALIZED TREATMENT OF COGNITIVE AND FUNCTIONAL DEFICITS ASSOCIATED WITH CHRONIC AND NEURODEGENERATIVE DESEASES

# Deliverable D10.1 – Descrizione del set up di acquisizione e del modello virtuale 3D

| | | |
|---|---|---|
| **Codice Progetto** | : | LQ8FBY0 |
| **Titolo Progetto** | : | PERSON |

| | | |
|---|---|---|
| **Numero Deliverable** | : | D10.1 |
| **Titolo Deliverable** | : | Descrizione del set up di acquisizione e del modello virtuale 3D |
| **Natura del Deliverable** | : | Report |
| **Mese di Rilascio da Capitolato** | : | Gennaio 2016 |
| **Livello di Disseminazione** | : | Pubblico |
| **Versione** | : | 1.1 |
| **Data Deliverable** | : | V1.0 rilasciato il 5/2/2016; V1.1 rilasciato il 14/4/2016 |


# Abstract

Dopo aver presentato le funzionalità che il set up di visione artificiale deve espletare, vengono passati in rassegna gli approcci comunemente usati per la ricostruzione tridimensionale del moto, e fra questi scelto il workflow su cui basare il set up. Viene quindi descritto l'hardware scelto per la realizzazione del prototipo di tale set up.

Infine vengono passati in rassegna vari modelli, presenti in letteratura, utilizzati nella ricostruzione del moto, e nello specifico della postura e del movimento della mano. Traendo ispirazione da alcuni di questi, viene definito il modello virtuale 3D che si è deciso di adottare.



# Keyword list

Descrizione del moto; stereo camera; marker; ricostruzione 3D;


# Tracking delle Versioni

| Version | Changes | Author(s) |
|---|---|---|
| **1.0** | **First Draft** | **Maria Francesca de Ruvo, Marco Colaprico** |
| **1.1** | **Modifica fig. 10**<br>**Inserimento Tesla K40**<br>**Aggiunta figg. 24 e 25** | **Marco Colaprico** |

# Indice

**1. OBIETTIVO ..... 7**

**2. APPROCCI ALLA RICOSTRUZIONE 3D DEL MOTO E SCELTA DELLA METODICA DA ADOTTARE ..... 10**

**3. COMPONENTI HARDWARE DEL SDVA ..... 12**

3.1. MARKERS ..... 12
3.2. TELECAMERE ..... 12
3.3. OTTICA ..... 12
3.4. ILLUMINATORI ..... 13
3.5. FILTRI ..... 13
3.6. SCHEDA DI SINCRONIZZAZIONE ..... 14
3.7. CAVETTERIA ..... 14
3.8. HARDWARE DI CALCOLO ..... 14
*3.8.1. Piattaforma FPGA-based ..... 14*
*3.8.2. GPU ..... 15*

**4. IL PROBLEMA DELLA RICOSTRUZIONE 3D DEL MOTO ..... 17**

4.1. LA STIMA DEI PARAMETRI DELLO SCHELETRO ..... 17
4.2. RICONOSCIMENTO DEI GESTI CON UNA SOLA TELECAMERA E SENZA MARKERS ..... 18
4.3. RICONOSCIMENTO DEI GESTI CON SISTEMI DI MOTION CAPTURE BASATI SUI MARKERS ..... 21

**5. DEFINIZIONE DEL MODELLO SCELTO ..... 25**

**BIBLIOGRAFIA ..... 28**

# Indice figure

FIGURA 1. SCHEMATIZZAZIONE DELLA CONFIGURAZIONE PROTOTIPALE DEL SISTEMA PERSON, IN CUI IL GAME SARÀ FRUITO SU UN NORMALE SCHERMO LCD. ..... 7
FIGURA 2. IPOTESI DI INSTALLAZIONE DI PERSON IN UNO SCENARIO 3D DI TIPO IMMERSIVO, FACENTE USO DI VIDEOPROIETTORI. ..... 8
FIGURA 3. SISTEMA DI RIFERIMENTO RELATIVO ALLA INDICIZZAZIONE DEI PIXEL DI UN MONITOR. ..... 9
FIGURA 4. MARCATORE PASSIVO DOTATO DI RIVESTIMENTO RIFLETTENTE 3M 7610 E FILETTATURA INTERNA PER IL FISSAGGIO. ..... 12
FIGURA 5. TELECAMERA JAI GO-5000-PGE. ..... 12
FIGURA 6. OBIETTIVI TUSS KV0814M2. ..... 13
FIGURA 7. OVERDRIVE RING LIGHT 850NM IR - EZ MOUNT RING LIGHT. ..... 13
FIGURA 8. CARATTERISTICA "TRASMISSIONE=F(□)" DEL FILTRO BN850-27 . ..... 14
FIGURA 9. VIDEO PROCESSOR MEIV-VQ4GE DELLA SILICON SOFTWARE. ..... 15
FIGURA 10. SCHEDE VIDEO NVIDIA: (A) GEFORCE GTX TITAN, (B) TESLA K40. ..... 16
FIGURA 11. FIGURE 1 E 3 DEL LAVORO [8] E RELATIVE DIDASCALIE. ..... 17
FIGURA 12. FIGURA 1 DEL LAVORO [9] E RELATIVA DIDASCALIA. ..... 18
FIGURA 13. FIGURE 4 DEL LAVORO [18] E RELATIVE DIDASCALIE. ..... 19
FIGURA 14. FIGURE 2, 3 E 5 DEL LAVORO [11] E RELATIVE DIDASCALIE. ..... 19
FIGURA 15. FIGURE 3, 4, 5, E 7 DEL LAVORO [19] E RELATIVE DIDASCALIE. ..... 20
FIGURA 16. FIGURA 1 DEL LAVORO [23] E RELATIVA DIDASCALIA. ..... 21
FIGURA 17. FIGURA 1 DEL LAVORO [24] E RELATIVA DIDASCALIA. ..... 21
FIGURA 18. FIGURE 1 E 2 DEL LAVORO [26] E RELATIVE DIDASCALIE. ..... 22
FIGURA 19. FIGURE 1, 2 E 7 DEL LAVORO [27] E RELATIVE DIDASCALIE. ..... 22
FIGURA 20. FIGURE 1 E 2 DEL LAVORO [28] E RELATIVE DIDASCALIE. ..... 23
FIGURA 21. FIGURA 4 DEL LAVORO [29] E RELATIVA DIDASCALIA. ..... 24
FIGURA 22. FIGURE 3, 4 E 8 DEL LAVORO [35] E RELATIVE DIDASCALIE. ..... 24
FIGURA 23. CONFIGURAZIONE SCELTA. NUMERO E POSIZIONE DEI MARCATORI, COME APPAIONO MONTATI SUL SIMULACRO ATTIVO USATO DURANTE I TEST PRELMINARI. ..... 25
FIGURA 24. CONFIGURAZIONE SCELTA. SIMULACRO ATTIVO. ..... 26
FIGURA 25. GESTUALITÀ. ..... 27

## Glossario, acronimi e abbreviazioni

| Item | Descrizione |
|---|---|
| SDVA | Set up di visione artificiale |
| 3D | Tridimensionale |
| (Oxyz) | Sistema di riferimento cartesiano 3D di centro O e assi x, y, z |
| LED | Light Emitting Diode |
| IR | Infrarosso |
| ROI | Region Of Interest, cioè la regione di interesse rispetto all'intero frame |
| CMOS | Complementary Metal-Oxide Semiconductor, è un tipo di tecnologia utilizzata in elettronica per la produzione di circuiti integrati |
| µm | Micron (1 µm = $10^{-6}$ m) |
| nm | Nanometro (1 nm = $10^{-9}$ m) |
| GigE | Standard di interfaccia Giga Ethernet Vision |
| MTBF | Mean Time Between Failures, cioè valore atteso dell'intervallo temporale tra un guasto ed il successivo |
| fps | Frame per second, cioè fotogrammi al secondo |
| Mpixel | Megapixel |
| $\lambda$ | Lunghezza d'onda |
| FPGA | Field Programmable Gate Array, è un circuito integrato le cui funzionalità sono programmabili via software |
| MB/s | Mega bytes al secondo |
| GB | Giga bytes |
| GPU | Graphical Processing Unit |
| GP-GPU | General Purpose GPU |
| CUDA | Compute Unified Device Architecture, architettura hardware di tipo SIMD per l'elaborazione parallela creata da NVIDIA |
| SIMD | Single Instruction Multiple Data, paradigma di architettura di calcolo parallela che prevede che lo stesso flusso di istruzioni sia operato su flussi multipli di dati |
| PCIe | Peripheral Component Interconnect Express, standard di interfaccia d'espansione a bus seriale per computer usato per connettere periferiche alla scheda madre |
| GDDR5 | GDDR, acronimo per Graphics Double Data Rate, è una tecnologia di memoria RAM specifica per schede video. Una delle caratteristiche di questa tipologia di memoria è l'utilizzo di due segnali di clock (uno per l'invio dei comandi e l'altro per le operazioni di lettura e scrittura) permette di inviare dati a più di 5 Gbps minimizzando il rumore |
| FLOPS | FLoating point Operations Per Second e indica il numero di operazioni in virgola mobile eseguite in un secondo dall'unità do calcolo |
| HSV | Hue Saturation Value, spazio di rappresentazione dei colori in cui un |

| Item | Descrizione |
|---|---|
| | colore è descritto attraverso la tonalità (hue), saturazione (saturation) e valore o intensità (value). |

# 1. Obiettivo

Obiettivo principale del SDVA è essenzialmente quello di realizzare un'interfaccia fra il player e il game che possa tradurre e interpretare la gestualità del paziente stesso come interazioni fra esso e le scene e le situazioni di gioco.

