

# HumanPose3D

*Monocular Markerless 3D Pose Analyse voor de Gezondheidszorg*

Datum:

27-01-2026

Minor:

Zorg & Technologie

Cursus:

*Public Health Engineering – Inleiding*

GKO-PHE.1-13

Docent:

Evelien Dijk

[Evelien.dijk@hu.nl](mailto:Evelien.dijk@hu.nl)

Opdrachtgever:

Jaap Jansen

[jaap.jansen@hu.nl](mailto:jaap.jansen@hu.nl)

Auteurs:

Joey Kardous (1890363)

[Joey.kardous@student.hu.nl](mailto:Joey.kardous@student.hu.nl)

Max Jansen (1848868)

[Max.jansen3@student.hu.nl](mailto:Max.jansen3@student.hu.nl)

Woordenaantal:

7789



Dit rapport beschrijft de ontwikkeling van *HumanPose3D*, een *open-source pipeline* voor *markerless 3D* bewegingsanalyses. Het systeem transformeert smartphone-video naar *ISB-compliant* gewrichtshoeken, zonder dure *motion-capture* apparatuur of fysieke markers.

Het gerealiseerde product is een gebruiksvriendelijke desktop-applicatie die monoculaire video's verwerkt naar een *3D pose*-reconstructie. De *pipeline* genereert 59 van de 64 biomechanische markers via *Pose2Sim* augmentatie (Pagnon et al., 2021) en berekent 12 *ISB-compliant* gewrichtsgroepen volgens de internationale standaarden (Wu et al., 2002, 2005). Het gebruik van deze ISB-standaarden waarborgt dat de gegenereerde data direct vergelijkbaar is met bestaande klinische literatuur en compatibel met gangbare biomechanische analysesoftware, wat de bruikbaarheid voor zorgprofessionals aanzienlijk vergroot. Het systeem draait volledig lokaal op een standaard laptop, met een verwerkingsijd van ongeveer 60 seconden per video.

Naast de *baseline pipeline* is een experimentele POF-module ontwikkeld (*Part Orientation Fields*, gebaseerd op Xiang et al., 2019). Deze module bereikt  $\sim 7^\circ$  *limb*-oriëntatiefout op validatiedata, vergeleken met  $\sim 16^\circ$  voor de *MediaPipe baseline*. De POF-uitbreiding is echter nog niet volledig geïntegreerd in de productie-*pipeline*.<sup>1</sup>

De *baseline pipeline* is functioneel en voldoet aan de projectdoelstellingen. De POF-uitbreiding toont veelbelovende resultaten maar vereist verdere integratie voordat de nauwkeurigheidswinst beschikbaar is voor eindgebruikers. Het systeem is geschikt voor *screening* en *monitoring* voor kritische beslissingen in de zorg blijft aanvullende validatie noodzakelijk.

---

<sup>1</sup> Voor technische details en integratiestatus: zie sectie 5.5. Voor een wiskundige formulering van de POF-aanpak: zie sectie 5.4.

# Inhoudsopgave

Inleiding.....	6
1.1 Aanleiding.....	6
1.2 Klantwensen .....	6
1.3 Probleemstelling .....	7
1.4 Opdrachtomschrijving .....	7
1.5 Terminologie .....	8
Methode.....	10
2.1. Technische Aanpak .....	10
2.2. Validatiestrategie .....	11
Alternatieve methodes en inspiratie.....	12
3.1 Overzicht Bestaande Oplossingen.....	12
3.2 <i>Marker-Based Motion Capture</i> .....	12
3.3 Beperkingen voor zorgtoepassingen.....	12
3.4 MediaPipe Pose.....	13
3.5 VideoPose3D.....	13
3.6 ElePose.....	13
3.7 Pose2Sim.....	14
3.8 Kennislacune.....	14
Eisen en Randvoorwaarden.....	15
4.1 Kernvereisten.....	15
4.2 Functionele Eisen.....	15
4.3 Niet-Functionele Eisen.....	15
4.4 Input/Output Specificaties.....	16
4.5 Realisatie .....	16
Bevindingen en Ontwerp/Ontwikkelkeuzes.....	17
5.1 Architectuurkeuzes .....	17
5.1.1. Waarom Modulaire Architectuur? .....	17
5.1.2. Waarom Part Orientation Fields (POF)?.....	17
5.1.3. Waarom Pose2Sim LSTM? .....	17
5.1.4. Waarom Neurale Joint Constraints?.....	17
5.1.5. Applicatie ontwerp .....	18
5.2.1. Pipeline Evolutie.....	19
5.2.2. Van Depth Refinement naar POF.....	19

5.3 Technische Afwegingen.....	19
5.4 Wiskundige Formulering POF.....	20
5.4.1 Probleemstelling: Monoculaire 3D Reconstructie.....	20
5.4.2 Orthografische Projectie Model .....	20
5.4.3 Part Orientation Fields (POF) Representatie.....	20
5.4.4 Neurale Netwerk Voorspelling.....	21
5.4.5 <i>Least-Squares</i> Diepte <i>Solver</i> .....	21
5.4.6 Hiërarchische Oplosvolgorde .....	22
5.4.7 <i>Bone Length Constraints</i> .....	22
5.4.8 Metrische Schaal Herstel .....	23
5.5 Onopgeloste Technische Uitdagingen.....	23
5.5.1 POF Reconstructie Instabiliteit .....	23
5.5.2 Temporale Context Faalt bij <i>Inference</i> .....	24
5.5.3 Camera Space vs World Space .....	24
5.5.4 Dataset Bias: Alleen Dansbewegingen .....	24
5.5.5 Joint Constraint Model: Afhankelijkheid van POF .....	25
5.6 Visualisaties .....	26
Pipeline Flowchart .....	26
Marker Progressie .....	27
Gewichtshoek Visualisatie .....	27
Resultaten .....	28
6.1 Baseline Pipeline .....	28
6.2 POF Model Performance .....	28
6.3 Joint Constraint Model Status .....	29
6.4 Vergelijking met Eisen .....	29
6.5 Visualisaties .....	30
3D Skelet Reconstructie .....	30
MediaPipe vs POF Vergelijking .....	30
Output Structuur .....	31
Discussie .....	32
Conclusie .....	33
8.1 Beantwoording Onderzoeksraag .....	33
8.2 Kernresultaten .....	33
8.3 Eindconclusie .....	34

Aanbevelingen.....	35
9.1 Technische Prioriteiten .....	35
9.2 Dataset Uitbreidig .....	35
9.3 Gebruikeronderzoek.....	36
9.4 Interpretatielaag .....	36
9.5 Van demo naar productie .....	36
9.6 Marktintroductie .....	37
Bibliografie .....	38
10.1 Core Methods & Architecture .....	38
10.2 ISB Biomechanics Standards.....	38
10.3 Pose Detection & Marker Augmentation.....	38
10.4 Monocular 3D Pose Estimation.....	39
10.5 Survey Articles .....	39
10.6 Anthropometric References.....	39
10.7 Datasets.....	40
Bijlagen .....	41
Bijlage A: Volledig Programma van Eisen.....	41
A.1 Frontend Requirements.....	41
A.2 Backend Requirements .....	41
A.3 Pipeline Requirements .....	42
A.4 Performance Requirements.....	42
Bijlage B: Module Overzicht .....	42
C.1 TRC Header (voorbeeld) .....	43
C.2 Joint Angles CSV (voorbeeld).....	43

# Hoofdstuk 1

## Inleiding

### 1.1 Aanleiding

Bewegingsanalyse speelt een cruciale rol in de gezondheidszorg, van revalidatie en fysiotherapie tot sportprestatie-optimalisatie. De huidige gouden standaard (i.e., *marker-based motion capture* systemen zoals Vicon en OptiTrack) biedt op de submillimeter nauwkeurigheid. Hier staan echter hoge kosten (€50.000+) en met praktische beperkingen tegenover. Zo moeten markers fysiek op het lichaam worden geplaatst en beschikt slechts een beperkte groep gespecialiseerde laboratoria over de apparatuur om dergelijke bewegingsanalyses te maken.

Consumentgerichte *monocular markerless* oplossingen zoals *MediaPipe* (Lugaresi et al., 2019) bieden een laagdrempelig alternatief. Deze systemen kunnen *pose-landmarks* detecteren vanuit reguliere video, maar leveren onbetrouwbare 3D diepte-informatie met een gemiddelde fout van ongeveer 16°. Dit beperkt hun bruikbaarheid in de gezondheidszorg voor toepassingen waar accurate gewrichtshoeken essentieel zijn.

### 1.2 Klantwensen

In afstemming met de opdrachtgever zijn de volgende kernwensen geïdentificeerd:

- **Markerless:** Geen fysieke markers nodig op het lichaam van de patiënt
- **Betaalbaar:** Moet werken met standaard hardware (smartphone en laptop)
- **Toegankelijk:** Gebruiksvriendelijke desktopapplicatie, geschikt voor niet-technische gebruikers
- **Gestandaardiseerd:** Gewrichtshoeken volgens ISB-standaarden (Wu et al., 2002, 2005)
- **Privacy-bewust:** Volledig lokale verwerking, geen cloud-upload vereist
- **Monocular:** input vanuit 1 camera

### **1.3 Probleemstelling**

*MediaPipe Pose* levert accurate 2D *landmark*-detectie, maar de ingebouwde 3D diepschatting vertoont systematische fouten. Dit komt vooral voor bij ledematen die van de camera af of naar de camera toe bewegen. In de academische literatuur betreffende dit onderwerp wordt hiernaar verwezen als het '*foreshortening*-probleem'. Een arm die naar de camera wijst lijkt in 2D kort, maar of deze naar voren of naar achteren beweegt is door *MediaPipe Pose* niet betrouwbaar te achterhalen.

#### **Centrale onderzoeksvraag:**

Hoe kunnen we betrouwbare 3D pose-reconstructie bereiken vanuit monoculaire smartphone-video, met nauwkeurigheid die geschikt is voor bewegingsanalyse in de gezondheidszorg?

### **1.4 Opdrachtomschrijving**

Het doel van dit project is het ontwikkeling van een toegankelijke desktop-applicatie voor *markerless* bewegingsanalyse. De applicatie moet video-input kunnen verwerken tot biomechanisch bruikbare gewrichtshoeken, zonder dat de gebruiker technische kennis nodig heeft.

