



GRADO EN INGENIERÍA EN TECNOLOGÍAS DE TELECOMUNICACIÓN

TRABAJO FIN DE GRADO

Evaluación Comparativa de Algoritmos para
Posicionamiento en 3D en Simulaciones de
Accidentes

Autor: Alejandro Pais Pena

Director: Luis Francisco Sánchez Merchante

Madrid

Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título
Evaluación Comparativa de Algoritmos para Posicionamiento en 3D en Simulaciones de
Accidentes

en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el

curso académico 2025/26 es de mi autoría, original e inédito y

no ha sido presentado con anterioridad a otros efectos.

El Proyecto no es plagio de otro, ni total ni parcialmente y la información que ha sido

tomada de otros documentos está debidamente referenciada.

Fdo.: Alejandro Pais Pena

Fecha: 21/ 05/ 2025

Autorizada la entrega del proyecto

EL DIRECTOR DEL PROYECTO

Fdo.: Luis Francisco Sánchez Merchante

Fecha://

Firmado electronicamente



GRADO EN INGENIERÍA EN TECNOLOGÍAS DE TELECOMUNICACIÓN

TRABAJO FIN DE GRADO

Evaluación Comparativa de Algoritmos para Posicionamiento en 3D en Simulaciones de Accidentes

Autor: Alejandro Pais Pena

Director: Luis Francisco Sánchez Merchante

Madrid

EVALUACIÓN COMPARATIVA DE ALGORITMOS PARA POSICIONAMIENTO EN 3D EN SIMULACIONES DE ACCIDENTES

Autor: Pais Pena, Alejandro.

Director: Sánchez Merchante, Luis Francisco.

Entidad Colaboradora: ICAI – Universidad Pontificia Comillas

RESUMEN DEL PROYECTO

Este TFG explora la viabilidad de algoritmos de estimación de pose humana 3D sin marcadores. Se evaluó su consistencia en mediciones biomecánicas, demostrando el impacto crítico del posicionamiento de cámaras y la efectividad del filtrado. Los resultados sugieren una herramienta prometedora para simulaciones rápidas, aunque se requiere validación futura para su precisión absoluta.

Palabras clave: Biomecánica, Captura de Movimiento, Pose2Sim, Visión por Computador, Filtrado.

1. Introducción

El presente Trabajo de Fin de Grado aborda la creciente necesidad de métodos accesibles y eficientes para la captura de movimiento humano en 3D, contrastándolos con los sistemas tradicionales basados en marcadores, como Vicon (Minh, 2020). Estos últimos, aunque altamente precisos y considerados el estándar de oro, conllevan elevados costes, intrusividad y complejidades operativas en entornos de investigación biomecánica. El proyecto se enfoca en Pose2Sim (Pagnon, Pose2Sim: An open-source Python package for multiview markerless kinematics, 2022), un flujo de trabajo de código abierto que aprovecha los avances en la visión por computador y el aprendizaje profundo para ofrecer una alternativa sin marcadores, buscando democratizar el acceso a la cinemática 3D y agilizar los procesos de análisis en diversas aplicaciones, incluyendo la simulación de impacto en laboratorios especializados.

2. Definición del proyecto

En el presente TFG, nos propusimos investigar la eficacia de los algoritmos de estimación de la posición humana (HPE) en 3D basándonos en estudios previos (Moeslund, 2006). Los objetivos específicos incluyeron evaluar la consistencia de las longitudes de segmentos óseos reconstruidos, visualizar el movimiento mediante la superposición de modelos esqueléticos y analizar el rendimiento del sistema en el desafiante entorno del laboratorio de biomecánica de impacto MOBIO. Se puso especial énfasis en la estimación de la posición de la cadera,

un punto crítico en simulaciones de accidentes. El trabajo busca proporcionar a MOBIO una base para futuras decisiones sobre la implementación de esta tecnología, valorando su precisión, eficiencia y simplicidad.

3. Descripción del modelo/sistema/herramienta

El sistema central de este trabajo es Pose2Sim, un flujo de trabajo de código abierto diseñado para la cinemática 3D sin marcadores. A grandes rasgos, su funcionamiento comienza con la grabación de vídeos desde múltiples cámaras. Estas grabaciones, una vez capturadas, se procesan mediante algoritmos avanzados de estimación de pose 2D, como OpenPose o RTMPose (Contributors), que detectan los keypoints articulares del cuerpo humano en cada fotograma, transformando la información visual en ficheros de datos 2D.

Posteriormente, estos datos 2D se combinan con la información de una calibración precisa de las cámaras. Esta fase de calibración es crucial, ya que permite al sistema comprender la posición y orientación de cada cámara en el espacio real. Con esta información, Pose2Sim es capaz de realizar la triangulación, reconstruyendo la posición tridimensional de cada keypoint y generando así ficheros de datos 3D. Estos datos tridimensionales son la base sobre la cual se construye el análisis de movimiento.

Finalmente, los ficheros 3D obtenidos pueden ser integrados con un modelo esquelético genérico en programas como OpenSim, que escala el modelo a las dimensiones del sujeto real. Esto permite realizar la cinemática inversa para calcular ángulos articulares y, junto con herramientas de visualización como Blender, superponer el modelo esquelético al vídeo original, ofreciendo una representación gráfica clara y detallada del movimiento. De esta manera, el sistema convierte la secuencia de imágenes bidimensionales en un análisis cinemático tridimensional completo y visualmente intuitivo.

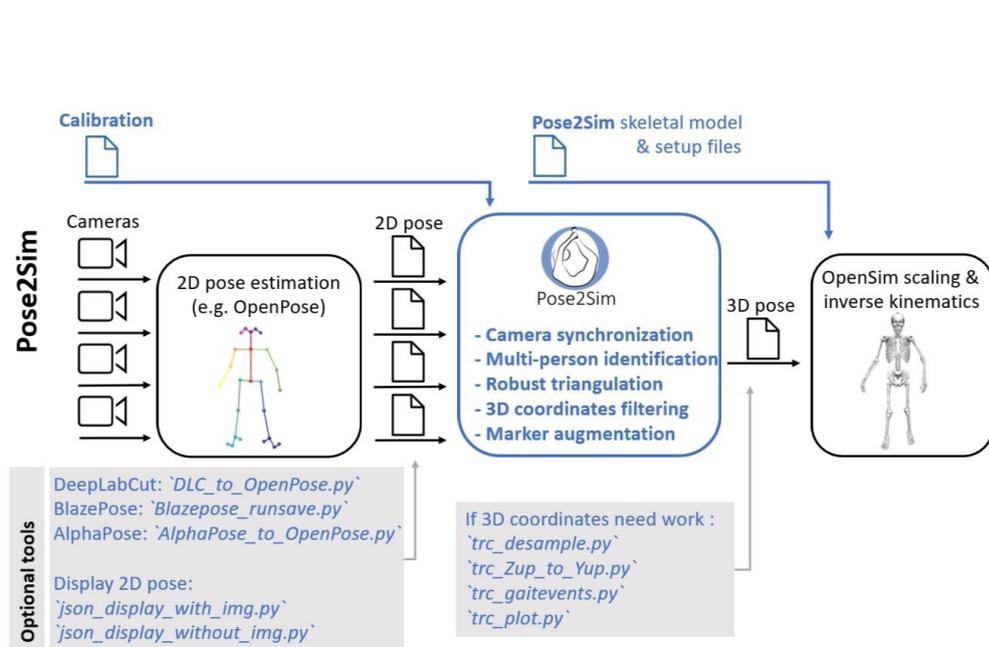


Ilustración 1 - Diagrama de Flujo Pose2Sim Fuente: Pagon,D.(2021)

4. Resultados

Los resultados cuantitativos obtenidos del análisis de la grabación en la universidad destacaron varios puntos clave:

- **Impacto del Número de Cámaras:** Contrario a la intuición de que más cámaras siempre mejoran la precisión, la consistencia de las longitudes óseas (húmero y tibia) no siguió una relación lineal. Para el húmero, la variabilidad fue mayor con 3 cámaras que con 4, y luego disminuyó con 2. Para la tibia, la consistencia mejoró consistentemente al reducir el número de cámaras. Esto sugiere que la geometría y el posicionamiento de las cámaras son tan críticos como su cantidad, pudiendo optimizar los resultados incluso con un setup minimalista.
- **Efectividad del Filtrado:** La aplicación del filtro Butterworth fue crucial. Las mediciones sin filtrar mostraban un alto nivel de ruido y movimientos abruptos. Con el filtro, se logró una reducción tangible en la desviación típica, varianza y coeficiente de variación, lo que confirmó que el filtrado es esencial para obtener datos más estables, consistentes y biomecánicamente plausibles.

Las pruebas en el laboratorio MOBIOs lamentablemente fallaron debido a condiciones desfavorables, incluyendo cámaras fijas en posiciones no adecuadas para el caso de estudio de este TFG, dado que están planteadas únicamente para ver en cámara lenta la reacción del cuerpo ante el accidente, severas oclusiones del sujeto por el equipo de simulación y desafíos significativos en la calibración. Esto demostró la importancia crítica de la visibilidad y una

calibración precisa para la fiabilidad de los datos de entrada en cualquier sistema de estimación de pose 3D.

En general, las mediciones obtenidas resultaron aceptables para la representación sencilla en 3D, aunque subóptimas para un estudio musculoesquelético preciso, con una variabilidad aproximada del 10% en el tamaño del hueso respecto al valor medio en términos de Coeficiente de Variación (cociente de la varianza respecto de la media) para las longitudes óseas en las mejores condiciones. Por lo tanto, Pose2Sim se revela como una buena herramienta para realizar simulaciones rápidas y gráficas. No obstante, sería necesario seguir estudiando o desarrollando el sistema para poder utilizarlo en estudios académicos o estudios de accidentes.

5. Conclusiones

El TFG ha demostrado que Pose2Sim es una herramienta prometedora y accesible para la cinemática 3D sin marcadores, capaz de generar datos consistentes y visualizaciones de alta calidad en entornos controlados. Los resultados han puesto de manifiesto que la cantidad de cámaras no garantiza por sí misma la precisión óptima, siendo la geometría del setup y la visibilidad de los keypoints factores determinantes. La aplicación de filtros, como el Butterworth, es indispensable para suavizar el ruido inherente a los datos triangulados, haciendo las mediciones más fiables para el análisis biomecánico.

Aunque Pose2Sim se enfrenta a limitaciones en entornos complejos con oclusiones y configuraciones de cámara restringidas, su potencial para complementar los sistemas tradicionales y simplificar los flujos de trabajo es considerable. No busca reemplazar la precisión absoluta de sistemas como Vicon sin una validación rigurosa, pero ofrece una alternativa eficiente para la visualización y un análisis preliminar.

Las futuras líneas de investigación se centran en la validación rigurosa con un sistema Vicon simultáneo, la exploración del marker augmentation, la optimización de pruebas en entornos desafiantes y el posible desarrollo de algoritmos de pose personalizados para aplicaciones específicas, como las simulaciones de accidentes. Este trabajo sienta las bases para futuras investigaciones que permitan la adopción más amplia de la cinemática sin marcadores en diversos campos.

6. Referencias

Pagnon, D. a. (2022). Pose2Sim: An open-source Python package for multiview markerless kinematics. *Journal of Open Source Software*.

Moeslund, T. B. (2006). A survey of advances in vision-based human motion capture and analysis. *Computer vision and image understanding*, 90--126.

Minh, V. T. (2020). Model Predictive Control for Modeling Human Gait Motions Assisted by Vicon Technology. . Journal Europeen des Systemes Automatises.

Contributors, M. (s.f.). OpenMMLab Pose Estimation Toolbox and Benchmark. OpenMMLab Pose Estimation Toolbox and Benchmark.

COMPARATIVE EVALUATION OF ALGORITHMS FOR 3D POSITIONING IN ACCIDENT SIMULATIONS

Author: Pais Pena, Alejandro.

Supervisor: Sánchez Merchante, Luis Francisco.

Collaborating Entity: ICAI – Universidad Pontificia Comillas

ABSTRACT

This Bachelor's Thesis explores the viability of markerless 3D human pose estimation algorithms. Their consistency in biomechanical measurements was evaluated, demonstrating the critical impact of camera positioning and the effectiveness of filtering. The results suggest a promising tool for rapid simulations, although future validation is required for its absolute accuracy.

Keywords: Biomechanics, Motion Capture, Pose2Sim, Computer Vision, Filtering.

1. Introduction

This Bachelor's Thesis addresses the growing need for accessible and efficient methods for 3D human motion capture, contrasting them with traditional marker-based systems, such as Vicon (Minh, 2020). The latter, although highly precise and considered the gold standard, involve high costs, intrusiveness, and operational complexities in biomechanical research environments. The project focuses on Pose2Sim (Pagnon, Pose2Sim: An open-source Python package for multiview markerless kinematics, 2022), an open-source workflow that leverages advances in computer vision and deep learning to offer a markerless alternative, seeking to democratize access to 3D kinematics and streamline analysis processes in various applications, including impact simulation in specialized laboratories.

2. Project Definition

In this Bachelor's Thesis, we set out to investigate the effectiveness of 3D human pose estimation (HPE) algorithms based on previous studies (Moeslund, 2006). Specific objectives included evaluating the consistency of reconstructed bone segment lengths, visualizing movement by superimposing skeletal models, and analyzing the system's performance in the challenging environment of the MOBIOUS impact biomechanics laboratory. Special emphasis was placed on hip position estimation, a critical point in

accident simulations. The work seeks to provide MOBIOUS with a basis for future decisions on the implementation of this technology, assessing its accuracy, efficiency, and simplicity.

3. Description of the Model

The central system of this work is Pose2Sim, an open-source workflow designed for markerless 3D kinematics. Broadly speaking, its operation begins with recording videos from multiple cameras. These recordings, once captured, are processed using advanced 2D pose estimation algorithms, such as OpenPose or RTMPose (Contributors), which detect the joint keypoints of the human body in each frame, transforming the visual information into 2D data files.

Subsequently, these 2D data are combined with information from a precise camera calibration. This calibration phase is crucial, as it allows the system to understand the position and orientation of each camera in real space. With this information, Pose2Sim is able to perform triangulation, reconstructing the three-dimensional position of each keypoint and thus generating 3D data files. These three-dimensional data are the basis upon which movement analysis is built.

Finally, the obtained 3D files can be integrated with a generic skeletal model in programs like OpenSim, which scales the model to the dimensions of the real subject. This allows for inverse kinematics to calculate joint angles and, together with visualization tools like Blender, to superimpose the skeletal model onto the original video, offering a clear and detailed graphical representation of the movement. In this way, the system converts the sequence of two-dimensional images into a complete and visually intuitive three-dimensional kinematic analysis.

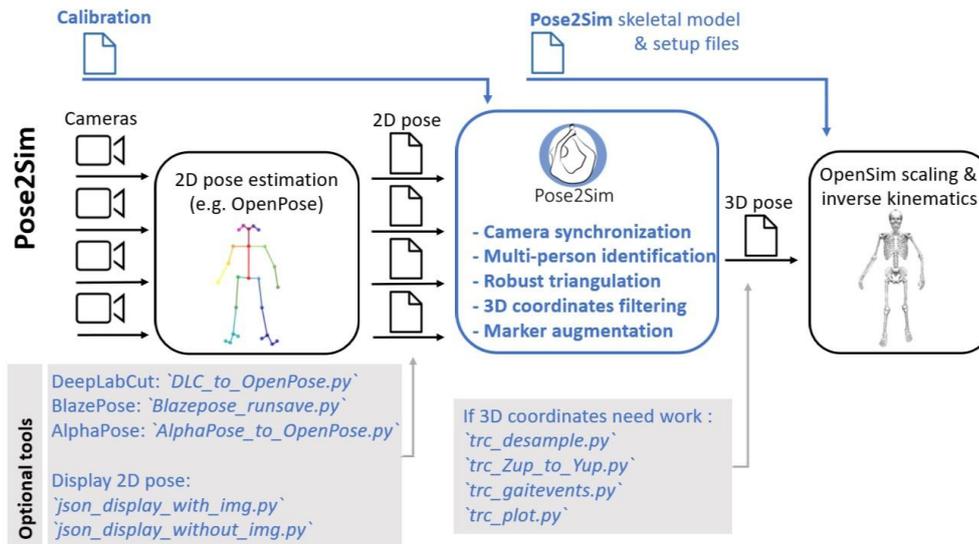


Ilustración 2 - Pose2Sim Flowchart (Pagnon, Diagrama de funcionamiento de Pose2Sim)

4. Results

The quantitative results obtained from the university recording analysis highlighted several key points:

- **Impact of Camera Number:** Contrary to the intuition that more cameras always improve accuracy, the consistency of bone lengths (humerus and tibia) did not follow a linear relationship. For the humerus, variability was greater with 3 cameras than with 4, and then decreased with 2. For the tibia, consistency consistently improved by reducing the number of cameras. This suggests that the geometry and positioning of the cameras are as critical as their quantity, potentially optimizing results even with a minimalist setup.
- **Effectiveness of Filtering:** The application of the Butterworth filter was crucial. Unfiltered measurements showed a high level of noise and abrupt movements. With the filter, a tangible reduction in standard deviation, variance, and coefficient of variation was achieved, which confirmed that filtering is essential for obtaining more stable, consistent, and biomechanically plausible data.

The tests in the MOBIOS laboratory unfortunately failed due to unfavorable conditions, including cameras fixed in positions not suitable for the case study of this Bachelor's Thesis, given that they are designed only to view the body's reaction to the accident in slow motion. There were also severe occlusions of the subject by the simulation equipment and significant challenges in calibration. This demonstrated the critical importance of visibility and precise calibration for the reliability of input data in any 3D pose estimation system.

In general, the measurements obtained were acceptable for simple 3D representation, although suboptimal for a precise musculoskeletal study, with an approximate variability of 10% in bone size relative to the mean value in terms of Coefficient of Variation (ratio of the variance to the mean) for bone lengths under the best conditions. Therefore, Pose2Sim proves to be a good tool for performing quick and graphical simulations. However, it would be necessary to continue studying or developing the system to be able to use it in academic studies or accident studies.

5. Conclusions

This Bachelor's Thesis has demonstrated that Pose2Sim is a promising and accessible tool for markerless 3D kinematics, capable of generating consistent data and high-quality visualizations in controlled environments. The results have shown that the number of cameras does not in itself guarantee optimal accuracy, with setup geometry and keypoint visibility being determining factors. The application of filters, such as Butterworth, is indispensable to smooth out the noise inherent in triangulated data, making the measurements more reliable for biomechanical analysis.

Although Pose2Sim faces limitations in complex environments with occlusions and restricted camera configurations, its potential to complement traditional systems and simplify workflows is considerable. It does not seek to replace the absolute precision of systems like Vicon without rigorous validation, but it offers an efficient alternative for visualization and preliminary analysis.

Future lines of research focus on rigorous validation with a simultaneous Vicon system, the exploration of marker augmentation, the optimization of tests in challenging environments, and the possible development of customized pose algorithms for specific applications, such as accident simulations. This work lays the foundations for future research that will allow for the wider adoption of markerless kinematics in various fields.

6. Referencias

Pagnon, D. a. (2022). Pose2Sim: An open-source Python package for multiview markerless kinematics. *Journal of Open Source Software*.

Moeslund, T. B. (2006). A survey of advances in vision-based human motion capture and analysis. *Computer vision and image understanding*, 90--126.

Minh, V. T. (2020). Model Predictive Control for Modeling Human Gait Motions Assisted by Vicon Technology. . Journal Europeen des Systemes Automatises.

Contributors, M. (s.f.). OpenMMLab Pose Estimation Toolbox and Benchmark. OpenMMLab Pose Estimation Toolbox and Benchmark.