Il contesto per il quale il game è pensato prevede una utenza che molto probabilmente non è abituata (o addirittura non è capace) di interagire attraverso joystick o mouse per effettuare le operazioni di puntamento utili a progredire durante il gioco, pertanto l'idea è quella di sostituire tali interazioni con un gesto più naturale come quello di indicare manualmente. È evidente che questo tipo di interazione può essere gestita idoneamente con un monitor di tipo touch-screen; tuttavia, l'evoluzione che si pensa di realizzare in un prossimo futuro per il sistema PERSON è quella di prevedere un'ambientazione immersiva, in cui la realtà aumentata venga fruita dal paziente più che con l'uso di un singolo monitor, adoperando un ambiente 3D sulle cui pareti vengano proiettate le scene del gioco. In questo scenario, l'adozione di mega schermi di tipo touch diverrebbe impraticabile.

Pertanto l'idea alla base dell'Attività 10 è quella di realizzare un set up di visione artificiale che ricostruisca la gestualità del paziente, in particolar modo determini in tempo reale e al troughput di 50 rilevazioni al secondo, quale è il pixel che il player sta indicando con la propria mano, e fornisca quindi al sensor hub le coordinate di detto pixel, per permettere alle altre componenti del game una fluida interazione con il player. Questo sistema potrà quindi essere usato nel prototipo del sistema, che adotterà uno schermo di tipo convenzionale (Figura 1), ma potrà consentire un'immediata migrazione del game ad una installazione che consenta un gaming più coinvolgente, che immerga cioè il player in uno scenario 3D di tipo immersivo, facendo uso di video proiettori (Figura 2).

**Figura 1. Schematizzazione della configurazione prototipale del sistema PERSON, in cui il game sarà fruito su un normale schermo LCD.**

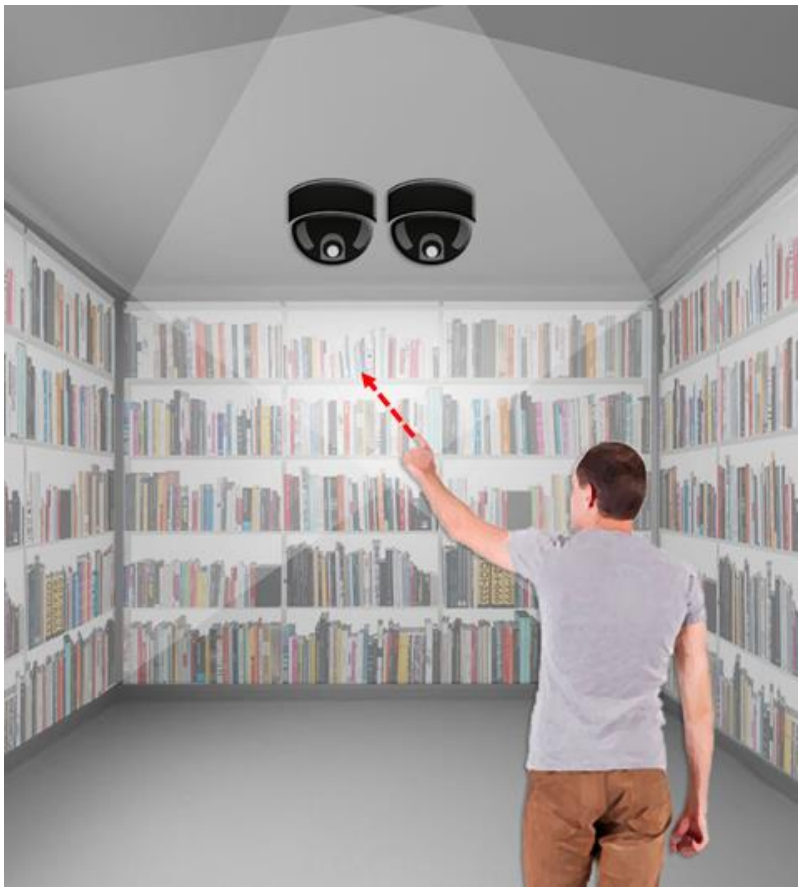


**Figura 2. Ipotesi di installazione di PERSON in uno scenario 3D di tipo immersivo, facente uso di videoproiettori.**

Perché questo sia quindi realizzabile, il sistema deve essere capace di:

- ricostruire e mappare nelle coordinate di un sistema di riferimento 3D (Oxyz) i singoli pixel dello schermo (o delle pareti su cui avverrebbe la proiezione[1]);
- ricostruire e mappare nelle coordinate di (Oxyz) i punti del corpo utili allo studio della interazione player-game, e nello specifico, quindi, alcuni punti salienti della mano sinistra o destra usata dal player per indicare;
- valutare autonomamente quando il player sta usando la mano per indicare piuttosto che mantenerla "a riposo";
- determinare la direzione indicata dalla mano;
- individuare le coordinate 3D ($x_P$, $y_P$, $z_P$) del punto P coincidente con la intersezione di tale direzione con il piano dello schermo (o delle parati su cui avverrebbe la proiezione);
- tradurre la terna ($x_P$, $y_P$, $z_P$) in coordinate ($i_P$, $j_P$) del pixel pixel dello schermo (o delle pareti su cui avverrebbe la proiezione, nella stessa accezione di cui al precedente punto 1), secondo il sistema di riferimento di Figura 3;

[1] Per "coordinate 3D dei pixel delle pareti" si vogliono intendere le coordinate 3D dei punti della parete su cui vengono proiettato i pixel del/dei video proiettore/i.

- trasferire quindi al sensor hub tali coordinate ($i_P$, $j_P$).

**Figura 3. Sistema di riferimento relativo alla indicizzazione dei pixel di un monitor.**

# 2. Approcci alla ricostruzione 3D del moto e scelta della metodica da adottare

Attualmente, la tecnica più comunemente usata per la motion capture impiega metodi ottici per registrare il movimento di un soggetto. Un insieme di markers sono fissati al soggetto, che è osservato da due o più telecamere.

Il sistema di acquisizione deduce, attraverso tecniche stereoscopiche basate sulla triangolazione la posizione spaziale di ciascun marker, al variare del tempo. Su grandi volumi di acquisizione, per ridurre al minimo la perdita di segnale a causa di occlusioni e per aumentare la precisione del sistema, si impiegano degli array ridondanti di una pluralità di telecamere, in modo tale da avere la certezza che esistano, per ogni marker di cui è richiesta la ricostruzione 3D, almeno una coppia di telecamere in cui detto marker sia visibile e non occluso.

Per facilitare l'operazione di segmentazione dei markers, si cerca di aumentare il contrasto tra marcatore e sfondo. I sistemi passivi più comuni lavorano con i markers di materiale fortemente retroriflettente e una fonte di illuminazione collocata in modo solidale ad ogni telecamera. I sistemi attivi, usano marcatori LED impulsati in modo sincrono con lo shutter delle telecamere digitali. In entrambi i tipi di sistemi, le telecamere possono essere dotate di filtri ottici centrati sulla lunghezza d'onda della sorgente di illuminazione (in caso di marker attivi) o di emissione (marker attivi) LED.

Una differenza significativa tra i sistemi è che i marcatori attivi possono essere identificati in modo univoco grazie alla modulazione dei loro impulsi, mentre i sistemi passivi devono dedurre l'identità del marker dall'osservazione continua di questi ultimi. Entrambi i tipi di sistemi ottici generano in modo affidabile la posizione esatta dei marcatori.

I sistemi di motion capture restituiscono la posizione dei markers al variare del tempo. Questo formato non descrive la struttura complessiva del soggetto catturato. Per semplificare la rappresentazione del moto si utilizza la struttura di uno scheletro che viene sovrapposto ai dati acquisiti. Parametrizzare i dati con un tale modello rende più semplice la visualizzazione del moto del soggetto.

Un altro approccio è quello alla base di dispositivi commerciali, come il Microsoft Kinect. Questo approccio non fa uso di marcatori posizionati sul corpo, ma proietta sulla scena degli spot in uno spettro non visibile (tipicamente IR) e effettua la ricostruzione 3D in forma di cloud puntiforme attraverso l'osservazione di come tali punti vengono ad adagiarsi sulle superfici. Questo approccio, però proprio perché si occupa di ricostruire la posizione degli spot proiettati, non garantisce la ricostruzione dello stesso punto del corpo istante dopo instante, poiché non è detto che tale punto sia sempre impattatato da uno spot proiettato.