De technische aanpak combineert bestaande open-source componenten met eigen ontwikkeling:

- **Pose-detectie:** *MediaPipe* (Lugaresi et al., 2019) of *RTMPose* voor 2D/3D *landmark*-extractie
- **Marker-augmentatie:** *Pose2Sim* (Pagnon et al., 2021, 2022) voor uitbreiding naar 64 biomechanische markers
- **Gewrichtshoeken:** *ISB-compliant* berekening volgens Wu et al. (2002, 2005)
- **Diepte-verbetering:** Experimentele POF-module gebaseerd op Xiang et al. (2019) voor toekomstige nauwkeurigheidsverbetering
- **Gewrichtshoek-correctie:** *Neural Joint Constraint* model dat zachte correcties leert in plaats van harde anatomische limieten, essentieel voor detectie van abnormale bewegingspatronen in de gezondheidszorg

De *baseline pipeline*, zonder POF en *Joint Constraint*, is functioneel en vormt het kernproduct. De experimentele modules zijn ontwikkeld als basis voor een volgende versie met verbeterde diepte-reconstructie en gewrichtshoek-correctie, maar vereisen nog verdere integratie voordat deze productierijp zijn.

## 1.5 Terminologie

In dit rapport worden diverse technische termen gehanteerd. Onderstaand worden de belangrijkste begrippen gedefinieerd om eenduidigheid te waarborgen. In de rest van dit rapport zal naar het Monocular Markerless 3D Pose Analyse Systeem gerefereerd worden als het **MMPA-systeem**.

Augmentatie verwijst naar het proces waarbij een beperkte set gedetecteerde punten wordt uitgebreid naar een grotere set. In dit project wordt de term specifiek gebruikt voor de uitbreiding van 22 gedetecteerde landmarks naar 64 biomechanische markers middels een LSTM-netwerk.

Baseline is de referentie-implementatie waartegen verbeteringen worden afgemeten. In dit rapport betreft dit de combinatie van MediaPipe pose-detectie met Pose2Sim marker-augmentatie zonder aanvullende diepte-correctie.

Biomechanisch betreft de toepassing van mechanische principes op biologische systemen. In de context van dit rapport verwijst dit naar de analyse van menselijke beweging volgens anatomische en fysiologische conventies.

Harde en zachte limieten beschrijven twee benaderingen voor het begrenzen van gewichtshoeken. Harde limieten kappen waarden direct af op een vaste grens, terwijl zachte limieten geleidelijke correcties toepassen die extreme waarden ontmoedigen zonder ze volledig te elimineren.

Inference is het proces waarbij een getraind neuraal netwerk voorspellingen genereert op nieuwe, ongeziene data. Dit in tegenstelling tot de trainingsfase waarin het model leert van gelabelde voorbeelden.

ISB-standaarden verwijzen naar de richtlijnen van de International Society of Biomechanics voor het definiëren van gewichtscoördinatensystemen en bewegingsbeschrijvingen (Wu et al., 2002, 2005). Deze standaarden waarborgen internationale vergelijkbaarheid van biomechanische data.

Least-squares solver is een wiskundige optimalisatiemethode die de beste oplossing vindt door de som van gekwadrateerde afwijkingen te minimaliseren. In dit project wordt deze methode gebruikt om 3D-reconstructies consistent te maken met 2D-observaties.

Open source betekent dat de broncode vrij beschikbaar is voor gebruik, aanpassing en verspreiding. Voor de definitie wordt de Open Source Initiative definitie gehanteerd, die vereist dat software vrij herdistribueerbaar is en de broncode toegankelijk.

Part Orientation Fields (POF) is een methode voor 3D pose-schatting waarbij voor elk ledemaat een oriëntatievector relatief aan de camera in de 3D-ruimte wordt voorgespeld (Xiang et al., 2019).

Real-time verwijst naar verwerking waarbij de output beschikbaar is binnen een tijdsframe dat directe interactie mogelijk maakt. In de context van pose-detectie betekent dit doorgaans verwerking met minimaal 15-30 frames per seconde.

Semantic Graph Convolutions (SemGCN) is een neural network architectuur die expliciet gebruik maakt van de grafenstructuur van het menselijk skelet (Zhao et al., 2019). De gewichten vormen knopen en de botten vormen verbindingen, waardoor het model anatomische relaties kan benutten.

Temporale context verwijst naar het meenemen van informatie uit meerdere opeenvolgende frames bij het maken van voorspellingen. Dit in tegenstelling tot per-frame analyse waarbij elk beeld onafhankelijk wordt verwerkt.

Transformer model is een neural network architectuur gebaseerd op het attention-mechanisme (Vaswani et al., 2017). Transformers kunnen lange-afstandsrelaties in data modelleren en worden breed toegepast in zowel natuurlijke taalverwerking als computervisie.

## Hoofdstuk 2

# Methode

Het systeem is ontwikkeld volgens een iteratieve aanpak. Modules zijn gebouwd met directe visuele feedback: 3D skelet-visualisaties maken fouten direct zichtbaar (e.g., verkeerde diepten, gespiegelde ledematen, onrealistische hoeken). Deze snelle visuele feedback-loop stuurt de ontwikkeling. De ontwikkeling volgde deze kernprincipes:

- **Modulaire architectuur:** Scheiding van concerns tussen modules waar praktisch
- **Visuele validatie:** 3D reconstructies direct beoordeeld op anatomische plausibiliteit
- **Resultaat gerichte ontwikkeling:** Werkende code eerst, documentatie volgde

### 2.1. Technische Aanpak

De *codebase* is modulair opgezet om het doen van experimenten mogelijk te maken. Componenten kunnen onafhankelijk worden vervangen (e.g., het POF-model wisselen zonder pipeline-aanpassingen aan te brengen). Dit is om praktische redenen noodzakelijk gezien de experimentele aard van het project. De *pipeline* is opgebouwd uit componenten die sequentieel data transformeren:

Video → MediaPipe → POF 3D → TRC → Pose2Sim → Joint Angles → Output

## 2.2. Validatiestrategie

Validatie vond plaats op meerdere niveaus:

De primaire validatiemethode berust op visuele inspectie: fouten in 3D reconstructie (e.g., verkeerde diepte, gespiegelde ledematen) zijn met het blote oog waarneembaar. De nauwkeurigheid van de berekende gewrichtshoeken is kwalitatief gevalideerd door de output aan de opdrachtgever voor te leggen, die heeft geverifieerd dat de waarden overeenkomen met verwachte bewegingspatronen.

Een formele kwantitatieve validatie tegen *marker-based ground truth* in een klinische setting is niet uitgevoerd. Dergelijke validatie vereist toegang tot een bewegingslaboratorium met optische motion capture-apparatuur, wat buiten de scope en middelen van dit project viel. Dit blijft een aanbeveling voor vervolgonderzoek (zie Discussie).

### Validatie overzicht

Niveau	Methode	Toelichting
<b>POF model</b>	Kwantitatief ( <i>train/val loss</i> )	~7° error op AIST++ validatie set
<b>3D reconstructie</b>	Visuele inspectie	Fouten direct zichtbaar in <i>3D viewer</i>
<b>Gewrichtshoeken</b>	Validatie met opdrachtgever	Vergelijking met verwachte bewegingspatronen
<b>End-to-end</b>	Praktijktest	Verwerking van eigen opgenomen video's

# Alternatieve methodes en inspiratie

Drie dimensionale bewegingsanalyse vereist het nauwkeurig reconstrueren van lichaamsposities in de ruimte. Traditioneel wordt dit gerealiseerd met marker-based motion capture systemen, die weliswaar zeer nauwkeurig zijn maar ook duur en onpraktisch voor veel zorgtoepassingen. De afgelopen jaren zijn diverse markerless alternatieven ontwikkeld die bewegingsanalyse toegankelijker maken. Dit hoofdstuk bespreekt de belangrijkste bestaande oplossingen, hun sterke en zwakke punten, en identificeert de kennislacune die dit project adresseert.

## 3.1 Overzicht Bestaande Oplossingen

Oplossing	Type	Sterkte	Zwakte
Vicon/OptiTrack	Marker-based	Gouden standaard nauwkeurigheid (<1mm)	Duur (€50K+), markers nodig, lab-setting
MediaPipe Pose (Lugaresi et al., 2019)	Markerless, monocular	Gratis, <i>real-time</i> , 33 landmarks	Onbetrouwbare 3D diepte-schatting
VideoPose3D (Pavllo et al., 2019)	2D-to-3D lifting	Temporale consistentie, <i>state-of-the-art</i>	Vereist voorgetraind 2D model
ElePose (Wandt et al., 2022)	Unsupervised 3D	Geen gelabelde data nodig	Beperkte anatomische constraints
Pose2Sim (Pagnon et al., 2022)	Multi-view workflow	Open-source, biomechanisch model	Vereist meerdere camera's

## 3.2 Marker-Based Motion Capture

Systemen zoals Vicon en OptiTrack gebruiken infrarood camera's om reflecterende markers te volgen. Deze aanpak biedt submillimeter nauwkeurigheid en wordt beschouwd als de gouden standaard voor biomechanische analyse.

## 3.3 Beperkingen voor zorgtoepassingen

Hoge aanschafkosten (€50.000 - €500.000), vereist gespecialiseerde ruimte, *markers* kunnen beweging beïnvloeden en is niet geschikt voor thuisgebruik of veldmetingen.

### **3.4 MediaPipe Pose**

Google's MediaPipe (Lugaresi et al., 2019) biedt *real-time pose-detectie* vanuit enkele camerabeelden. Het systeem detecteert 33 landmarks en levert zowel 2D- als 3D-coördinaten. Het systeem is gratis, open-source, draait *real-time* op CPU en biedt robuuste 2D-detectie.

De 3D diepte-schatting is echter gebaseerd op statistische aannames uit trainingsdata in plaats van werkelijke diepte-informatie. Het model fit de gedetecteerde 2D-poses naar een intern mensmodel met een aangenomen standaardhoogte, wat leidt tot systematische fouten bij bewegingen in de diepterrichting. In onze tests resulteerde dit in een gemiddelde oriëntatiefout van 16.3°. Daarnaast kent MediaPipe geen anatomische constraints, waardoor fysiologisch onmogelijke gewichtsposities kunnen voorkomen.

