

Inhoudsopgave

HumanPose3D: Markerless 3D Pose Analyse voor de Gezondheidszorg	3
Inhoudsopgave	3
Samenvatting	3
Gerealiseerd Product	4
Experimentele Uitbreiding	4
Conclusie	4
1. Inleiding	4
1.1 Aanleiding	4
1.2 Klantwensen	5
1.3 Probleemstelling	5
1.4 Opdrachtoomschrijving	5
2. Methode	6
2.1 Ontwikkelmethode	6
2.2 Technische Aanpak	6
2.3 Validatiestrategie	6
2.4 Tooling	7
3. Huidige Oplossingen (State of the Art)	7
3.1 Overzicht Bestaande Oplossingen	7
3.2 Marker-Based Motion Capture	7
3.3 MediaPipe Pose	7
3.4 VideoPose3D	8
3.5 ElePose	8
3.6 Pose2Sim	8
3.7 Kennislacune	8
4. Eisen en Randvoorwaarden	8
4.1 Projectcontext	8
4.2 Kernvereisten	9
4.3 Functionele Eisen	9
4.4 Niet-Functionele Eisen	10
4.5 Input/Output Specificaties	10
4.6 Realisatie	10
5. Bevindingen en Ontwerp/Ontwikkelkeuzes	10
5.1 Architectuurkeuzes	10
Waarom Modulaire Architectuur?	10
Waarom Part Orientation Fields (POF)?	11
Waarom Pose2Sim LSTM?	11
Waarom Neurale Joint Constraints?	11
5.2 Pipeline Evolutie	12
Van Depth Refinement naar POF	12
5.3 Technische Afwegingen	13
5.4 Wiskundige Formulering POF	13
5.4.1 Probleemstelling: Monoculaire 3D Reconstructie	13

5.4.2 Orthografische Projectie Model	13
5.4.3 Part Orientation Fields (POF) Representatie	14
5.4.4 Neurale Netwerk Voorspelling	14
5.4.5 Least-Squares Diepte Solver	14
5.4.6 Hiërarchische Oplosvolgorde	15
5.4.7 Bone Length Constraints	15
5.4.8 Metrische Schaal Herstel	16
5.5 Onopgeloste Technische Uitdagingen	16
5.5.1 POF Reconstructie Instabiliteit	16
5.5.2 Temporale Context Faalt bij Inference	17
5.5.3 Camera Space vs World Space	17
5.5.4 Dataset Bias: Alleen Dansbewegingen	17
5.5.5 Joint Constraint Model: Afhankelijkheid van POF	18
5.6 Visualisaties	18
Pipeline Flowchart	18
Marker Progressie	18
Gewrichtshoek Visualisatie	18
6. Resultaten	20
6.1 Baseline Pipeline	20
6.2 POF Model Performance	22
6.3 Joint Constraint Model Status	22
6.4 Vergelijking met Eisen	22
6.5 Visualisaties	23
3D Skelet Reconstructie	23
MediaPipe vs POF Vergelijking	23
Output Structuur	23
7. Discussie	25
7.1 Sterke Punten	25
7.2 Beperkingen en Huidige Status	25
Productstatus	25
Inherente Beperkingen	26
Validatiebeperkingen	26
7.3 Implicaties voor Gebruik	26
8. Conclusie	27
8.1 Beantwoording Onderzoeksvraag	27
8.2 Kernresultaten	27
8.3 Eindconclusie	27
9. Aanbevelingen	28
9.1 Doorontwikkeling	28
9.2 Vervolgonderzoek	28
9.3 Marktintroductie	29
10. Literatuurlijst	29
Core Methods & Architecture	29
ISB Biomechanics Standards	29

Pose Detection & Marker Augmentation	30
Monocular 3D Pose Estimation	30
Survey Articles	30
Anthropometric References	30
Datasets	31
11. Bijlagen	31
Bijlage A: Volledig Programma van Eisen	31
A.1 Frontend Requirements	31
A.2 Backend Requirements	32
A.3 Pipeline Requirements	32
A.4 Performance Requirements	33
Bijlage B: Plan van Aanpak	33
Bijlage C: Pipeline Flowchart	33
Bijlage D: Module Overzicht	35
Bijlage E: Voorbeeld Output Files	35
E.1 TRC Header (voorbeeld)	35
E.2 Joint Angles CSV (voorbeeld)	36
Bijlage F: Development Guidelines	36

HumanPose3D: Markerless 3D Pose Analyse voor de Gezondheidszorg

Minor Zorg en Technologie Auteurs: Joey Kardolus, Max Jansen **Opdrachtgever:** Jaap Jansen (Hogeschool Utrecht) **Datum:** Januari 2026

Inhoudsopgave

1. [Inleiding](#)
 2. [Methode](#)
 3. [Huidige Oplossingen \(State of the Art\)](#)
 4. [Eisen en Randvoorwaarden](#)
 5. [Bevindingen en Ontwerp/Ontwikkelkeuzes](#)
 6. [Resultaten](#)
 7. [Discussie](#)
 8. [Conclusie](#)
 9. [Aanbevelingen](#)
 10. [Literatuurlijst](#)
 11. [Bijlagen](#)
-

Samenvatting

Dit rapport beschrijft de ontwikkeling van HumanPose3D, een open-source pipeline voor markerless 3D bewegingsanalyse. Het systeem transformeert smartphone-video

naar ISB-compliant gewrichtshoeken, zonder dure motion capture apparatuur of fysieke markers.

Gerealiseerd Product

Een gebruiksvriendelijke desktop-applicatie die: - Monoculaire video verwerkt naar 3D pose-reconstructie - 59/64 biomechanische markers genereert via Pose2Sim augmentatie (Pagnon et al., 2021) - 12 ISB-compliant gewrichtsgroepen berekent (Wu et al., 2002, 2005) - Volledig lokaal draait op een standaard laptop (~60 seconden per video)

Experimentele Uitbreiding

Een POF-module (Part Orientation Fields, gebaseerd op Xiang et al., 2019) die: - $\sim 7^\circ$ limb-oriëntatiefout bereikt op validatiedata (vs. $\sim 16^\circ$ voor MediaPipe baseline) - Nog niet volledig geïntegreerd is in de productie-pipeline (zie sectie 5.5)

Conclusie

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.