Ciò fa anche si che l'estrapolazione della direzione puntata dall'indice del player partendo dal cloud di punti 3D che verrebbe ricostruito con tecniche di questo tipo non garantirebbe una precisione accurata come invece ci si prefigge di fare, e pertanto il workflow che si è deciso di considerare per lo sviluppo è quello che fa uso di marcatori. Infatti, anche se nell'optare per l'adozione di marcatori da applicare al corpo va certamente considerata la loro invasività si è valutato che il sistema PERSON prevede comunque che il player indosserà durante il gaming un guanto aptico utile a determinare altri parametri dell'attività diagnostica, e pertanto, l'apposizione di marcatori sul guanto stesso è tale da non costituire un aggravio dell'invasività a cui il player sarà sottoposto.

Nella fattispecie, consultandosi con il partner SENSICHIPS, responsabile della progettazione e realizzazione del guanto aptico, si è optato per marcatori passivi, piuttosto che per quelli attivi, per sgravare il guanto stesso da problematiche legate all'alimentazione dei LED.

# 3. Componenti hardware del SDVA

Alla luce delle considerazioni presentate nel paragrafo precedente, il SDVA che si è deciso di realizzare sarà basato sulla componentistica di seguito descritta.

## 3.1. Markers

Come detto si opterà per marcatori passivi, per non gravare il sensore aptico con problematiche legate all'alimentazione di marcatori attivi. Nello specifico, si sono scelti dei marcatori sferici rivestiti da Reflective Tape "3M 7610" del diametro di ¼ di pollice (6.4 mm), che consentono il fissaggio sul guanto aptico nel numero e nelle posizioni che verranno descritte nella sezione "5. Definizione del modello scelto", attraverso opportune viti, grazie ad una predisposizione filettata, come visibile in Figura 4.

**Figura 4. Marcatore passivo dotato di rivestimento riflettente 3M 7610 e filettatura interna per il fissaggio.**

## 3.2. Telecamere

Dato il volume oggetto dell'osservazione, inferiore ad 1 $m^3$, si ritiene che una sola coppia di telecamere, configurate in modalità stereo, sia sufficiente allo scopo. Esse dovranno avere compattezza, adeguata sensitività e risoluzione, oltre che consentire un frame rate di almeno 50 fps. Dopo un'attenta analisi dei dispositivi in commercio, si è optati per le GO-5000-PGE, della casa costruttrice JAI (Figura 5). Le caratteristiche principali di tali dispositivi sono riassunte di seguito.

- Sensore CMOS con global shutter
- Risoluzione 5-megapixel (2560 x 2048)
- Pixel quadrati da 5.0 µm
- 60 dB linear dynamic range
- CMOS pattern correction implementata a bordo
- 8/10/12-bit output GE
- MTBF > 200.000 ore
- Alimentabile direttamente tramite interfaccia GigE
- Ottica attacco C-mount

**Figura 5. Telecamera JAI GO-5000-PGE.**

## 3.3. Ottica

Nella scelta dell'ottica occorre tenere presente sia la risoluzione della telecamera, che la dimensione del sensore, nonché lo standard di attacco del corpo camera, oltre al campo di vista e la distanza di lavoro. Poiché le GO-5000-PGE verranno configurate per operare con una ROI 1920x1080, che garantisce sia adeguata risoluzione, che un throughput di 56 fps (quindi superiore a quello di 50 fps fissato come target) di 56 (1080p HD video), da un calcolo della misura della diagonale della matrice parzializzata risultante sul CMOS (1920x1080 pixel da 5.0 µm) risulta che obiettivi da 2/3 di pollice sono idonei ad ottenere frame senza incappare nella vignettatura.

Si è pertanto optato per obiettivi attacco C, da 2/3 di pollice e risoluzione di 2 Mpixel, come i KV0814M2 della TUSS, con focale di 8 mm (Figura 6).

**Figura 6. Obiettivi TUSS KV0814M2.**

## 3.4. Illuminatori

Avendo optato per un approccio basato su marcatori passivi, è opportuno demandare l'illuminazione a sorgenti terze rispetto ai marcatori stessi, affinché essi siano comunque ben nitidi nelle immagini da elaborare.

Allo scopo di non creare disturbo al player e rendere il più neutro possibile il SDVA, si è considerata una lunghezza d'onda nella banda IR, e specificatamente quella ad 850 nm. Si è pertanto scelto di adoperare gli illuminatori ad anello "OverDrive Ring Light 850nm IR - EZ Mount Ring Light" della Smart Vision Light (Figura 7).

**Figura 7. OverDrive Ring Light 850nm IR - EZ Mount Ring Light.**

## 3.5. Filtri

Allo scopo di ridurre l'interferenza con sorgenti di luminosità terze, rendendo stabile il SDVA rispetto alla luminosità dell'ambiente, si doteranno le telecamere di filtri centrati su $\lambda$ pari a 850 nm, come i BN850-

27 della MIDWEST OPTICAL, che ha una trasmissione a 850 nm superiore all'85%, come risulta dalla caratteristica riportata in Figura 8.

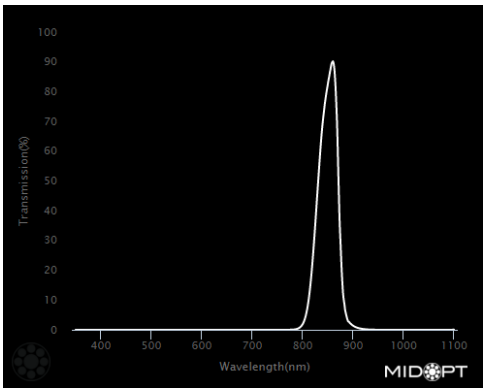



**Figura 8. Caratteristica "Trasmissione=f(□)" del filtro BN850-27 .**

## 3.6. Scheda di sincronizzazione

Perché la ricostruzione 3D ottenuta da una coppia di immagini stereoscopiche sia corretta, occorre che le due immagini siano state acquisite contemporaneamente. Se così non fosse, il matching dei marker nelle due distinte immagini, essendo queste riferite ad istanti di acquisizione diversi, porterebbe, in caso di moto, a derivare coordinate 3D non corrispondenti alla realtà. È pertanto necessario sincronizzare le telecamere attraverso uno stesso segnale di trigger. Allo scopo il SDVA comprenderà la scheda ME IV-OPTO-TRIGGER I/O della Silicon Software.

## 3.7. Cavetteria

Le varie componenti appena descritte saranno cablate ai fini dell'alimentazione e del trasferimento dati, attraverso opportuni cavi.

Nello specifico, il trasferimento dati dalle telecamere al sensor hub avverrà attraverso cavi gigabit ethernet categoria 6 RJ45, i segnali di sincronismo alle stesse viaggeranno su cavi MVA-50-3-X-3M, mentre l'alimentazione ai ring IR sarà provvista tramite cavi Smart Vision Light 5 Pin.

## 3.8. Hardware di calcolo

Infine occorre considerare i moduli di calcolo su cui gli algoritmi di video analisi dovranno essere implementati per garantire un processing in tempo reale.

Come descritto nel Piano di Attuazione, verranno portate avanti, arrivando ad una valutazione comparativa, due strategie parallele: una basata su implementazione in hardware su dispositivi FPGA based, un'altra esplorando paradigmi di parallel computing che farà riferimento a GPU.

### 3.8.1. Piattaforma FPGA-based

Come piattaforma di calcolo basata su FPGA su cui svolgere le attività A.10.3, si è scelto il video processor MEIV-VQ4GE della Silicon Software (Figura 9). Esso, oltre ad essere perfettamente interfacciabile con la scheda di sincronizzazione di cui al paragrafo 3.6, consente l'interfacciamento con quattro telecamere GigE (quindi rendendo possibile anche l'espansione dall'attuale configurazione, ad un set up che comprenda 2 coppie stereo) ed è provvista di un FPGA Spartan 3 della Xilinx accessibile all'utilizzatore per essere configurato in modo da svolgere a bordo della scheda stessa una efficiente elaborazione video.

Dal punto di vista della connessione, essa richiede un alloggiamento PCIe 4x sulla scheda madre del sensor hub. Di seguito vengono riassunte le principali caratteristiche.