Índice de la memoria

Capítulo 1. Introducción	7
1.1 Motivación del proyecto.....	8
Capítulo 2. Descripción de las Tecnologías.....	10
2.1 Entorno de Programación y Librerías.....	10
2.2 Herramientas de Detección de Pose y Simulación.....	11
2.3 Herramientas de Edición y Postproducción Audiovisual.....	12
2.4 Equipamiento de Grabación y Hardware	12
2.5 Instalaciones y Recursos de Mobios.....	13
Capítulo 3. Estado de la Cuestión	15
3.1 Introducción a la Captura de Movimiento.....	15
3.2 Sistemas de Captura de Movimiento Basados en Marcadores	16
3.3 Avances en la Detección de Pose sin Marcadores.....	17
3.4 Pose2Sim en el Contexto de la Investigación Actual.....	18
3.5 Brecha de Investigación y Justificación del Presente Trabajo	19
3.6 Aclaración con respecto a OpenPose	20
Capítulo 4. Definición del Trabajo	23
4.1 Justificación.....	23
4.2 Objetivos del Proyecto	24
<i>Objetivos Principales:</i>	24
<i>Objetivos Secundarios:</i>	25
4.3 Metodología de Trabajo	25
4.4 Planificación y Estimación Económica.....	27
<i>Planificación y desviaciones</i>	27
<i>Costes directos del Proyecto</i>	28
<i>Valoración de recursos No Monetizados</i>	28
Capítulo 5. Sistema/Modelo Desarrollado.....	31
5.1 Análisis del Sistema	31
<i>Introducción al Flujo de Trabajo de Pose2Sim</i>	31
<i>Componentes Clave del Sistema</i>	35

<i>Entorno de Desarrollo y Librerías</i>	37
5.2 <i>Diseño del Sistema</i>	38
<i>Estructura de Carpetas y Gestión de Archivos</i>	38
<i>Parámetros de Configuración (config.toml)</i>	43
<i>Fases del Algoritmo Pose2Sim: Un Recorrido Detallado</i>	54
<i>Visualización y Exportación de Resultados</i>	85
5.3 <i>Diseño del experimento</i>	90
<i>Experimentos en Entorno Controlable</i>	91
<i>Experimento en Entorno Realista (Grabación Satisfactoria en ICAI)</i>	91
<i>Experimentos en Entorno de Interés (Grabaciones Fallidas en MOBIOs)</i>	92
<i>Configuración General de la Adquisición de Datos</i>	92
Capítulo 6. <i>Análisis de Resultados</i>	94
6.1 <i>Introducción al Análisis de Resultados</i>	94
<i>Condiciones Experimentales y Configuración de la Grabación Exitosa en la Universidad</i> ...	94
6.2 <i>Metodología del Análisis</i>	96
<i>Herramientas y Selección de Datos para el Análisis Cuantitativo</i>	96
<i>Métricas de Evaluación Cuantitativa</i>	97
<i>Configuraciones Experimentales a Comparar</i>	98
6.3 <i>Resultados Cuantitativos de la Grabación en la Universidad</i>	99
<i>Análisis de Longitudes Óseas por Configuración de Cámaras</i>	99
<i>Impacto del Filtrado Butterworth (Caso 4 Cámaras)</i>	108
<i>Comparación con Medidas Antropométricas de Referencia</i>	112
6.4 <i>Análisis de las Pruebas Fallidas en MOBIOs</i>	112
<i>Descripción de las Condiciones y Desafíos en el Laboratorio MOBIOs</i>	112
<i>Desafíos en la calibración</i>	114
<i>Propagación de Errores por Oclusión en los Datos 3D</i>	115
6.5 <i>Discusión General y Conclusiones del Análisis</i>	118
<i>Impacto del Número de Cámaras</i>	118
<i>Efectividad del Filtrado</i>	120
<i>Importancia de las Condiciones de Grabación y Visibilidad</i>	121
<i>Limitaciones del Estudio y Futuras Líneas</i>	122
Capítulo 7. <i>Conclusiones y Trabajos Futuros</i>	125

Capítulo 8. Bibliografía..... 127

ANEXO I: ALINEACIÓN DEL PROYECTO CON LOS ODS 129

Índice de figuras

Ilustración 1 - Diagrama de Flujo Pose2Sim Fuente: Pagon,D.(2021)	9
Ilustración 2 - Pose2Sim Flowchart (Pagnon, Diagrama de funcionamiento de Pose2Sim)	14
Ilustración 4 - Diagrama de flujo Pose2Sim (Pagnon, Diagrama de funcionamiento de Pose2Sim).....	35
Ilustración 5 - Imagen del patrón utilizado 4x7.....	58
Ilustración 6 - Imagen del proceso de calibración intrínseca	58
Ilustración 7 - Diagrama del patrón a dibujar para la calibración	59
Ilustración 8 - Sobre el tablero, calibración extrínseca de una cámara. Grabación en Casa	61
Ilustración 9 - Sobre el tablero, calibración extrínseca de otra cámara. Grabación en Casa	61
Ilustración 10 - Puntos utilizados para la triangulación	65
Ilustración 11 - Keypoints del modelo completo Body_with_feet (Contributors).....	66
Ilustración 12 - Snapshot, estimación de pose 2D.....	68
Ilustración 13 - Ejemplo de detección errónea del cuerpo	69
Ilustración 14 - Ejemplo de intento de detección aceptable	70
Ilustración 15 - Ejemplo de manejo bastante aceptable ante una posición difícil.....	71
Ilustración 16 – Misma posición, Cámara 1	72
Ilustración 17 - Misma posición, Cámara 2.....	72
Ilustración 18 - Misma posición, Cámara 3.....	73
Ilustración 19 - Ejemplo 1, visualización de ficheros .trc	87
Ilustración 20 - Ejemplo 2, visualización de ficheros .trc	88
Ilustración 21 - Ejemplo 1, visualización con Pose2Sim	89
Ilustración 22 - Ejemplo 2, visualización con Pose2Sim	89
Ilustración 23 - Ubicación de las Cámaras grabación ICAI	95
Ilustración 24 - Imagen Calibración Extrínseca Cámara 1	96
Ilustración 25 - Distribución de Longitud del Húmero derecho (4 Cámaras).....	101
Ilustración 26 - Distribución de longitud de la Tibia derecha (4 Cámaras)	101

Ilustración 27 - Distribución de Longitud del Húmero derecho (3 Cámaras).....	103
Ilustración 28 - Distribución de longitud de la tibia derecha (3 Cámaras).....	104
Ilustración 29 - Distribución de Longitud del Húmero derecho (2 Cámaras).....	106
Ilustración 30 - Distribución de Longitud de la tibia derecha (2 Cámaras)	106
Ilustración 31 - Comparación de valores filtrados RKnee	109
Ilustración 32 - Comparación de valores filtrados RElbow	109
Ilustración 33 - Distribución de Longitud del Húmero derecho (4 Cámaras sin filtro)	110
Ilustración 34 - Distribución de longitud de la Tibia derecha (4 Cámaras sin filtro).....	110
Ilustración 35 - Diagrama de Posición de las cámaras de MOBIOs.....	113
Ilustración 36 - Visión de la cámara cenital	116
Ilustración 37 - Vista de una cámara lateral	116
Ilustración 38 - Imagen grabada de la última prueba	117

Índice de tablas

Tabla 1 - Estimación económica	30
Tabla 2 - Modelos de estimación 2D.....	64
Tabla 3 - Resultados ensayo con 4 cámaras	101
Tabla 4 - Resultados del ensayo con 3 cámaras	103
Tabla 5 - Resultados del ensayo con 2 cámaras	106
Tabla 6 - - Resultados del ensayo con 4 cámaras sin filtrado.....	110
Tabla 7 - Tabla comparativa de valores estadísticos en función del número de cámaras .	118
Tabla 8 - Tabla comparativa del efecto del filtrado.....	120

Capítulo 1. INTRODUCCIÓN

El Comillas MOBIOS LAB, inaugurado en marzo de 2023, se ha establecido como un laboratorio de referencia en la Universidad Pontificia Comillas dentro del ámbito de la Movilidad, Biomecánica y Salud. Sus principales objetivos radican en la investigación aplicada y la transferencia de conocimiento, abordando desafíos emergentes en los escenarios futuros de la movilidad. Es particularmente singular en Europa por su enfoque en la biomecánica de lesiones, con aplicaciones en medicina, deporte y, de manera crucial, en la seguridad vial y la dinámica de impacto, contando con una catapulta inversa de última generación para ensayos de choque.

En su labor de investigación, MOBIOS LAB depende en gran medida de sistemas de modelado 3D, que tradicionalmente operan mediante cámaras y puntos blancos de referencia, conocidos como marcadores. Estos marcadores, si bien son el estándar de oro en precisión, pueden resultar problemáticos y tediosos en el contexto de sus estudios de impacto. La preparación y colocación de estos puntos consume un tiempo considerable, generando altos costes operativos. Además, su naturaleza intrusiva y la dificultad para situarlos en áreas críticas, como la cadera cuando el sujeto lleva un cinturón de seguridad, limitan la completitud del análisis. Estas deficiencias ralentizan el proceso de investigación y la disponibilidad de datos cruciales para la seguridad.

Ante esta realidad, se planteó la necesidad de explorar soluciones más eficientes y menos intrusivas. Fue así como surgió la idea de integrar y evaluar los algoritmos de Estimación de la Posición Humana (HPE). Estas nuevas aplicaciones, que prescinden del uso de marcadores, prometen optimizar los flujos de trabajo, reducir los tiempos de preparación y permitir el análisis de movimientos en condiciones donde los marcadores tradicionales son inviables, ofreciendo una vía potencial para superar las limitaciones actuales de la captura de movimiento en MOBIOS LAB.

En este proyecto, se analizará una solución tecnológica que permite detectar las articulaciones sin necesidad de emplear marcadores reflectantes. En este proyecto utilizaremos Pose2Sim, que utiliza un flujo de trabajo de cinemática 3D sin marcadores, como alternativa a los métodos convencionales de captura de movimiento basados en marcadores físicos.

El objetivo principal del presente proyecto fue determinar la capacidad de los algoritmos de estimación de la posición humana (HPE) en 3D para generar datos fiables y consistentes, y evaluar su viabilidad como herramienta complementaria o sustituta de los sistemas tradicionales. Específicamente, se buscó analizar la consistencia de las longitudes de segmentos óseos reconstruidos, superponer modelos esqueléticos para una visualización intuitiva del movimiento y estudiar el comportamiento del sistema en el desafiante entorno del laboratorio de biomecánica de impacto MOBIO. Este trabajo pretende ofrecer una base para que MOBIO decida la inversión de tiempo en esta tecnología, valorando no solo la precisión sino también la eficiencia y simplicidad que ofrece frente a los métodos existentes.

1.1 MOTIVACIÓN DEL PROYECTO

El proyecto surge como necesidad de buscar una alternativa a los imperfectos y costosos sistemas tradicionales de marcadores. Si los algoritmos basados en las detecciones de los puntos de referencia a partir del procesamiento de las imágenes supera los estándares del sistema que se utiliza actualmente, podrá ser usado en lugar de este para los diferentes estudios de MOBIO. Actualmente el sistema de captura de movimiento tiene varios inconvenientes como es la caída de los marcadores reflectantes durante las simulaciones. Otro de los inconvenientes y que supone innumerables horas al personal del laboratorio, es identificar manualmente cada uno de los marcadores reflectantes durante el vídeo de la simulación. Otro problema que podría solventar es la dificultad de obtener una ubicación precisa de la cadera de la persona accidentada ya que la presencia del cinturón de seguridad

dificulta la posibilidad de poner marcadores anclados a los huesos de la cadera. Surge por tanto por parte del laboratorio la necesidad de buscar soluciones alternativas y se pretende testear si esta solución resulta válida o no para su labor de investigación.

Capítulo 2. DESCRIPCIÓN DE LAS TECNOLOGÍAS

El presente Trabajo de Fin de Grado ha sido desarrollado mediante la aplicación y el uso de una diversidad de tecnologías, abarcando tanto herramientas de software para el análisis de datos y la postproducción audiovisual, como equipamiento de hardware para la captura de vídeo y las instalaciones especializadas del laboratorio MOBIOS de la Universidad Pontificia Comillas. Este capítulo detalla de forma exhaustiva los medios técnicos que han posibilitado la consecución de los objetivos del proyecto.

2.1 ENTORNO DE PROGRAMACIÓN Y LIBRERÍAS

El análisis de datos, la ejecución de algoritmos avanzados de detección de pose y la visualización de los resultados se han llevado a cabo exclusivamente en un entorno de programación basado en Python. Este entorno se ha gestionado a través de Jupyter, que forma parte de la distribución Anaconda, proporcionando un ambiente interactivo para el desarrollo y la experimentación.

- Python: Como lenguaje de programación principal, Python ha proporcionado la flexibilidad y el soporte necesario para implementar los procesos de análisis de datos, así como para interactuar con las herramientas de detección de pose. Su vasta comunidad y la riqueza de sus librerías han sido fundamentales para el proyecto.
- Jupyter (Anaconda): El desarrollo y la ejecución de los scripts de Python se han realizado en entornos de cuadernos Jupyter, gestionados a través de la distribución Anaconda. Esto ha permitido un desarrollo iterativo, la visualización en línea de resultados y una gestión eficiente de las dependencias de las librerías.
- Pandas: Esta librería de código abierto para Python ha sido empleada para la estructuración, manipulación y limpieza de los conjuntos de datos, incluyendo la lectura del fichero .trc y la preparación de los datos para el análisis comparativo.

- NumPy: Esencial para la computación numérica, NumPy ha permitido realizar operaciones eficientes sobre grandes volúmenes de datos, facilitando los cálculos matemáticos inherentes al análisis de las coordenadas articulares y la simulación.
- Plotly: La librería Plotly ha sido utilizada para la creación de visualizaciones interactivas de los datos y los resultados, lo que ha permitido una comprensión más profunda de la cinemática en 3D y la representación de las detecciones de Pose2Sim.

2.2 HERRAMIENTAS DE DETECCIÓN DE POSE Y SIMULACIÓN

El núcleo de este proyecto reside en la evaluación comparativa de algoritmos de detección de pose sin marcadores. Para ello, se han utilizado las siguientes herramientas clave, destacando su rol específico:

- Pose2Sim: Este proyecto ha sido el objeto principal de estudio y evaluación. Pose2Sim ofrece un flujo de trabajo de cinemática 3D sin marcadores, integrando otras herramientas para convertir datos de coordenadas 2D a modelos de movimiento 3D. Ha sido fundamental para detectar las articulaciones a partir de las imágenes y vídeos capturados, y su viabilidad ha sido analizada a fondo en el contexto de este trabajo.
- OpenPose (RTMPose): Integrado dentro del flujo de trabajo de Pose2Sim, OpenPose es un modelo de aprendizaje profundo que se utiliza como punto de partida para identificar las posiciones articulares en 2D a partir de vídeos. Sus datos de detección de keypoints del cuerpo humano sirven como entrada para el procesamiento posterior en Pose2Sim.
- OpenSim: Este software de simulación permite modelar movimientos en 3D. En este proyecto, OpenSim ha sido empleado para generar las visualizaciones de los modelos esqueléticos superpuestos a las imágenes de vídeo, permitiendo analizar el efecto del impacto en las estructuras óseas y facilitando la representación de los movimientos detectados por Pose2Sim.

- Blender: Esta herramienta de código libre ha permitido la representación en 3D de los movimientos y los modelos esqueléticos generados a partir de los keypoints detectados, facilitando la visualización y el análisis de los resultados.

2.3 HERRAMIENTAS DE EDICIÓN Y POSTPRODUCCIÓN AUDIOVISUAL

La gestión y preparación del material de vídeo era una parte importante del proyecto, requiriendo software especializado para garantizar la calidad y la sincronización de las grabaciones.

- DaVinci Resolve: Este software ha sido empleado para la sincronización de los vídeos grabados con las diferentes cámaras. La capacidad de DaVinci Resolve para alinear múltiples fuentes de vídeo en base al audio ha sido esencial para asegurar que todas las grabaciones comenzaran y terminaran al mismo tiempo, lo cual es un requisito indispensable para el posterior análisis cinemático y la correcta aplicación de Pose2Sim.

2.4 EQUIPAMIENTO DE GRABACIÓN Y HARDWARE

La adquisición de datos visuales ha involucrado una variedad de dispositivos de grabación, tanto de uso común como especializados, así como hardware de apoyo para la estabilidad y el procesamiento.

- Smartphones (iPhone 15, iPhone 13 y teléfonos Android): Se han utilizado teléfonos inteligentes, incluyendo un iPhone 15 y un iPhone 13, junto con varios dispositivos Android, para la captura de vídeo. Estos han proporcionado una solución flexible y accesible para las grabaciones de simulacros en el laboratorio.
- iPad: Una tableta iPad se ha utilizado para tareas de apoyo durante las grabaciones, como visualización o control, dependiendo de los requisitos específicos de cada sesión. También en momentos se ha utilizado su cámara para la grabación.

- Cámaras del equipo de MOBIOs: Se han empleado cámaras pertenecientes al equipo de MOBIOs para las grabaciones en el laboratorio. Si bien estas cámaras son de alta velocidad para otras necesidades del laboratorio, en el contexto de este proyecto las grabaciones se realizaron a velocidad normal para la implementación y evaluación de Pose2Sim.
- Trípodes: Han sido elementos indispensables para asegurar la estabilidad de todas las cámaras y smartphones durante las grabaciones y el proceso de calibración. Su inmovilidad era crítica, ya que el más mínimo movimiento podría invalidar la calibración de las cámaras, afectando directamente la precisión del sistema de detección de pose.
- Ordenadores: Para el desarrollo y la ejecución de los algoritmos se ha utilizado un portátil MacBook Pro M1 con 16GB de RAM. Adicionalmente, se ha contado con un ordenador de sobremesa personal (especificaciones a añadir: [ej: Procesador Intel Core i5-8600k, Tarjeta Gráfica Radeon RX580, 16 GB RAM]) para tareas que requerían mayor capacidad de procesamiento, aunque finalmente se acabó utilizando como ordenador principal por un problema que explicaremos más adelante. También se han utilizado los ordenadores del laboratorio de MOBIOs, los cuales estaban directamente conectados a las cámaras para la adquisición y gestión de las grabaciones.

2.5 INSTALACIONES Y RECURSOS DE MOBIOs

El laboratorio Comillas MOBIOs LAB ha proporcionado el entorno y los recursos necesarios para la realización de los experimentos y el desarrollo del proyecto.

- Instalaciones de MOBIOs: Las instalaciones de este laboratorio de investigación sobre movilidad, mecánica y salud han sido el lugar principal para la realización de las grabaciones de los simulacros de accidentes. Dichas instalaciones cuentan con el equipamiento necesario, incluyendo los ordenadores conectados a las cámaras y la infraestructura para realizar el trigger de los vídeos.

- Datos de Grabaciones Propias: Debido a limitaciones legales y de compatibilidad con los datos históricos de MOBIOS, el proyecto ha procedido a generar sus propias grabaciones de simulacros en formato de vídeo en las instalaciones del laboratorio. Estos vídeos propios han servido como la base para la evaluación comparativa y la determinación de la precisión de los algoritmos sin marcadores.

La integración de todas estas tecnologías, desde los entornos de programación y las librerías especializadas hasta el equipamiento de grabación y el soporte de las instalaciones de MOBIOS, ha sido determinante para llevar a cabo la evaluación comparativa de los algoritmos de posicionamiento en 3D sin marcadores y determinar su viabilidad como alternativa o complemento a los sistemas tradicionales.

Capítulo 3. ESTADO DE LA CUESTIÓN

El presente capítulo tiene como objetivo establecer el marco teórico y contextualizar el trabajo de investigación dentro del conocimiento existente en el ámbito de la captura de movimiento y la estimación de pose humana. Se revisarán los enfoques tradicionales y las innovaciones recientes, identificando la brecha de conocimiento que este proyecto busca abordar.

3.1 INTRODUCCIÓN A LA CAPTURA DE MOVIMIENTO

La captura de movimiento (MoCap) es una tecnología que permite registrar y analizar el movimiento de personas u objetos en el espacio tridimensional. Su objetivo principal es transformar el movimiento del mundo real en datos digitales que pueden ser utilizados para diversas aplicaciones. Aunque sus orígenes se remontan a principios del siglo XX con estudios de biomecánica y cronofotografía, ha experimentado un crecimiento exponencial con los avances tecnológicos, especialmente en las últimas décadas. (Moeslund, 2006)

Tradicionalmente, la captura de movimiento se ha basado en sistemas ópticos con marcadores, donde pequeños puntos reflectantes o emisores se colocan en el cuerpo del sujeto. Cámaras especializadas detectan la posición de estos marcadores en el espacio, y algoritmos de triangulación reconstruyen su trayectoria 3D. Este enfoque, aunque altamente preciso y considerado el "estándar de oro" en muchos campos como la biomecánica clínica o el cine, conlleva ciertas desventajas, como el elevado coste del equipo, la necesidad de un entorno de laboratorio controlado, y la intrusividad y el tiempo que requiere la preparación del sujeto.