Voor technische details zie de individuele secties. De wiskundige formulering van de POF-aanpak staat in sectie 5.4, de onopgeloste uitdagingen in sectie 5.5.

1. Inleiding

1.1 Aanleiding

Bewegingsanalyse speelt een cruciale rol in de gezondheidszorg, van revalidatie en fysiotherapie tot sportprestatie-optimalisatie. De huidige gouden standaard—marker-based motion capture systemen zoals Vicon en OptiTrack—biedt submillimeter nauwkeurigheid, maar tegen hoge kosten (€50.000+) en met praktische beperkingen: markers moeten fysiek op het lichaam worden geplaatst, en opnames zijn beperkt tot gespecialiseerde laboratoria.

Consumentgerichte 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 desktop-applicatie, geschikt voor niet-technische gebruikers
- **Gestandaardiseerd:** Gewrichtshoeken volgens ISB-standaarden (Wu et al., 2002, 2005)
- **Privacy-bewust:** Volledig lokale verwerking, geen cloud-upload vereist

1.3 Probleemstelling

MediaPipe Pose levert accurate 2D landmark-detectie, maar de ingebouwde 3D diepteschatting vertoont systematische fouten. Dit komt vooral voor bij ledematen die naar of van de camera bewegen—het zogenaamde “foreshortening” probleem. Een arm die naar de camera wijst lijkt in 2D kort, maar of deze naar voren of naar achteren wijst is zonder aanvullende informatie niet te bepalen.

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 de 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:** Neuraal 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.

2. Methode

2.1 Ontwikkelmethode

Het project is ontwikkeld volgens een iteratieve, pragmatische aanpak. Modules zijn incrementeel gebouwd met directe visuele feedback: 3D skelet-visualisaties maakten fouten direct zichtbaar (verkeerde diepte, gespiegelde ledematen, onrealistische hoeken). Deze snelle visuele feedback-loop stuurde 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
- **Incrementele ontwikkeling:** Werkende code eerst, documentatie volgde

2.2 Technische Aanpak

De codebase is modulaair opgezet om experimenten mogelijk te maken. Componenten kunnen onafhankelijk worden vervangen (bijv. POF-model wisselen zonder pipeline-aanpassingen). Dit was praktisch 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

2.3 Validatiestrategie

Validatie vond plaats op meerdere niveaus:

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

De primaire validatiemethode was visuele inspectie: fouten in 3D reconstructie (verkeerde diepte, gespiegelde ledematen) zijn met het blote oog waarneembaar. Voor gewrichtshoeken werd de output besproken met de opdrachtgever om te verifiëren dat de waarden overeenkwamen met verwachte bewegingspatronen.

Formele kwantitatieve validatie tegen marker-based ground truth in een klinische setting is niet uitgevoerd. Dit blijft een aanbeveling voor vervolgonderzoek (zie Discussie).

2.4 Tooling

Tool	Doel	Referentie
Python 3.11	Primaire programmeertaal	-
PyTorch	Neural network training	-
MediaPipe	2D/3D pose detection	Lugaresi et al. (2019)
Pose2Sim	Marker augmentatie	Pagnon et al. (2022)
OpenCV	Video I/O	-
NumPy/SciPy	Numerieke berekeningen	-

3. Huidige Oplossingen (State of the Art)

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, monoclair	Gratis, real-time, 33 landmarks	Onbetrouwbare 3D diepte-schatting
VideoPose3D (Pavlo et al., 2019)	2D-to-3D lifting	Temporele consistentie, state-of-the-art	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.

Beperkingen voor zorgtoepassingen: - Hoge aanschafkosten (€50.000 - €500.000)
- Vereist gespecialiseerde ruimte - Markers kunnen beweging beïnvloeden - Niet geschikt voor thuisgebruik of veldmetingen

3.3 MediaPipe Pose

Google's MediaPipe (Lugaresi et al., 2019) biedt real-time pose-detectie vanuit enkele camera-beelden. Het systeem detecteert 33 landmarks en levert zowel 2D als 3D coördinaten.

Sterke punten: - Gratis en open-source - Real-time performance op CPU - Robuuste 2D detectie

Zwakke punten: - 3D diepte gebaseerd op heuristieken, niet op werkelijke diepte-informatie - Gemiddelde oriëntatiefout van 16.3° in onze tests - Geen anatomische constraints

3.4 VideoPose3D

Pavlo et al. (2019) presenteren een aanpak die temporele convoluties gebruikt om 2D poses te “liften” naar 3D. Door meerdere frames te analyseren wordt temporele consistentie afgedwongen.

Relevantie voor ons project: - Demonstreert waarde van temporele context - Getraind op Human3.6M dataset (Ionescu et al., 2014) - Architectuur inspiratie voor onze POF-aanpak

3.5 ElePose

Wandt et al. (2022) introduceren ElePose, dat camera-elevatie voorspelt en normalizing flows leert op 2D poses. Dit maakt unsupervised training mogelijk zonder gelabelde 3D data.

Beperking: De globale camera-aanpak lost het per-limb diepte-probleem niet op.

3.6 Pose2Sim

Pagnon et al. (2022) bieden een open-source workflow voor multi-camera markerless kinematics: - Marker augmentatie via LSTM ($22 \rightarrow 64$ markers) - OpenSim integratie voor biomechanische analyse - ISB-compliant gewrichtshoeken (Wu et al., 2002, 2005)

Onze integratie: We gebruiken de Pose2Sim marker-augmentatie module als onderdeel van onze single-camera pipeline.