### **3.5 VideoPose3D**

Pavllo et al. (2019) presenteren een methode die temporale convoluties gebruikt om 2D poses te *liften* naar 3D. Door meerdere frames te analyseren wordt temporale consistentie afgedwongen. Dit is van meerwaarde voor het door ons ontwikkelde systeem omdat het waarde demonstreert van temporale context. Verder is *VideoPose3D* getraind aan de hand van een *Human3.6M* dataset (Ionescu et al., 2014).

### **3.6 ElePose**

Het onderzoek van Wandt et al. (2022) introduceert ElePose, een unsupervised methode voor 3D pose-schatting. Waar de meeste 3D pose-modellen afhankelijk zijn van grote datasets met gelabelde 3D ground truth, beoogt ElePose dit te omzeilen door camera-elevaties te voorspellen en probabilistische modellen (normalizing flows) te trainen op uitsluitend 2D-posedata. Dit maakt training mogelijk zonder kostbare 3D-annotaties.

Het model kent echter beperkingen voor onze toepassing. ElePose schat een globale camera-oriëntatie, maar lost daarmee niet het per-segment diepteprobleem op: de vraag of individuele ledematen naar voren of naar achteren bewegen ten opzichte van de romp blijft onbeantwoord. Juist deze per-segment diepte-ambigüiteit is de kernuitdaging bij monoculaire bewegingsanalyse.

### **3.7 Pose2Sim**

Het onderzoek van Pagnon et al. (2022) biedt een open-source workflow voor multi-camera markerless kinematics. De pipeline omvat marker-augmentatie via een LSTM-netwerk dat 22 gedetecteerde landmarks uitbreidt naar 64 biomechanische markers, en integratie met OpenSim voor inverse kinematics en gewrichtshoekberekening.

Hoewel Pose2Sim oorspronkelijk is ontworpen voor multi-camera opstellingen, is de marker-augmentatiemodule ook toepasbaar op monoculaire input. In het door ons ontwikkelde systeem wordt deze module ingezet om de MediaPipe-landmarks uit te breiden naar een volledige biomechanische marker-set.

### **3.8 Kennislacune**

Geen van de bestaande oplossingen combineert monoculaire input (zoals een enkele smartphone-camera) met accurate per-segment 3D-reconstructie, een volledige biomechanische marker-set én ISB-compliant gewrichtshoeken. Dit project beoogt deze lacune te adresseren.

## Hoofdstuk 4

# Eisen en Randvoorwaarden

Het MMPA-systeem is ontwikkeld voor bewegingsanalyse in de gezondheidszorg. Binnen de zorgsector worden verschillende randvoorwaarden opgelegd aan dergelijke systemen. Zo vereist AVG/GDPR-wetgeving dat er secuur wordt omgegaan met zaken omtrent patiënt privacy. Videodata moet dan ook zonder een *cloud-upload* lokaal verwerkt worden. De doelgroep dat wij beogen te ondersteunen zijn zorgprofessionals zonder IT-achtergrond. Het MMPA-systeem moet dus op gebruiksvriendelijke wijze ontwikkeld worden en moet kunnen draaien op standaard hardware dat beschikbaar is in alle zorginstellingen.

Voor toekomstige inzet van het MMPA-systeem als medisch hulpmiddel is certificering volgens de *Medical Device Regulation* (MDR) een vereiste. De huidige versie betreft vooralsnog een onderzoekstool en valt hier zodoende nog niet onder.

### 4.1 Kernvereisten

Vereiste	Specificatie	Status
Lokale verwerking	Volledig op laptop, geen cloud-upload	✓
Webapp interface	Gebruiksvriendelijke GUI voor niet-technische gebruikers	✓
ISB-compliant output	Gewrichtshoeken volgens Wu et al. (2002, 2005)	✓
Standaard hardware	Werkt op gemiddelde laptop (geen dedicated GPU vereist)	✓
Eenvoudige input	Enkele video van smartphone-kwaliteit (720p+)	✓

### 4.2 Functionele Eisen

ID	Eis	Prioriteit	Status
F1	Video upload via webapp	Must	✓
F2	3D pose reconstructie uit 2D keypoints	Must	✓
F3	Marker augmentatie (22 → 64 markers)	Must	✓
F4	Gewrichtshoeken berekening (12 gewrichtsgroepen)	Must	✓
F5	TRC export voor interoperabiliteit met OpenSim/Visual3D	Must	✓
F6	Visualisatie van gewrichtshoeken (CSV + grafieken)	Should	✓

### 4.3 Niet-Functionele Eisen

ID	Eis	Specificatie	Status
N1	Verwerkingstijd	< 2 minuten per video	✓ (~60s)
N2	Geheugengebruik	< 8 GB RAM	✓
N3	POF nauwkeurigheid	< 15° limb-oriëntatiefout	✓ (~7° gerealiseerd)
N4	GPU fallback	Automatische CPU-modus indien geen GPU	✓

#### **4.4 Input/Output Specificaties**

**Input:** Videoformaten: MP4, AVI, MOV (standaard codecs), minimale kwaliteit: 720p (smartphone-camera voldoende), Perspectief: frontaal of schuin aanzicht, belichting: normale binnenverlichting.

**Output:** TRC-bestanden (Track Row Column) voor biomechanica-software, CSV-bestanden met gewrichtshoeken per frame, PNG-visualisaties van gewrichtshoekverloop, Ruwe landmark-data voor debugging en verder onderzoek.

#### **4.5 Realisatie**

Alle must-have eisen zijn gerealiseerd. Het systeem draait lokaal op een gemiddelde laptop, biedt een webapp interface en genereert ISB-compliant gewrichtshoeken. De experimentele POF-module overtreft de gestelde nauwkeurigheidseis ruimschoots ( $\sim 7^\circ$  versus de eis van  $< 15^\circ$ ), al vereist volledige productie-integratie nog aanvullend werk zoals beschreven in sectie 5.5.

## Hoofdstuk 5

# Bevindingen en Ontwerp/Ontwikkelkeuzes

Dit hoofdstuk beschrijft de belangrijkste architectuur en ontwerpkeuzes die zijn gemaakt tijdens de ontwikkeling van het MMPA-systeem. Per onderdeel wordt toegelicht welk probleem de keuze adresseert en waarom voor deze specifieke oplossing is gekozen.

## 5.1 Architectuurkeuzes

### 5.1.1. Waarom Modulaire Architectuur?

De pipeline is opgebouwd uit onafhankelijke modules die via gedefinieerde interfaces communiceren. Deze modulaire opzet biedt drie belangrijke voordelen. Ten eerste verhoogt het de onderhoudbaarheid, aangezien wijzigingen in één module andere modules niet beïnvloeden. Ten tweede verbetert het de testbaarheid doordat elke module afzonderlijk getest kan worden. Ten derde vergroot het de herbruikbaarheid, waardoor modules in toekomstige projecten kunnen worden ingezet.

### 5.1.2. Waarom Part Orientation Fields (POF)?

Het kernprobleem met monoculaire 3D is diepte-ambiguïteit. Een verkort ledemaat in 2D kan zowel naar de camera gericht zijn als van de camera af. *MediaPipe* (Lugaresi et al., 2019) lost dit op met heuristieken die vaak falen.

POF-aanpak (Xiang et al., 2019): Voorspel 3D oriëntatievector per ledemaat (14 *limbs*), Neural network leert diepte-richting uit 2D *appearance en temporale context*, *Least-squares solver* zorgt voor consistentie met 2D observaties

Dit is fundamenteel anders dan globale camera-hoek voorspelling zoals *ElePose* (Wandt et al., 2022), omdat elk ledemaat onafhankelijk zijn diepte-richting krijgt.

### 5.1.3. Waarom Pose2Sim LSTM?

De *marker-augmentatie* van *Pose2Sim* (Pagnon et al., 2022) is bewezen technologie: Expandeert 22 markers naar 64 biomechanische markers, GPU-geaccelereerd voor snelle verwerking, Getraind op grote *motion capture* datasets

### 5.1.4. Waarom Neurale Joint Constraints?

Een eerste aanpak voor gewrichtshoek validatie was het toepassen van harde anatomische limieten: een elleboog kan niet hyperextenderen voorbij 0°, een knie niet verder dan ~150° flexie. Dit bleek ongeschikt voor de beoogde zorgtoepassingen.

**Het fundamentele probleem:** In de gezondheidszorg zijn juist de afwijkende bewegingspatronen relevant. Hypermobiliteit, compensatiestrategieën bij pijn, of pathologische bewegingspatronen overschrijden per definitie de “normale” anatomische limieten. Een systeem dat deze afkapt op vaste grenzen verliest precies de informatie die zorgprofessionals nodig hebben.

**De neurale oplossing:** In plaats van harde *constraints* wordt een *SemGCN*-model (vergelijkbaar met het POF-model) gebruikt dat zachte correcties leert. Het model vergelijkt gewrichtshoeken uit de *pipeline* met *ground truth* uit *motion capture* data en leert de systematische afwijkingen te voorspellen. Dit behoudt de mogelijkheid om abnormale bewegingen te detecteren terwijl structurele fouten worden gecorrigeerd.

**Architectuur:** SemGCN met *cross-joint attention* (916K parameters), Input: 12 gewichtsgroepen × 3 Euler-hoeken (flexie, abductie, rotatie), Output: correctie-delta per gewrichtshoek, Getraind op AIST++ *motion capture* data (660K samples)

**Afhankelijkheid van POF:** Het *Joint Constraint* model kan pas effectief getraind worden wanneer de basis 3D-reconstructie voldoende nauwkeurig is. Met de huidige *MediaPipe baseline* ( $\sim 16^\circ$  error) is de ruis te hoog om zinvolle correcties te leren. Het model zou voornamelijk ruis voorspellen. De POF-module ( $\sim 7^\circ$  error) biedt de benodigde basis, maar de volledige integratie is nog niet afgerond. Dit maakt het *Joint Constraint* model een essentieel onderdeel van de volgende versie.

### 5.1.5. Applicatie ontwerp

De applicatie omvat zich in een webinterface waarin het MMPA-systeem via drie pagina's: upload, resultaten en analyse. De applicatie draait op Django met Bootstrap voor styling.

Op de uploadpagina selecteert de gebruiker een video en configureert verwerkingsparameters. Eerste gebruik toont instructies voor opname, privacybeleid en eventuele modeldownloads. De 15+ beschikbare parameters: landmark-visualisatie, LSTM-iteraties, analysetoggles, zijn standaard ingeklappt onder "Advanced settings" om de interface toegankelijk te houden. Tijdens verwerking geeft een overlay met voortgangsbalk real-time feedback.