Las aplicaciones de la captura de movimiento son vastas y multidisciplinarias. En el ámbito de la biomecánica, permite un análisis detallado de la marcha, la postura y la ejecución de movimientos específicos, siendo fundamental en la investigación médica, la rehabilitación y el rendimiento deportivo. En la industria del entretenimiento, es indispensable para la

animación de personajes en videojuegos y películas. También encuentra utilidad en la realidad virtual y aumentada, la robótica, el diseño ergonómico y la seguridad, entre otros campos.

La estimación de pose humana (HPE), un subcampo dentro de la captura de movimiento sin marcadores, se ha convertido en un área de investigación muy activa. Los avances en redes neuronales convolucionales han permitido que los algoritmos de HPE alcancen niveles de precisión impresionantes a partir de imágenes o vídeos estándar, abriendo nuevas posibilidades para aplicaciones que requieren un seguimiento del movimiento en entornos no controlados o sin la preparación intensiva que demandan los sistemas tradicionales. (Ce Zheng, 2020)

3.2 SISTEMAS DE CAPTURA DE MOVIMIENTO BASADOS EN MARCADORES

Tradicionalmente, los sistemas de estimación de pose humana (Human Pose Estimation, HPE) han dependido de un conjunto de cámaras especializadas y puntos reflectantes adheridos al cuerpo del sujeto. Estos marcadores son capturados por las cámaras, permitiendo la reconstrucción precisa del movimiento del objeto analizado en un espacio tridimensional (Smirnova, 2022). Un ejemplo prominente de esta tecnología es el sistema Vicon, ampliamente utilizado en laboratorios como el MOBIOS de la Universidad Pontificia Comillas, el cual, según la observación directa y la información interna del laboratorio, emplea hasta 10 cámaras para capturar la dinámica de impacto en simulaciones de accidentes. Este tipo de técnicas de análisis de movimiento basado en marcadores, son un estándar de la industria de la investigación y ampliamente utilizados en diversas funcionalidades, también en el estudio de accidentes. (Yoganandan, 1998).

Si bien estos sistemas basados en marcadores ofrecen una alta precisión y son considerados el estándar de oro en muchos campos, presentan una serie de limitaciones significativas. Entre ellas se incluyen el elevado coste del equipamiento, el tiempo considerable requerido para la preparación del sistema (colocación y calibración de marcadores), la intrusividad

para el usuario y la necesidad de operar en entornos de laboratorio controlados. Además, como se ha observado en el contexto de este proyecto, la caída de marcadores durante simulaciones (Federolf, 2013) o la dificultad para identificar manualmente cada uno de ellos en vídeos largos, pueden generar ineficiencias y errores en el proceso de análisis. La imposibilidad de colocar marcadores en ciertas áreas (como la cadera con el cinturón de seguridad) también limita su aplicación en escenarios específicos. Existen alternativas como las que presenta el siguiente estudio utilizando sistemas casi infrarrojos. (Metzler, 2007)

3.3 AVANCES EN LA DETECCIÓN DE POSE SIN MARCADORES

En respuesta a las limitaciones de los sistemas basados en marcadores, ha surgido y evolucionado rápidamente la estimación de pose humana sin marcadores. Este enfoque se centra en explorar algoritmos capaces de detectar las articulaciones del cuerpo humano directamente a partir de imágenes estáticas o vídeos capturados por una o varias cámaras convencionales (Zheng, 2023). La principal ventaja de estos métodos radica en su no intrusividad, la reducción de costes de equipamiento y la mayor flexibilidad para su aplicación en diversos entornos.

Las técnicas de estimación de pose sin marcadores se pueden clasificar en dos enfoques principales: estimación de poses en 2D y estimación de poses en 3D, con subcategorías específicas para cada caso (Zheng, 2023).

- **Métodos de regresión:** Estos enfoques predicen directamente las coordenadas de las articulaciones a partir de las imágenes de entrada. DeepPose es un ejemplo notable de este método, que utiliza redes neuronales profundas para mapear píxeles de imagen a coordenadas de *keypoints* (Toshev, 2014). Otros trabajos relevantes en esta línea incluyen las contribuciones de (Mao, 2021)
- **Métodos basados en mapas de calor:** Predicen mapas de calor que indican la probabilidad de que una articulación se encuentre en una ubicación específica de la imagen. La *Stacked Hourglass Network* (Newell, 2016) es un ejemplo influyente de

este método, que ha demostrado una gran eficacia en la detección precisa de *keypoints*.

- **Métodos de levantamiento de 2D a 3D:** Estos enfoques primero estiman la pose en 2D y, posteriormente, la convierten al espacio 3D utilizando modelos geométricos o redes neuronales adicionales. Trabajos destacados en esta área incluyen las investigaciones de Julieta Martínez (Martínez, 2017) y Bugra Tekin (Tekin, 2017), que exploran cómo inferir la profundidad y la orientación 3D a partir de proyecciones 2D.

3.4 POSE2SIM EN EL CONTEXTO DE LA INVESTIGACIÓN ACTUAL

En este marco de avances, el proyecto Pose2Sim ha emergido como una solución integral para la cinemática 3D sin marcadores. Desarrollado por David Pagnon, Mathieu Domalain y Lionel Reveret, Pose2Sim es un flujo de trabajo de código abierto que integra herramientas como OpenPose y OpenSim para convertir datos de coordenadas 2D (obtenidos de vídeos) en modelos de movimiento 3D (Pagnon, Pose2Sim: an end-to-end workflow for 3D markerless sports kinematics—part 1: robustness, 6530)

- **OpenPose:** Como componente clave dentro de Pose2Sim, OpenPose es un modelo de aprendizaje profundo que se utiliza para la detección inicial de los *keypoints* articulares en 2D a partir de los fotogramas de vídeo. Su capacidad para identificar con precisión las posiciones de las articulaciones humanas en imágenes es fundamental para el procesamiento posterior de Pose2Sim (Cao, 2019)
- **OpenSim:** Este software de simulación musculoesquelética es utilizado por Pose2Sim para modelar movimientos en 3D y generar visualizaciones. Permite superponer un modelo esquelético a las imágenes de vídeo, facilitando el análisis del efecto del impacto en las estructuras óseas y la representación de los movimientos detectados.

Existen publicaciones relevantes que describen en detalle la robustez y la precisión de Pose2Sim. Un estudio inicial se centró en la robustez del algoritmo, evaluando su capacidad

para mantener la consistencia bajo diversas condiciones (Pagnon, Pose2Sim: an end-to-end workflow for 3D markerless sports kinematics—part 1: robustness, 6530). Una segunda parte de la investigación profundizó en la precisión de Pose2Sim, comparando sus resultados con los de sistemas de captura de movimiento basados en marcadores (Pagnon, Pose2Sim: an end-to-end workflow for 3D markerless sports kinematics—part 2: accuracy, 2022). Estos estudios han sido de gran ayuda para comprender la estructura y las capacidades de Pose2Sim.

3.5 BRECHA DE INVESTIGACIÓN Y JUSTIFICACIÓN DEL PRESENTE TRABAJO

A pesar de los avances significativos en la estimación de pose sin marcadores y la existencia de herramientas prometedoras como Pose2Sim, la aplicación y evaluación de estas tecnologías en contextos específicos, como las simulaciones de accidentes y la biomecánica de impacto, aún requiere una validación exhaustiva. Si bien los estudios existentes sobre Pose2Sim han demostrado su robustez y precisión en ámbitos como el deporte, es crucial determinar si su rendimiento es suficiente para reemplazar o complementar los métodos tradicionales en las exigentes condiciones de un laboratorio como MOBIOS, donde la precisión es un factor crítico para el desarrollo de sistemas de seguridad.

Este proyecto se justifica al abordar la necesidad de evaluar la precisión de Pose2Sim utilizando grabaciones propias de simulacros realizadas en el laboratorio de MOBIOS. A diferencia de estudios previos que podrían basarse en conjuntos de datos públicos o en entornos deportivos, este trabajo ofrece una evaluación práctica en un escenario relevante para la investigación en seguridad y movilidad. La imposibilidad de utilizar datos de referencia históricos de MOBIOS, debido a limitaciones legales y de compatibilidad, ha reforzado la necesidad de generar un conjunto de datos propio, lo que a su vez permite una evaluación más controlada y adaptada a las particularidades del laboratorio.

En resumen, el presente trabajo contribuye al campo de la captura de movimiento al proporcionar una evaluación empírica de una alternativa sin marcadores en un contexto de

aplicación crítica. El objetivo principal de este análisis preliminar es ofrecer al laboratorio MOBIOs una base sólida para decidir si dedicar recursos adicionales a la investigación y desarrollo de esta tecnología. Independientemente de su capacidad para sustituir completamente a los sistemas Vicon en términos de precisión, este sistema ofrece un valor significativo al permitir la creación de ilustraciones y visualizaciones de alta calidad de manera considerablemente más sencilla y rápida, superando la complejidad y el tiempo de procesamiento de varios días que requieren los datos de los sistemas Vicon.

3.6 ACLARACIÓN CON RESPECTO A OPENPOSE

A lo largo del presente Trabajo de Fin de Grado (TFG), se ha hecho referencia a OpenPose como una de las soluciones pioneras y más influyentes en el campo de la estimación de pose 2D sin marcadores. Su relevancia radica en haber sido uno de los primeros frameworks en lograr una estimación robusta y en tiempo real de múltiples personas, con una amplia adopción tanto en la investigación como en aplicaciones prácticas. Debido a su impacto histórico y a la familiaridad que el término "OpenPose" ha adquirido en el ámbito de la cinemática sin marcadores, a menudo se utiliza como un término genérico o un punto de referencia para describir este tipo de tecnología.

Sin embargo, el framework Pose2Sim, utilizado en este proyecto, es una plataforma que integra y es compatible con múltiples modelos de estimación de pose, no solo con OpenPose. Aunque OpenPose puede ser una opción válida dentro de Pose2Sim (como se observa en las opciones de `output_format` y comentarios en el `config.toml` que hacen referencia a sus esqueletos `BODY_25B`, `COCO`, etc.), el `config.toml` del proyecto especifica el uso de RTMPose.

¿Por qué la elección de RTMPose en lugar de OpenPose para este proyecto?

- **Evolución Tecnológica y Rendimiento:** La estimación de pose es un campo en constante evolución. RTMPose, desarrollado por OpenMMLab, es un modelo más reciente y, en general, ofrece mejoras significativas en términos de velocidad,

precisión y robustez en comparación con OpenPose, especialmente en arquitecturas de hardware modernas (como las GPU). Esto se debe a avances en las arquitecturas de redes neuronales, métodos de entrenamiento y optimización de inferencia. Para los objetivos de este TFG, que buscan una alta precisión en la reconstrucción 3D y cinemática, el uso de un modelo más avanzado como RTMPose es ventajoso.

- Integración Nativa en Pose2Sim: Como se indica en los comentarios del config.toml, "Only RTMPose is natively embeded in Pose2Sim. For all other pose estimation methods, you will have to run them yourself, and then refer to the documentation to convert the output files if needed". Esto significa que RTMPose está directamente optimizado e integrado en el flujo de trabajo de Pose2Sim, simplificando su implementación y reduciendo la necesidad de pasos intermedios de conversión de datos. Esta integración nativa facilita el despliegue del proyecto y la reproducibilidad de los resultados.
- Flexibilidad de Pose2Sim: La mención de OpenPose en la configuración y la documentación de Pose2Sim (output_format = 'openpose') subraya la flexibilidad del framework. Aunque el motor de estimación de pose principal en este proyecto sea RTMPose, Pose2Sim puede interoperar con formatos de datos generados por otros sistemas, incluyendo OpenPose. Esto es útil para:
 - Compatibilidad con datos existentes: Si se tuvieran datos de pose 2D previamente generados con OpenPose, Pose2Sim podría importarlos y utilizarlos en sus fases posteriores (sincronización, triangulación, etc.).
 - Normalización de formatos: Al permitir la exportación en formato OpenPose, Pose2Sim facilita la compatibilidad con otras herramientas o flujos de trabajo que sí esperan dicho formato.

Por todo ello, para el lector no familiarizado con la terminología específica del campo de la Visión por Computador, es importante aclarar la distinción entre el campo de la Estimación de Pose Humana (HPE, Human Pose Estimation) y las implementaciones algorítmicas concretas dentro del mismo. El HPE se refiere a la disciplina y el objetivo general de detectar

y localizar los keypoints corporales en imágenes o vídeos. Dentro de este campo, han surgido numerosos algoritmos y frameworks.

En el presente TFG, si bien se reconoce la relevancia histórica y el impacto de OpenPose como una de las soluciones pioneras y más influyentes en HPE, y se ha optado por el modelo RTMPose para la estimación de pose 2D debido a sus características avanzadas y su integración nativa en Pose2Sim, a lo largo de este documento se podrán encontrar referencias indistintas a "OpenPose" y "RTMPose". Esta práctica busca simplificar la lectura y comprensión del flujo de trabajo, enfatizando que ambos términos, en el contexto de este trabajo, hacen referencia a la función principal de la detección de puntos clave del cuerpo humano a partir de datos visuales, necesaria para el análisis cinemático sin marcadores. De esta manera, el término HPE conservará su sentido más amplio como el campo de estudio, mientras que "OpenPose" y "RTMPose" se emplearán para describir la herramienta algorítmica utilizada para la detección específica en este proyecto.

Capítulo 4. DEFINICIÓN DEL TRABAJO

Este capítulo establece la definición precisa del proyecto, detallando su justificación técnica y de mercado, los objetivos específicos que se persiguen, la metodología de trabajo adoptada y una visión general de la planificación y los recursos económicos asociados.

4.1 JUSTIFICACIÓN

A diferencia de la motivación inicial, que aborda la relevancia general de las soluciones sin marcadores, la justificación de este proyecto se centra en un análisis crítico de los trabajos previos y en la necesidad técnica y operativa de desarrollar esta investigación. El laboratorio MOBIOS de la Universidad Pontificia Comillas busca optimizar sus recursos y mejorar la eficiencia en la obtención de datos para sus patrocinadores y financiadores. En este contexto, los sistemas tradicionales de captura de movimiento basados en marcadores, como Vicon, presentan inconvenientes significativos que el presente proyecto busca abordar:

Coste y Eficiencia Operativa: El procesamiento de datos de los sistemas Vicon es un proceso que consume una cantidad considerable de tiempo y recursos humanos. Habitualmente, se requiere la dedicación de un becario o aprendiz durante días o incluso semanas para la colocación manual y el nombramiento de los puntos en los marcadores, lo que ralentiza drásticamente la obtención de resultados de un único ensayo de accidente. Este elevado coste en horas de personal y el retraso en la disponibilidad de los datos no es práctico para la dinámica de un laboratorio de investigación que necesita presentar resultados de manera eficiente.

Problemas de Detección Específicos: La precisión en la detección de ciertas partes del cuerpo, como la cadera, se ve comprometida en los sistemas con marcadores debido a la presencia de elementos externos como los cinturones de seguridad en simulaciones de

accidentes. Esto dificulta la colocación de marcadores anclados a los huesos de la cadera, limitando la capacidad de obtener una visión completa de la reacción del cuerpo al impacto.

Costes Asociados: Aunque el proyecto no ha requerido un presupuesto formal, se han incurrido en costes directos de transporte para los desplazamientos al laboratorio MOBIOS y en la impresión de un tablero de calibración de alta calidad, ascendiendo a un total aproximado de 75€ (60€ para el tablero y 15€ para gasolina). Estos costes, aunque modestos, se suman a las horas de dedicación del personal del laboratorio que colaboró en las grabaciones y al desgaste inherente del hardware utilizado.

Al inicio del proyecto, no se identificaron alternativas directas a Pose2Sim que pudieran realizar el trabajo de estimación de pose en 3D con la misma integración y funcionalidad. Si bien existen soluciones muy distintas con algoritmos diferentes para la estimación de pose en 2D, el algoritmo predeterminado de Pose2Sim resultó ser más que suficiente para los objetivos planteados. Se dedicaron aproximadamente cinco horas a analizar un plugin para Blender que prometía una interfaz de usuario más sencilla para Pose2Sim, eliminando la necesidad de usar la consola de comandos. Sin embargo, este plugin presentaba errores, carecía de continuidad y soporte por parte de su desarrollador original, lo que llevó a su descarte.

4.2 OBJETIVOS DEL PROYECTO

Los objetivos de este proyecto se mantienen como se definieron inicialmente, buscando una evaluación pragmática de la tecnología de estimación de pose sin marcadores en el contexto del laboratorio MOBIOS:

OBJETIVOS PRINCIPALES:

Evaluar la precisión de Pose2Sim: Determinar si la precisión de los algoritmos implementados en Pose2Sim es suficientemente buena en comparación con los sistemas de captura de movimiento tradicionales. La métrica de "suficientemente buena" se evaluará mediante un umbral de error bajo, utilizando el error cuadrático medio o métricas similares,

aunque sin un valor específico predefinido, sino buscando una aproximación razonable al comportamiento real del sujeto y a los valores reales de los huesos humanos.

Incrustar un modelo esquelético: Superponer un modelo esquelético a las imágenes de vídeo utilizando los keypoints detectados para realizar visualizaciones del movimiento. El propósito es analizar visualmente el efecto del impacto en las estructuras óseas a través de las representaciones generadas por OpenSim y Blender, sin implicar un análisis biomecánico complejo adicional.

OBJETIVOS SECUNDARIOS:

Calcular la posición de la cadera: Determinar la posición de la cadera a partir del modelo esquelético incrustado. Este objetivo es relevante debido a la dificultad inherente de detectar esta articulación con los sistemas Vicon, y permitirá observar la reacción completa del cuerpo durante un accidente.

Estudiar la posibilidad de sustituir o complementar el sistema Vicon: Analizar la viabilidad de que el sistema basado en Pose2Sim pueda, en el futuro, sustituir o complementar el sistema actual de captura de movimiento basado en cámaras Vicon en el laboratorio MOBIO. Este TFG se concibe como un análisis preliminar para que el laboratorio pueda tomar una decisión informada sobre la dedicación de recursos a esta tecnología.

4.3 METODOLOGÍA DE TRABAJO

Para el desarrollo de este proyecto, se adoptó una metodología ágil, con un enfoque pragmático en la gestión de tareas y la resolución de problemas a medida que surgían. Aunque se planteó inicialmente una versión reducida de Scrum con reuniones quincenales y gestión en Microsoft Planner/Teams, la aplicación de estas herramientas fue más bien flexible, adaptándose a las necesidades del proyecto. Las metodologías SMART y los "siete

pasos de McKinsey" se utilizaron como herramientas de organización y resolución de problemas, aplicándose de manera flexible en función de los desafíos encontrados.

Las principales tareas y el flujo de trabajo seguido fueron los siguientes:

- Investigación del estado del arte: Se realizó una investigación exhaustiva sobre los algoritmos de estimación de pose humana, sus posibilidades y funcionamiento, sentando las bases teóricas del proyecto.
- Instalación, configuración y replicación de ejemplos de Pose2Sim: Se procedió a instalar y configurar el entorno de Pose2Sim. Se replicaron los tutoriales y ejemplos proporcionados por la herramienta para comprender sus pasos y su funcionamiento, lo que permitió adaptarlos posteriormente a los vídeos propios.
- Generación de vídeos en MOBIO: Dada la imposibilidad de utilizar los datos de referencia históricos de MOBIO (debido a limitaciones legales y de compatibilidad, ya que nunca fueron recibidos), se realizaron grabaciones propias de simulacros en las instalaciones del laboratorio. Esta situación ralentizó el trabajo y requirió múltiples desplazamientos para la captura de datos.
- Calibración de cámaras para Pose2Sim: Se calibraron los dispositivos de grabación (teléfonos móviles y cámaras de MOBIO) específicamente para su uso con Pose2Sim. A diferencia de los sistemas Vicon que utilizan ficheros de calibración específicos y no graban vídeo, se realizó una calibración adaptada a las necesidades de Pose2Sim, garantizando que los trípodes permanecieran completamente inmóviles durante todo el proceso, ya que el más mínimo movimiento invalidaría la calibración.
- Sincronización de vídeos con DaVinci Resolve: Los vídeos grabados con las diferentes cámaras se sincronizaron utilizando DaVinci Resolve para asegurar que todas las grabaciones comenzaran y terminaran simultáneamente, un paso crítico para el análisis cinemático posterior.