- Throughput massimo 750 MB/s
- 8 input trigger
- 8 output trigger
- 512 MB buffer
- 4 porte RJ45
- conversione diretta Bayer-RGB

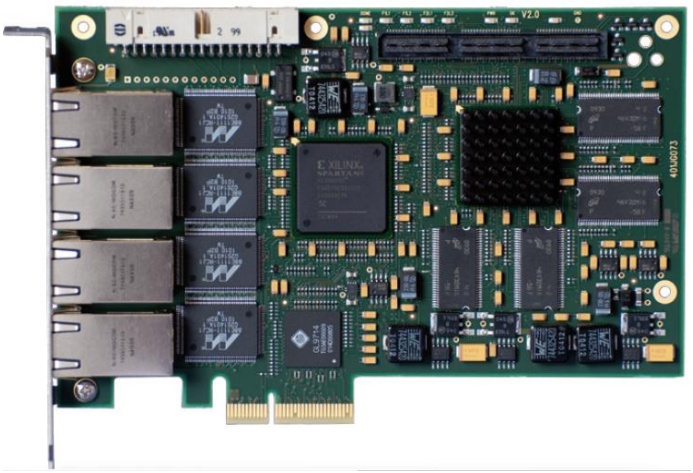


**Figura 9. Video processor MEIV-VQ4GE della Silicon Software.**

### 3.8.2. GPU

Le crescenti quote di mercato che hanno recentemente interessato l'home entertainment hanno spinto lo sviluppo delle schede video fino a renderle sempre più performanti e disponibili a costi competitivi. Questo scenario ha destato l'attenzione di un gran numero di ricercatori interessati al parallel computing, che hanno visto la possibilità di usare questi dispositivi come base di sviluppo di generici processi (General Purpose Graphic Processing Units, GP-GPU), piuttosto che come deputati alle elaborazioni grafiche da visualizzare su monitor. Dal loro canto, i produttori di schede video (NVIDIA su tutti), intravedendo la chiave di accesso ad un nuovo mercato, hanno rapidamente offerto un consistente supporto tramite lo sviluppo di ambienti di programmazione come CUDA.

In questo scenario, l'attività A.10.2 si prefigge di esplorare la possibilità di implementare gli algoritmi di analisi e ricostruzione 3D proprio sfruttando il paradigma di parallel computing di tipo SIMD tramite GP-GPU, e per avvalersi della tecnologia CUDA, si è scelto di adoperare come piattaforma delle GPU NVIDIA, nella fattispecie, una Geforce GTX Titan equipaggiata con processore GK110 con architettura Kepler (capability 3.5) (Figura 10.A), che presenta le seguenti caratteristiche:

- 2688 CUDA core
- Clock 876 Mhz
- 6 GB di memoria GDDR5 con banda passante di 288 GB/s
- Bus PCIe 3.0

- Perfomance di picco in virgola mobile sono di 4.5 Tflops (singola precisione) e 1.5 Tflops (doppia precisione)

e una Tesla K40 equipaggiata con processore GK110 con architettura Kepler (capability 3.5).

- 2880 CUDA core
- Clock 875 Mhz
- 12 GB di memoria GDDR5 con banda passante pari a 288 GB/s
- Bus PCIe 3.0
- Perfomance di picco in virgola mobile sono di 4.29 Tflops (singola precisione) e 1.43 Tflops (doppia precisione)

Entrambe le GPU sono basate su architettura Kepler, che offre streaming multiprocessor più potenti rispetto alla precedente architettura (Fermi) e la possibilità di eseguire simultaneamente fino a 32 connessioni (gestite in hardware) tra CPU e GPU.

(A)

(B)

**Figura 10. Schede video NVIDIA: (A) Geforce GTX Titan, (B) Tesla K40.**

# 4. Il problema della ricostruzione 3D del moto

In questa sezione passiamo brevemente in rassegna alcune strategie che sono state utilizzate nella ricostruzione del moto, e nello specifico della postura e del movimento della mano.

Riferendoci quindi ad alcuni articoli scientifici, ne descriviamo succintamente gli approcci a cui ci siamo ispirati per definire il modello virtuale 3D che adotteremo e che verrà descritto nella sezione successiva.

## 4.1. La stima dei parametri dello scheletro

Molti ricercatori hanno studiato il problema della stima dei parametri dello scheletro a partire da dati ottenuti da diversi tipi di sistemi di motion capture. Ad esempio O'Brien [1] e i suoi colleghi hanno escogitato un algoritmo che stima lo scheletro da dati ottenuti da un sistema di motion capture magnetico. Alcuni esempi di metodi di stima dello scheletro a partire da dati ottenuti con dei sistemi di motion capture ottici si possono trovare in [2, 3, 4]. Problemi simili sono stati studiati anche in letteratura biomeccanica e robotica, alcuni specifici esempi per inferire informazioni relative all'anatomia dello scheletro di un soggetto umano a partire dal moto di markers montati sulle ossa o sulla pelle, può essere trovata in [5, 6, 7].

Nel lavoro [8], da cui sono tratte le figure mostrate in Figura 11, è presentato un algoritmo per la stima automatica dei parametri dello scheletro di un soggetto, a partire dai dati provenienti da dispositivi di motion capture ottici attivi. L'algoritmo proposto raggruppa i marcatori in gruppi di segmenti, determina poi la connettività topologica tra questi gruppi, e individua le posizioni dei giunti di collegamento.

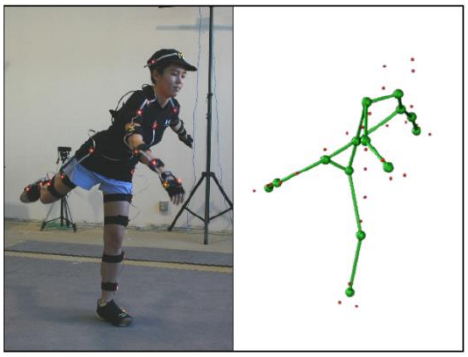


Figure 1. Automatic skeletal reconstruction for a human subject captured with active markers: The image on the left shows a photograph of the subject during the capture session. The image on the right shows the reported marker positions corresponding to when the picture was taken along with the kinematic skeleton automatically constructed by our system.

Figure 3. Reconstructed skeleton overlaid on video frame. This figure allows rough visual assessment showing that the correct topology has been recovered and that joint locations are approximately correct.

**Figura 11. Figure 1 e 3 del lavoro [8] e relative didascalie.**

Nel lavoro [9], da cui è estratta l'immagine mostrata in Figura 12, è proposto un metodo completamente automatizzato per inizializzare e monitorare la configurazione dello scheletro di esseri umani a partire da dati ottici di motion capture, senza la necessità di alcun intervento da parte dell'utente.

Il metodo applica un'inizializzazione con una posizione a T, che risulta essere molto flessibile in quanto funziona con una vasta gamma di posizionamenti dei markers. La configurazione dello scheletro è stimata

in modo robusto attraverso l'ottimizzazione ai minimi quadrati e si sfrutta la struttura dello scheletro per l'etichettatura completamente automatica dei marcatori.

I sistemi di motion capture soffrono per l'occlusione dei markers che porta ad una perdita di informazione utile. Andreas Aristidou e Joan Lasenby [10] hanno affrontato il problema della localizzazione delle articolazioni dello scheletro umano in presenza di dati mancanti e hanno proposto un framework che predice la posizione dei markers occlusi utilizzando un Variable Turn Model all'interno di un Unscented Kalman filter.

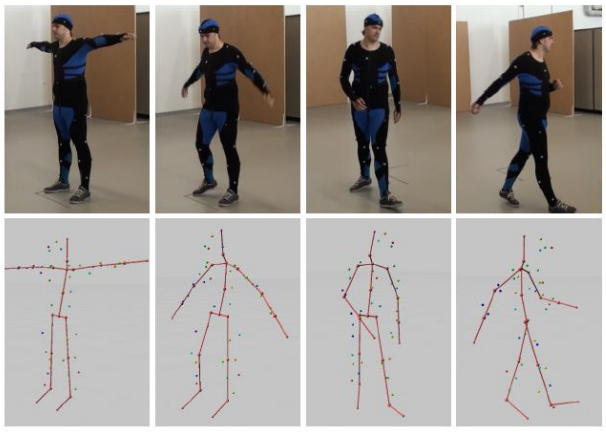


Fig. 1. Skeleton pose estimation. Top: Human with passive optical markers attached to the body. Bottom: Observed markers and the skeleton configuration as estimated by our fully automated method.

**Figura 12. Figura 1 del lavoro [9] e relativa didascalia.**

## 4.2. Riconoscimento dei gesti con una sola telecamera e senza markers

Focalizzando la nostra attenzione sulla mano, e nello specifico sul riconoscimento della gestualità, citiamo alcuni lavori che non sono ricorsi all'uso dei marker.

La maggior parte dei sistemi di riconoscimento dei gesti si può considerare composto da tre stadi: il rilevamento della mano, il tracking e il riconoscimento della gestualità.

Il primo stadio di rilevamento della mano è cruciale nei sistemi che non fanno uso di markers, in quanto isola la mano dallo sfondo prima di passarla agli stadi successivi di tracking e di riconoscimento. In letteratura sono stati proposti molti metodi per la segmentazione della mano.

Ad esempio il lavoro [11], utilizza una telecamera ad infrarossi per una rilevazione affidabile delle mani dell'utente.

La segmentazione del colore della pelle è stata utilizzata da diversi approcci. Per la segmentazione sono stati proposti diversi spazi di colore: RGB, RGB normalizzato, HSV, YCrCb, YUV, ecc. In genere si preferiscono gli spazi di colore che separano in modo efficiente la cromaticità dalla componente di luminanza. Questo è dovuto al fatto che impiegando componenti che dipendono solo dal colore, si riesce a rendere il sistema più robusto alle variazioni di illuminazione della scena. Per aumentare la robustezza alle variazioni di illuminazione [12], alcuni metodi lavorano nello spazio HSV [13] o YUV [14], in modo da approssimare la cromaticità della pelle piuttosto che il suo colore apparente.