De resultatenpagina presenteert de originele video naast een skelet-overlay preview, met daaronder een downloadtabel voor alle gegenereerde bestanden: .trc motion capture data, grafieken en CSV-exports.

De analysepagina combineert interactieve grafieken met video en 3D-visualisatie. Via drie dropdowns selecteert de gebruiker markers of gewrichtshoeken; voor elke selectie toont een Chart.js-grafiek de vrijheidsgraden (flexie/extensie, abductie/adductie, rotatie) over tijd met real-time numerieke waarden. Naast de grafieken staan een videospeler en Plotly.js-skeletweergave met rotatiebesturing—de actieve marker licht op in het 3D-model. Een timeline-dock onderaan houdt alle componenten frame-accuraat gesynchroniseerd.

De architectuur gebruikt bewust vanilla JavaScript—geen buildproces, minimale overhead, eenvoudig onderhoud. Server-polling (per seconde) vervangt WebSockets omwille van simpelheid. Alle verwerking blijft lokaal, essentieel voor privacy bij medische beelddata.

### 5.2.1. Pipeline Evolutie

Chronologisch overzicht van belangrijke beslissingen:

Datum	Beslissing	Reden
2025-11	MediaPipe + Pose2Sim basis	Snelle prototype voor proof-of-concept
2026-01-10	Pelvis angle fixes	Validatie toonde 180° flips bij bepaalde poses
2026-01-13	Neural refinement toegevoegd	Structurele fouten vereisten lerende correctie
2026-01-15	Harde constraints → neuraal	Zorgtoepassingen vereisen detectie van abnormale bewegingen
2026-01-21	POF-only architectuur	Depth module verwijderd; POF simpeler en nauwkeuriger

### 5.2.2. Van Depth Refinement naar POF

De oorspronkelijke aanpak gebruikte een *depth delta* correctie: voorspel hoeveel de *MediaPipe* diepte afwijkt en corrigeer. Dit bleek problematisch, *MediaPipe* diepte bevat systematische fouten die lastig te leren zijn, *Cumulatieve errors* bij slechte input

De POF-aanpak (Xiang et al., 2019) reconstrueert 3D volledig opnieuw vanuit 2D, waardoor *MediaPipe*'s dieptefouten worden omzeild.

## 5.3 Technische Afwegingen

Keuze	Alternatieven	Beslissing	Rationale
POF vs. Depth Refinement	Delta correcties	POF	~7° vs ~16° error
SemGCN vs. Transformer	MLP, CNN	SemGCN-Temporal	Anatomische bias + temporale context
Augmentation cycles	1, 5, 10, 20	20 cycles	Balans nauwkeurigheid/rekentijd
ISB standaard	Custom angles	ISB (Wu et al., 2002, 2005)	Interoperabiliteit in de gezondheidszorg
Z-sign classificatie	Regression only	Classification + regression	Expliciete diepte-richting supervisie
Harde vs. zachte joint limits	Anatomische clipping	Neurale correctie	Behoud detectie abnormale bewegingen

## 5.4 Wiskundige Formulering POF

Dit hoofdstuk beschrijft de wiskundige basis van de POF-reconstructie. Het doel is inzicht te geven in hoe 3D posities worden berekend uit 2D observaties en waarom bepaalde problemen optreden.

### 5.4.1 Probleemstelling: Monoculaire 3D Reconstructie

**Gegeven:** - 2D keypoints  $p_i$  voor  $i = 1, \dots, 17$  (COCO joints) - Lichaamslengte  $h$  in meters

**Gevraagd:** - 3D posities  $P_i$  in metrische schaal

**Fundamentele ambiguïteit:** Een 2D projectie correspondeert met oneindig veel 3D poses langs de kijkrichting. Zonder stereo-informatie is diepte onbepaald.

### 5.4.2 Orthografische Projectie Model

Bij typische opname-afstanden ( $>2m$ ) en menselijke afmetingen ( $\sim 0.5m$  dieptevariantie) is perspectief-vervorming klein. We approximeren:

$$x_{2d} \approx X_{3d}, \quad y_{2d} \approx Y_{3d}$$

**Implicatie:** X,Y uit 2D observaties zijn direct bruikbaar als X,Y in 3D. Alleen de Z-coördinaat (diepte) moet worden opgelost.

### 5.4.3 Part Orientation Fields (POF) Representatie

Voor elk ledemaat  $l$  (14 limbs totaal) definiëren we de oriëntatievector als een eenheidsvector die de 3D richting aangeeft:

$$\mathbf{o}_l = \frac{P_{child} - P_{parent}}{\|P_{child} - P_{parent}\|}$$

**Decompositie:** We splitsen in XY-vlak component en Z-component:

$$\mathbf{o} = (o_x, o_y, o_z), \quad o_x^2 + o_y^2 + o_z^2 = 1$$

#### 5.4.4 Neurale Netwerk Voorspelling

Het SemGCN-Temporal model voorspelt twee outputs per *limb*:

1. **Z-magnitude:**  $|z| \in [0,1]$  — hoeveel het ledemaat naar/van camera wijst ( $0 =$  zijwaarts,  $1 =$  recht naar camera)
2. **Z-sign:**  $s_z \in \{-1, +1\}$  — naar camera (-1) of van camera af (+1)

Oriëntatie reconstructie uit 2D observatie  $\delta_{2d}$  en netwerk outputs:

Uit de unit vector constraint  $|o_{xy}|^2 + z^2 = 1$  volgt:

$$|o_{xy}| = \sqrt{1 - |z|^2}$$

De XY-richting komt uit de 2D observatie (betrouwbaar):

$$o_{xy} = |\delta_{2d}| \cdot \frac{\delta_{2d}}{|\delta_{2d}|}$$

De Z-component uit het model (moet worden geleerd):

$$o_z = s_z \cdot |z|$$

**Cruciaal inzicht:** De XY-richting is geometrisch bepaald uit 2D, maar de Z-component (diepte-richting) kan niet geometrisch worden afgeleid — het model moet dit leren uit visuele context (i.e., *foreshortening cues*).

#### 5.4.5 Least-Squares Diepte Solver

**Probleem:** Gegeven oriëntatievector  $\mathbf{o}_l$  en 2D observatie  $\delta_{2d}$ , vind de schaalfactor  $\lambda$  zodanig dat de 3D displacement:

$$\lambda \cdot \mathbf{o}_l = (\Delta x, \Delta y, \Delta z)$$

projecteert naar de 2D observatie:  $(\Delta x, \Delta y) = \delta_{2d}$

**Least-squares oplossing via normale vergelijking:**

$$\lambda^* = \frac{o_{xy} \cdot \delta_{2d}}{|o_{xy}|^2}$$

Dit minimaliseert de projectie-error:  $|\lambda \cdot o_{xy} - \delta_{2d}|^2$

**Diepte propagatie door de keten:**

$$Z_{child} = Z_{parent} + \lambda^* \cdot o_z$$

**Edge case:** Wanneer  $|o_{xy}| < 0.05$  (ledemaat wijst recht naar camera), is de 2D observatie onbepaald. Fallback: gebruik botlengte als schaal.

#### 5.4.6 Hiërarchische Oplosvolgorde

De solver propageert diepte door de kinematische keten in vaste volgorde:

pelvis (Z = 0, referentie)  
 hip\_width → R\_hip Z offset (kritiek voor side-view!)  
 L\_torso → L\_shoulder  
 R\_torso → R\_shoulder  
 upper\_arm → elbow → forearm → wrist  
 L\_thigh → L\_knee → L\_shin → L\_ankle  
 R\_thigh → R\_knee → R\_shin → R\_ankle

#### 5.4.7 Bone Length Constraints

Botlengtes worden geschat als fracties van lichaamslengte, gebaseerd op antropometrische referentiedata (Winter, 2009; NASA, 1978; Drillis & Contini, 1966).

Botlengtes als fractie van lichaamslengte:

Segment	Proportie
Bovenarm	0.172
Onderarm	0.157
Bovenbeen	0.245
Onderbeen	0.246
Torso-zijde	0.288
Schouder-breedte	0.259
Heup-breedte	0.191

**Constraint enforcement:** Gegeven bekende  $\Delta x, \Delta y$  uit 2D en gewenste botlengte  $L$ :

$$\Delta z = \pm \sqrt{L^2 - \Delta x^2 - \Delta y^2}$$

Teken bepaald door  $sign(o_z)$  uit POF. Alleen toegepast bij significante *foreshortening* ( $|\delta_{2d}| < 0.7L$ ).

#### 5.4.8 Metrische Schaal Herstel

Training vindt plaats in genormaliseerde ruimte (pelvis-gecentreerd, torso = 1). Herstel naar meters:

$$torso_{metric} = \frac{h}{3.4}$$

$$P_{metric} = P_{norm} \cdot torso_{metric} + p_{pelvis}$$

**Voordeel:** Lost monoculaire schaal-ambiguïteit op met alleen bekende lichaamslengte, zonder camera-*intrinsics*.

### 5.5 Onopgeloste Technische Uitdagingen

Ondanks de goede resultaten op validatiedata kent de POF-integratie nog fundamentele problemen die volledige productie-inzet blokkeren. Deze sectie documenteert de bekende problemen en mogelijke oplossingsrichtingen voor toekomstig werk.

#### 5.5.1 POF Reconstructie Instabiliteit

**Probleem:** Het POF-model bereikt  $\sim 7^\circ$  fout op validatiedata, maar de 3D reconstructie is in praktijk onstabiel.

**Wiskundige oorzaak:** De *solver* propageert fouten door de kinematische keten. Voor  $n$  segmenten:

$$Z_n = Z_0 + \sum_{i=1}^n \lambda_i \cdot o_{z,i}$$

Een kleine fout  $\epsilon$  in elke  $o_z$  accumuleert tot  $n \cdot \epsilon$  aan het einde van de keten.

**Symptomen:** Frame-to-frame flikkering (jitter), occasionele diepte-flips (ledemaat springt van voor naar achter), asymmetrische reconstructie (links vs. rechts).

**Mogelijke oplossingen (toekomstig):** Global skeleton optimization (alle joints simultaan), multi-hypothesis tracking, confidence-weighted depth propagation.

### 5.5.2 Temporele Context Faalt bij *Inference*

**Probleem:** Het SemGCN-Temporal model gebruikt POF van vorig frame als input, maar dit werkt niet bij inference.

**Training vs. Inference discrepantie:** Training: *ground truth* POF van frame  $t - 1$  - *Inference*: voorspelde (foutieve) POF van frame  $t - 1$ .