- Ejecución de los algoritmos para la detección de keypoints: Se aplicó Pose2Sim a los vídeos de los ensayos grabados en MOBIO para detectar los keypoints articulares.
- Inserción de un modelo esquelético y visualización: Con los keypoints obtenidos de los vídeos, se generó un modelo esquelético que se pudo representar en programas como Blender y OpenSim, permitiendo la visualización de los movimientos en 3D.
- Análisis de la precisión: La comparación de las coordenadas de los keypoints detectados se realizó midiendo la longitud de los huesos. Dado que no se pudo acceder a ensayos reales con datos Vicon de referencia, la verificación se centró en asegurar que la longitud de un hueso específico se mantuviera constante a lo largo de todos los fotogramas y que, de manera aproximada, el tamaño coincidiera con el de la persona real grabada.
- Redacción del TFG: Paralelamente a las fases técnicas, se llevó a cabo la redacción del Trabajo de Fin de Grado.

4.4 PLANIFICACIÓN Y ESTIMACIÓN ECONÓMICA

Este apartado detalla la estimación económica del presente Trabajo de Fin de Grado, considerando tanto los costes directos como la valoración de los recursos no monetizados, como el equipamiento informático y las horas de dedicación.

PLANIFICACIÓN Y DESVIACIONES

La planificación inicial del proyecto se representó mediante un diagrama de Gantt. Sin embargo, la ejecución real experimentó desviaciones significativas. La principal de ellas fue la imposibilidad de ejecutar el modelo Pose2Sim en el portátil MacBook Pro personal en Madrid. Este imprevisto obligó a trasladar el trabajo a un ordenador de sobremesa ubicado en Coruña, generando un retraso de dos a tres meses en la fase de implementación. Esta demora se debió a la dificultad para ubicar la falla del programa, descubriendo finalmente que funcionaba correctamente en el sobremesa con sistema operativo Windows. Consecuentemente, este problema retrasó el resto de las fases del proyecto en

aproximadamente un mes. Adicionalmente, la tarea de estudiar los datos de referencia de MOBIOs fue eliminada del cronograma al no poder acceder a ellos.

COSTES DIRECTOS DEL PROYECTO

El proyecto no requirió un presupuesto formal extenso, y los costes económicos directos se limitaron a la adquisición de materiales específicos, los gastos de transporte y el desgaste del vehículo personal:

- Materiales: 60€ para la impresión de un tablero de calibración en cartón pluma de alta calidad.
- Transporte (Combustible): Aproximadamente 15€ en gasolina para los desplazamientos al laboratorio MOBIOs.
- Desgaste de vehículo (Moto): Considerando una moto con un coste de 5700€ y una vida útil estimada de 100.000 km, el desgaste por kilómetro es de 0.057€/km. Para los 150 km recorridos para el proyecto, el coste estimado por desgaste es de $0.057€/km \times 150 km = 8,55€$.

VALORACIÓN DE RECURSOS NO MONETIZADOS

Para ofrecer una estimación más completa del valor del proyecto, se ha realizado una valoración de los recursos no monetizados, incluyendo la amortización del equipamiento informático y la estimación de las horas de trabajo.

4.4.1.1 Equipamiento informático

Para la implementación y desarrollo del proyecto, se ha utilizado el siguiente material informático personal:

- Ordenador portátil: Modelo MacBook Pro M1, adquirido de segunda mano por 1000€.
- Ordenador de sobremesa: Ensamblado con piezas en 2018, con un coste total aproximado de 1600€.

Considerando una vida útil estimada de 5 años (60 meses) para ambos equipos, y el tiempo de uso efectivo dedicado al proyecto:

Amortización del portátil MacBook Pro M1:

- Coste mensual: $1000\text{€}/60 \text{ meses} = 16,67\text{€/mes}$
- Uso efectivo en el proyecto: 10 meses
- Coste imputable al proyecto: $16,67\text{€/mes} \times 10 \text{ meses} = 166,70\text{€}$

Amortización del ordenador de sobremesa:

- Coste mensual: $1600\text{€}/60 \text{ meses} = 26,67\text{€/mes}$
- Uso efectivo en el proyecto: 6 meses
- Coste imputable al proyecto: $26,67\text{€/mes} \times 6 \text{ meses} = 160,02\text{€}$

Amortización total del equipamiento informático imputable al proyecto:

$$166,70\text{€}(\text{MacBook Pro}) + 160,02\text{€}(\text{Sobremesa}) = 326,72\text{€}$$

4.4.1.2 Estimación de Horas de Trabajo

El proyecto ha requerido una dedicación aproximada de 330 horas, correspondientes a los 12 créditos ECTS asignados al Trabajo de Fin de Grado.

Tomando como referencia el salario medio aproximado de un ingeniero junior en España (15€/hora) (Glassdoor, s.f.), el coste equivalente en términos de capital humano se estima en:

$$330 \text{ horas} \times 15\text{€/hora} = 4.950,00\text{€}$$

4.4.1.3 Resumen de Costes estimados

A continuación, se presenta un resumen de los principales costes estimados del proyecto, considerando tanto los costes directos como la valoración de los recursos no monetizados.

Concepto	Valor Estimado
Amortización de equipos informáticos	326,72€
Valoración horas de trabajo	4.950€
Costes directos (materiales, transporte y desgaste de vehículo)	83,55€
Total Estimado	5.360€

Tabla 1 - Estimación económica

Capítulo 5. SISTEMA/MODELO DESARROLLADO

Este capítulo detalla la arquitectura del sistema utilizado para la estimación de pose humana sin marcadores, centrándose en el flujo de trabajo de Pose2Sim y los componentes clave que lo integran. Se explicará el funcionamiento de cada fase del algoritmo, la configuración del entorno de trabajo, la gestión de archivos y las opciones de personalización.

5.1 ANÁLISIS DEL SISTEMA

INTRODUCCIÓN AL FLUJO DE TRABAJO DE POSE2SIM

En esta sección, se presenta una visión de alto nivel del funcionamiento de Pose2Sim, delineando los pasos fundamentales que componen su proceso de estimación de pose 3D. Un análisis más detallado de cada una de estas fases se abordará en el apartado 0Fases del Algoritmo Pose2Sim: Un Recorrido Detallado.0

La explicación se fundamenta en el diagrama de flujo de Pose2Sim, representado en la Ilustración 3. Esta figura, de origen externo y publicada en el repositorio oficial de GitHub de Pose2Sim, ilustra de manera concisa la secuencia de operaciones del algoritmo.

El proceso se inicia con la captura de grabaciones de vídeo realizadas desde múltiples cámaras. Pose2Sim permite la utilización de un número variable de cámaras, siendo dos el mínimo indispensable para la triangulación 3D y tres el mínimo recomendado para asegurar una triangulación robusta y de alta calidad, lo que contribuye a una mayor fiabilidad en la reconstrucción espacial.

Posteriormente, a cada una de estas grabaciones se le aplica un algoritmo de Estimación de Pose Humana (HPE) en 2D. Estos algoritmos, que se detallarán en el apartado 5.2.1.16 Estimación de Pose 2D, son capaces de identificar con precisión los keypoints (puntos clave) del cuerpo humano en cada fotograma de vídeo. El resultado de esta fase es una serie de ficheros JSON, uno por cada fotograma, que contienen las coordenadas 2D de estos

keypoints detectados en cada vista de cámara. Estos ficheros son necesarios para el procesamiento subsiguiente, ya que codifican la información espacial bidimensional que será la base para la reconstrucción 3D.

Una vez completada la estimación de pose 2D para todas las cámaras, estos ficheros JSON, junto con el fichero de calibración de las cámaras (que contiene los parámetros intrínsecos y extrínsecos de cada dispositivo de captura), se introducen en el sistema central de Pose2Sim. En esta etapa, el algoritmo realiza una serie de procesos para transformar los datos 2D en una representación 3D coherente y utilizable:

- Sincronización de cámaras (opcional): Esta fase permite alinear temporalmente los datos provenientes de las diferentes cámaras. Pose2Sim asegura que los fotogramas correspondientes de cada vídeo se procesen de forma conjunta, lo que es beneficioso para una triangulación precisa y para evitar inconsistencias temporales en la reconstrucción del movimiento. Esta sincronización es opcional si las cámaras ya han sido previamente sincronizadas de forma externa. Una sincronización deficiente puede conducir a errores en la posición 3D estimada de los keypoints. El método empleado es descartar los frames que no sean comunes desde todas las cámaras.
- Identificación multipersona (opcional): Cuando en las grabaciones aparecen múltiples individuos, Pose2Sim es capaz de reconocer y seguir a cada persona de forma independiente a lo largo de los fotogramas. Este proceso de asociación de keypoints a individuos específicos es relevante para garantizar que la triangulación 3D se realice correctamente para cada sujeto, evitando la mezcla de datos entre diferentes personas y facilitando el análisis de escenarios complejos. Si solo hay un sujeto en el vídeo, esta función puede desactivarse, y el algoritmo seleccionará automáticamente el valor más fiable.
- Triangulación robusta: Este es el paso más significativo en la reconstrucción 3D. Utilizando las coordenadas 2D de los keypoints detectados desde al menos dos vistas de cámara y los parámetros de calibración, Pose2Sim calcula la posición tridimensional de cada keypoint en el espacio. La robustez de este proceso contribuye a minimizar el impacto del ruido y las imprecisiones inherentes a la detección 2D.

- Filtrado de coordenadas 3D (opcional): Tras la triangulación, los datos 3D pueden contener cierto nivel de ruido o pequeñas inconsistencias. Pose2Sim ofrece la posibilidad de aplicar diferentes filtros (como filtros de paso bajo o suavizado) para mejorar la calidad de la salida. Este filtrado ayuda a reducir las fluctuaciones no deseadas en las trayectorias de los keypoints, resultando en un movimiento más suave y biomecánicamente plausible, lo que es útil para un análisis preciso.
- Marker augmentation (opcional): Esta función avanzada permite estimar la posición de hasta 43 marcadores adicionales que no fueron detectados directamente por el algoritmo HPE. El marker augmentation es particularmente útil cuando se utilizan menos de cuatro cámaras, ya que puede proporcionar una salida más estable y completa del modelo esquelético, aunque es importante señalar que esta interpolación puede, en algunos casos, reducir ligeramente la precisión intrínseca de los keypoints originales. Su principal beneficio radica en la generación de un modelo más denso y apto para ciertos análisis posteriores.

Una vez completado el procesamiento de los ficheros 2D, donde la triangulación con el fichero de calibración constituye el paso fundamental y el resto son opciones recomendadas para optimizar la salida, se obtiene un fichero con extensión .trc (TRaCking file). Este archivo contiene las posiciones 3D de cada uno de los marcadores (ya sean los keypoints originales triangulados o los aumentados) a lo largo del tiempo, representando la cinemática reconstruida del movimiento. El formato .trc es un estándar en biomecánica, lo que facilita su compatibilidad con diversas herramientas de análisis.

El fichero .trc resultante es directamente utilizable para el siguiente paso en el flujo de trabajo: la integración con modelos esqueléticos en OpenSim. Este proceso implica el escalado del modelo de OpenSim para adaptarlo a las dimensiones del sujeto a partir de las coordenadas trianguladas, sin necesidad de un ensayo estático. Posteriormente, se ejecuta la cinemática inversa, donde los marcadores del modelo de OpenSim siguen las posiciones de los marcadores del fichero .trc respetando las restricciones cinemáticas del modelo, lo que permite computar los ángulos articulares.

Como entradas para estas fases finales se utiliza el `config_dict`, un diccionario generado a partir del fichero de calibración `.toml`. Las salidas de este proceso incluyen un modelo `.osim` escalado para cada persona, ficheros de datos de ángulos articulares (`.mot`) para cada individuo, y opcionalmente, los ficheros de configuración de escalado y cinemática inversa de OpenSim guardados en el directorio de cinemática. Además, se generan registros (logs) de Pose2Sim y OpenSim que se guardan en archivos para su revisión y depuración. Esta visualización no solo facilita la interpretación de los datos cinemáticos, sino que también ofrece una herramienta potente para la ilustración y el análisis cualitativo de la interacción del modelo esquelético con el movimiento real del sujeto.

Finalmente, una vez obtenidos los modelos escalados y los datos cinemáticos, es posible visualizar el movimiento del modelo esquelético reconstruido en programas como OpenSim o Blender. Estas herramientas permiten reproducir la secuencia temporal del movimiento, superponiendo el esqueleto a las grabaciones de vídeo originales o presentándolo de forma independiente en un entorno 3D, lo que ofrece una potente capacidad de análisis cualitativo y de ilustración de los resultados cinemáticos obtenidos. Para utilizar blender, es necesario un plugin de Pose2Sim que será indicado posteriormente.

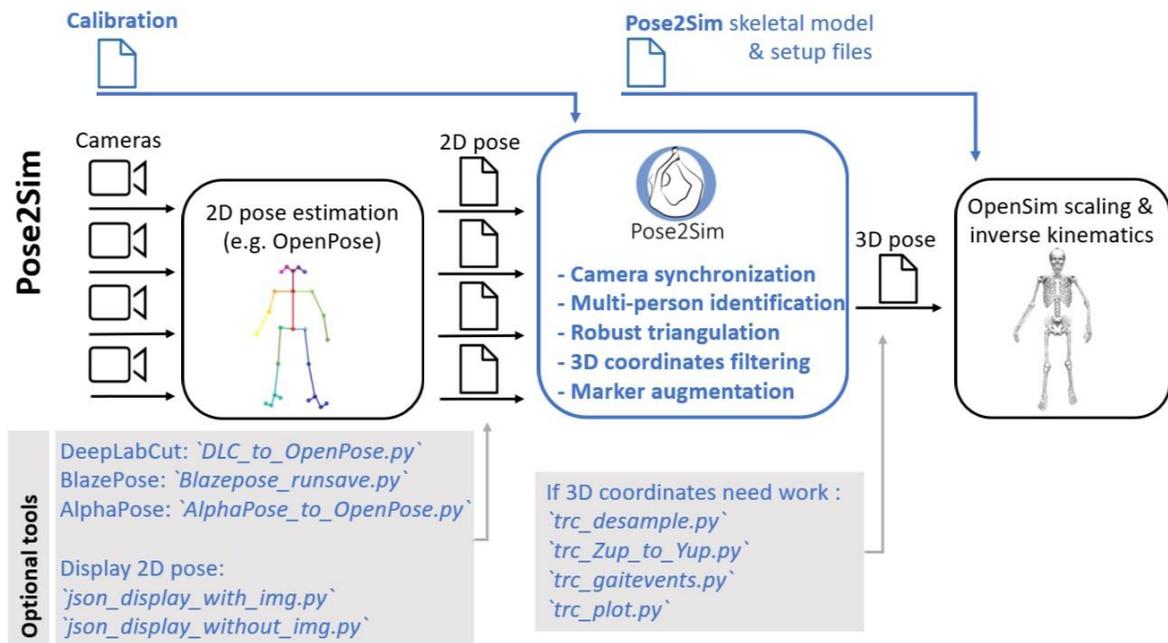


Ilustración 3 - Diagrama de flujo Pose2Sim (Pagnon, Diagrama de funcionamiento de Pose2Sim)

COMPONENTES CLAVE DEL SISTEMA

El sistema de estimación de pose humana sin marcadores, centrado en el uso de Pose2Sim, integra diversas herramientas y plataformas que trabajan de forma conjunta para lograr la reconstrucción y el análisis del movimiento 3D. A continuación, se describen los componentes principales que conforman este ecosistema, aclarando el papel de cada uno en el flujo de trabajo:

- Pose2Sim: Este proyecto, desarrollado por David Pagnon, Mathieu Domalain y Lionel Reveret, es el componente central del sistema. Actúa como un flujo de trabajo integral ("end-to-end") para la cinemática deportiva 3D sin marcadores. Su función principal es integrar y orquestar el procesamiento de las poses 2D obtenidas de las grabaciones de vídeo (utilizando algoritmos como OpenPose), realizar la triangulación para generar datos 3D y facilitar la integración con modelos biomecánicos para el análisis de movimiento. Pose2Sim ha sido objeto de estudios de robustez y precisión, lo que respalda su viabilidad en el ámbito de la biomecánica.

- **OpenPose:** Se emplea como el algoritmo principal para la estimación de pose en 2D a partir de los vídeos capturados. Desarrollado por la Universidad Carnegie Mellon, OpenPose es un modelo de aprendizaje profundo capaz de detectar y localizar puntos clave del cuerpo humano (articulaciones, rasgos faciales, manos) en imágenes y secuencias de vídeo. Los datos de coordenadas 2D generados por OpenPose son la entrada fundamental para el flujo de trabajo de Pose2Sim, siendo utilizado por este último como la base para la posterior reconstrucción tridimensional.
- **OpenSim:** Este software de simulación biomecánica, de código abierto, es crucial para el análisis de movimiento en 3D. Una vez que Pose2Sim ha generado las coordenadas 3D de los keypoints en formato .trc, OpenSim se utiliza para escalar un modelo esquelético genérico a las dimensiones del sujeto estudiado y para realizar la cinemática inversa. Este proceso permite calcular los ángulos articulares y simular el movimiento del modelo 3D, lo que a su vez posibilita la comparación de los resultados con los obtenidos por sistemas de captura de movimiento tradicionales basados en marcadores.
- **Blender:** Como herramienta de software libre, Blender se utiliza para la visualización y representación en 3D de los modelos esqueléticos y los movimientos reconstruidos. A través de un addon específico, Blender permite importar y mostrar los datos 3D generados por OpenSim, superponiendo el modelo articulado sobre los vídeos originales o presentándolo en un entorno virtual, lo que ofrece una potente capacidad de visualización que complementa el análisis cuantitativo de los datos cinemáticos.
- **Python y sus librerías:** Todo el sistema se apoya en el lenguaje de programación Python. Pose2Sim está fundamentalmente basado en Python, lo que facilita la integración de sus componentes y el desarrollo de scripts personalizados para el procesamiento y análisis de datos. Librerías como NumPy o Pandas son esenciales para la manipulación y el cálculo con las grandes cantidades de datos generados.
- **GitHub:** Esta plataforma de desarrollo colaborativo es utilizada para la gestión de versiones y el acceso al código fuente de Pose2Sim y otros algoritmos de código abierto. Su naturaleza de repositorio facilita la colaboración con la comunidad de

desarrolladores y asegura el acceso a las últimas actualizaciones y recursos del proyecto.

Estos componentes, al trabajar en conjunto, forman un sistema robusto para la estimación de pose humana sin marcadores, ofreciendo una alternativa innovadora a los métodos tradicionales de captura de movimiento.

ENTORNO DE DESARROLLO Y LIBRERÍAS

5.1.1.1 Entorno para la Ejecución de Pose2Sim

La ejecución de los algoritmos de Pose2Sim se realizó en un entorno virtual de Python gestionado a través de Anaconda. Anaconda, como distribución de Python y R, facilitó la creación de un entorno aislado donde se instaló la versión específica de Python requerida y, posteriormente, el propio paquete de Pose2Sim. Esta configuración aseguró la compatibilidad de las dependencias y la estabilidad del sistema durante la ejecución de los comandos de Pose2Sim. Es importante destacar que, para esta fase, la interacción con Pose2Sim se realizó directamente a través de la línea de comandos, sin la utilización de Jupyter Notebook, y no se realizaron modificaciones en el código fuente de Pose2Sim.

5.1.1.2 Entorno para el Análisis de Datos y Visualización

Para el análisis posterior de los ficheros .trc generados por Pose2Sim, así como para el cálculo de distancias, errores y la creación de gráficos y visualizaciones, se empleó Jupyter Notebook. Este entorno interactivo, también gestionado a través de Anaconda, permitió un desarrollo flexible y la visualización directa de los resultados del análisis.

En esta fase, se utilizaron las siguientes librerías de Python, fundamentales para el procesamiento y la interpretación de los datos:

- NumPy: Esencial para la computación numérica, NumPy proporcionó las herramientas para el manejo eficiente de arrays y matrices multidimensionales. Su capacidad para realizar operaciones matemáticas de alto rendimiento fue crucial en el cálculo de distancias y otras métricas a partir de las coordenadas 3D de los keypoints.
- Pandas: Esta librería fue indispensable para la manipulación y el análisis de los datos tabulares contenidos en los ficheros .trc. Pandas facilitó la lectura, estructuración y procesamiento de grandes volúmenes de datos, permitiendo operaciones como la extracción de información específica de los keypoints y la preparación para el análisis.
- Plotly: Se empleó para la creación de visualizaciones de datos interactivas y de alta calidad. Plotly permitió generar gráficos que ilustraban las trayectorias de los keypoints, las variaciones en las longitudes de los segmentos óseos, y otras representaciones visuales de los resultados del análisis, contribuyendo a una mejor comprensión y comunicación de los hallazgos del proyecto.