3.7 Kennislacune

Geen van de bestaande oplossingen combineert: 1. Monoculaire input (enkele smartphone-camera) 2. Accurate per-limb 3D reconstructie 3. Volledige biomechanische marker-set 4. ISB-compliant gewrichtshoeken

4. Eisen en Randvoorwaarden

4.1 Projectcontext

Het systeem is ontwikkeld voor bewegingsanalyse in de gezondheidszorg, waar specifieke randvoorwaarden gelden. Patiëntprivacy vereist dat videodata lokaal verwerkt wordt zonder cloud-upload, conform AVG/GDPR-wetgeving. De doelgroep—zorgprofessionals zonder IT-achtergrond—bepaalt dat het systeem gebruiksvriendelijk moet zijn en moet draaien op standaard hardware die typisch beschikbaar is in zorginstellingen.

Voor toekomstige inzet als medisch hulpmiddel zal certificering volgens de Medical Device Regulation (MDR) onderzocht moeten worden. De huidige versie is gepositieerd als onderzoekstool en valt daarom niet onder CE-markeringsvereisten.

4.2 Kernvereisten

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

4.3 Functionele Eisen

ID	Eis	Prioriteit	Status
F1	Video upload via webapp	Must	<input type="checkbox"/>
F2	3D pose reconstructie uit 2D keypoints	Must	<input type="checkbox"/>
F3	Marker augmentatie (22 → 64 markers)	Must	<input type="checkbox"/>
F4	Gewrichtshoeken berekening (12 gewrichtsgroepen)	Must	<input type="checkbox"/>
F5	TRC export voor interoperabiliteit met OpenSim/Visual3D	Must	<input type="checkbox"/>

ID	Eis	Prioriteit	Status
F6	Visualisatie van gewrichtshoeken (CSV + grafieken)	Should	☐

4.4 Niet-Functionele Eisen

ID	Eis	Specificatie	Status
N1	Verwerkingstijd	< 2 minuten per video	☐ (~60s)
N2	Geheugengebruik	< 18 GB RAM	☐
N3	POF nauwkeurigheid	< 15° limb-oriëntatiefout	☐ (~7° gerealiseerd)
N4	GPU fallback	Automatische CPU-modus indien geen GPU	☐

4.5 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

4.6 Realisatie

Alle must-have requirements 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 (~7° versus de eis van <15°), al vereist volledige productie-integratie nog aanvullend werk zoals beschreven in sectie 5.5.

5. Bevindingen en Ontwerp/Ontwikkelkeuzes

5.1 Architectuurkeuzes

Waarom Modulaire Architectuur?

De pipeline is opgebouwd uit onafhankelijke modules die via gedefinieerde interfaces communiceren:

MediaPipe

POF 3D

TRC

Detector Reconstruct Converter

Joint Joint Pose2Sim
Output Constraint Augment
 Model

Joint
Angles

Voordelen: - Onderhoudbaarheid: wijzigingen in één module beïnvloeden andere niet
- Testbaarheid: elke module apart testbaar - Herbruikbaarheid: modules kunnen in andere projecten worden ingezet

Waarom Part Orientation Fields (POF)?

Het kernprobleem met monoculaire 3D is diepte-ambigüiteit: 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 (gebaseerd op Xiang et al., 2019): - Voorspel 3D oriëntatievector per ledemaat (14 limbs) - Neural network leert diepte-richting uit 2D appearance - 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.

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

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 klinisch 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 gebruiken we een SemGCN-model (vergelijkbaar met het POF-model) 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 gewrichtsgroepen \times 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.

5.2 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 \rightarrow neuraal	Zorgtoepassingen vereisen detectie van abnormale bewegingen
2026-01-21	POF-only architectuur	Depth module verwijderd; POF simpeler en nauwkeuriger

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	$\sim 7^\circ$ vs $\sim 16^\circ$ error
SemGCN vs. Transfor- mer	MLP, CNN	SemGCN-Temporal	Anatomische bias + temporele context
Augmentation cycles	1, 5, 10, 20	20 cycles	Balans nauwkeu- righeid/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 onderbepaald.

5.4.2 Orthografische Projectie Model

Bij typische opname-afstanden ($> 2\text{m}$) en menselijke afmetingen ($\sim 0.5\text{m}$ dieptevarian-
tie) 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 Network 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 δ_{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):

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

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 (foreshortening cues).