Dit is *exposure bias*, het model leert met perfecte temporale context, maar krijgt bij *inference* zijn eigen fouten terug. Fouten propageren en versterken.

**Mogelijke oplossingen:** *Scheduled sampling* (mix GT en predicted tijdens training) - Recurrent architectuur met *hidden state*.

### 5.5.3 Camera Space vs World Space

**Probleem:** POF reconstrueert in camera-space (Y omlaag), maar Pose2Sim LSTM verwacht world-space (Y omhoog).

Systeem	Y-as	Oorsprong
Camera-space (POF)	Omlaag	Camera
World-space (LSTM)	Omhoog	Ground plane

**Huidige workaround:** Pelvis-centering vóór augmentatie.

**Waarom onvoldoende:** - LSTM getraind met consistente gravity-richting, camera-space oriëntatie varieert met camera-hoek, pelvis-centering lost positie op, niet oriëntatie.

**Mogelijke oplossingen:** 1. LSTM hertrainen op camera-space data 2. Ground plane detectie → camera rotatie schatten 3. Gravity-vector voorspelling als extra model output.

### 5.5.4 Dataset Bias: Alleen Dansbewegingen

**Probleem:** Het POF-model is uitsluitend getraind op de AIST++ dataset (Li et al., 2021), bestaande uit 1.2M frames van professionele dansers.

**Implicaties:** Bewegingspatronen: snel/expressief vs. Dagelijks, posen: extreme houdingen overgerekend, demografie: jonge, fit dansers.

**Verwachte failure modes:** Revalidatie-oefeningen (langzaam, subtiel), ouderen met beperkte mobiliteit, zittende activiteiten

**Mogelijke oplossingen:** Dataset uitbreiding met zorg-specifieke bewegingen, validatiestudie in zorgsetting

### 5.5.5 Joint Constraint Model: Afhankelijkheid van POF

**Probleem:** Het Joint Constraint model is ontworpen om gewrichtshoek-correcties te leren, maar kan niet effectief getraind worden met de huidige MediaPipe baseline.

#### Wiskundige onderbouwing:

Het model leert de mapping  $f: \theta_{pipeline} \rightarrow \Delta\theta$  waar  $\Delta\theta = \theta_{GT} - \theta_{pipeline}$ . Voor effectief leren moet de fout  $\Delta\theta$  structureel en voorspelbaar zijn. Bij MediaPipe ( $\sim 16^\circ$  limb-oriëntatiefout) domineert stochastische ruis:

$$SNR = \frac{\text{systematische fout}}{\text{stochastische ruis}} \ll 1$$

Het model leert dan voornamelijk ruis te voorspellen—overfitting zonder generalisatie.

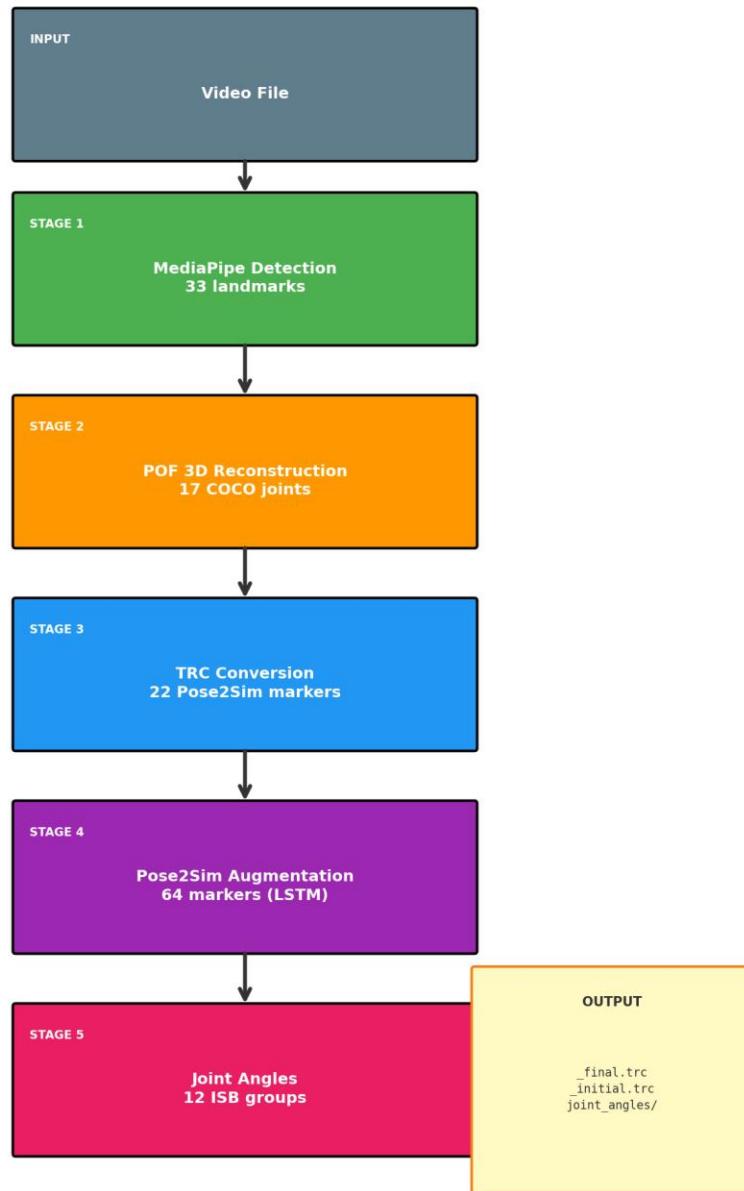
**Vereiste voor effectieve training:** POF-baseline met  $\sim 7^\circ$  error biedt voldoende SNR, systematische fouten (zoals Euler gimbal lock bij extreme hoeken) worden dan leerbaar, model kan zinvolle correcties generaliseren

**Huidige status:** Het joint constraint model is geïmplementeerd en getraind op AIST++ data (660K samples), maar kan pas gevalideerd worden na volledige POF-integratie. De theoretische basis is solide; de praktische effectiviteit moet nog worden aangetoond.

**Rationale voor zachte constraints:** Harde anatomische limieten zouden abnormale bewegingspatronen maskeren die juist klinisch relevant zijn. Het neurale model corrigeert systematische fouten terwijl het de mogelijkheid behoudt om hypermobilitéit, compensatiestrategieën, en pathologische bewegingen te detecteren.

## 5.6 Visualisaties

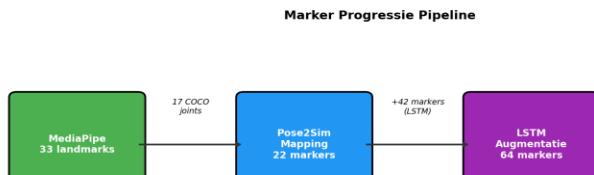
Pipeline Flowchart



Pipeline Flowchart

*Figuur 5.1: Schematisch overzicht van de verwerkingspipeline van video-input tot gewrichtshoek-output.*

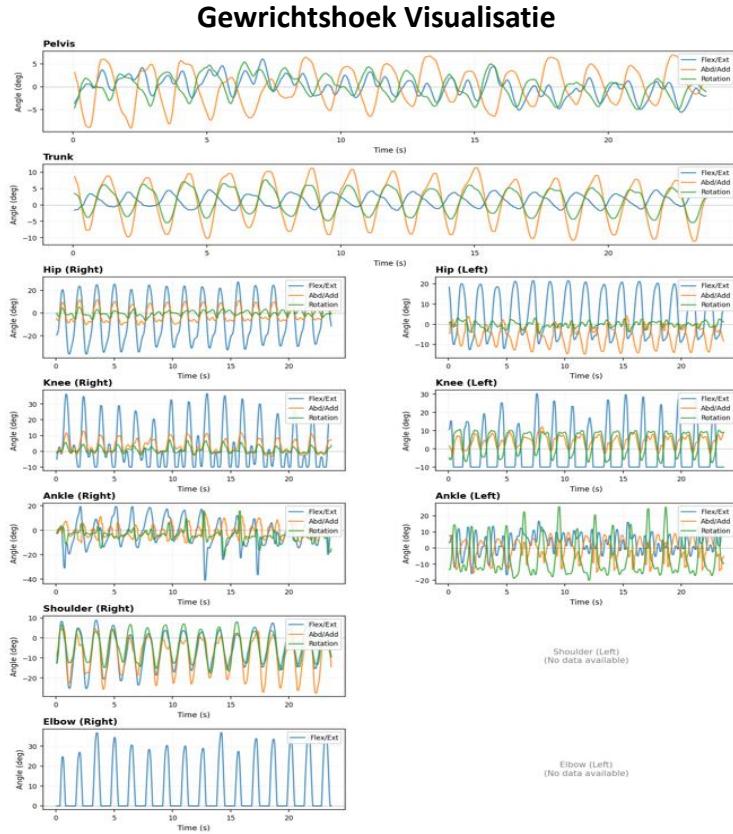
## Marker Progressie



33 MediaPipe landmarks → 22 Pose2Sim markers → 64 biomechanische markers

## Marker Progressie

Figuur 5.2: Progressie van 33 MediaPipe landmarks naar 64 biomechanische markers via de Pose2Sim LSTM-augmentatie.



## Joint Angles

Figuur 5.3: Voorbeeld van automatisch gegenereerde gewichtshoek-plots voor alle 12 ISB-compliant gewrichtsgroepen (linkerschouder en ellenboog waren bij input video niet goed zichtbaar en konden zo niet berekend worden).

## Hoofdstuk 6

# Resultaten

Dit hoofdstuk presenteert de resultaten van het ontwikkelde MMPA-systeem. Achtereenvolgens worden de prestaties van de baseline pipeline, het experimentele POF-model en het Joint Constraint model besproken. Afsluitend wordt een vergelijking gemaakt met de gestelde eisen uit het programma van eisen.

### 6.1 Baseline Pipeline

De functionele baseline pipeline, MediaPipe pose-detectie gecombineerd met Pose2Sim marker-augmentatie (Pagnon et al., 2021, 2022) levert consistente resultaten op standaard hardware. Een typische video van 535 frames wordt verwerkt in ongeveer 60 seconden, ruim binnen de gestelde eis van twee minuten. De verwerking vindt volledig lokaal plaats op een laptop zonder dedicated GPU, conform de privacy vereisten voor zorgtoepassingen.