5.2 DISEÑO DEL SISTEMA

ESTRUCTURA DE CARPETAS Y GESTIÓN DE ARCHIVOS

Una comprensión clara de la estructura de carpetas y la gestión de archivos es esencial para trabajar eficientemente con Pose2Sim. El algoritmo organiza sus componentes y los datos de entrada y salida de manera lógica, facilitando el flujo de trabajo desde la calibración hasta la obtención de los resultados finales.

5.2.1.1 Estructura General del Directorio de Pose2Sim

El directorio raíz de Pose2Sim, denominado Pose2Sim, contiene todos los ficheros del entorno del algoritmo. Se organiza en ficheros ejecutables principales, ficheros de utilidades y carpetas de demostración que sirven como plantillas para nuevos proyectos.

- **Ficheros Ejecutables Principales:** Estos scripts de Python representan las distintas fases del flujo de trabajo de Pose2Sim y pueden ser ejecutados individualmente para tareas específicas:
 - `calibration.py`: Gestiona los procesos de calibración intrínseca y extrínseca de las cámaras.
 - `filtering.py`: Aplica filtros a los datos 3D triangulados para mejorar su suavidad y precisión.
 - `kinematics.py`: Realiza el escalado del modelo y la cinemática inversa con OpenSim.
 - `markerAugmentation.py`: Estima y añade marcadores virtuales adicionales a los keypoints detectados.
 - `personAssociation.py`: Asocia los keypoints detectados a individuos específicos, útil en escenarios multipersona.
 - `poseEstimation.py`: Ejecuta los algoritmos de estimación de pose 2D sobre los vídeos.
 - `synchronization.py`: Sincroniza los datos de pose 2D de múltiples cámaras.
 - `triangulation.py`: Reconstruye las coordenadas 3D de los keypoints a partir de sus proyecciones 2D.
- **Ficheros de Utilidades Comunes:**
 - `common.py`: Contiene funciones y utilidades que son compartidas y utilizadas por varios de los scripts principales.
 - `pose2sim.py`: Permite ejecutar de forma secuencial y automatizada todas o varias de las fases del flujo de trabajo de Pose2Sim, evitando la necesidad de ejecutar cada script individualmente.
 - `skeletons.py`: Incluye las definiciones de los modelos esqueléticos utilizados por los diferentes algoritmos de estimación de pose 2D, estandarizando la representación de los keypoints.
- **Carpetas de Utilidades Adicionales:**
 - `utils`: Contiene utilidades generales que apoyan el funcionamiento del sistema.

- marker_augmentation_utils: Incluye funciones específicas para el proceso de marker augmentation.
- Carpetas de Demostración (Demos): Pose2Sim proporciona ejemplos preconfigurados en carpetas como Demo_SinglePerson, Demo_MultiPerson y Demo_Batch. Estas carpetas actúan como plantillas, y su estructura interna debe ser replicada o duplicada para organizar los ficheros de los proyectos propios.

5.2.1.2 Estructura de un Directorio de Proyecto (Ejemplo de Carpeta Demo)

Cada proyecto o ensayo en Pose2Sim se organiza en un directorio específico que sigue una estructura estandarizada. Esta organización es fundamental para que el algoritmo pueda localizar y procesar correctamente los datos en cada una de sus fases. A continuación, se detalla la estructura típica de una de estas carpetas de proyecto:

```
[Nombre_del_Proyecto]/
├── Calibration/
│   ├── Extrinsic/
│   │   ├── ext_cam01_img/
│   │   │   ├── img_0001.jpg
│   │   │   └── ...
│   │   ├── ext_cam02_img/
│   │   │   └── ...
│   │   └── ...
│   ├── Intrinsic/
│   │   ├── int_cam01_img/
│   │   │   ├── img_0001.jpg
│   │   │   └── ...
│   │   ├── int_cam02_img/
│   │   │   └── ...
│   │   └── ...
│   ├── calib_board.toml (Generado tras la calibración)
│   └── object_points.trc (Fichero 3D con las coordenadas de los puntos del
tablero de calibración)
├── Config.toml (Fichero de configuración principal del proyecto)
├── Kinematics/ (Contiene los ficheros .mot, .osim y otros resultados
de la cinemática inversa)
├── Pose/
│   ├── cam01.video1_json/
│   │   ├── cam01.video1_000001.json (Fichero JSON de pose 2D por fotograma)
│   │   └── ...
│   ├── cam01.video1_pos.mp4 (Video con la estimación de pose superpuesta)
│   ├── cam02.video1_json/
│   │   └── ...
└── ...
```

```
|   | cam02.video1_pos.mp4
|   | ...
|---| Pose-3d/           (Almacena los ficheros 3D triangulados y, si aplica,
|   |                   filtrados o aumentados)
|---| Pose-associated/  (Contiene los ficheros JSON de pose 2D tras la
|   |                   asociación de personas)
|---| Pose-sync/       (Almacena los ficheros JSON de pose 2D sincronizados,
|   |                   si se aplica la sincronización interna)
|---| Videos/         (Contiene los videos brutos de las grabaciones)
|   |   | video_cam01.mp4
|   |   | video_cam02.mp4
|   |   | ...
```

Descripción de las Carpetas y Ficheros Clave dentro de un Proyecto: (/ indica carpeta, sin / indica fichero)

- Calibration/: Esta carpeta alberga toda la información relacionada con la calibración de las cámaras.
 - o Extrinsic/: Inicialmente vacía, se debe poblar con subcarpetas para cada cámara (ej., ext_cam01_img, ext_cam02_img), donde se almacenan las imágenes utilizadas para la calibración extrínseca (posición y orientación de las cámaras en el espacio).
 - o Intrinsic/: De manera similar, contiene subcarpetas por cámara (ej., int_cam01_img) con las imágenes para la calibración intrínseca (parámetros internos de la cámara, como la distancia focal y la distorsión de la lente).
 - o calib_board.toml: Este fichero es generado automáticamente al finalizar los procesos de calibración intrínseca y extrínseca, y contiene los parámetros calibrados.
 - o object_points.trc: Fichero que contiene las coordenadas 3D de los puntos de referencia del tablero de calibración, utilizado por el algoritmo para la reconstrucción espacial.
- Config.toml: Es el fichero de configuración principal para cualquier proyecto de Pose2Sim. Debe ser modificado por el usuario para especificar rutas de archivos, parámetros de las cámaras, opciones de los modelos esqueléticos y configuraciones para cada fase del procesamiento. Su contenido se detallará en el siguiente apartado.

- **Kinematics/**: Una vez ejecutado el comando `kinematics.py`, esta carpeta se pobla con los ficheros resultantes del escalado del modelo y la cinemática inversa, como los modelos `.osim` escalados y los ficheros de datos de ángulos articulares (`.mot`), listos para su uso en OpenSim.
- **Pose/**: Tras la ejecución de `poseestimation.py`, esta carpeta contiene subcarpetas con el formato `camX.videoY_json` (ej., `cam01.video1_json`), donde se almacenan los ficheros JSON individuales para cada fotograma (ej., `cam01.video1_000001.json`) con las coordenadas 2D de los keypoints detectados. Además, dentro de esta carpeta, se genera un fichero de vídeo con el formato `camX.videoY_pos.mp4` (ej., `cam01.video1_pos.mp4`), que es el vídeo original de la cámara con la estimación de la pose superpuesta visualmente.
- **Pose-3d/**: Aquí se almacenan los ficheros `.trc` con las coordenadas 3D de los keypoints después de la triangulación. Si se aplican pasos opcionales como el filtrado o el marker augmentation, los resultados se guardarán en esta misma carpeta, generalmente con nombres de archivo diferenciados.
- **Pose-associated/**: Cuando se ejecuta `personassociation.py`, esta carpeta se utiliza para volcar los ficheros JSON donde los keypoints han sido asociados a personas específicas, o donde se han eliminado personas no deseadas de los JSON iniciales.
- **Pose-sync/**: Generada al ejecutar `synchronization.py`. Si los vídeos de entrada no estaban previamente sincronizados, esta carpeta contendrá los ficheros JSON de pose 2D ya temporalmente alineados.
- **Videos/**: Carpeta destinada a almacenar los vídeos brutos de las grabaciones. El formato de los vídeos debe ser especificado en el fichero `Config.toml`.

Esta estructura modular y bien definida permite a Pose2Sim gestionar de manera eficaz los múltiples tipos de datos y los resultados intermedios generados a lo largo de su flujo de trabajo.

PARÁMETROS DE CONFIGURACIÓN (CONFIG.TOML)

El archivo config.toml es el fichero de configuración central de Pose2Sim, actuando como el cerebro del proyecto al definir y controlar el comportamiento de cada fase del flujo de trabajo. Su estructura jerárquica permite establecer parámetros globales para una sesión completa y, si es necesario, sobrescribirlos para participantes o ensayos específicos, ofreciendo una gran flexibilidad y personalización. La modificación de este archivo es indispensable para adaptar el algoritmo a las características particulares de cada conjunto de datos y objetivos de análisis.

A continuación, se presenta un análisis exhaustivo de las secciones principales del config.toml y los parámetros más relevantes, incluyendo la justificación de las configuraciones elegidas para el presente proyecto.

5.2.1.3 Sección [project]

Esta sección define los parámetros generales del proyecto, afectando la gestión global de los datos y el procesamiento:

- multi_person: Este parámetro booleano controla si el análisis se realiza para múltiples participantes (true) o para un único individuo (false).
 - o Configuración del Proyecto: Se ha establecido en false, ya que los ensayos realizados para este TFG involucran a una única persona, simplificando el proceso de identificación y seguimiento. Si se configura como false, Pose2Sim analiza automáticamente a la persona con el menor error de proyección.
- participant_height: Permite especificar la altura del participante en metros. Puede ser 'auto', un valor flotante (ej., 1.72), o una lista de flotantes para múltiples participantes.
 - o Configuración del Proyecto: Se ha mantenido en 'auto', permitiendo que el algoritmo intente estimar la altura a partir de los datos. Este parámetro se utiliza principalmente para el marker augmentation y el escalado del modelo.

- participant_mass: Define la masa del participante en kilogramos. Similar a la altura, puede ser un valor flotante o una lista.
 - Configuración del Proyecto: Se ha establecido en 60.0. Este valor es relevante para el marker augmentation y el escalado, pero no tiene un impacto directo en los resultados cinemáticos si el objetivo no es calcular fuerzas.
- frame_rate: La tasa de fotogramas por segundo (fps). Puede ser un entero o 'auto'.
 - Configuración del Proyecto: Se ha configurado en 'auto', lo que permite a Pose2Sim detectar automáticamente la tasa de fotogramas del vídeo de entrada. En caso de trabajar con imágenes, por defecto se establece en 60 fps.
- frame_range: Define el rango de fotogramas a analizar. Se especifica como una lista (ej., [10, 300]) o vacía ([]) para procesar todos los fotogramas. Si las cámaras no están sincronizadas, este rango designa el intervalo de la cámara con la grabación más corta.
 - Configuración del Proyecto: Se ha dejado como [], indicando que se procesará la totalidad de los vídeos capturados. Para muchas de las pruebas, se comenzó analizando únicamente unos pocos frames [10,100] para agilizar el proceso y detectar errores.
- exclude_from_batch: Una lista de directorios de ensayos que deben ser excluidos de un análisis por lotes.
 - Configuración del Proyecto: Se ha mantenido como [], ya que no se ha realizado un análisis por lotes que requiera exclusiones específicas.

5.2.1.4 Sección [pose]

Esta sección controla los parámetros relacionados con la fase de estimación de pose 2D:

- vid_img_extension: La extensión de los ficheros de vídeo o imagen de entrada (ej., 'mp4', 'png').
 - Configuración del Proyecto: Variaba según las fuentes. Al sincronizarlo en davinci se exportaban en Mp4

- pose_model: Define el modelo de estimación de pose 2D a utilizar. Pose2Sim soporta varios modelos a través de diferentes librerías (RTMLib, MMPose, OpenPose, MediaPipe, AlphaPose, DeepLabCut).
 - Configuración del Proyecto: Se ha utilizado 'Body_with_feet', que corresponde al modelo HALPE_26 por defecto de RTMLib. Este modelo es adecuado para la detección de keypoints corporales completos, incluyendo los pies, lo cual es relevante para el análisis cinemático. Es importante señalar que solo RTMPose está integrado de forma nativa en Pose2Sim; otros métodos requieren su ejecución externa y la conversión de sus ficheros de salida.
- mode: Permite seleccionar el modo de operación del modelo de pose, con opciones como 'lightweight', 'balanced' o 'performance'. También puede ser un diccionario para una selección manual de modelos de detección y estimación de pose.
 - Configuración del Proyecto: Se ha configurado como 'balanced', buscando un equilibrio entre la velocidad de procesamiento y la precisión de la detección.
- det_frequency: Controla la frecuencia con la que se ejecuta la detección de personas (enfoque top-down). Si es mayor que 1, la detección de personas se realiza cada N fotogramas, y entretanto se rastrean las cajas delimitadoras previamente detectadas, lo que puede acelerar el proceso a costa de una posible menor precisión en entornos muy concurridos.
 - Configuración del Proyecto: Se ha fijado en 1, lo que significa que la detección de personas se ejecuta en cada fotograma, asegurando la máxima precisión en el seguimiento del sujeto.
- device: Especifica el dispositivo de hardware a utilizar para la computación (ej., 'CPU', 'CUDA', 'MPS', 'ROCM').
 - Configuración del Proyecto: Se ha dejado en 'auto', permitiendo que Pose2Sim seleccione automáticamente el dispositivo más adecuado disponible (GPU si está presente y configurada, o CPU).

- backend: Define el backend de inferencia a utilizar (ej., 'openvino', 'onnxruntime', 'opencv').
 - o Configuración del Proyecto: Se ha configurado en 'auto', dejando que Pose2Sim elija el backend óptimo.
- display_detection: Booleano que activa la visualización en tiempo real de las detecciones de pose.
 - o Configuración del Proyecto: Se ha establecido en true para facilitar el seguimiento visual del proceso de detección.
- overwrite_pose: Si se establece en false, Pose2Sim no recalculará la estimación de pose si ya se ha realizado previamente para los mismos vídeos.
 - o Configuración del Proyecto: Se ha configurado como false para evitar recálculos innecesarios y acelerar el flujo de trabajo en caso de reejecuciones.
- save_video: Controla cómo se guarda el vídeo de salida con la pose superpuesta ('to_video', 'to_images', 'none', o una lista de ellos).
 - o Configuración del Proyecto: Se ha especificado 'to_video', generando un único archivo de vídeo con la pose visualizada.
- output_format: El formato de salida de los ficheros de pose 2D (ej., 'openpose', 'mmpose').
 - o Configuración del Proyecto: Se ha establecido en 'openpose', ya que es el formato compatible por defecto con el flujo de trabajo interno de Pose2Sim.

5.2.1.5 Sección [synchronization]

Esta sección contiene los parámetros para la sincronización temporal de los vídeos de múltiples cámaras:

- display_sync_plots: Booleano para mostrar gráficos de sincronización.
 - o Configuración del Proyecto: Se mantiene en false salvo para mostrar los gráficos y explicarlos.
- keypoints_to_consider: Permite seleccionar los keypoints a considerar para la sincronización. 'all' para todos, o una lista de nombres (ej., ['RWrist', 'RElbow']) para movimientos específicos.

- Configuración del Proyecto: Se ha mantenido en 'all'.
- approx_time_maxspeed: Tiempo aproximado (en segundos) de un evento vertical claro en cada cámara. 'auto' considera toda la captura.
 - Configuración del Proyecto: Se ha dejado en 'auto'.
- time_range_around_maxspeed: Rango de búsqueda para la correlación alrededor del approx_time_maxspeed.
 - Configuración del Proyecto: 2.0.
- likelihood_threshold: Umbral de confianza de los keypoints. Los puntos con una probabilidad inferior se filtran.
 - Configuración del Proyecto: 0.4.
- filter_cutoff: Frecuencia de corte para el suavizado de series temporales.
 - Configuración del Proyecto: 6.
- filter_order: Orden del filtro para el suavizado.
 - Configuración del Proyecto: 4.

5.2.1.6 Sección [calibration]

Esta sección es fundamental para configurar el proceso de calibración de las cámaras, tanto intrínseca como extrínseca.

- calibration_type: Define el tipo de calibración a realizar: 'calculate' (calcular desde cero) o 'convert' (convertir desde un fichero de calibración de otro sistema).
 - Configuración del Proyecto: Se ha establecido en 'calculate'. Esta elección se justifica por la imposibilidad de utilizar ficheros de calibración de sistemas externos (como Vicon) debido a las limitaciones de acceso a los datos de referencia históricos de MOBIOS, lo que obligó a realizar una calibración propia desde cero.
- [calibration.convert]: Subsección para la conversión de calibraciones de otros sistemas (ej., qualisys, vicon, optitrack).
 - Configuración del Proyecto: No se ha utilizado en este proyecto, ya que la calibración se realizó desde cero.
- [calibration.calculate]: Subsección para el cálculo de calibraciones.

- [calibration.calculate.intrinsics]: Parámetros para la calibración intrínseca (propiedades internas de la cámara).
 - `overwrite_intrinsics`: Si es `false`, no recalcula los parámetros intrínsecos si ya existen.
 - Configuración del Proyecto: `false`, para evitar recálculos si la calibración intrínseca ya se había realizado con éxito.
 - `show_detection_intrinsics`: Muestra el resultado de la detección intrínseca.
 - Configuración del Proyecto: `true`.
 - `intrinsics_extension`: Extensión de las imágenes de calibración intrínseca.
 - Configuración del Proyecto: `'png'`.
 - `extract_every_N_sec`: Si se utiliza un vídeo para la calibración, extrae fotogramas cada N segundos.
 - Configuración del Proyecto: `1`.
 - `intrinsics_corners_nb`: Número de esquinas interiores del tablero de calibración en formato [altura, ancho].
 - Configuración del Proyecto: `[4,7]`.
 - `intrinsics_square_size`: Tamaño de los cuadrados del tablero de calibración en milímetros.
 - Configuración del Proyecto: `90`.
- [calibration.calculate.extrinsics]: Parámetros para la calibración extrínseca (posición y orientación de las cámaras en el espacio).
 - `calculate_extrinsics`: Booleano para activar el cálculo de extrínsecos.
 - Configuración del Proyecto: `true`.
 - `extrinsics_method`: Método para la calibración extrínseca: `'board'` (tablero), `'scene'` (puntos conocidos en la escena), `'keypoints'` (estimación automática de pose).
 - Configuración del Proyecto: `'board'`. Este método fue seleccionado debido a la disponibilidad de un tablero de

- calibración físico impreso y la necesidad de una calibración precisa en un entorno de laboratorio.
- `moving_cameras`: Indica si las cámaras están en movimiento.
 - Configuración del Proyecto: `false`, ya que las cámaras permanecieron fijas durante las grabaciones.
 - `[calibration.calculate.extrinsics.board]`: Parámetros específicos para el método de calibración con tablero.
 - `show_reprojection_error`: Muestra el error de reproyección.
 - Configuración del Proyecto: `true`.
 - `extrinsics_extension`: Extensión de las imágenes de calibración extrínseca.
 - Configuración del Proyecto: `'png'`.
 - `extrinsics_corners_nb`: Número de esquinas del tablero en formato [altura, ancho].
 - Configuración del Proyecto: `[4,7]`.
 - `extrinsics_square_size`: Tamaño de los cuadrados del tablero en milímetros.
 - Configuración del Proyecto: `90`.
 - `[calibration.calculate.extrinsics.scene]`: Parámetros para el método de calibración por escena.
 - `object_coords_3d`: Lista de coordenadas 3D de puntos conocidos en la escena, que se etiquetan manualmente en las imágenes. Estas coordenadas deben estar en metros.
 - Configuración del Proyecto: Esta subsección no fue utilizada en el presente TFG, ya que se optó por el método de calibración con tablero (`'board'`).
 - `[calibration.calculate.extrinsics.keypoints]`: Subsección para el método de calibración por keypoints.
 - Configuración del Proyecto: Este método aún no está implementado en la versión de Pose2Sim utilizada.

5.2.1.7 Sección [personAssociation]

Controla los parámetros para la asociación de keypoints a individuos en escenarios multipersona.