Quando la detection della mano è basata sulla segmentazione del colore della pelle, possono però verificarsi degli errori se nella scena sono presenti oggetti con un colore simile a quello della pelle. Questo problema viene risolto negli approcci basati sulla sottrazione del background [15]. Comunque i metodi di sottrazione del background si basano sull'assunzione che il background sia statico. Per risolvere i problemi che si verificano quando lo sfondo varia nel tempo, alcuni ricercatori hanno utilizzato metodi di correzione dinamica o di compensazione del background[16].

Una volta che, nell'immagine di input, è stata identificata la regione della mano, questa è analizzata per calcolarne la posizione e l'orientamento globale, o per stimare la posizione delle punta delle dita. Per esempio, nel metodo proposto da Maggioni [17], per determinare la posizione delle punta delle dita è utilizzato il contorno della regione della mano estratta. Analogamente Segen e Kummar [18] partono dal contorno della regione estratta, individuano il numero di picchi e di valli (Figura 13) del contorno e ne deducono il tipo di gesture.

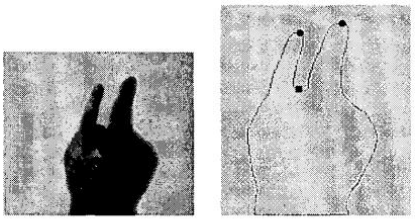


Figure 4: Two peaks and a valley.

**Figura 13. Figure 4 del lavoro [18] e relative didascalie.**

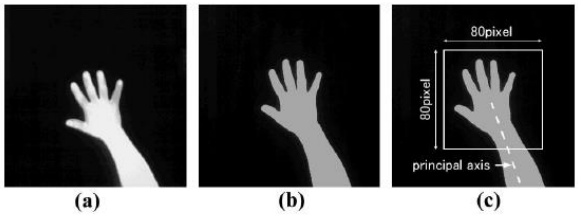



Figure 2. Extraction of hand region

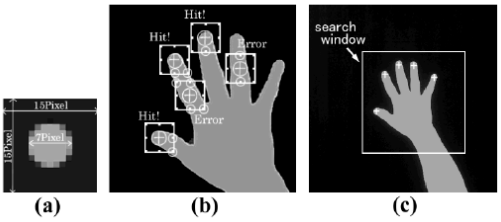



Figure 3. Template matching for fingertips

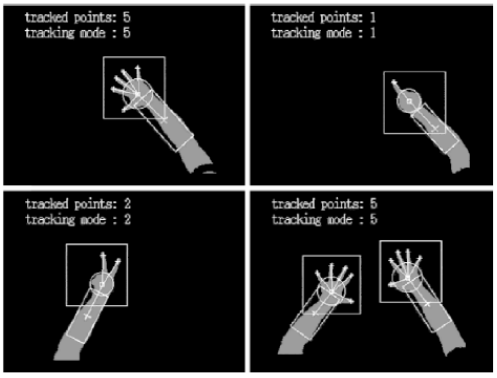



Figure 5. several examples of tracking results

**Figura 14. Figure 2, 3 e 5 del lavoro [11] e relative didascalie.**

Il lavoro [11], da cui sono tratte le immagini di Figura 14, utilizza invece una strategia di matching template per trovare le punte delle dita e propone un metodo veloce e robusto per il tracking della posizione dei centri della mano sinistra e della mano destra.

La pubblicazione [19], da cui sono tratte le immagini di Figura 15, propone un sistema per il riconoscimento dei gesti, in cui il rilevamento delle mani viene eseguito utilizzando l'algoritmo di Viola-Jones. Inoltre, è utilizzato il Continuously Adaptive Mean Shift Algorithm (CAMShift) [20] per tenere traccia della posizione della mano rilevata in ciascun fotogramma del video. Il contorno della mano viene estratto utilizzando un algoritmo di Border Following, che è preceduto da una sogliatura sul colore della pelle, eseguita nello spazio di colore HSV. Infine viene introdotta una rete probabilistico-semantica, che utilizza un modello ontologico dei gesti, e viene utilizzata una rete Bayesiana per il riconoscimento dei gesti.

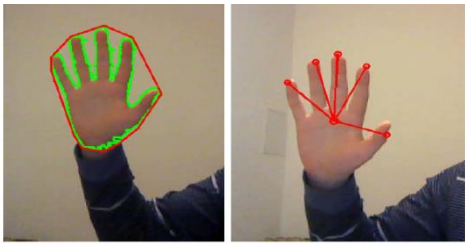


Figure 5. Example of finger tips extraction

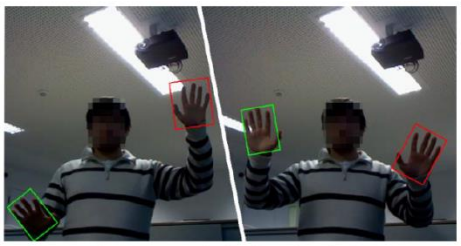


Figure 3. Example of tracking of two hands.

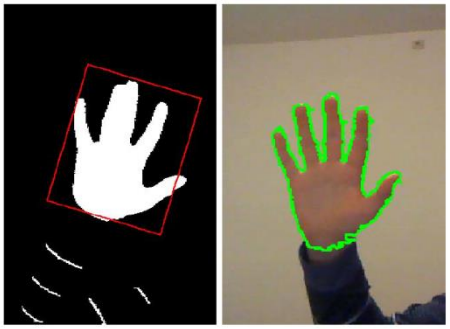


Figure 4. Example of skin-colour detection and contour extraction

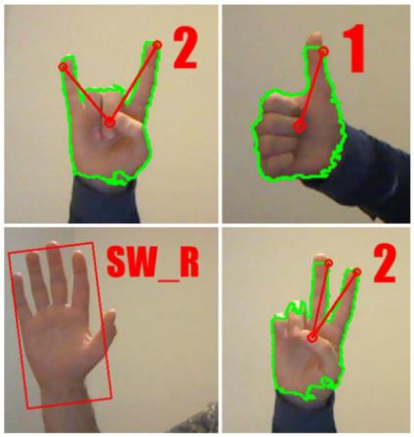



Figure 7. Results of gesture recognition.

**Figura 15. Figure 3, 4, 5, e 7 del lavoro [19] e relative didascalie.**

Stenger [21, 22], per inseguire il moto della mano, ha proposto dei metodo statistici, l'Unscented Kalman Filter e il Hierarchical Bayesian Filter. Tali metodi sono però molto lontani dall'essere real-time e quindi il loro uso è limitato.

Wang e Popovic [23] e Fredriksson [24], propongono un metodo che utilizza un guanto ordinario su cui è stato stampato un pattern custom (Figura 16 e 17). La stampa del pattern colorato sul guanto semplifica notevolmente il problema della stima della posizione della mano e rende tale soluzione semplice e computazionalmente economica.

**Figure 1:** *We describe a system that can reconstruct the pose of the hand from a single image of the hand wearing a multi-colored glove. We demonstrate our system as a user-input device for desktop virtual reality applications.*

**Figura 16. Figura 1 del lavoro [23] e relativa didascalia.**

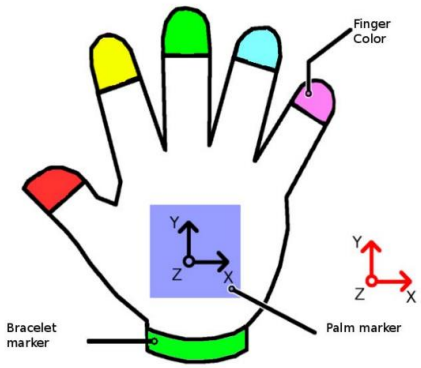



**Figure 1. Finger colors are painted while bracelet and palm markers (back and front) are solid.**

**Figura 17. Figura 1 del lavoro [24] e relativa didascalia.**

## 4.3. Riconoscimento dei gesti con sistemi di motion capture basati sui markers

Benché interessanti, gli approcci che non fanno uso di marker come quelli descritti nel precedente paragrafo, risentono di una intrinseca imprecisione dovuta all'accuratezza con cui i contorni vengono estratti, né del resto è possibile ottenere indicazioni tridimensionali ricorrendo ad una sola telecamera, quando questa non è affiancata a un proiettore di pattern come per esempio avviene nella profilometria. Nell'applicazione d'interesse del progetto PERSON, non è sufficiente riconoscere la gestualità ma, come detto nella sezione 1 di questo documento, è fondamentale individuare il pixel puntato dal player, con grande precisione ed in tempo reale. Pertanto, ci rifaremo ad un approccio simile a quello visto nella sezione 4.1, relativo all'intero scheletro umano, ma che focalizza l'attenzione sulla sola mano. Diversi metodi sono stati proposti per ottenere il movimento della mano a partire da sistemi di motion capture che fanno uso di markers.

Lien e Huang [25] propongono un modello della mano associato ad una soluzione in forma chiusa di cinematica inversa per il processo di fitting delle dita. La posizione 3D dei markers è stata ottenuta utilizzando markers colorati e la visione stereo, e la posizione delle dita è ottenuta con un metodo di ricerca che trova la posizione migliore tra tutte le possibili posizioni.