De marker-augmentatie expandeert de initiële 22 Pose2Sim-compatibele markers naar gemiddeld 59 van de 64 mogelijke biomechanische markers. De ontbrekende worden voornamelijk veroorzaakt door onstabiele markers voorspelt door media pipe of RTM pose.

De gewichtshoekberekening volgt de ISB-standaard (Wu et al., 2002, 2005). De pipeline berekent twaalf gewichtsgroepen: pelvis, heup, knie en enkel, trunk en elleboog (links en rechts waar mogelijk). Elk gewicht levert drie Euler-hoeken volgens de anatomische conventies—flexie/extensie, abductie/adductie, en interne/externe rotatie—with uitzondering van de elleboog die als scharniergegewicht alleen flexie/extensie kent.

De nauwkeurigheid van de baseline wordt beperkt door MediaPipe's diepte-schatting, die een gemiddelde limb-oriëntatiefout van ongeveer 16° vertoont (Lugaresi et al., 2019). Deze fout is waar passend acceptabel voor screening en voortgangsmonitoring, maar onvoldoende voor nauwkeurige diagnostiek.

### 6.2 POF Model Performance

Het experimentele POF-model, gebaseerd op de Part Orientation Fields aanpak van Xiang et al. (2019), toont significante verbeteringen op validatiedata. De SemGCN-Temporal architectuur bereikt een gemiddelde limb-oriëntatiefout van ongeveer 7° meer dan een halvering ten opzichte van de MediaPipe baseline.

De training op de AIST++ dataset (Li et al., 2021), bestaande uit 1.2 miljoen frames van professionele dansers, verliep efficiënt. Het model bereikte al na één epoch een validatie-error van 7.62°, wat direct beter was dan de transformer-variant met 11° error. Na vijftig epochs stabiliseerde de performance rond 6° met een Z-sign classificatie-nauwkeurigheid van ongeveer 95%. Deze Z-sign voorspelt of een ledemaat naar de camera of van de camera af wijst, dit is cruciaal voor het oplossen van de diepte-ambiguïteit.

De modelarchitectuur is compact met ongeveer 1.7 miljoen parameters, waardoor inference snel genoeg is voor praktisch gebruik. De combinatie van Semantic Graph Convolutions met temporale context blijkt effectiever dan pure transformer-architecturen voor deze taak, vermoedelijk door de ingebouwde anatomische structuur van het skelet-graaf.

Een belangrijke kanttekening is dat deze resultaten zijn gemeten op data uit dezelfde distributie als de trainingsdata. Hoe het model presteert op zorg-specifieke bewegingen—langzame revalidatie-oefeningen, bewegingen van ouderen, of pathologische patronen—is nog niet gevalideerd. De AIST++ dataset bevat uitsluitend snelle, expressieve dansbewegingen van jonge, fitte performers.

### 6.3 Joint Constraint Model Status

Het Joint Constraint model is geïmplementeerd en getest op 660.000 samples uit de AIST++ dataset. Het model leert zachte correcties voor gewrichtshoeken in plaats van harde anatomische limieten toe te passen. Deze aanpak is essentieel voor zorgtoepassingen waar abnormale bewegingspatronen: hypermobilité, compensatiestrategieën, pathologische bewegingen, juist gedetecteerd moeten worden.

De effectiviteit van dit model kan echter nog niet worden gevalideerd. Zoals beschreven in sectie 5.5.5, vereist zinvolle training een basis met voldoende lage ruis. De MediaPipe baseline met  $\sim 16^\circ$  error produceert te veel stochastische variatie; het model zou voornamelijk ruis leren voorspellen in plaats van systematische correcties. Pas wanneer de POF-integratie is afgerond en de basis-error naar  $\sim 7^\circ$  is teruggebracht, kan het Joint Constraint model effectief worden getraind en gevalideerd.

### 6.4 Vergelijking met Eisen

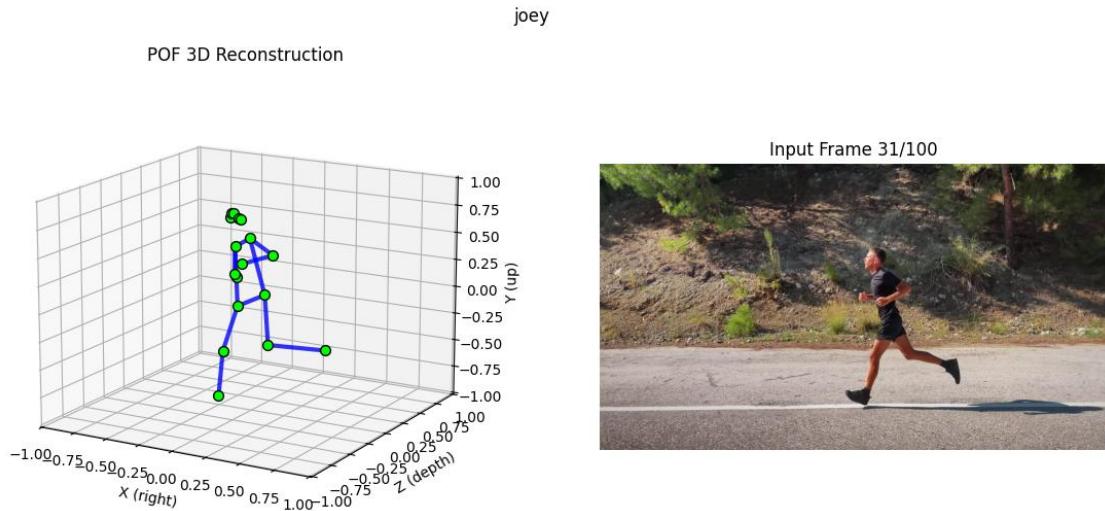
De baseline pipeline voldoet aan alle must-have eisen uit sectie 4. De verwerkingstijd van ongeveer 60 seconden blijft ruim onder de gestelde limiet van twee minuten. Het aantal gereconstrueerde markers gemiddeld 59 van 64, overtreft de minimumeis van 50. Alle twaalf gewichtsgroepen worden berekend volgens ISB-standaard.

De experimentele POF-module overtreft de gestelde nauwkeurigheidseisen ruimschoots:  $\sim 7^\circ$  gemeten error versus de eis van maximaal  $15^\circ$ . Deze resultaten zijn echter gemeten op validatiedata en niet in productie-omstandigheden. De volledige integratie van POF in de pipeline, en daarmee de realisatie van deze nauwkeurigheidswinst voor eindgebruikers, vereist nog aanvullend werk aan de camera-space naar world-space transformatie en temporale stabiliteit.

## 6.5 Visualisaties

De pipeline genereert automatisch visualisaties van de gewrichtshoeken over tijd (figuur 5.3). Voor elk gewricht wordt een grafiek geproduceerd met de drie Euler-componenten, samen met een CSV-bestand voor verdere analyse in externe software. De 3D skelet-reconstructie kan worden geïnspecteerd via een interactieve viewer die deel uitmaakt van de development tools.

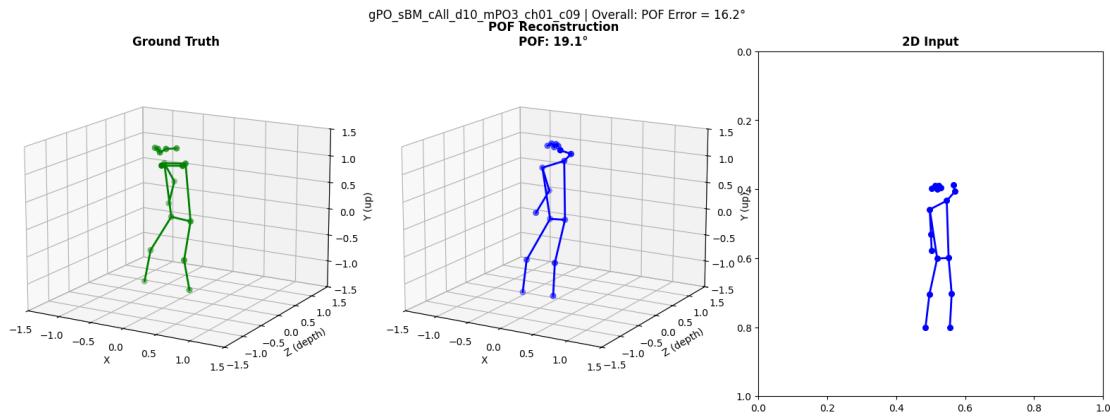
### 3D Skelet Reconstructie



### 3D Skeleton

Figuur 6.1: 3D skelet-reconstructie met POF-model, gevisualiseerd vanuit meerdere hoeken.

### MediaPipe vs POF Vergelijking



### POF Comparison

Figuur 6.2: Vergelijking van MediaPipe baseline (links) en POF-reconstructie (rechts) tegen ground truth. De POF-reconstructie toont betere diepte-schatting.

## Output Structuur

```
data/output/pose-3d/<video>/  
|  
|   |-- <video>_final.trc      # 59-64 augmented markers  
|   |-- <video>_initial.trc    # 22 initial markers  
|   |-- <video>_raw_landmarks.csv # Raw MediaPipe output  
|  
|   \-- joint_angles/  
|       |-- all_joint_angles.csv  # Combined angles  
|       |-- all_joint_angles.png  # Multi-panel plot  
|  
|       |-- pelvis_angles.csv    # Per-joint CSV  
|       |-- pelvis_angles.png    # Per-joint plot  
|  
|       |-- left_hip_angles.csv  
|       |-- left_hip_angles.png  
|  
|       \-- ... (12 joint groups)
```

## Output Structure

*Figuur 6.3: Overzicht van de automatisch gegenereerde output-bestanden per video.*

## Hoofdstuk 7

# Discussie

Het kernresultaat van dit project is toegankelijkheid: waar Vicon-systeem €50.000+ en dedicated ruimtes vereisen, draait HumanPose3D op een laptop met smartphone-video als input. De Part Orientation Fields-aanpak (Xiang et al., 2019) pakt monoculaire diepteschatting fundamenteel anders aan. Elk ledemaat krijgt een eigen oriëntatievector, wat situaties oplöst waar traditionele methoden falen (arm naar camera, been opzij). De volledige open-source stack (MediaPipe, Pose2Sim, PyTorch) elimineert vendor lock-in, en ISB-compliant output (Wu et al., 2002, 2005) integreert direct met bestaande biomechanica-software.