- likelihood_threshold_association: Umbral de confianza para la asociación de keypoints.
 - Configuración del Proyecto: 0.3.
- [personAssociation.single_person]: Parámetros para el seguimiento de una sola persona.
 - reproj_error_threshold_association: Umbral de error de reproyección para la asociación.
 - Configuración del Proyecto: 20 px.
 - tracked_keypoint: Keypoint utilizado para el seguimiento de la persona de interés (ej., 'Neck').
 - Configuración del Proyecto: 'Neck'.
- [personAssociation.multi_person]: Parámetros para el seguimiento de múltiples personas.
 - No se ha utilizado

5.2.1.8 Sección [triangulation]

Define los parámetros para la reconstrucción 3D de los keypoints a partir de las poses 2D.

- reproj_error_threshold_triangulation: Umbral de error de reproyección para la triangulación.
 - Configuración del Proyecto: 15 px.
- likelihood_threshold_triangulation: Umbral de confianza para los keypoints en la triangulación.
 - Configuración del Proyecto: 0.3.

- `min_cameras_for_triangulation`: Número mínimo de cámaras necesarias para la triangulación.
 - o Configuración del Proyecto: 2.
- `interpolation`: Tipo de interpolación para rellenar keypoints faltantes ('linear', 'slinear', 'quadratic', 'cubic', o 'none').
 - o Configuración del Proyecto: 'linear'.
- `interp_if_gap_smaller_than`: No interpola si la brecha de keypoints es mayor que este valor.
 - o Configuración del Proyecto: 10.
- `fill_large_gaps_with`: Cómo rellenar grandes brechas ('last_value', 'nan', o 'zeros').
 - o Configuración del Proyecto: 'last_value'.
- `show_interp_indices`: Booleano para mostrar los índices de los fotogramas interpolados.
 - o Configuración del Proyecto: true.
- `handle_LR_swap`: Opción para manejar el intercambio de extremidades izquierda-derecha.
 - o Configuración del Proyecto: false. Es más útil con pocas cámaras o riesgo de intercambio de miembros.
- `undistort_points`: Si se deben corregir los puntos por distorsión de la imagen.
 - o Configuración del Proyecto: false, asumiendo que las distorsiones de las imágenes son bajas o se han corregido previamente.
- `make_c3d`: Guarda los datos triangulados también en formato .c3d.
 - o Configuración del Proyecto: true.

5.2.1.9 Sección [filtering]

Contiene los parámetros para suavizar los datos 3D triangulados.

- `type`: Tipo de filtro a aplicar ('butterworth', 'kalman', 'gaussian', 'LOESS', 'median', 'butterworth_on_speed').
 - o Configuración del Proyecto: 'butterworth'.
- `display_figures`: Booleano para mostrar figuras de los resultados del filtrado.

- Configuración del Proyecto: true.
- make_c3d: Guarda los datos filtrados también en formato .c3d.
 - Configuración del Proyecto: true.
- [filtering.butterworth]: Parámetros específicos para el filtro Butterworth.
 - order: Orden del filtro.
 - Configuración del Proyecto: 4.
 - cut_off_frequency: Frecuencia de corte en Hz.
 - Configuración del Proyecto: 6 Hz.
- Otras subsecciones para otros tipos de filtros (Kalman, Gaussian, LOESS, Median, Butterworth_on_speed) no se detallan aquí, ya que el filtro Butterworth fue el seleccionado.

5.2.1.10 Sección [markerAugmentation]

Define los parámetros para la adición de marcadores virtuales al modelo.

- make_c3d: Guarda los datos aumentados también en formato .c3d.
 - Configuración del Proyecto: true.

5.2.1.11 Sección [kinematics]

Controla los parámetros para el escalado del modelo y la cinemática inversa con OpenSim.

- use_augmentation: Si se deben usar los marcadores aumentados para el modelo.
 - Configuración del Proyecto: true, para aprovechar el modelo más denso generado por el marker augmentation.
- right_left_symmetry: Booleano para asumir simetría derecha-izquierda del participante.
 - Configuración del Proyecto: true, asumiendo que el participante es simétrico.
- default_height: Altura por defecto en metros si el cálculo automático falla.
 - Configuración del Proyecto: 1.7.

- `remove_individual_scaling_setup`: Si se deben eliminar los ficheros de configuración de escalado individuales.
 - Configuración del Proyecto: `true`, para mantener el directorio de resultados limpio.
- `remove_individual_IK_setup`: Si se deben eliminar los ficheros de configuración de cinemática inversa individuales.
 - Configuración del Proyecto: `true`, para mantener el directorio de resultados limpio.
- `fastest_frames_to_remove_percent`: Porcentaje de fotogramas más rápidos a eliminar (considerados outliers).
 - Configuración del Proyecto: `0.1`.
- `close_to_zero_speed_m`: Umbral de velocidad cercano a cero en m/fotograma para eliminar fotogramas (posibles outliers).
 - Configuración del Proyecto: `0.2`.
- `large_hip_knee_angles`: Ángulos de cadera y rodilla por debajo de este valor se consideran imprecisos.
 - Configuración del Proyecto: `45`.
- `trimmed_extrema_percent`: Proporción de los valores extremos de segmento a eliminar antes de calcular su media para el escalado.
 - Configuración del Proyecto: `0.5`.

5.2.1.12 Sección [logging]

Parámetros para la configuración del registro de eventos.

- `use_custom_logging`: Si se utiliza un sistema de registro personalizado.
 - Configuración del Proyecto: `false`.

5.2.1.13 Sección [pose.CUSTOM]

Esta sección es una plantilla para definir esqueletos personalizados. Si se utiliza un modelo de pose con keypoints diferentes o en un orden distinto a los predefinidos, se puede crear una nueva definición de esqueleto aquí. También se indica que, para la cinemática inversa con un esqueleto personalizado, se necesita un modelo OpenSim compatible con las ubicaciones de los keypoints triangulados.

- Configuración del Proyecto: Esta sección no se ha modificado Implementación del Sistema

FASES DEL ALGORITMO POSE2SIM: UN RECORRIDO DETALLADO

5.2.1.14 Grabación y Sincronización de Vídeos

La fase inicial del flujo de trabajo de Pose2Sim es la adquisición de los datos de vídeo, seguida de su sincronización, un paso necesario para la posterior reconstrucción 3D. Este proceso se llevó a cabo utilizando una combinación de dispositivos y técnicas, adaptándose a los recursos disponibles y a las particularidades del entorno de grabación.

5.2.1.14.1 Proceso de Grabación:

Para la captura de los vídeos, se emplearon diversos dispositivos, incluyendo un iPhone 15, un iPhone 13, teléfonos Android prestados para aumentar el número de cámaras, un iPad y las cámaras integradas del laboratorio MOBIOS. Aunque la calidad de vídeo intrínseca de estos dispositivos variaba, la resolución y las condiciones de grabación se mantuvieron consistentes con las utilizadas durante la fase de calibración. Esto implicaba asegurar que el formato de vídeo fuera el mismo que el de las imágenes de calibración (por ejemplo, 16:9, a diferencia del 4:3 común en fotografías de móviles) y que, en el caso de los iPhones, se utilizara la misma lente (por ejemplo, 1x en lugar de 0.5x). La tasa de fotogramas (framerate) también era un factor relevante, manteniéndose constante en todos los vídeos, ya que el algoritmo de Pose2Sim opera fotograma a fotograma.

Los dispositivos se colocaron sobre trípodes o, en ocasiones, se improvisaron sistemas de soporte con objetos como libros para asegurar su estabilidad. Era fundamental que las cámaras permanecieran completamente inmóviles tras la calibración extrínseca; cualquier movimiento, por mínimo que fuera, invalidaría la calibración y, consecuentemente, los vídeos resultantes no serían útiles para el análisis. La altura y distancia de las cámaras no se estandarizaron en cada configuración. Se procuró utilizar siempre el mayor número de cámaras posible en cada grabación, con el fin de evaluar posteriormente cómo el número de vistas afectaba la calidad y precisión de los resultados en el software.

Durante las grabaciones, se observó que la visibilidad completa del cuerpo del sujeto era un factor influyente para el éxito de la estimación de pose. Cuando había objetos que obstruían la vista del cuerpo o cuando solo se veía una parte del mismo, el algoritmo de detección de pose no era capaz de realizar una estimación precisa, lo que invalidaba el resultado de la triangulación 3D. La iluminación no se controlaba de forma específica, utilizando la luz ambiental de cada ubicación. Todas las cámaras se utilizaron en modo horizontal, ya que los intentos iniciales de grabar en vertical generaron problemas en el procesamiento que llevaron a descartar esa orientación.

5.2.1.14.2 Entornos de Grabación y Desafíos:

Las grabaciones se realizaron en diferentes entornos para probar la adaptabilidad del sistema:

- Pruebas iniciales en un entorno doméstico: Se realizaron vídeos de prueba preliminares en un entorno no controlado para familiarizarse con el proceso.
- Aula de iCAI: Se llevaron a cabo grabaciones de mayor calidad y con más cámaras en un aula de la Universidad Pontificia Comillas. Aunque se encontraron algunos problemas de oclusión que impedían que todas las cámaras vieran el cuerpo entero en ciertos momentos, los resultados fueron satisfactorios.
- Laboratorio MOBIO: Las grabaciones en el laboratorio MOBIO presentaron desafíos significativos. La posición de las cámaras era a menudo subóptima, resultando en vistas donde solo se apreciaba la mitad del cuerpo. En una de las pruebas, el asiento de la simulación obstruía parcialmente la vista de las piernas del

sujeto, lo que generó numerosos valores erróneos en la estimación de pose y dificultó el análisis preciso de esa región corporal.

5.2.1.14.3 Sincronización de Vídeos con DaVinci Resolve:

La sincronización temporal de los vídeos capturados con las diferentes cámaras se realizó utilizando DaVinci Resolve. Este software de edición de vídeo fue elegido por su facilidad de uso y por la necesidad de acortar los segmentos de vídeo relevantes para el análisis.

El proceso de sincronización se llevó a cabo mediante la herramienta de sincronización automática por waveform. Esta función analiza los patrones de audio de cada vídeo y los alinea temporalmente, lo que es una forma eficiente de conseguir una sincronización precisa. Para facilitar este proceso, durante las grabaciones se utilizaron señales de audio claras, como aplausos o, en el caso del laboratorio MOBIO, pitidos generados por el propio equipamiento. Una vez sincronizados, los vídeos se exportaban uno a uno en el formato deseado, generalmente en 16:9 H.264, aunque también se utilizaron formatos MOV o MP4 sin que esto afectara la relevancia de los datos para el algoritmo. Cabe destacar que, si bien la sincronización interna de Pose2Sim también funciona de manera eficiente, el uso de DaVinci Resolve permitió integrar la sincronización con el recorte de los vídeos, optimizando el flujo de trabajo.

Si en lugar de utilizar la sincronización externa se decide utilizar el propio sistema de Pose2Sim, correspondería a ejecutar el comando `Pose2Sim.synchronization` después de ejecutar la `poseEstimation()`.

5.2.1.15 Calibración de Cámaras (Intrínseca y Extrínseca)

Este apartado corresponde a ejecutar el comando `Pose2Sim.calibration()`.

La calibración de cámaras es un paso esencial en cualquier sistema de visión por computador que busque reconstruir información tridimensional a partir de imágenes 2D. Su propósito principal es determinar los parámetros intrínsecos y extrínsecos de cada cámara, así como

los coeficientes de distorsión de la lente. Esta información es indispensable para corregir las deformaciones que las lentes introducen en las imágenes y, lo que es más importante para Pose2Sim, para establecer la relación entre las coordenadas de un punto en el espacio 3D real y su proyección en el plano de la imagen 2D de cada cámara. Una calibración precisa es el pilar sobre el que se asienta la exactitud de la triangulación 3D posterior.

Pose2Sim ofrece la posibilidad de realizar la calibración de cámaras de dos maneras principales, tal como se configura en el parámetro `calibration_type` del `config.toml`:

- Conversión ('convert'): Permite importar y convertir ficheros de calibración generados por otros sistemas de captura de movimiento (ej., Qualisys, Vicon). En el presente proyecto, esta opción no fue utilizada, ya que no se dispuso de ficheros de calibración externos compatibles.
- Cálculo ('calculate'): Este método, basado en la calibración con tablero de ajedrez (checkerboard), es el que se ha empleado en este TFG. Se fundamenta en los principios de la calibración de cámaras de OpenCV, una librería de visión por computador ampliamente reconocida.

El proceso de calibración se divide en dos fases principales:

5.2.1.15.1 Calibración Intrínseca:

La calibración intrínseca se enfoca en determinar las propiedades internas de cada cámara, las cuales son inherentes al dispositivo y no cambian a menos que se modifique la lente o su configuración. Estos parámetros incluyen la distancia focal, el centro óptico y los coeficientes de distorsión de la lente. Las distorsiones, como la radial (que hace que las líneas rectas parezcan curvas) y la tangencial (debido a la desalineación de la lente), deben corregirse para obtener mediciones precisas.

Para realizar la calibración intrínseca en Pose2Sim, se utilizan múltiples imágenes (o un vídeo) de un patrón conocido, como un tablero de ajedrez, capturadas desde diferentes

ángulos y distancias. El algoritmo detecta las esquinas interiores de este patrón y, a partir de ellas, calcula los parámetros de la cámara.

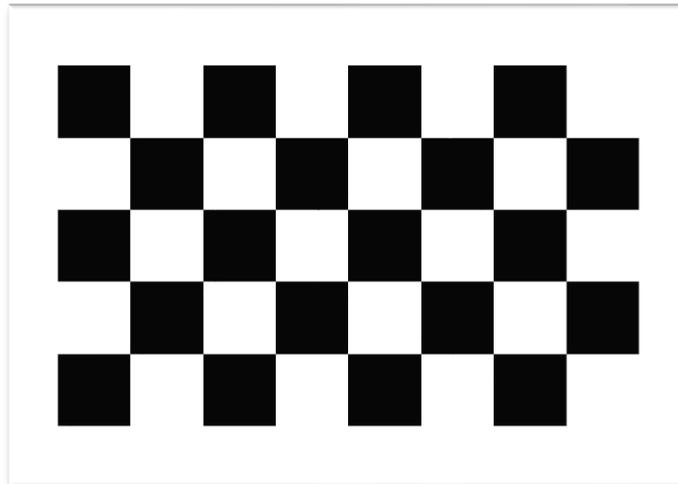


Ilustración 4 - Imagen del patrón utilizado 4x7

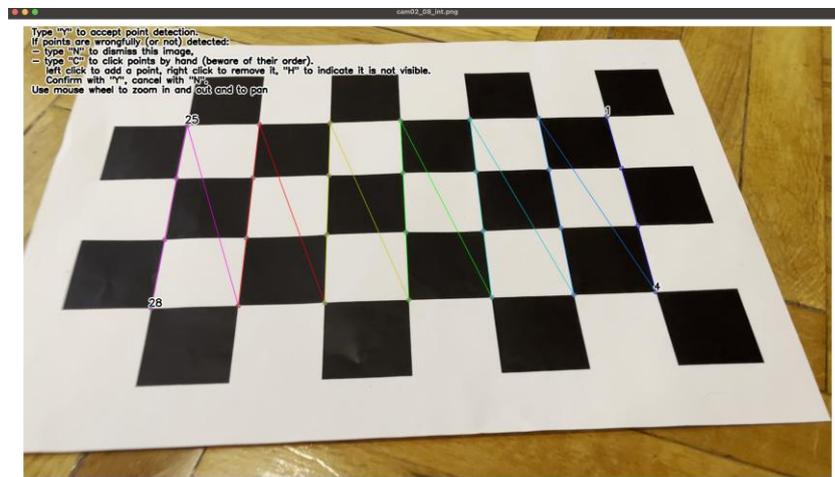


Ilustración 5 - Imagen del proceso de calibración intrínseca

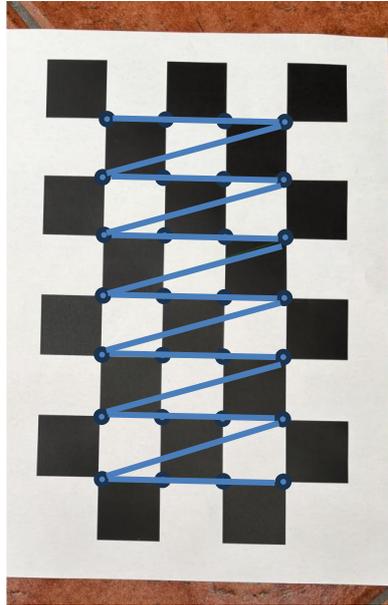


Ilustración 6 - Diagrama del patrón a dibujar para la calibración

Parámetros relevantes en config.toml ([calibration.calculate.intrinsics]):

- `overwrite_intrinsics = false`: Se configuró en `false` para evitar recalcular los parámetros intrínsecos si ya se habían obtenido con éxito en una sesión previa, optimizando el tiempo de procesamiento.
- `show_detection_intrinsics = true`: Permite visualizar las esquinas detectadas en el tablero durante el proceso de calibración, lo cual es útil para verificar la calidad de la detección.
- `intrinsics_extension = 'png'`: Indica la extensión de las imágenes utilizadas para la calibración intrínseca.
- `extract_every_N_sec = 1`: Si se usa un vídeo para la calibración, extrae un fotograma cada segundo para el análisis.
- `intrinsics_corners_nb = [4,7]`: Define el número de esquinas interiores del tablero de calibración en formato [altura, ancho].
- `intrinsics_square_size = 90`: Especifica el tamaño de los cuadrados del tablero en milímetros, un dato crucial para la reconstrucción a escala real.

5.2.1.15.2 Calibración Extrínseca:

La calibración extrínseca determina la posición y orientación de cada cámara en el espacio 3D con respecto a un sistema de coordenadas global (el "mundo real"). Esta fase es crítica para la triangulación 3D, ya que permite a Pose2Sim saber dónde se encuentra cada cámara y hacia dónde "mira" en relación con las demás y con el área de interés.

Pose2Sim soporta diferentes métodos para la calibración extrínseca, configurados en `extrinsics_method`:

- 'board': Utiliza un tablero de ajedrez colocado en una posición conocida en el espacio. Este fue el método empleado en el proyecto, aprovechando el tablero físico impreso.
- 'scene': Implica la selección manual de puntos con coordenadas 3D conocidas en la escena.
- 'keypoints': Un método más automatizado que utiliza la estimación de pose de una persona moviéndose en la escena (aún en desarrollo).

Una vez realizada la calibración extrínseca, es fundamental que las cámaras permanezcan completamente inmóviles. Cualquier desplazamiento, por mínimo que sea, invalidaría los parámetros extrínsecos calculados, haciendo que los vídeos grabados con esa configuración de cámaras no sean útiles para una reconstrucción 3D precisa.



Ilustración 7 - Sobre el tablero, calibración extrínseca de una cámara. Grabación en Casa



Ilustración 8 - Sobre el tablero, calibración extrínseca de otra cámara. Grabación en Casa

Parámetros relevantes en `config.toml` (`[calibration.calculate.extrinsics]` y `[calibration.calculate.extrinsics.board]`):

- `calculate_extrinsics = true`: Se activó para realizar el cálculo de los parámetros extrínsecos.
- `extrinsics_method = 'board'`: Se seleccionó este método debido a la disponibilidad de un tablero de calibración físico y la necesidad de una calibración controlada en el entorno del laboratorio.

- `show_reprojection_error = true`: Muestra el error de reproyección, una métrica clave para evaluar la precisión de la calibración. Un error de reproyección cercano a cero indica una mayor exactitud de los parámetros obtenidos.
- `extrinsics_extension = 'png'`: Extensión de las imágenes usadas para la calibración extrínseca.
- `extrinsics_corners_nb = [4,7]`: Número de esquinas del tablero.
- `extrinsics_square_size = 90`: Tamaño de los cuadrados del tablero en milímetros.

Como entradas para el módulo de calibración, se utilizan las imágenes de calibración (ubicadas en las carpetas `calibration/intrinsics` y `calibration/extrinsics`) y el archivo `Config.toml` del proyecto. La salida principal de este proceso es un fichero de calibración en formato `.toml`, que contiene todos los parámetros intrínsecos y extrínsecos calculados, y que será utilizado en las fases posteriores de triangulación 3D.

5.2.1.16 Estimación de Pose 2D

Este apartado, corresponde a ejecutar el comando `Pose2Sim.poseEstimation()`.