5.4.5 Least-Squares Diepte Solver

Probleem: Gegeven oriëntatievector \mathbf{o}_l en 2D observatie δ_{2d} , vind de schaalfactor λ 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
```

Kritieke afhankelijkheid: Heup-breedte moet eerst worden opgelost om bij zij-aanzichten de links/rechts diepte-scheiding vast te leggen. Zonder dit “klapt” het skelet plat in Z.

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-ambigüiteit op met alleen bekende lichaams-lengte, 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 issues 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 ϵ 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 Temporale 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 temporele 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

Code locaties: - src/pof/reconstruction.py:97 - src/coordinates/transforms.py:37-105
- src/markeraugmentation/markeraugmentation.py:78-90

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 overgerepresenteerd - 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 hypermobilititeit, compensatiestrategieën, en pathologische bewegingen te detecteren.

5.6 Visualisaties

Pipeline Flowchart

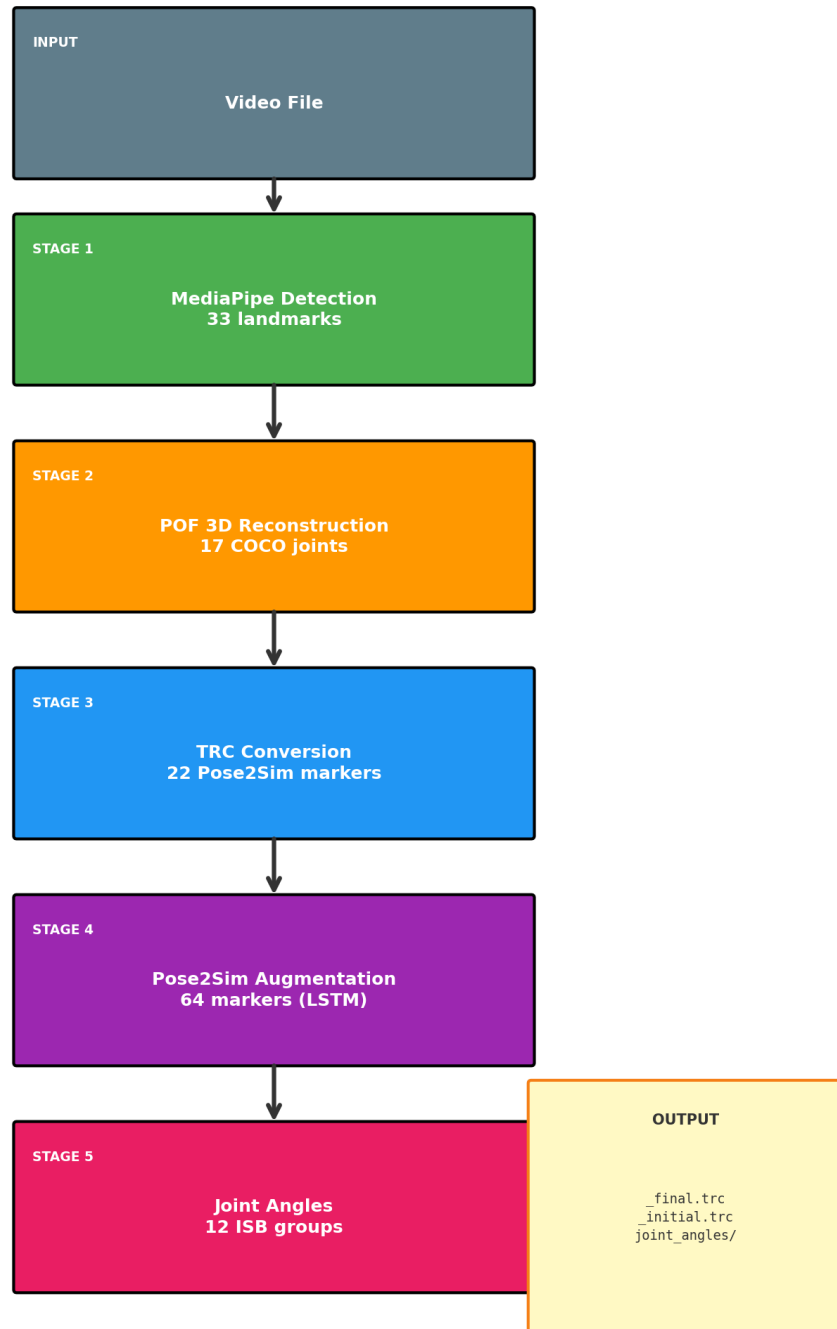
Figuur 5.1: Schematisch overzicht van de verwerkingspipeline van video-input tot gewrichtshoek-output.

Marker Progressie

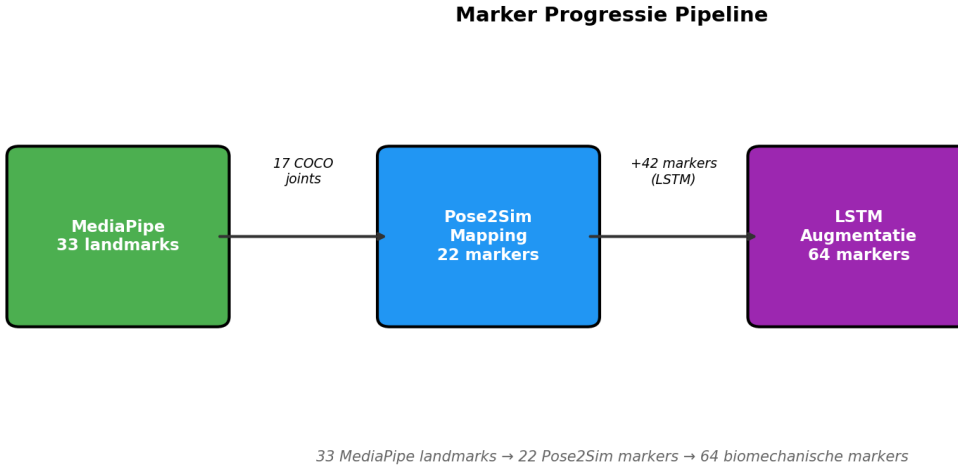
Figuur 5.2: Progressie van 33 MediaPipe landmarks naar 64 biomechanische markers via de Pose2Sim LSTM-augmentatie.