Qua productstatus is de baseline pipeline: pose-detectie, marker-augmentatie, gewichtshoekberekening, stabiel en functioneel. De POF-uitbreiding ( $\sim 16^\circ$  naar  $\sim 7^\circ$  error) blijft experimenteel; sectie 5.5 beschrijft de openstaande integratieproblemen. Het Joint Constraint model gebruikt zachte, geleerde correcties in plaats van harde anatomische limieten, essentieel voor zorgtoepassingen waar juist afwijkende patronen (hypermobiliteit, compensatiestrategieën, pathologie) klinisch relevant zijn. Dit model is geïmplementeerd maar wacht op POF-validatie.

De gekozen aanpak brengt inherente beperkingen met zich mee. Single-view occlusie betekent dat bij verdwijnende ledematen simpelweg informatie ontbreekt, en  $\sim 60$  seconden verwerkingstijd sluit real-time biofeedback uit. Ook de validatie kent methodologische zwaktes: het model is uitsluitend getraind op AIST++ dansdata (snelle, expressieve bewegingen), fundamenteel anders dan revalidatie-oefeningen of looppatronen van ouderen. Formele validatie tegen marker-based ground truth in zorgsettings ontbreekt; de gerapporteerde nauwkeurigheid is gemeten op trainingsdistributie.

Tot slot is de oorsprong als schoolopdracht zichtbaar in de architectuur: snelle iteraties, geen schaalbaarheid, minimale aandacht voor niet-technische gebruikers. Als demonstrator functioneert het systeem; voor productiegebruik vereist het herontwerp vanaf de basis.

Het systeem is daarmee geschikt voor screening, eerste beoordeling en trendmonitoring, mits gebruikers de meetonzekerheid meewegen. Het is ongeschikt voor kritische klinische beslissingen of precisietoepassingen.

## Hoofdstuk 8

# Conclusie

Dit hoofdstuk beantwoordt de centrale onderzoeksvervraag, vat de kernresultaten samen en formuleert een eindconclusie over de bruikbaarheid van het ontwikkelde MMPA-systeem voor bewegingsanalyse in de gezondheidszorg.

### 8.1 Beantwoording Onderzoeksvervraag

De centrale vraag van dit project was of betrouwbare 3D pose-reconstructie mogelijk is vanuit monocolaire smartphone-video, met voldoende nauwkeurigheid voor bewegingsanalyse in de gezondheidszorg.

Het antwoord is genuanceerd. We hebben een functionerende pipeline gebouwd die monocolaire video transformeert naar ISB-compliant gewrichtshoeken (Wu et al., 2002, 2005). De combinatie van open-source 2D/3D detectie via MediaPipe (Lugaresi et al., 2019) en RTMPose, Pose2Sim marker-augmentatie (Pagnon et al., 2021, 2022), en gestandaardiseerde gewrichtshoekberekening levert bruikbare output op een standaard laptop.

Daarnaast is een experimentele POF-module ontwikkeld, gebaseerd op de Part Orientation Fields aanpak van Xiang et al. (2019), als basis voor een volgende versie met verbeterde diepte-reconstructie. Deze module bereikt  $\sim 7^\circ$  error op validatiedata—bijna een halvering ten opzichte van de MediaPipe baseline—maar vereist nog verdere integratie.

### 8.2 Kernresultaten

Het kernresultaat van deze versie is een gebruiksvriendelijke, toegankelijke desktop-applicatie die monocolaire 3D pose-analyse uitvoert inclusief ISB-compliant gewrichtshoekberekeningen. De applicatie is lokaal installeerbaar, vereist geen technische kennis, en levert gestructureerde output die direct bruikbaar is in bestaande biomechanica-workflows.

De huidige nauwkeurigheid met de MediaPipe baseline ( $\sim 16^\circ$  limb-oriëntatiefout) is voldoende voor screening en monitoring, maar nog niet voor diagnostiek in de gezondheidszorg. De POF-module legt de technische basis voor een volgende versie met hogere nauwkeurigheid, maar de integratie-uitdagingen beschreven in sectie 5.5 moeten eerst worden opgelost.

### **8.3 Eindconclusie**

HumanPose3D overbrugt de kloof tussen dure marker-based systemen en onnauwkeurige consumentenoplossingen. Met een smartphone en laptop kunnen gebruikers nu bewegingsanalyse uitvoeren die voorheen specialistische apparatuur vereiste. De baseline pipeline is functioneel en voldoet aan de initiële projectdoelstellingen.

Het systeem is in de huidige vorm geschikt voor screening, voortgangsmonitoring, thuisgebruik door patiënten, en als open-source basis voor verder onderzoek. Voor kritische beslissingen in de zorg blijft aanvullende validatie noodzakelijk totdat de POF-integratie voltooid is en formele validatiestudies in zorgsettings zijn uitgevoerd.

De volgende stap is het oplossen van de technische uitdagingen rond POF-integratie—met name de camera-naar-world transformatie en temporale stabiliteit—om de volledige nauwkeurigheidswinst ook in productie te realiseren.

## Hoofdstuk 9

# Aanbevelingen

### 9.1 Technische Prioriteiten

**Error-propagatie in kinematische keten:** Kleine fouten in oriëntatie accumuleren door de keten. Een global skeleton optimization die alle joints simultaan optimaliseert in plaats van sequentieel, zou stabiever zijn.

**Exposure bias aanpakken:** Het temporale model gebruikt ground truth POF van het vorige frame tijdens training, maar voorspelde (foutieve) POF tijdens inference. Dit veroorzaakt error-accumulatie. Scheduled sampling—een mix van ground truth en predicted POF tijdens training—kan het model robuuster maken tegen eigen fouten.

**Camera-space naar world-space transformatie:** De POF-module reconstrueert in camera-space (Y omlaag), maar de Pose2Sim LSTM verwacht world-space (Y omhoog). De huidige pelvis-centering lost de positie op, maar niet de oriëntatie. Oplossingsrichtingen zijn: (1) ground plane detectie om camera-rotatie te schatten, (2) gravity-vector voorspelling als model output, of (3) de LSTM hertrainen op camera-space data.

De eerste prioriteit is het oplossen van de technische uitdagingen die in sectie 5.5 zijn beschreven. Zonder deze fixes blijft de POF-module experimenteel en kan de nauwkeurigheidswinst niet in productie worden gerealiseerd.

### 9.2 Dataset Uitbreidung

Een zorg-specifieke dataset zou de generaliseerbaarheid naar de beoogde toepassingen significant verbeteren. Samenwerking met revalidatiecentra of fysiotherapiepraktijken zou hiervoor een logische route zijn.

- Dagelijkse activiteiten: opstaan uit stoel, traplopen, reiken
- Looppatronen van verschillende populaties (ouderen, mensen met beperkingen)
- Revalidatie-oefeningen (langzaam, subtiel, vaak zittend)

### **9.3 Gebruikersonderzoek**

Dit gebruikersonderzoek is essentieel om de stap van technisch prototype naar bruikbaar product te maken.

**Usability testing met zorgprofessionals:** Fysiotherapeuten en bewegingswetenschappers de applicatie laten gebruiken met echte patiëntcasussen. Vragen: Is de output interpreteerbaar? Sluit het aan bij bestaande workflows? Welke informatie mist?

**Kwantitatieve validatie:** Vergelijking van HumanPose3D output met marker-based ground truth (bijv. Vicon) in een zorgsetting. Dit geeft inzicht in de werkelijke nauwkeurigheid bij de bewegingen die er toe doen, niet alleen bij dansbewegingen.

Formele validatie met eindgebruikers ontbreekt nog.

### **9.4 Interpretatielaag**

Op termijn kan machine learning patronen herkennen in de gewrichtshoekdata die correleren met specifieke aandoeningen of risicofactoren. Dit is een ambitieus doel dat pas zinvol wordt nadat de basisnauwkeurigheid is gevalideerd.

Dit vereist domeinkennis: welke waarden zijn normaal, welke afwijkingen zijn klinisch relevant, en hoe formuleer je dit begrijpelijk. Samenwerking met klinisch experts is hiervoor noodzakelijk.

- "Elleboogextensie >180° — mogelijk hypermobiliteit"
- "Heupabductie bij lopen toont compensatiepatroon: piek aan einde standfase"
- "Knieflexie rechts bereikt maximaal 95°, links 120° — asymmetrie van 25°"

### **9.5 Van demo naar productie**

De huidige applicatie is een werkende proof-of-concept, maar productie-inzet vereist aanvullend onderzoek en herontwikkeling. De deployment-keuze kent trade-offs: lokale installatie biedt datacontrole en offline-werking maar schaalt moeilijk, cloud-oplossingen bieden schaalbaarheid maar zijn duur en vereisen strenge cybersecurity, en edge devices (zoals Jetson Orin Nano) combineren lokale verwerking met lagere kosten maar hebben beperkte capaciteit. Daarbij variëren hardware, netwerk en beveiliging per instelling—GPU-versnelling verbetert prestaties significant maar is niet overal beschikbaar. Voordat verdere ontwikkeling zinvol is, moet in kaart worden gebracht hoe zorgprofessionals de applicatie daadwerkelijk gebruiken en of interface en output begrijpelijk zijn voor behandelbeslissingen. Eindgebruikers zonder technische achtergrond hebben bovendien duidelijke documentatie en toegankelijke support nodig. Een hybride architectuur lijkt het meest kansrijk: lokale applicatie op werkstations gekoppeld aan een edge device per instelling voor AI-verwerking, wat datacontrole combineert met consistente performance onafhankelijk van gebruikershardware.

## **9.6 Marktintroductie**

Voor marktintroductie zijn fysiotherapiepraktijken de meest voor de hand liggende doelgroep. Zij hebben behoefte aan objectieve voortgangsmonitoring maar beschikken zelden over budget voor *marker-based* systemen. Revalidatiecentra vormen een tweede doelgroep, waar bewegingsanalyse zonder markers de belasting voor patiënten vermindert. Sportanalyse biedt mogelijkheden voor techniekanalyse en blessurepreventie.

Het businessmodel verdient zorgvuldige overweging. Een *SaaS-model* biedt recurring revenue en centraal beheer, maar roept privacy-vragen op bij verwerking van patiëntdata in de cloud. On-premise licenties houden data lokaal en passen beter bij zorgorganisaties, tegen een eenmalige betaling. De open-source basis kan behouden blijven voor community building en transparantie, met commerciële ondersteuning als inkomstenbron.