La fase de estimación de pose 2D es el primer paso donde los algoritmos de visión por computador analizan directamente el contenido de los vídeos para detectar los keypoints del cuerpo humano. Este proceso se realiza de forma independiente para cada fotograma de cada cámara, generando una representación bidimensional de la pose.

5.2.1.16.1 Modelo de Estimación de Pose y Procesamiento:

En el presente proyecto, el modelo de estimación de pose 2D utilizado para la reconstrucción 3D fue RTMPose con la configuración `'Body_with_feet'`, tal como se especifica en el archivo `config.toml`. La elección de RTMPose se fundamenta en su avanzada arquitectura y su capacidad para ofrecer un rendimiento superior en términos de velocidad y precisión en comparación con modelos predecesores, como OpenPose. RTMPose se basa en un enfoque

top-down, donde primero se detectan las personas en la imagen y luego se estima la pose de cada individuo dentro de sus respectivas cajas delimitadoras. Su optimización para la inferencia en tiempo real y su robustez en diversas condiciones de iluminación y oclusión lo convierten en una opción eficiente para la detección de keypoints. Aunque al inicio se realizaron pruebas con otros modelos de estimación de pose 2D, para las estimaciones 3D y el análisis posterior, se mantuvo consistentemente el modelo 'Body_with_feet' debido a su fiabilidad y a la integración optimizada dentro de Pose2Sim. No se profundizará en el detalle de las pruebas con otros modelos para mantener el enfoque del TFG, pero si veremos en alto nivel las opciones que ofrece RTMPose.

RTMPose es un proyecto dedicado a la implementación de modelos y algoritmos de estimación de pose, desarrollado por OpenMMLab. Dispone de varios modelos pre-entrenados diseñados para diferentes aplicaciones y niveles de detalle en la detección de keypoints. A continuación, se presenta una tabla con algunos de los modelos más relevantes incluidos en RTMPose, tal como se describen en la configuración de Pose2Sim:

Modelo RTMPose	Descripción
Body	Algoritmo estándar que detecta 17 keypoints principales del cuerpo (modelo COCO_17).
Body_with_feet	Modelo por defecto (HALPE_26) que detecta keypoints del cuerpo y los pies.
Whole_body / Whole_body_wrist	Modelos más exhaustivos con 133 keypoints, incluyendo puntos adicionales en la cara, pies y manos (modelo COCO_133).

Face	Especializado en la detección de 106 keypoints exclusivamente en el rostro humano (modelo FACE_106).
Hand	Enfocado en la detección de 21 keypoints únicamente en la mano (modelo HAND_21).
Animal	Modelo para la detección de pose en animales (ANIMAL2D_17).

Tabla 2 - Modelos de estimación 2D

En el caso de este proyecto, se utilizó específicamente el modelo 'Body_with_feet' de RTMPose, que proporciona un equilibrio entre la detección de keypoints del cuerpo y los pies, lo cual es adecuado para el análisis cinemático de movimientos completos. A continuación, una imagen de creación propia con los puntos que estudia el modelo. Como se puede apreciar en la imagen, cada punto tiene un identificador, y algunos identificadores no están, como pueden ser el 1,2,3 y 4. Esto se debe a que se utiliza un modelo más sencillo para la triangulación en 3D. Se añade también una imagen de cómo es el modelo completo. Imagen sacada del repositorio de RTMPose.

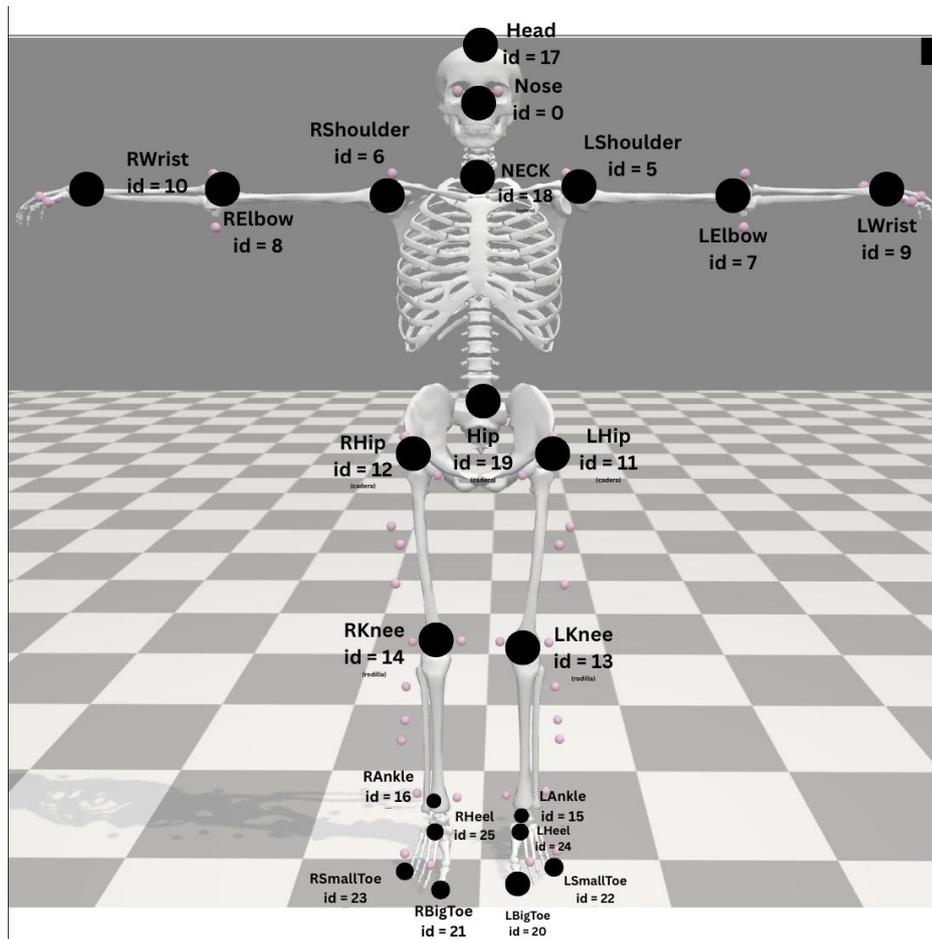


Ilustración 9 - Puntos utilizados para la triangulación

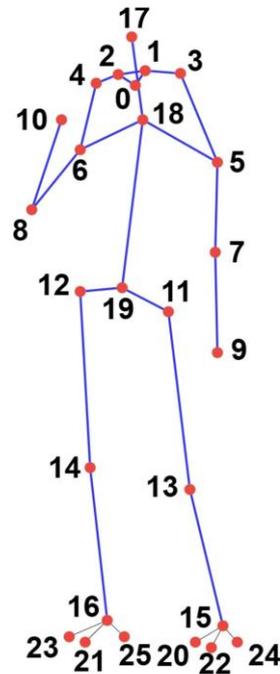


Ilustración 10 - Keypoints del modelo completo Body_with_feet (Contributors)

El procesamiento de los vídeos se realiza fotograma a fotograma. Para cada fotograma de cada vídeo de entrada, el modelo RTMPose analiza la imagen y genera un conjunto de keypoints que representan las articulaciones y puntos de interés del cuerpo humano. El output de esta fase son una serie de ficheros JSON, donde cada fichero corresponde a un fotograma específico de una cámara. Estos ficheros JSON contienen las coordenadas (X, Y) de cada keypoint detectado en el plano de la imagen, junto con una puntuación de confianza asociada a cada detección. La puntuación de confianza es un valor entre 0 y 1 que indica la probabilidad de que el keypoint haya sido detectado correctamente, siendo 1 la máxima confianza. Esta información es crucial para las fases posteriores, ya que los keypoints con baja confianza pueden ser filtrados o tratados como datos faltantes.

A continuación, se presenta un ejemplo de la estructura de un fichero JSON de salida, ilustrando cómo se organizan las coordenadas y la confianza de los keypoints:

```
["version": 1.3, "people": [{"person_id": -1, "pose_keypoints_2d":  
999.5418828328452, 185.31122398376465, 0.9360028505325317, 1012.6458587646484,  
177.82323789596558, 0.9968372583389282, 990.1819000244141, 175.9512413740158,  
1.031402587890625, 1035.1098175048828, 189.05521702766418, 1.0577402114868164,  
978.9499206542969, 189.05521702766418, 0.9684553146362305, 1068.8057556152344,  
269.5510674715042, 0.8983489274978638, 947.1259791056316, 262.06308138370514,  
0.8499956130981445, 1076.293741861979, 365.0228900909424, 0.9115113019943237,  
887.2220891316731, 316.3509805202484, 0.8628891706466675, 1053.8297831217446,  
439.9027509689331, 0.9102861881256104, 842.2941716512044, 381.8708587884903,  
0.9203320145606995, 1033.2378209431968, 432.41476488113403, 0.7714303731918335,  
969.5899378458657, 430.54276835918427, 0.780997097492218, 1035.1098175048828,  
561.582524895668, 0.9059980511665344, 948.9979756673176, 555.9665353298187,  
0.8520834445953369, 1033.2378209431968, 672.0303196907043, 0.907092809677124,  
965.8459447224936, 668.2863266468048, 0.8546547293663025, 1008.9018656412759,  
140.38330745697021, 0.9238814115524292, 1007.0298690795898, 228.3671439886093,  
0.9319547414779663, 1001.4138793945312, 428.6707718372345, 1.1719615459442139,  
1012.6458587646484, 707.5982536077499, 0.8745652437210083, 947.1259791056316,  
705.7262570858002, 0.8500854969024658, 1033.2378209431968, 707.5982536077499,  
0.8948899507522583, 935.8939997355144, 698.2382709980011, 0.8458097577095032,  
1036.9818140665689, 688.8782883882523, 0.7852720022201538, 977.0779240926108,  
683.262298822403, 0.8304267525672913}], "face_keypoints_2d": [],  
"hand_left_keypoints_2d": [], "hand_right_keypoints_2d": [], "pose_keypoints_3d":  
[], "face_keypoints_3d": [], "hand_left_keypoints_3d": [],  
"hand_right_keypoints_3d": []}]
```

5.2.1.16.2 Visualización y Rendimiento:

Durante el proceso de estimación de pose 2D, se utilizó la opción `display_detection = true` en el `config.toml`. Esta configuración activó la visualización en tiempo real de los keypoints detectados y las conexiones entre ellos (el "esqueleto" 2D) superpuestos sobre los vídeos originales. Esta característica es de gran utilidad para la verificación cualitativa inmediata de la calidad de la detección y el seguimiento del algoritmo, permitiendo identificar rápidamente problemas como oclusiones o detecciones erróneas. Los vídeos resultantes con la pose estimada superpuesta se generan automáticamente y se almacenan en la carpeta `Pose/` con el formato `camX.videoY_pos.mp4`, y se pueden encontrar ejemplos de estos en el repositorio de GitHub del proyecto. A continuación, una snapshot de uno de estos vídeos.

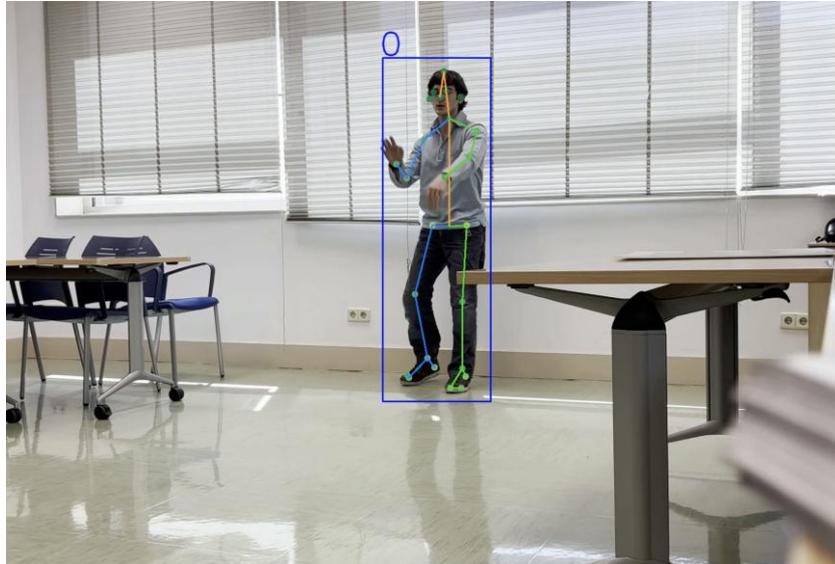


Ilustración 11 - Snapshot, estimación de pose 2D.

Para la ejecución de esta fase, que puede ser computacionalmente intensiva, se utilizó el ordenador de sobremesa con los recursos previamente mencionados en apartados anteriores. Es importante destacar que, aunque la estimación de pose 2D requiere una capacidad de procesamiento considerable, la configuración utilizada (CPU con valores predeterminados) demostró ser suficiente para ejecutar la tarea de manera efectiva en el contexto de este proyecto.

5.2.1.16.3 Manejo de Oclusiones y Errores en la Detección 2D:

Uno de los principales desafíos en la estimación de pose 2D es la presencia de oclusiones, es decir, cuando partes del cuerpo del sujeto están tapadas por objetos, mobiliario o incluso otras partes del propio cuerpo. Estas situaciones representan una pérdida de información visual que el algoritmo debe intentar inferir o manejar. En estos casos, el algoritmo de detección de keypoints puede verse comprometido, manifestándose de varias maneras:

- Detección incorrecta: El algoritmo puede "inventar" keypoints en posiciones donde no hay ninguna parte del cuerpo, o asignar coordenadas erróneas a keypoints parcialmente visibles, basándose en patrones aprendidos que no se aplican correctamente a la situación de oclusión.

- Baja confianza: Los keypoints detectados bajo oclusión a menudo presentan puntuaciones de confianza muy bajas, lo que indica la incertidumbre del modelo sobre la veracidad de esa detección.
- Ausencia de detección: En casos de oclusión severa o prolongada, el algoritmo simplemente no detecta ciertos keypoints, dejando un "hueco" en la información de la pose.

Estos errores en la fase 2D tienen un impacto directo y significativo en la posterior triangulación 3D. Un keypoint mal detectado en una o más cámaras puede llevar a una reconstrucción 3D errónea o incluso hacer imposible la triangulación para ese punto en particular.

A continuación, se presenta una imagen que ilustra un caso de oclusión y la detección incorrecta de keypoints resultante, evidenciando los desafíos que estas situaciones plantean al algoritmo:

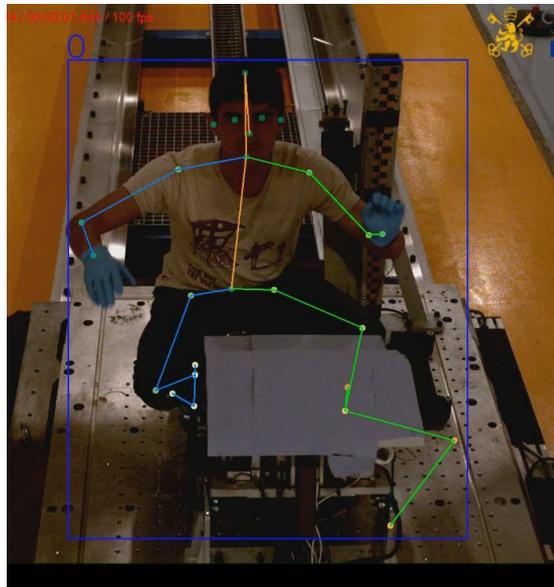


Ilustración 12 - Ejemplo de detección errónea del cuerpo

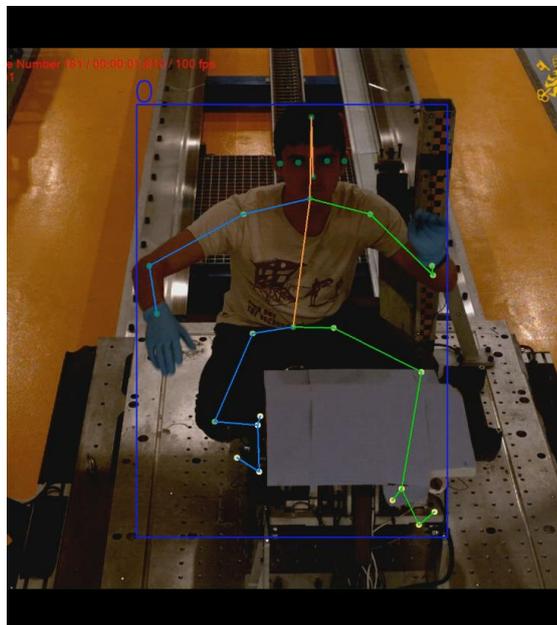


Ilustración 13 - Ejemplo de intento de detección aceptable

Pose2Sim incorpora mecanismos para manejar errores y gaps en los datos 3D (como la interpolación), pero estos son efectivos principalmente cuando los errores son pequeños y aislados. En casos de oclusión significativa, donde los keypoints son incorrectos o faltan en varios fotogramas consecutivos (debido a que los fotogramas anteriores y siguientes también presentan errores), la interpolación no es suficiente para corregir la información. Esto subraya la importancia de una buena planificación de la grabación y una visibilidad óptima del sujeto para minimizar estos problemas y asegurar la calidad de los datos de entrada para la estimación de pose 2D.

Para ilustrar el rendimiento del algoritmo en diferentes escenarios de grabación y la importancia de las condiciones de captura, se presentan a continuación ejemplos de estimaciones de pose realizadas:

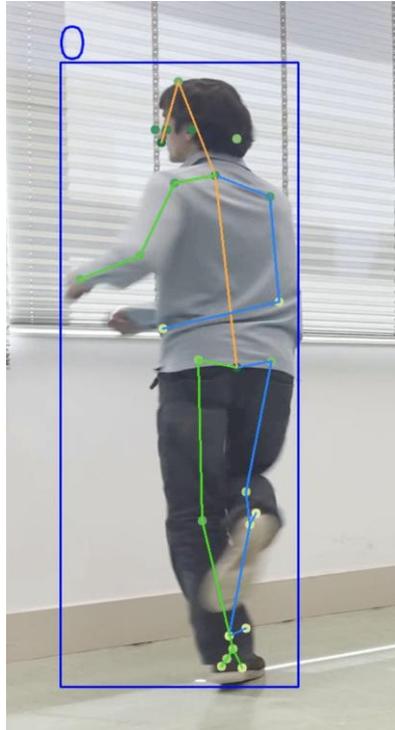


Ilustración 14 - Ejemplo de manejo bastante aceptable ante una posición difícil

A continuación, la misma pose desde diferentes cámaras

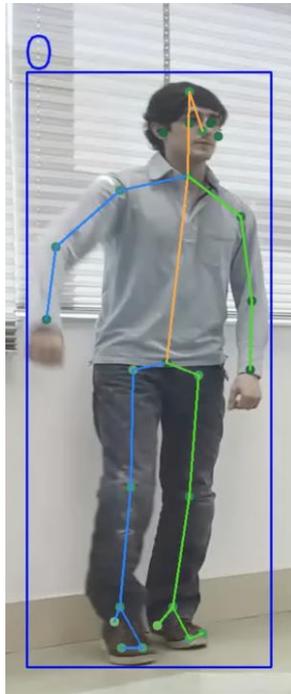


Ilustración 15 – Misma posición, Cámara 1

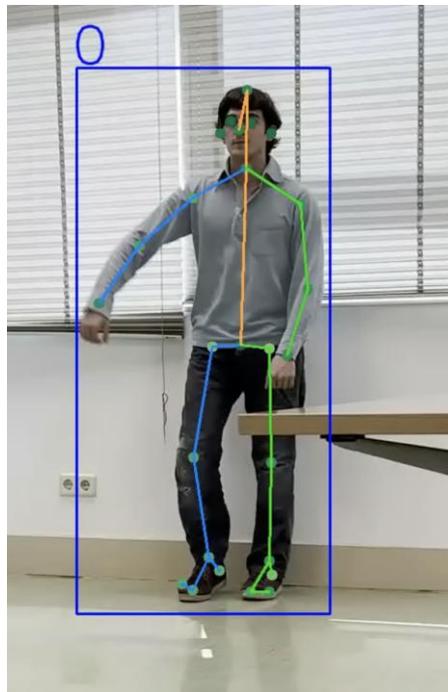


Ilustración 16 - Misma posición, Cámara 2

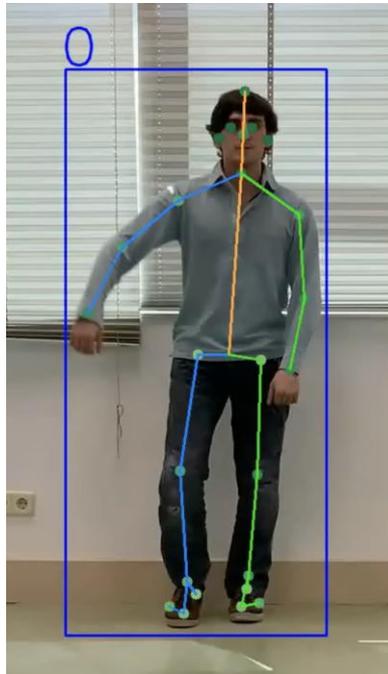


Ilustración 17 - Misma posición, Cámara 3

5.2.1.17 Asociación de Personas y Seguimiento (Tracking)

Este apartado corresponde a ejecutar el comando `Pose2Sim.personAssociation()`.