Gewrichtshoek Visualisatie

Figuur 5.3: Voorbeeld van automatisch gegenereerde gewrichtshoek-plots voor alle 12 ISB-compliant gewrichtsgroepen.



Figuur 1: Pipeline Flowchart



Figuur 2: Marker Progressie

6. Resultaten

6.1 Baseline Pipeline

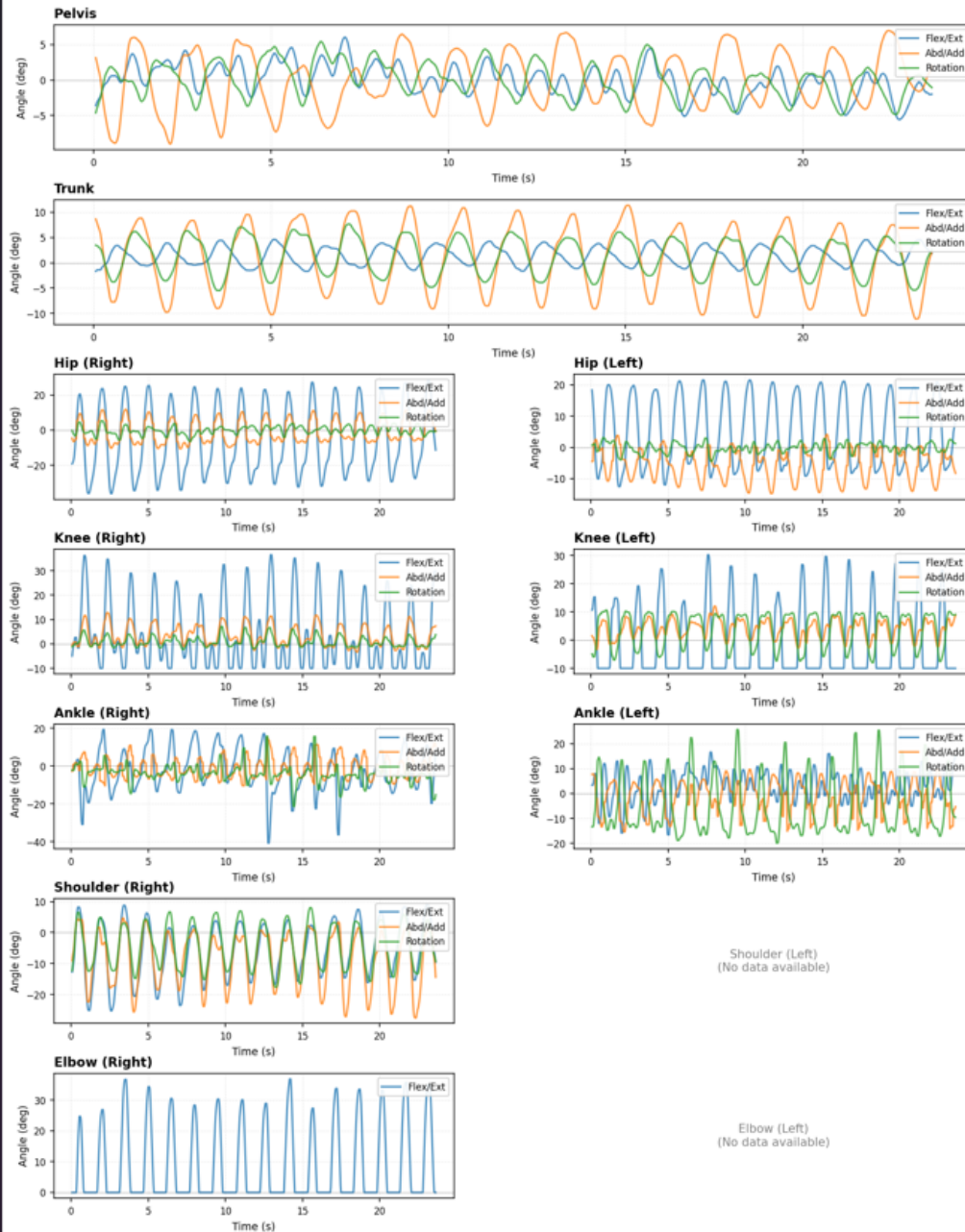
De functionele baseline pipeline—MediaPipe pose-detectie gecombineerd met Pose2Sim marker-augmentatie (Pagnon et al., 2021, 2022)—levert consistente resultaten op standaard hardware. Een typische video van 535 frames wordt verwerkt in ongeveer 60 seconden, ruim binnen de gestelde eis van twee minuten. 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 COCO markers voorspeld door media pipe of RTM pose.

De gewrichtshoekberekening volgt de ISB-standaard (Wu et al., 2002, 2005) en de Grood & Suntay conventie (1983) voor knie-analyse. De pipeline berekent twaalf gewrichtsgroepen: pelvis, linker en rechter heup, knie en enkel, trunk, linker en rechter schouder, en elleboog. Elk gewricht levert drie Euler-hoeken volgens de anatomische conventies—flexie/extensie, abductie/adductie, en interne/externe rotatie—met uitzondering van de elleboog die als scharniergewricht 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 acceptabel voor screening en voortgangsmoitoring, maar onvoldoende voor nauwkeurige diagnostiek.

MicrosoftTeams-video — Comprehensive Joint Angles (ISB)



Figuur 3: Joint Angles