Il metodo proposto da Qiushi Fu [26] prevede il tracking della cinematica della mano utilizzando il filtro di kalman esteso (EKF). La Figura 18 mostra il modello della mano utilizzato, che prevede 29 gradi di libertà ed è descritto da 24 markers, di cui 4 sull'avambraccio distale, 15 sulla superficie dorsale delle articolazioni

delle dita e 5 sulle punta delle dita. Per ridurre il carico computazionale lo spazio degli stati è stato suddiviso in quattro sotto-spazi, ciascuno dei quali è stimato con un EKF.

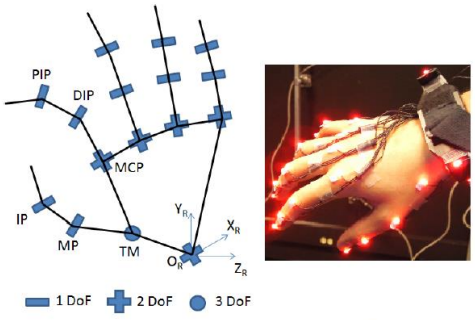



Fig. 1. (a) Kinematic structure of the hand model, and (b) marker protocol.

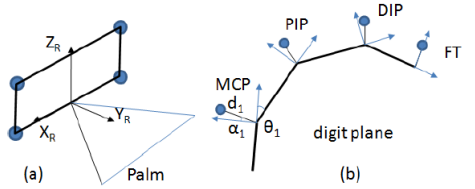



Fig. 2. Illustration of marker model: (a) markers placed on bracelet for tracking hand global posture, and (b) markers placed on the joint surface for tracking of digit movements.

**Figura 18. Figure 1 e 2 del lavoro [26] e relative didascalie.**

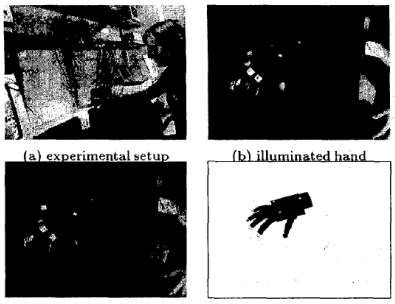



Figure 1: Hand reconstruction procedure

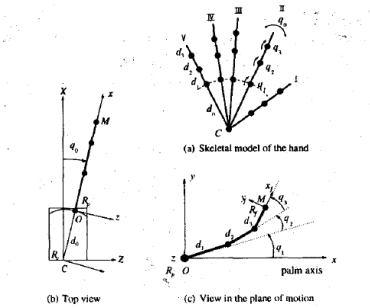



Figure 2: Kinematic model of fingers II to V

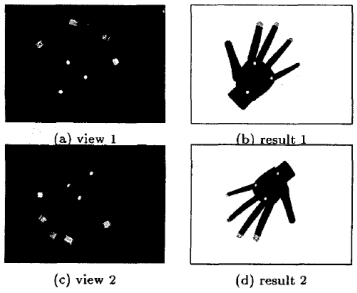



Figure 7: Two views of the same hand posture and the resulting reconstructions

**Figura 19. Figure 1, 2 e 7 del lavoro [27] e relative didascalie.**

In maniera simile, nel lavoro [27], da cui sono tratte le immagini di Figura 19, veniva proposto un modello della mano con 26 gradi di libertà, 6 gradi di libertà per il polso e 4 gradi di libertà per ogni dito.

Per ottenere una ricostruzione accurata della posizione della mano esso utilizza 8 markers. Tre markers dello stesso colore sono posti sul palmo e servono per ottenere la posizione e l'orientamento globale della mano. Sulle punta delle dita è posto un marker di colore unico, in quanto la posizione della punta del dito è sufficiente per calcolare gli angoli dell'articolazione, noto l'angolo di abduzione. I colori dei markers sono stati scelti in modo tale da massimizzare la distanza tra i clusters nello spazio dei colori HSV. Il bianco è stato scelto per i markers del palmo e per le dita sono stati scelti il rosso, l'arancione, il ciano, il verde e il giallo.

L'approccio di Cerveri [28], Figura 20, prevede 24 markers e un modello cinematico della mano che consiste in una catena gerarchica con 22 gradi di libertà più la roto-traslazione globale. Da una posizione statica, una procedura completamente automatica di calibrazione, calcola gli assi e i centri di rotazione, che sono poi utilizzati come base di una animazione interattiva in tempo reale del modello della mano. Il tracking del moto, basato su un labelling automatico dei markers e su un filtro predittivo, è reso più robusto dall'introduzione di vincoli relativi alla posizione funzionale delle dita.

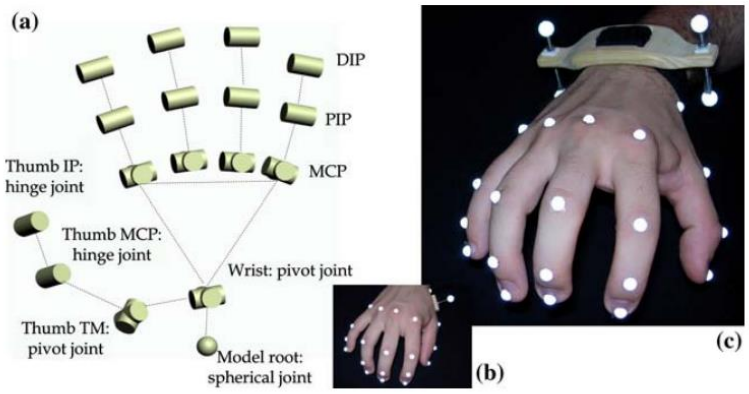



**FIGURE 1. Kinematic model of the complete hand structure: the model root is characterized by a spherical joint endowed with three rotational DoFs and three translations. The wrist is modeled with a pivot joint (two orthogonal rotation axes—flexion-extension and adduction–abduction). The TM joint is represented by two coupled hinge joints; IP and DIP joints of the thumb are modeled with hinge joints (flexion-extension); the MCP of the fingers is modeled with a pivot joint (flexion-extension and adduction–abduction); the PIP and DIP joints of the fingers are modeled with hinge joints (flexion-extension) (a). Additional marker used to calibrate thumb intrinsic pronation (c): it is positioned on a rigid structure bound to the first metacarpal segment and directed outward from a flat open palm. With the two markers on TM and MCP, it defines a fit plane whose normal characterizes the flexion *X*-axis; the adduction *Z*-axis is orthogonal to *X*-axis, belongs to the plane and passes through the TM marker. The pronation axis is therefore estimated as *Z*–*X* cross product (b). Configuration of 24 markers used in the marker protocol: four markers, rigidly linked on a structure locked on the wrist, account for the global translation and orientation of the hand. Each finger is endowed with four markers placed in correspondence of joints. The four markers in correspondence of the metacarpo-phalangeal joints allow the palm motion reconstruction (c).**

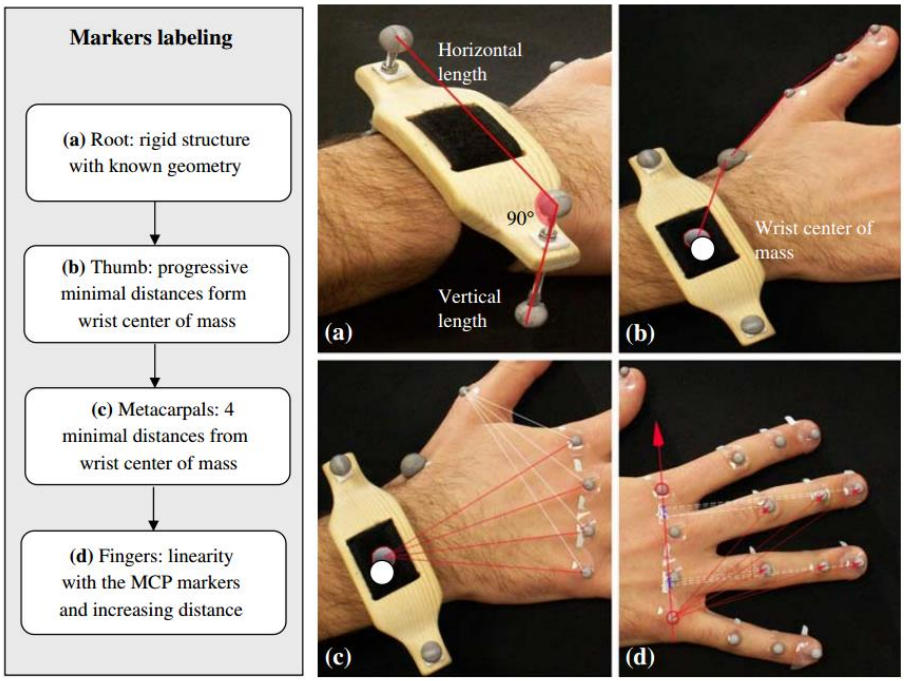
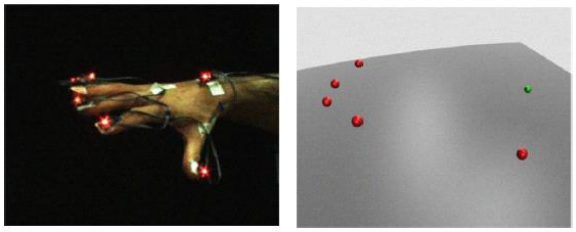



**FIGURE 2. Chart of the automatic labeling procedure. The markers on the wrist are first classified using geometrical constraints originated from the structure geometry (a). Then thumb markers are recovered using progressive distance constraints from wrist markers (b). Markers at the metacarpo-phalangeal joints are classified using minimal distances from wrist markers (c). At last, each finger is labeled using geometric constraints (d).**

**Figura 20. Figure 1 e 2 del lavoro [28] e relative didascalie.**

Andreas Aristidou [29] ha posizionato i markers in punti strategici, e ha utilizzato un risolutore di cinematica inversa [30, 31, 32] per sovrapporre il resto delle articolazioni al modello della mano. Il modello è fortemente limitato con vincoli di rotazione e di orientazione, consentendo solo un numero limitato di movimenti.