Voor toepassing in de gezondheidszorg is certificering onvermijdelijk. CE-markering vereist classificatie van het product volgens de Medical Device Regulation en een conformiteitsbeoordeling. Dit traject is tijdsintensief maar noodzakelijk voor brede adoptie. Het open-source karakter en de afwezigheid van hardware lock-in kunnen als concurrentievoordeel worden gepositioneerd ten opzichte van gesloten commerciële

# Bibliografie

## 10.1 Core Methods & Architecture

Xiang, D., Joo, H., & Sheikh, Y. (2019). Monocular total capture: Posing face, body, and hands in the wild. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10965-10974. <https://doi.org/10.1109/CVPR.2019.01122>

Keller, M., Werling, K., Shin, S., Delp, S., Pujades, S., Liu, C. K., & Black, M. J. (2024). MANIKIN: Biomechanically accurate neural inverse kinematics. *Proceedings of the European Conference on Computer Vision (ECCV)*. [https://doi.org/10.1007/978-3-031-73232-4\\_5](https://doi.org/10.1007/978-3-031-73232-4_5)

## 10.2 ISB Biomechanics Standards

Wu, G., Siegler, S., Allard, P., Kirtley, C., Leardini, A., Rosenbaum, D., Whittle, M., D'Lima, D. D., Cristofolini, L., Witte, H., Schmid, O., & Stokes, I. (2002). ISB recommendation on definitions of joint coordinate system of various joints for the reporting of human joint motion—part I: Ankle, hip, and spine. *Journal of Biomechanics*, 35(4), 543-548. [https://doi.org/10.1016/S0021-9290\(01\)00222-6](https://doi.org/10.1016/S0021-9290(01)00222-6)

Wu, G., van der Helm, F. C. T., Veeger, H. E. J., Makhsoos, M., Van Roy, P., Anglin, C., Nagels, J., Karduna, A. R., McQuade, K., Wang, X., Werner, F. W., & Buchholz, B. (2005). ISB recommendation on definitions of joint coordinate systems of various joints for the reporting of human joint motion—part II: Shoulder, elbow, wrist and hand. *Journal of Biomechanics*, 38(5), 981-992. <https://doi.org/10.1016/j.jbiomech.2004.05.042>

Grood, E. S., & Suntay, W. J. (1983). A joint coordinate system for the clinical description of three-dimensional motions: Application to the knee. *Journal of Biomechanical Engineering*, 105(2), 136-144. <https://doi.org/10.1115/1.3138397>

## 10.3 Pose Detection & Marker Augmentation

Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C.-L., Yong, M. G., Lee, J., Chang, W.-T., Hua, W., Georg, M., & Grundmann, M. (2019). MediaPipe: A system for building perception pipelines. *arXiv preprint arXiv:1906.08172*. <https://doi.org/10.48550/arXiv.1906.08172>

Pagnon, D., Domalain, M., & Reveret, L. (2021). Pose2Sim: An end-to-end workflow for 3D markerless sports kinematics—part 2: Accuracy. *Sensors*, 21(19), 6530. <https://doi.org/10.3390/s21196530>

Pagnon, D. (2022). Pose2Sim: An open-source Python package for multiview markerless kinematics. *Journal of Open Source Software*, 7(77), 4362. <https://doi.org/10.21105/joss.04362>

#### **10.4 Monocular 3D Pose Estimation**

Pavllo, D., Feichtenhofer, C., Grangier, D., & Auli, M. (2019). 3D human pose estimation in video with temporal convolutions and semi-supervised training. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 7753-7762.  
<https://doi.org/10.1109/CVPR.2019.00794>

Wandt, B., Rudolph, M., Zell, P., Rhodin, H., & Rosenhahn, B. (2022). ElePose: Unsupervised 3D human pose estimation by predicting camera elevation and learning normalizing flows on 2D poses. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 6635-6645. <https://doi.org/10.1109/CVPR52688.2022.00653>

#### **10.5 Survey Articles**

Chen, Y., Wang, Z., Peng, Y., Zhang, Z., Yu, G., & Sun, J. (2024). Enhancing 3D human pose estimation with bone length adjustment. *arXiv preprint arXiv:2410.20731*.  
<https://doi.org/10.48550/arXiv.2410.20731>

Toshpulatov, M., Lee, W., Lee, S., & Yoon, K. (2025). 3D human pose estimation using deep learning: A systematic literature review. *Sensors*, 25(8), 2409.  
<https://doi.org/10.3390/s25082409>

Hu, M., Yin, W., Zhang, C., Cai, Z., Long, X., Chen, H., Wang, K., Yu, G., Shen, C., & Shen, S. (2025). Metric3D v2: A versatile monocular geometric foundation model for zero-shot metric depth and surface normal estimation. *arXiv preprint arXiv:2501.11841*.  
<https://doi.org/10.48550/arXiv.2501.11841>

#### **10.6 Anthropometric References**

Winter, D. A. (2009). *Biomechanics and motor control of human movement* (4th ed.). John Wiley & Sons. <https://doi.org/10.1002/9780470549148>

National Aeronautics and Space Administration. (1978). *Anthropometric source book: Volume I: Anthropometry for designers* (NASA Reference Publication 1024). NASA Scientific and Technical Information Office.

Drillis, R., & Contini, R. (1966). *Body segment parameters* (Technical Report No. 1166-03). New York University School of Engineering and Science.

## 10.7 Datasets

Li, R., Yang, S., Ross, D. A., & Kanazawa, A. (2021). AI choreographer: Music conditioned 3D dance generation with AIST++. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 13401-13412. <https://doi.org/10.1109/ICCV48922.2021.01315>

Ionescu, C., Papava, D., Olaru, V., & Sminchisescu, C. (2014). Human3.6M: Large scale datasets and predictive methods for 3D human sensing in natural environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(7), 1325-1339.  
<https://doi.org/10.1109/TPAMI.2013.248>

Joo, H., Simon, T., Li, X., Liu, H., Tan, L., Gui, L., Banerjee, S., Godisart, T., Nabbe, B., Matthews, I., Kanade, T., Nobuhara, S., & Sheikh, Y. (2019). Panoptic studio: A massively multiview system for social interaction capture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(1), 190-204. <https://doi.org/10.1109/TPAMI.2017.2782743>

# Bijlagen

## Bijlage A: Volledig Programma van Eisen

### A.1 Frontend Requirements

ID	Eis	Prioriteit	Rationale
F1	HTML voor structuur	Must	Semantische markup, toegankelijkheid
F2	CSS voor custom styling, Bootstrap voor basis-layout	Must	Onderhoudbaarheid, responsive design
F3	JavaScript voor client-side interacties en datavisualisatie	Must	Dynamische UI, real-time feedback
F4	Strikte OOP, Single Responsibility Principle, PEP8, documentatie	Must	Onderhoudbaarheid, duidelijkheid, testbaarheid
F5	Responsive design (mobile-friendly)	Should	Toegankelijkheid op tablets
F6	Accessibility (WCAG 2.1 AA)	Could	Inclusiviteit

### A.2 Backend Requirements

ID	Eis	Prioriteit	Rationale
B1	Django MVT architectuur	Must	Webframework verbindt frontend en backend, Python ecosystem
B2	Strikte OOP, Single Responsibility Principle, PEP8, documentatie	Must	Onderhoudbaarheid, duidelijkheid, testbaarheid
B3	Modulaire pipeline (losse componenten)	Must	Flexibiliteit, herbruikbaarheid
B4	GPU fallback naar CPU	Must	Brede hardware ondersteuning
B5	ISB-compliant gewrichtshoeken	Must	Standaard in de gezondheidszorg
B6	Applicatielogica alleen in src/application/	Must	Scheiding concerns
B7	Type hints op alle functies	Should	Code kwaliteit
B8	Logging op alle pipeline stappen	Should	Debugging, monitoring

### A.3 Pipeline Requirements

ID	Eis	Prioriteit	Rationale
P1	33 landmarks → 22 → 64 markers progressie	Must	Volledige biomechanische marker-set
P2	POF 3D reconstructie (<15° error)	Must	Kernfunctionaliteit
P3	Neural joint refinement	Should	Verbeterde nauwkeurigheid
P4	Automatische output organisatie	Must	Gebruiksvriendelijkheid
P5	TRC export voor biomechanische analyse	Must	Interoperabiliteit
P6	CSV export gewichtshoeken	Must	Analyse in spreadsheets
P7	PNG visualisaties per gewricht	Should	Snelle visuele inspectie

### A.4 Performance Requirements

ID	Eis	Prioriteit	Target
R1	Verwerking per video	Should	<120s (30s video, 30fps)
R2	Memory gebruik	Must	<8GB RAM
R3	Stabiele output	Must	Geen crashes tijdens verwerking
R4	GPU memory	Should	<4GB VRAM

### Bijlage B: Module Overzicht

Module	Pad	Verantwoordelijkheid
mediastream	src/mediastream/	Video I/O via OpenCV
posedetector	src/posedetector/	MediaPipe inference, landmark mapping
datastream	src/datastream/	CSV/TRC conversie, marker estimation
markeraugmentation	src/markeraugmentation/	Pose2Sim integratie, GPU acceleratie
kinematics	src/kinematics/	ISB joint angles, Euler decomposition
visualizedata	src/visualizedata/	3D plotting, skeleton connections
pof	src/pof/	POF models, LS solver, metric scale
joint_refinement	src/joint_refinement/	Neural joint constraint model
pipeline	src/pipeline/	Orchestratie, cleanup
application	src/application/	Django web interface

### **C.1 TRC Header (voorbeeld)**

```
PathFileType 4 (X/Y/Z) output.trc
DataRate CameraRate NumFrames NumMarkers Units OrigDataRate
OrigDataStartFrame OrigNumFrames
30.0 30.0 535 59 m 30.0 1 535
Frame# Time Nose RShoulder LShoulder ...
    X Y Z X Y Z X Y Z ...
1 0.000 0.123 1.456 0.789 ...
```

### **C.2 Joint Angles CSV (voorbeeld)**

```
frame,time,pelvis_flex_deg,pelvis_abd_deg,pelvis_rot_deg,left_hip_flex_deg, ...
0,0.000,5.2,-2.1,3.4,15.7, ...
1,0.033,5.3,-2.0,3.5,16.1, ...
```