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 voorspelling—of een ledemaat naar de camera of van de camera af wijst—is cruciaal voor het oplossen van de diepte-ambigüiteit.

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 temporele 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 getraind 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—hypermobiliteit, 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 requirements uit sectie 4. De verwerkingstijd van ongeveer 60 seconden blijft ruim onder de gestelde limiet van twee minuten. Het aantal gereconstrueerde markers—59 van 64—overtreft de minimumeis van 50. Alle twaalf gewrichtsgroepen worden berekend volgens ISB-standaard.

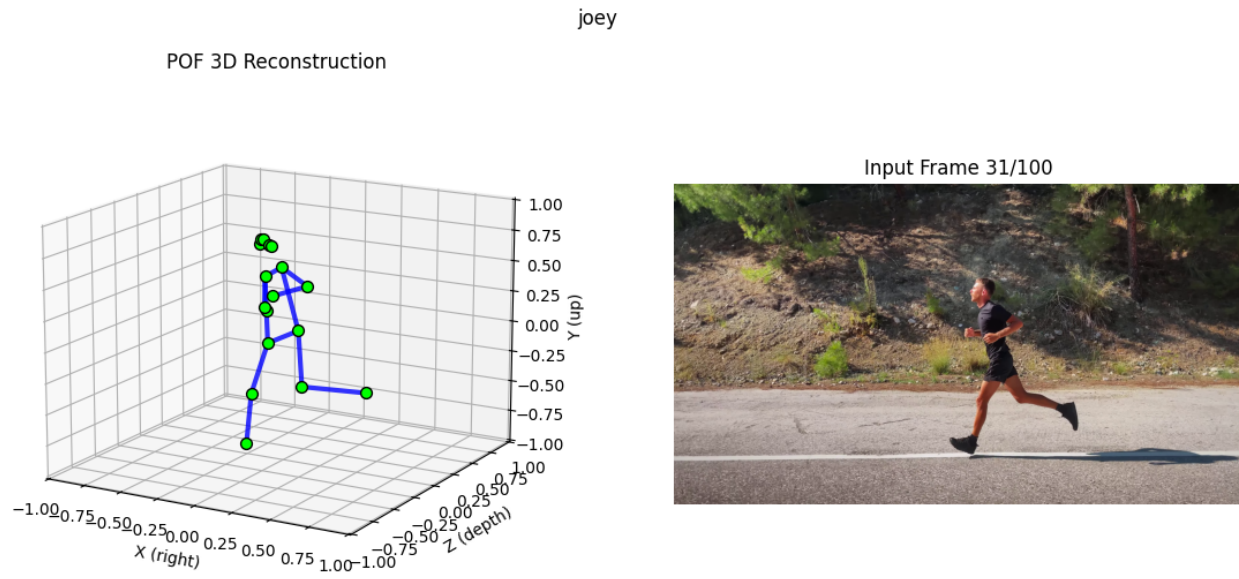
De experimentele POF-module overtreft de gestelde nauwkeurigheidseis ruimschoots: $\sim 7^\circ$ gemeten error versus de eis van maximaal 15° . Deze resultaten zijn echter geme-

ten 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 temporele stabiliteit.

6.5 Visualisaties

De pipeline genereert automatisch visualisaties van de gewrichtshoeken over tijd. 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



Figuur 4: 3D Skeleton

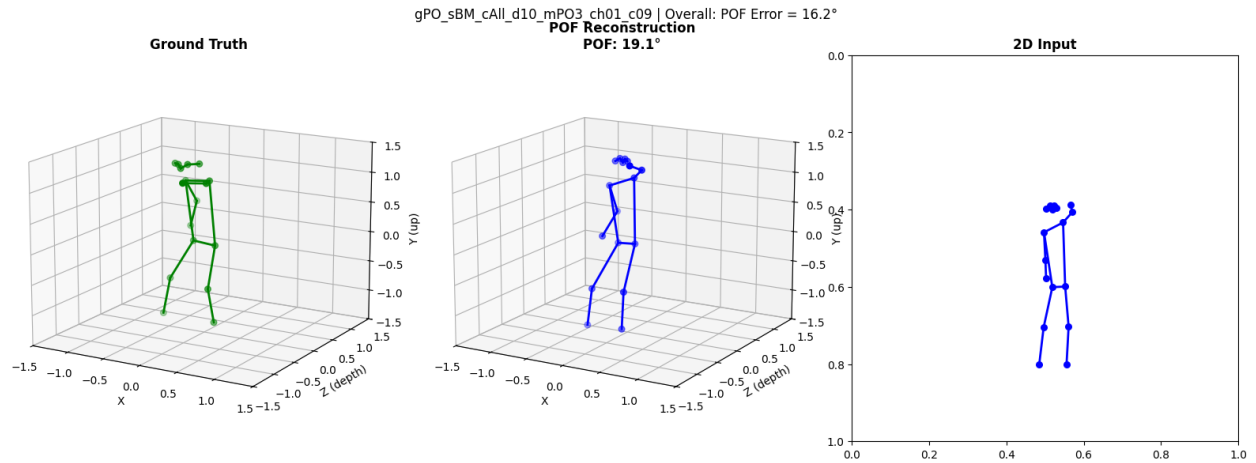
Figuur 6.1: 3D skelet-reconstructie met POF-model, gevisualiseerd vanuit meerdere hoeken.

MediaPipe vs POF Vergelijking

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

Output Structuur

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



Figuur 5: POF Comparison

Output Directory Structure