(a) (b)

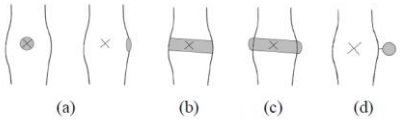
Fig. 4. (a) The true hand pose, (b) the markers as seen from the motion capture system.

**Figura 21. Figura 4 del lavoro [29] e relativa didascalia.**

Raviraj Nataraj e Zong-Ming Li [33] presentano un metodo per determinare la cinematica del pollice e dell'indice utilizzando un set minimo di markers. Per il calcolo della posizione 3D e dell'orientamento dei segmenti delle dita sono applicate delle tecniche di cinematica inversa e di ottimizzazione.

Specialmente nel tracking della mano, i problemi di occlusione ( spesso auto-occlusione dovuta alle altre dita) rendono il post processing molto difficile e dispendioso dal punto di vista del tempo di calcolo. Jonathan Maycock e Tobias Rohlig [34] introducono un metodo di tracking completamente automatico che utilizza un approccio di cinematica inversa. Il metodo Ungherese è utilizzato per calcolare in modo efficiente l'associzione tra i markers del modello e i markers del sistema di motion capture. Il loro algoritmo opera in modalità online e in modalità offline che utilizza la conoscenza dell'intera traiettoria per risolvere il problema delle occlusioni.

Citiamo infine il lavoro [35], cui si riferiscono le immagini in Figura 22, ed in cui è stato effettuato uno studio su diversi tipi di markers. Viene dimostrato che i markers caratterizzati da una superficie piana sono ben visibili solo se perfettamente di fronte alla telecamera. Quando i marcatori hanno una superficie convessa, una porzione della superficie riflettente del marker sarà sempre orientata verso la telecamera, diminuendo le occlusioni. I markers che soffrono meno dell'auto-occlusione delle dita sono risultati essere delle palline retroriflettenti posizionate sulle articolazioni delle dita. Nello stesso lavoro, il modello 3D del dito utilizzato è una catena cinematica delle articolazioni del dito mappata direttamente sui marcatori.



(a) (b) (c) (d)

Figure 3. Shape of markers - in contrast to round blips on the surface (a) that do not always represent the joint position (cross) well if rotated away from the camera, flat rings (b) are always centered at the joint, while convex rings (c) improve upon flat rings in that they have better retroreflective properties. Our final choice are displaced balls (d) that suffer the least from self-occlusion of the fingers.

(a) (b)

Figure 4. Gloves fitted with retroreflective markers

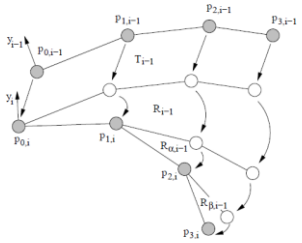


Figure 8. Marker transformation

**Figura 22. Figure 3, 4 e 8 del lavoro [35] e relative didascalie.**

# 5. Definizione del modello scelto

Dopo aver esaminato diversa letteratura scientifica inerente la tematica in questione, si sono operate delle scelte progettuali sulla strategia da implementare nel modello.

In primis si è considerato che un approccio che faccia uso di marker è senz'altro preferibile ad un markerless, sia perché la precisione ottenibile con questo tipo di approccio è notevolmente maggiore di quello privo di marcatori, sia perché lo svantaggio dell'indossare marker è praticamente nullo nella nostra applicazione, appurato che la mano del paziente-player già per altre motivazioni a carattere diagnostico è soggetta all'indossamento di un guanto aptico, e pertanto equipaggiare detto guanto con marker non rappresenta alcun appesantimento all'infrastruttura a cui deve sottostare il giocatore, che tra l'altro, deve anche prestarsi ad indossare un caschetto per il rilevamento dell'elettroencefalogramma.

Successivamente si è valutato il numero di marcatori e dove posizionarli.

Per questa attività si è ritenuto opportuno sperimentare diverse configurazioni, e per farlo si è realizzato un semplice simulacro di mano in cartone (quindi piegabile in maniera coerente al normale movimento della mano) su cui sono stati cablati diversi LED, in diverse posizioni. Sono stati quindi effettuati alcuni test, variando in ciascuno i LED alimentati, allo scopo di verificare diverse situazioni.

Considerate le funzionalità che il SDVA deve svolgere all'interno di PERSON si è quindi verificato che la configurazione ottimale per accuratezza e numero minimo di marker è quella mostrata in Figura 23.

**Figura 23. Configurazione scelta. Numero e posizione dei marcatori, come appaiono montati sul simulacro attivo usato durante i test prelminari.**

Essa prevede otto marcatori: uno sulla punta di ogni dito (dal mignolo al pollice A, B, C, D, E), e tre sul dorso della mano (uno sulle nocche del mignolo e dell'indice, rispettivamente F e G, uno sul dorso, centrato fra il polso e le nocche, H).

In questa configurazione si è verificata la possibilità di interpretare una gestualità che fosse relativa al numero di dita aperte, alla possibilità che la mano fosse chiusa o aperta, e soprattutto, come detto

individuare l'esatta direzione di puntamento. Mentre è chiara l'utilità e l'utilizzo dei marcatori A, B, C, D ed E, è meno intuitiva la necessità e l'utilizzo dei F, G e H.

In primis, teniamo presente che la distanza fra il segmento F-G e uno qualsiasi dei marcatori A, B, C, D ed E è decisiva per comprendere se una delle dita è piegata su se stessa o meno, e ancora più utile è lo studio di come, detto X uno qualunque dei marcatori A, B, C, D ed E, il coseno direttore del piano comprendente F, G e X sia approssimativamente parallelo al coseno direttore del piano comprendente F, G e H. Questa misura infatti si traduce in quanto il dito relativo al marcatore X sia "steso" nell'atto dell'indicare qualcosa o meno.

Inoltre, segnaliamo che l'analisi mirante a ricostruire della direzione puntata parte dalla ricostruzione della posizione dei marcatori nello spazio 3D, tuttavia, la direzione che si vuole calcolare è quella legata al puntamento del dito, e non dei marcatori, che fisicamente sono sovrapposti al dito, con un certo offset dovuto al loro montaggio sul guanto aptico. Per questo motivo, nel determinare il pixel puntato dai marcatori occorrerà correggere le coordinate del pixel ottenuto dal puntamento dei marcatori, spostandosi da questo di una misura pari alla componente parallela al piano del monitor dell'offset di montaggio, nella direzione del coseno direttore del piano comprendente F, G e X.

La realizzazione del simulacro di Figura 23, che fa uso di sensori attivi (LED) si è resa necessaria in attesa dell'approvvigionamento degli illuminatori infrarossi (vedi sez. 3.4) e filtri passabanda (vedi sez. 3.5). Una volta resi disponibili tali apparati è stato possibile realizzare un sensore passivo, fissando su un guanto di tela i markers riflettenti descritti in sez. 3.5.

Il sensore passivo è mostrato in Figura 24. In esso il triangolo verde è necessario per introdurre rigidità al triangolo "FGH", in quanto il tessuto del guanto è elastico. In questo modo il sensore può essere indossato da mani diverse, senza necessariamente tarare e calibrare il ssitema.

**Figura 24. Configurazione scelta. Simulacro attivo.**

Nella Figura 25 viene mostra la gestualità che il sistema dovrà interpretare, come concordata con i partner responsabili della implementazione del Serious Game (Grifo Multimedia) e del Sensor Hub (Reply).