```
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)
```

Figuur 6: Output Structure

7. Discussie

7.1 Sterke Punten

Het belangrijkste resultaat van dit project is dat bewegingsanalyse nu toegankelijk wordt voor een breder publiek. Waar traditionele motion capture systemen zoals Vicon investeringen van €50.000 of meer vereisen plus gespecialiseerde ruimtes, draait HumanPose3D op een standaard laptop met als enige input een smartphone-video. Dit verlaagt de drempel voor gebruik in fysiotherapiepraktijken, revalidatiecentra en thuissituaties aanzienlijk.

De technische aanpak biedt concrete voordelen. Door gebruik te maken van Part Orientation Fields (Xiang et al., 2019) wordt het fundamentele probleem van monoculaire diepte-schatting op een nieuwe manier aangepakt: in plaats van één globale camera-hoek te voorspellen, krijgt elk ledemaat zijn eigen oriëntatievector. Dit lost situaties op waar bijvoorbeeld een arm naar de camera wijst terwijl een been opzij staat—iets waar traditionele methoden zoals ElePose (Wandt et al., 2022) moeite mee hebben.

De keuze voor een volledig open-source stack betekent dat gebruikers niet afhankelijk zijn van één leverancier. MediaPipe (Lugaresi et al., 2019), Pose2Sim (Pagnon et al., 2021, 2022) en PyTorch zijn vrij beschikbaar en worden actief onderhouden door grote communities. Dit geeft vertrouwen in de continuïteit van de oplossing.

Praktisch gezien werkt het systeem zonder GPU, hoewel verwerking dan langer duurt. De output is direct bruikbaar in bestaande biomechanica-software doordat ISB-standaarden (Wu et al., 2002, 2005) worden gevolgd voor gewrichtshoekdefinities.

7.2 Beperkingen en Huidige Status

Productstatus

Om een eerlijk beeld te geven van wat wel en niet werkt, is het belangrijk onderscheid te maken tussen productierijpe en experimentele componenten.

De baseline pipeline—MediaPipe (Lugaresi et al., 2019) voor pose-detectie, Pose2Sim (Pagnon et al., 2021) voor marker-augmentatie, en ISB-compliant gewrichtshoekberekening (Wu et al., 2002, 2005)—is stabiel en functioneel. Gebruikers kunnen video's uploaden via de desktop-applicatie en krijgen betrouwbare output. Dit deel van het systeem voldoet aan de initiële projectdoelstellingen.

De POF-uitbreiding (Xiang et al., 2019), bedoeld om de diepte-nauwkeurigheid te verbeteren van $\sim 16^\circ$ naar $\sim 7^\circ$ error, bevindt zich nog in experimentele fase. Het model zelf presteert goed op validatiedata, maar de integratie met de rest van de pipeline kent nog onopgeloste problemen. Sectie 5.5 beschrijft deze technische uitdagingen in detail: error-propagatie door de kinematische keten, exposure bias bij temporele context, en de mismatch tussen camera-space en world-space coördinaten.

Het Joint Constraint model vormt een belangrijk onderdeel van de geplande volgende versie. In plaats van harde anatomische limieten—die abnormale bewegingspatronen

zouden maskeren—gebruikt dit model zachte, geleerde correcties. Dit is essentieel voor zorgtoepassingen waar juist de afwijkende bewegingen klinisch relevant zijn: hypermobiliteit, compensatiestrategieën bij pijn, of pathologische bewegingspatronen. Het model is geïmplementeerd en getraind, maar kan pas gevalideerd worden wanneer de POF-integratie is afgerond. De theoretische onderbouwing is sterk, maar praktische resultaten moeten nog volgen.

Inherente Beperkingen

Sommige beperkingen zijn inherent aan de gekozen aanpak. Single-view occlusie blijft een fundamenteel probleem: wanneer een ledemaat achter het lichaam verdwijnt, is er simpelweg geen informatie beschikbaar om de positie te reconstrueren. Multi-view ondersteuning zou dit kunnen oplossen, maar valt buiten de huidige scope.

De verwerkingstijd van ongeveer 60 seconden per video maakt real-time feedback onmogelijk. Voor toepassingen waar directe terugkoppeling essentieel is—zoals biofeedback tijdens oefeningen—is dit een significante beperking.

Validatiebeperkingen

De validatie kent methodologische beperkingen die eerlijk benoemd moeten worden. Het POF-model is uitsluitend getraind op de AIST++ dataset (Li et al., 2021), die bestaat uit professionele dansers die expressieve, snelle bewegingen uitvoeren. Dit is fundamenteel anders dan de typische bewegingen in zorgsettings: langzame revalidatie-oefeningen, subtiele looppatronen, of bewegingen van ouderen met beperkte mobiliteit.

Formele kwantitatieve validatie tegen marker-based ground truth in een zorgsetting is niet uitgevoerd. De gerapporteerde nauwkeurigheid ($\sim 7^\circ$ error) is gemeten op dezelfde distributie waarop het model is getraind. Hoe het systeem presteert op bewegingen buiten deze distributie is onbekend.

7.3 Implicaties voor Gebruik

Gegeven de huidige status is het systeem geschikt voor bepaalde toepassingen, maar niet voor andere. Voor screening en eerste beoordeling van bewegingspatronen biedt het waardevolle informatie tegen lage kosten. Voor het volgen van voortgang over tijd—bijvoorbeeld tijdens een revalidatietraject—kan het trends zichtbaar maken, mits de gebruiker rekening houdt met de inherente meetonzekerheid.

Het systeem is niet geschikt voor situaties waar hoge nauwkeurigheid vereist is. Kritische beslissingen in de zorg moeten niet gebaseerd worden op deze output zonder aanvullende validatie. Sub-millimeter precisie, zoals nodig voor chirurgische planning, is principieel onhaalbaar met deze technologie.

Gebruikers moeten zich bewust zijn van de beperkingen bij interpretatie van de output: een inherente fout van $\sim 16^\circ$ (baseline) tot $\sim 7^\circ$ (POF) in limb-oriëntaties, mogelijke frame-dropouts bij snelle bewegingen, en artefacten bij occlusie.

8. Conclusie

8.1 Beantwoording Onderzoeksvraag

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

Het antwoord is genuanceerd. We hebben een functionerende pipeline gebouwd die monoculaire 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 monoculaire 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 overbruggt 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 temporele stabiliteit—om de volledige nauwkeurigheidswinst ook in productie te realiseren.

9. Aanbevelingen

9.1 Doorontwikkeling

De belangrijkste technische prioriteit voor een volgende versie is real-time processing. De huidige verwerkingstijd van ongeveer 60 seconden per video sluit directe feedback uit. Voor toepassingen zoals biofeedback tijdens oefeningen of live coaching is real-time verwerking essentieel. Dit vereist GPU-optimalisatie, model pruning, en mogelijk een streaming-architectuur in plaats van batch-verwerking.

Een tweede prioriteit is multi-view ondersteuning. Het occlusie-probleem—ledematen die achter het lichaam verdwijnen—is fundamenteel onoplosbaar met één camera. Door een tweede camerahoek toe te voegen kunnen occlusies worden opgelost via triangulatie. Dit zou de betrouwbaarheid significant verhogen, vooral bij complexe bewegingen.

Een derde prioriteit is de validatie van het Joint Constraint model. Dit model is ontwikkeld om zachte correcties te leren in plaats van harde anatomische limieten—essentieel voor zorgtoepassingen waar abnormale bewegingspatronen juist gedetecteerd moeten worden. De implementatie is gereed, maar effectieve training en validatie vereisen eerst een stabiele POF-integratie als basis. Zodra de POF-module volledig geïntegreerd is, kan het Joint Constraint model worden gevalideerd en ingezet.

9.2 Vervolgonderzoek

Het vervolgonderzoek zou zich moeten richten op de volledige keten van 2D detectie tot gewrichtsinterpretatie.

De basis van alles is accurate 2D pose-detectie. MediaPipe heeft een gemiddelde fout van ongeveer 10 centimeter vergeleken met ground truth. Elke verbetering hier plant zich voort door de hele pipeline. Onderzoek naar nieuwere 2D detectiemethoden of fine-tuning op zorg-specifieke bewegingen zou directe winst opleveren.

Met betere 2D input kan het POF-model verder worden verbeterd. Het huidige doel van sub-5 centimeter gemiddelde error lijkt haalbaar door combinatie van betere input en architectuurverbeteringen. De onopgeloste problemen uit sectie 5.5—error-propagatie, exposure bias, camera-world transformatie—vereisen gericht onderzoek.

De trainingsdata zou moeten worden uitgebreid met meer variatie in cameraperspectieven en bewegingstypen. De huidige AIST++ dataset bevat voornamelijk frontale dansbewegingen. Zorg-specifieke bewegingen zoals looppatronen, ontstaan uit een stoel, of rekoefeningen ontbreken. Een dataset met deze bewegingen zou de generaliseerbaarheid naar zorgtoepassingen verbeteren.

Op langere termijn is een interpretatielaag waardevol: automatische duiding van gewrichtshoeken in begrijpelijke termen. Bijvoorbeeld: “elleboogflexie $>90^\circ$ bij deze beweging wijst mogelijk op hypermobiliteit” of “asymmetrie links-rechts heup van $>10^\circ$ bij lopen”. Dit zou de output direct bruikbaar maken voor zorgprofessionals zonder biomechanische achtergrond.

9.3 Marktintrductie

Voor marktintrductie 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 oplossingen.

10. Literatuurlijst

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

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., Makhsous, 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>

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 framework 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>

Monocular 3D Pose Estimation

Pavlo, 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>

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>

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.

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/ICCV48922.2021.01>

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>

11. Bijlagen

Bijlage A: Volledig Programma van Eisen

A.1 Frontend Requirements

ID	Eis	Prioriteit	Rationale
F1	Plain HTML, geen inline JavaScript	Must	Scheiding concerns, onderhoudbaarheid
F2	Bootstrap voor styling	Must	Responsive, moderne UI
F3	Responsive design (mobile-friendly)	Should	Toegankelijkheid op tablets
F4	Geen inline CSS, dedicated .css files	Must	Onderhoudbaarheid
F5	Accessibility (WCAG 2.1 AA)	Could	Inclusiviteit

A.2 Backend Requirements

ID	Eis	Prioriteit	Rationale
B1	Django	Must	Bewezen framework,
	MVT archi- tectuur		Python ecosystem
B2	Strikte	Must	Onderhoudbaarheid, testbaarheid
	OOP, Single Responsibi- lity		
B3	Principle Modulaire pipeline (losse com- ponenten)	Must	Flexibiliteit, herbruikbaarheid
B4	GPU fallback naar CPU	Must	Brede hardware ondersteuning
B5	ISB- compliant gewrichts- hoeken	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 re- constructie ($<15^\circ$ error)	Must	Kernfunctionaliteit

ID	Eis	Prioriteit	Rationale
P3	Neural joint refinement	Should	Verbeterde nauwkeurigheid
P4	Automatische output organisatie	Must	Gebruiksvriendelijkheid
P5	TRC export voor biomechanische analyse	Must	Interoperabiliteit
P6	CSV export gewrichtshoeken	Must	Analyse in spreadsheets
P7	PNG visualisaties per gewricht	Should	Snelle visuele inspectie

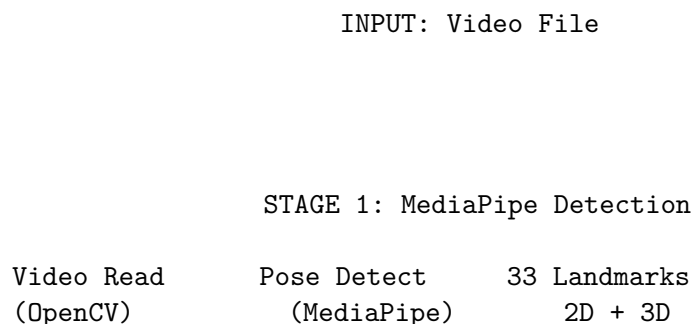
A.4 Performance Requirements

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: Plan van Aanpak

[In te vullen: oorspronkelijk projectplan met fasering en milestones]

Bijlage C: Pipeline Flowchart



STAGE 2: POF 3D Reconstruction

17 COCO Keypoints	SemGCN- Temporal	14 Limb Orientations
----------------------	---------------------	-------------------------

Least-Sq Solver	17 Joints 3D Metric
--------------------	------------------------

STAGE 3: TRC Conversion

17 COCO Joints	Mapping + Derived	22 Pose2Sim Markers
-------------------	----------------------	------------------------

STAGE 4: Pose2Sim Augmentation

22 Markers (TRC)	LSTM Model (GPU/CPU)	64 Markers Augmented
---------------------	-------------------------	-------------------------

STAGE 5: Joint Angle Computation

64 Markers	ISB Segment Coord Sys	12 Joint Groups
------------	--------------------------	--------------------

STAGE 6: Neural Joint Refinement

Raw Angles	Transformer Model	Refined Angles
------------	----------------------	-------------------

OUTPUT: Organized Files

```
data/output/pose-3d/<video>/
  <video>_final.trc          # 59-64 markers
  <video>_initial.trc        # Initial 22 markers
  joint_angles/
    all_joint_angles.csv     # Combined angles
    pelvis_angles.csv/png    # Per-joint files
    ...                      # 12 joint groups
```

Bijlage D: 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/	LSB 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

Bijlage E: Voorbeeld Output Files

E.1 TRC Header (voorbeeld)

```
PathFileType    4    (X/Y/Z)    output.trc
DataRate        CameraRate    NumFrames    NumMarkers    Units    OrigDataRate    OrigDataStartFrame
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	...						

E.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,...
```

Bijlage F: Development Guidelines

Zie AGENTS.md in de repository voor volledige development richtlijnen.

Kernprincipes: - Single Responsibility Principle - Applicatielogica gescheiden van domeinlogica - Type hints op alle functies - Imperative commit messages