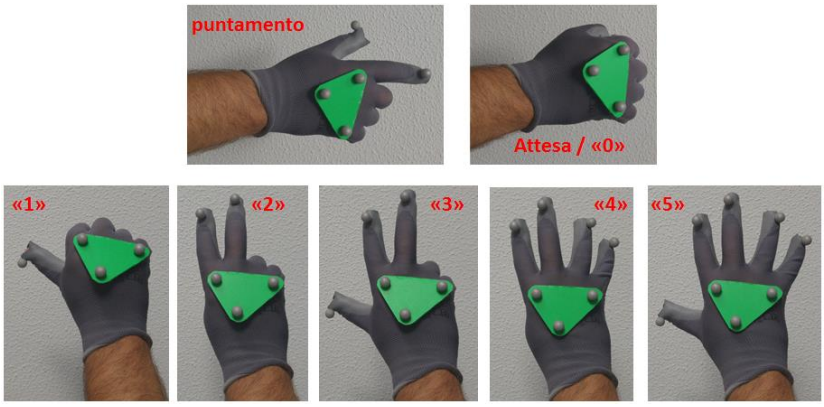



**Figura 25. Gestualità.**

# Bibliografia


[1] J. F. O'Brien, R. E. Bodenheimer, G. J. Brostow, and J. K. Hodgins, "Automatic joint parameter estimation from magnetic motion capture data. ", In Proceedings of Graphics Interface 2000, pages 53–60, May 2000

[2] M.-C. Silaghi, R. Plankers, R. Boulic, P. Fua, and D. Thalmann, "Local and global skeleton fitting techniques for optical motion capture". In N. M. Thalmann and D. Thalmann editors, Modelling and Motion Capture Techniques for Virtual Environments, volume 1537 of Lecture Notes in Artificial Intelligence, pages 26–40, 1998.

[3] M. Ringer and J. Lasenby, " A procedure for automatically estimating model parameters in optical motion capture". In BMVC02, page Motion and Tracking, 2002.

[4] E. de Aguiar, C. Theobalt, and H.-P. Seidel. "Automatic learning of articulated skeletons from 3D marker trajectories". In Second Int. Symposium on Advances in Visual Computing (ISVC), Part I, 2006.

[5] J. H. Challis. "A procedure for determining rigid body transformation parameters". Journal of Biomechanics, 28(6):733–737, 1995.

[6] J. J. Spiegelman and S. L.-Y. Woo. "A rigid-body method for finding centers of rotation and angular displacements of planar joint motion". Journal of Biomechanics, 20(7):715– 721, 1987.

[7] F. E. Veldpaus, H. J. Woltring, and L. J. M. G. Dortmans. "A least-squares algorithm for the equiform transformation from spatial marker co-ordinates". Journal of Biomechanics, 21(1):45–54, 1988.

[8] Adam G. Kirk, James F. O'Brien, David A. Forsyth, "Skeletal Parameter Estimation from Optical Motion Capture Data", Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) Vol. 2, pp. 782-788, 2005.

[9] J. Meyer, M. Kuderer, J. Muller, and W. Burgard, "Online Marker Labeling for Fully Automatic Skeleton Tracking in Optical Motion Capture," Proceedings of the IEEE International Conference on Robotics & Automation (ICRA), Hong Kong, China, pp. 5652–5657, 2014.

[10] Andreas Aristidou and Joan Lasenby. "Real-time marker prediction and CoR estimation in optical motion capture". Submitted to Image and Vision Computing, February 2010.

[11] Yoichi Sato, Yoshinori Kobayashi, Hideki Koike, "Fast Tracking of Hands and Fingertips in Infrared Images for Augmented Desk Interface", Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition 2000, p. 462-467, March 26-30, 2000

[12] J. Martin and J. Crowley. "An appearance-based approach to gesture-recognition. In Int. Conf. on Image Analysis and Processing", pages 340–347, Florence, Italy, 1997.

[13] D. Saxe and R. Foulds. "Toward robust skin identification in video images". In IEEE Int. Conf. on Automatic Face and Gesture Recognition, pages 379–384, 1996, 1996.

[14] A. A. Argyros and M. I. A. Lourakis. "Real-time tracking of multiple skin-colored objects with a possibly moving camera". In Proc. European Conference on Computer Vision, pages 368–379, Prague, Chech Republic, May 2004.

[15] J. Rehg and T. Kanade. Digiteyes: “Vision-based hand tracking for human-computer interaction”. In Workshop on Motion of Non-Rigid and Articulated Bodies, pages 16–24, Austin Texas, November 1994.

[16] A. Utsumi and J. Ohya. “Image segmentation for human tracking using sequential-image-based hierarchical adaptation”. In Proc. IEEE Computer Vision and Pattern Recognition (CVPR), pages 911–916, 1998.

[17] C. Maggioni and B. Kammerer, "GestureComputer - history, design and applications," Computer Vision for Human-Machine Interaction (R. Cipolla and A. Pentland, eds.), pp. 23-51, Cambridge University Press, 1998.

[18] J. Segen and S. Kumar, "Shadow gestures: 3D hand pose estimation using a single camera," Proc. 1999 IEEE Conference on Computer Vision and Pattern Recognition, pp. 479-485, June 1999.

[19] Mykyta Kovalenko, Svetlana Antoshchuk, Juergen Sieck, “Real-Time Hand Tracking and Gesture Recognition Using Semantic-Probabilistic Network”, UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, Cambridge, UK, pp. 269-274, 2014.

[20] G.R. Bradski, “Real Time Face and Object Tracking as a Component of a Perceptual User Interface,” Proc. IEEE Workshop Applications of Computer Vision, pp. 214-219, 1998.

[21] Bjorn Stenger, Paulo R. S. Mendona, and Roberto Cipolla. “Modelbased 3d tracking of an articulated hand”. In IEEE Conference on Computer Vision and Pattern Recognition, volume 2, pages 310-315, Los Alamitos, CA, USA, 2001. IEEE Computer Society.

[22] Bjorn Stenger, Arasanathan Thayananthan, Philip H.S. Torr, and Roberto Cipolla. “Model-based hand tracking using a hierarchical bayesian filter”. IEEE Trans. on Pattern Analysis and Machine Intelligence, 28(9):1372-1384, Sept. 2006.

[23] Robert Y. Wang and Jovan Popovic. “Real-time hand-tracking with a color glove”. In ACM SIGGRAPH 2009, pages 1-8, New York, NY, USA, 2009. ACM.

[24] Jonas Fredriksson, Sven Berg Ryen, and Morten Fjeld. “Real-time 3d hand-computer interaction: optimization and complexity reduction”. In Proceedings of the 5th Nordic conference on Human-computer interaction, pages 133-141, New York, NY, USA, 2008. ACM.

[25] Cheng-Chang Lien and Chung-Lin Huang. “Model-based articulated hand motion tracking for gesture recognition”. Image and Vision Computing, 16(2):121-134, 1998.

[26] Qiushi Fu, Marco Santello, “Tracking whole hand kinematics using extended Kalman filter”, 32nd Annual International Conference of the IEEE EMBS, Buenos Aires, Argentina, pp. 4606-4609, 2010.

[27] F. Lathuilière and J. Y. Hervé, “Visual hand posture tracking in a gripper guiding application,” in Proc. Int. Conf. Robotics and Automation (ICRA), San Francisco, US, vol. 2, pp. 1688-1694, 2000.

[28] Cerveri P, De Momi E, Lopomo N, Baud-Bovy G, Barros RML, Ferrigno G (2007) “Finger kinematic modeling and real-time hand motion estimation”. Ann Biomed Eng 35(11):1989–2002

[29] Andreas Aristidou, Student Member, IEEE, Joan Lasenby, “Motion Capture with Constrained Inverse Kinematics for Real-Time Hand Tracking”, Proceedings of the 4th International Symposium on Communications, Control and Signal Processing, ISCCSP 2010, Limassol, cyprus, 3-5 March 2010

[30] CHRIS WEIMAN. "INVERSE KINEMATICS AND GEOMETRIC CONSTRAINTS FOR ARTICULATED FIGURE MANIPULATION". MASTER DISSERTATION, SIMON FRASER UNIVERSITY, DEPARTMENT OF COMPUTER SCIENCE, 1993.

[31] ANDREAS ARISTIDOU AND JOAN LASENBY. "FABRIK: A FAST, ITERATIVE SOLVER FOR THE INVERSE KINEMATICS PROBLEM". SUBMITTED TO GRAPHICAL MODELS, FEBRUARY 2010.

[32] ANDREAS ARISTIDOU AND JOAN LASENBY. "INVERSE KINEMATICS: A REVIEW OF EXISTING TECHNIQUES AND INTRODUCTION OF A NEW FAST ITERATIVE SOLVER". TECHNICAL REPORT F-INFENGLFR. 632, CUED, 2009.

[33] L. Y. CHANG, AND , N. S. POLLARD, "CONSTRAINED LEAST-SQUARES OPTIMIZATION FOR ROBUST ESTIMATION OF CENTER OF ROTATION," J BIOMECH ENG 135(9), 091002 (JUL 10, 2013)

[34] JONATHAN MAYCOCK, TOBIAS ROHLIG , MATTHIAS SCHRODER , MARIO BOTSCH AND HELGE RITTER. "FULLY AUTOMATIC OPTICAL MOTION TRACKING USING AN INVERSE KINEMATICS APPROACH". 15TH INTERNATIONAL CONFERENCE ON HUMANOID ROBOTS, 2015

[35] KLAUS DORFMULLER-ULHAAS, DIETER SCHMALSTIEG, "FINGER TRACKING FOR INTERACTION IN AUGMENTED ENVIRONMENTS", PROCEEDINGS OF THE IEEE AND ACM INTERNATIONAL SYMPOSIUM ON AUGMENTED REALITY (ISAR'01), P.55-64, 2001.