Tras la estimación de pose 2D, el siguiente paso en el flujo de trabajo de Pose2Sim es la asociación de personas y el seguimiento de los individuos de interés a lo largo de los fotogramas y entre las diferentes vistas de cámara. Los algoritmos de estimación de pose 2D, como RTMPose (que se ha utilizado en este TFG), detectan todas las personas presentes en el campo de visión de cada cámara, lo que puede resultar en múltiples detecciones por fotograma y por vista. La fase de asociación es crucial para identificar a la misma persona a través de las distintas cámaras y para mantener su seguimiento a lo largo del tiempo.

El comportamiento de este módulo se rige principalmente por el parámetro `multi_person` en la sección `[project]` del archivo `config.toml`:

- `multi_person = false` (Análisis de una única persona):

Cuando se configura `multi_person` a `false`, el módulo de asociación de personas se encarga de identificar al individuo de interés principal en la escena. Para ello, el algoritmo intenta todas las triangulaciones posibles de un punto anatómico predefinido (especificado en `tracked_keypoint` en la sección `[personAssociation.single_person]` del `config.toml`, por ejemplo, el 'Neck'). De todas las triangulaciones posibles, se selecciona a la persona para la cual el error de reproyección es el más pequeño. Este enfoque garantiza que, incluso si hay otras personas en el campo de visión, el análisis se centre en el sujeto principal con la detección más fiable.

- `multi_person = true` (Análisis de múltiples personas):

Si se configura `multi_person` a `true`, el objetivo es triangular y seguir a varias personas a través de las diferentes vistas de cámara. En este caso, el módulo calcula la distancia entre las líneas epipolares (líneas que conectan el centro óptico de una cámara con un keypoint detectado y se proyectan en la otra cámara) para todas las personas detectadas en todas las vistas. Posteriormente, selecciona las mejores correspondencias entre las detecciones de las diferentes cámaras. El cálculo de la matriz de afinidad a partir de estas distancias se inspira en el enfoque de EasyMocap, lo que permite establecer de manera robusta qué keypoints pertenecen a la misma persona en distintas vistas. El seguimiento de estas personas a lo largo de los fotogramas se gestiona posteriormente en la fase de triangulación.

5.2.1.17.1 Entradas y Salidas del Módulo:

Las entradas para este módulo incluyen:

- Los ficheros JSON generados por la fase de estimación de pose 2D, que contienen las detecciones de múltiples personas por cada cámara.
- El archivo `Config.toml`, que define los parámetros de asociación.

Como salida, el módulo genera nuevos ficheros JSON para cada cámara, pero esta vez conteniendo la pose de una única persona de interés. Estos ficheros se almacenan en la carpeta `Pose-associated/`, listos para ser utilizados en la siguiente fase de triangulación 3D.

5.2.1.18 Triangulación 3D

La triangulación es una etapa fundamental en el flujo de trabajo de Pose2Sim, ya que transforma las coordenadas bidimensionales (keypoints 2D) detectadas en cada cámara a una representación tridimensional del movimiento. Su propósito principal es resolver el problema de estimar la posición real de un objeto en el espacio 3D a partir de sus proyecciones en dos o más imágenes capturadas desde diferentes perspectivas, un proceso que es imposible de lograr únicamente con datos 2D. Esta fase es, por tanto, la base para cualquier análisis cinemático espacial del movimiento humano.

5.2.1.18.1 Proceso de Triangulación en Pose2Sim:

El módulo de triangulación en Pose2Sim toma como entrada los ficheros JSON con las coordenadas 2D de los keypoints (tras la estimación de pose y, si se aplicó, la asociación de personas) y el fichero de calibración de las cámaras. El proceso general se basa en la intersección de los rayos proyectados desde los centros ópticos de las cámaras a través de los keypoints detectados en sus respectivos planos de imagen.

La precisión de la triangulación se ve influenciada por varios factores y parámetros configurables en el archivo config.toml:

- **Ponderación por Confianza:** La triangulación se pondera por el valor de confianza (likelihood) de cada keypoint 2D detectado. Los keypoints con una confianza por debajo del umbral establecido en `likelihood_threshold_triangulation = 0.3` son descartados, ya que su baja fiabilidad podría introducir errores significativos en la reconstrucción 3D.
- **Manejo de Errores de Reproyección:** Pose2Sim incorpora un mecanismo robusto para manejar errores. Si el error de reproyección de un keypoint (la diferencia entre la proyección 2D del punto 3D triangulado y su detección original 2D) supera el umbral `reproj_error_threshold_triangulation = 15 px`, el algoritmo intenta corregir la situación. Una de las estrategias es intentar un "intercambio de lados" (`handle_LR_swap = false` por defecto), donde se verifica si el keypoint pertenece

realmente a la extremidad contralateral. Si el error persiste tras este intento, Pose2Sim puede optar por eliminar temporalmente una de las cámaras para ese keypoint y fotograma en particular, hasta que el error de reproyección se sitúe por debajo del umbral. Si se eliminan más cámaras de las permitidas por el valor de `min_cameras_for_triangulation = 2`, la triangulación para ese keypoint y fotograma se omite.

- Interpolación de Valores Faltantes: A pesar de los mecanismos de corrección, pueden persistir gaps (valores faltantes) en los datos 3D. Pose2Sim interpola estos valores faltantes según la configuración del `config.toml`. Los parámetros que rigen este proceso son:
 - `interpolation = 'linear'`: Se utiliza una interpolación lineal para rellenar los huecos.
 - `interp_if_gap_smaller_than = 10`: Solo se interpolan los huecos de keypoints que sean menores o iguales a 10 fotogramas consecutivos.
 - `fill_large_gaps_with = 'last_value'`: Para huecos mayores a 10 fotogramas, los valores se rellenan con el último valor válido conocido.
 - `show_interp_indices = true`: Esta opción es útil para depuración, ya que devuelve los índices de los fotogramas que han sido interpolados para cada keypoint.

Otras Consideraciones:

- `handle_LR_swap = false`: En este proyecto, se mantuvo desactivado. Aunque es beneficioso en escenarios con pocas cámaras y riesgo de confusión entre extremidades (por ejemplo, si una cámara apunta al plano sagital), su activación puede ralentizar el procesamiento y ser menos preciso si no es estrictamente necesario.
- `undistort_points = false`: También se mantuvo desactivado. Si las imágenes presentan distorsiones significativas, esta opción las corrige durante la triangulación, pero añade tiempo de procesamiento; se asume que las distorsiones de las cámaras utilizadas eran bajas.

5.2.1.18.2 Contexto del Proyecto (multi_person = false):

Dado que en este TFG consistía en el estudio concreto para el caso de MOBIOS en donde se realiza un estudio sobre un único cuerpo, el parámetro multi_person en la sección [project] del config.toml se configuró como false, el módulo de triangulación se centra en la persona principal detectada y asociada en la fase anterior. Esto simplifica el proceso al evitar la complejidad de la asignación de keypoints a múltiples individuos en la reconstrucción 3D.

5.2.1.18.3 Entradas y Salidas del Módulo:

Las entradas para la triangulación incluyen:

- Un fichero de calibración en formato .toml.
- Los ficheros JSON de cada cámara que contienen la pose 2D de una única persona de interés (generados en la fase de estimación de pose y asociación).
- El archivo Config.toml del proyecto.
- El modelo de esqueleto utilizado.

La salida principal de este módulo es un fichero con extensión .trc (TRaCking file). Este archivo contiene las coordenadas 3D (X, Y, Z) de todos los keypoints triangulados a lo largo de todos los fotogramas, representadas en un sistema de coordenadas "Y-up". En el sistema "Y-up", el eje Y apunta hacia arriba (verticalmente), mientras que los ejes X y Z definen el plano horizontal. Esta convención es común en software de biomecánica como OpenSim, facilitando la compatibilidad y la visualización directa del movimiento. Además, si make_c3d = true está activado, los datos triangulados también se guardan en formato .c3d, otro estándar en la captura de movimiento.

5.2.1.18.4 Desafíos Encontrados en el Proyecto:

Durante la fase de triangulación, se enfrentaron desafíos significativos que impactaron la obtención de resultados coherentes. En el equipo MacBook Pro, la ejecución del módulo de triangulación presentaba problemas continuos, incluso tras ajustar los umbrales de error, lo que resultaba en datos 3D inconsistentes. Este inconveniente se solucionó al trasladar el

procesamiento al ordenador de sobremesa con sistema operativo Windows, donde el algoritmo se ejecutó sin fallos.

Otro desafío relevante provino de los problemas de oclusión experimentados en la fase de grabación y estimación de pose 2D. En situaciones de oclusión severa, la detección de keypoints en 2D resultaba en valores erráticos o completamente ausentes en múltiples fotogramas. Esta falta de coherencia entre las proyecciones 2D de las diferentes cámaras impedía al algoritmo realizar una triangulación precisa, ya que los keypoints simplemente no podían ser reconstruidos de forma fiable en 3D. Esto evidenció que, a pesar de los mecanismos de interpolación y corrección de errores de Pose2Sim, la calidad de la información de entrada 2D es un factor determinante para el éxito de la triangulación 3D.

5.2.1.19 Filtrado de Datos 3D

Una vez que las coordenadas 3D de los keypoints han sido trianguladas, es común que los datos presenten cierto nivel de ruido o inconsistencias, inherentes a los procesos de detección de pose 2D y triangulación. El filtrado de datos 3D es una fase opcional pero altamente recomendada en el flujo de trabajo de Pose2Sim, cuyo objetivo es suavizar las trayectorias de los keypoints, eliminar el ruido y generar un movimiento más biomecánicamente plausible y coherente. Un filtrado adecuado mejora la calidad de los datos para análisis posteriores de cinemática y dinámica.

5.2.1.19.1 Propósito del Filtrado:

Las estimaciones de pose, especialmente las basadas en visión por computador, pueden verse afectadas por factores como:

- Variaciones en la calidad de la imagen (iluminación, oclusión, desenfoque).
- Errores en la detección de keypoints por el algoritmo 2D.
- Imprecisiones en la calibración de las cámaras.
- Ruido en el proceso de triangulación.

Estos factores pueden resultar en trayectorias de keypoints irregulares o "temblorosas". El filtrado aplica operaciones matemáticas a las series temporales de los datos 3D para reducir estas fluctuaciones no deseadas sin eliminar la información relevante del movimiento.

5.2.1.19.2 Filtros Disponibles en Pose2Sim:

Pose2Sim ofrece una variedad de algoritmos de filtrado, cada uno con sus propias características y aplicaciones, configurables a través del parámetro `type` en la sección `[filtering]` del `config.toml`:

- **Butterworth:** Un filtro de paso bajo que permite pasar frecuencias por debajo de una frecuencia de corte y atenúa las frecuencias por encima de ella. Es ampliamente utilizado en biomecánica para suavizar datos de movimiento, ya que el ruido de alta frecuencia suele corresponder a artefactos no biológicos.
- **Butterworth on speed:** Similar al filtro Butterworth tradicional, pero aplicado a la velocidad de los keypoints, lo que puede ser útil en ciertos escenarios.
- **Kalman:** Un filtro recursivo que estima el estado de un sistema (en este caso, la posición y velocidad de un keypoint) a partir de una serie de mediciones ruidosas y un modelo dinámico del sistema. Permite ajustar el "nivel de confianza" entre las mediciones y el modelo de proceso.
- **Gaussian:** Aplica un suavizado gaussiano a los datos.
- **LOESS (Locally Estimated Scatterplot Smoothing):** Un método no paramétrico de suavizado local que puede ser útil para datos con patrones no lineales.
- **Median:** Un filtro no lineal que reemplaza cada punto de datos con la mediana de los puntos vecinos, siendo efectivo para eliminar el ruido tipo "sal y pimienta" o outliers extremos sin distorsionar demasiado los bordes.

Configuración y Aplicación en el Proyecto:

Para el presente proyecto, se utilizó el filtro Butterworth, configurado en el `config.toml` de la siguiente manera:

- `type = 'butterworth'`: Selecciona el filtro Butterworth como método de suavizado.
- `display_figures = true`: Permite visualizar los resultados del filtrado en forma de gráficos, lo cual es útil para evaluar la efectividad del filtro y ajustar sus parámetros.
- `make_c3d = true`: Además de generar un fichero `.trc` filtrado, también guarda los datos en formato `.c3d`, lo que aumenta la compatibilidad con otros softwares de biomecánica.

Los parámetros específicos para el filtro Butterworth fueron los predeterminados por el sistema:

- `order = 4`: Define el orden del filtro, que influye en la pendiente de la atenuación de las frecuencias. Un orden más alto proporciona una atenuación más pronunciada, pero también puede introducir distorsiones de fase si no se maneja adecuadamente.
- `cut_off_frequency = 6 # Hz`: Establece la frecuencia de corte en Hertzios. Las frecuencias del movimiento por debajo de este valor se mantendrán, mientras que las frecuencias por encima serán atenuadas. La selección de una frecuencia de corte adecuada es crucial y a menudo depende de la naturaleza del movimiento estudiado y la tasa de muestreo de los datos. Una frecuencia de 6 Hz es un valor común en biomecánica para movimientos humanos.

Entradas y Salidas del Módulo:

Las entradas para este módulo de filtrado son:

- Un fichero `.trc` con las coordenadas 3D de los keypoints triangulados (o triangulados y aumentados).
- Los parámetros de filtrado definidos en el archivo `Config.toml`.

La salida principal es un nuevo fichero `.trc` que contiene las coordenadas 3D de los keypoints ya filtrados y suavizados. Si la opción `make_c3d = true` está activada, también se generará una versión de este fichero en formato `.c3d`. Estos datos filtrados son de mayor calidad y están listos para las fases de análisis cinemático y biomecánico con OpenSim.

5.2.1.20 Marker Augmentation

El Marker Augmentation es una fase opcional dentro del flujo de trabajo de Pose2Sim que permite aumentar la cantidad de marcadores disponibles en el modelo tridimensional. Su objetivo principal es estimar la posición de marcadores adicionales a partir de los keypoints que ya han sido triangulados, incluso si estos no fueron detectados directamente por el algoritmo de estimación de pose 2D.

5.2.1.20.1 Propósito y Funcionamiento:

La estimación de pose 2D inicial detecta un número limitado de keypoints (por ejemplo, 26 en el modelo Body_with_feet). Sin embargo, para ciertos análisis biomecánicos o para mejorar la estabilidad del modelo esquelético en OpenSim, es beneficioso contar con un conjunto más denso de marcadores. El módulo de Marker Augmentation de Pose2Sim aborda esta necesidad estimando la posición de hasta 43 marcadores adicionales.

Este proceso es particularmente útil en escenarios donde se utilizan menos de cuatro cámaras, ya que puede compensar la falta de información visual al generar un modelo más completo y estable. Si bien la interpolación de estos marcadores adicionales puede introducir una ligera reducción en la precisión intrínseca de los keypoints originales, el beneficio de obtener un modelo más denso y apto para análisis posteriores a menudo compensa esta consideración.

Para que el Marker Augmentation funcione correctamente, el sistema requiere que se hayan detectado y triangulado un conjunto base de keypoints específicos. Según la configuración de Pose2Sim, estos marcadores necesarios incluyen: "Neck", "RShoulder", "LShoulder", "RHip", "LHip", "RKnee", "LKnee", "RAnkle", "LAnkle", "RHeel", "LHeel", "RSmallToe", "LSmallToe", "RBigToe", "LBigToe", "RElbow", "LElbow", "RWrist", y "LWrist".

5.2.1.20.2 Entradas y Salidas del Módulo:

Las entradas para este módulo son:

- Un fichero .trc que contiene las coordenadas 3D de los keypoints triangulados (y opcionalmente, ya filtrados).
- Los parámetros definidos en la sección [markerAugmentation] del archivo Config.toml.

La salida principal es un nuevo fichero .trc (y opcionalmente un .c3d) que contiene las coordenadas 3D de un conjunto ampliado de marcadores, incluyendo tanto los keypoints originales detectados como los 43 marcadores adicionales estimados. Este archivo, más denso, está ahora optimizado para las fases finales de análisis cinemático y biomecánico, como el escalado del modelo y la cinemática inversa en OpenSim.

5.2.1.21 Kinematics

La fase final del flujo de trabajo de Pose2Sim se centra en el procesamiento cinemático, que implica la integración de los datos de pose 3D con modelos biomecánicos para cuantificar el movimiento articular y preparar visualizaciones avanzadas. Este módulo ejecuta dos procesos clave utilizando OpenSim: el escalado del modelo y la cinemática inversa.

5.2.1.21.1 Propósito General:

El objetivo de esta fase es transformar las coordenadas 3D de los keypoints (o marcadores aumentados) en información biomecánica significativa, como los ángulos articulares. Esto se logra ajustando un modelo musculoesquelético virtual a las dimensiones del sujeto grabado y luego haciendo que este modelo siga el movimiento detectado, respetando sus propias restricciones anatómicas. Si bien no utilizaremos toda esta información al no ser expertos en biomecánica, es relevante resaltar esta funcionalidad del algoritmo y explicar para que sirve.

5.2.1.21.2 Escalado del Modelo (Scaling):

El escalado es el primer subproceso. Su función es adaptar un modelo musculoesquelético genérico de OpenSim a las dimensiones antropométricas específicas del individuo que fue

grabado. A diferencia de los sistemas tradicionales que a menudo requieren un ensayo estático específico para el escalado, Pose2Sim realiza este proceso directamente sobre las coordenadas trianguladas obtenidas del fichero .trc.

Para garantizar un escalado robusto y preciso, el algoritmo implementa una serie de pasos para el manejo de datos atípicos (outliers):

Se eliminan el 10% de los fotogramas con la mayor velocidad, ya que se consideran potenciales outliers o momentos de movimiento no representativos para el escalado.

Se descartan los fotogramas donde la velocidad de las coordenadas es nula, lo que sugiere que la persona pudo haber estado fuera del encuadre o inmóvil de forma no deseada.

Se eliminan el 40% de los valores más extremos calculados para cada segmento corporal, con el fin de mitigar la influencia de outliers en las mediciones.

Para cada segmento óseo, el escalado se calcula basándose en la media de los valores de longitud restantes, es decir, aquellos que no fueron eliminados en los pasos previos. Este enfoque mejora la fiabilidad de la estimación de las dimensiones del modelo.

5.2.1.21.3 Cinemática Inversa (Inverse Kinematics - IK):

Una vez que el modelo musculoesquelético ha sido escalado a las dimensiones del sujeto, se ejecuta la cinemática inversa. Este proceso consiste en calcular la configuración articular (ángulos de las articulaciones) del modelo para que sus marcadores virtuales sigan de la forma más aproximada posible a los keypoints triangulados (o aumentados) contenidos en el fichero .trc. Durante este ajuste, el algoritmo de cinemática inversa de OpenSim respeta las restricciones cinemáticas inherentes al propio modelo biomecánico (es decir, cómo se mueven las articulaciones de forma realista). El resultado de esta fase es la serie temporal de los ángulos articulares, que proporciona una descripción cuantitativa del movimiento del sujeto.

Parámetros de Configuración en config.toml ([kinematics]):

- `use_augmentation = true`: Se establece en true para utilizar los marcadores aumentados (generados en la fase de Marker Augmentation) en el modelo de OpenSim, proporcionando una base de datos más densa para el escalado y la cinemática inversa.
- `use_contacts_muscles = true`: Si se configura en true, se añadirán esferas de contacto y modelos musculares al modelo, lo que es útil para análisis dinámicos más avanzados. En el contexto de este proyecto, se ha mantenido esta opción activa.
- `right_left_symmetry = true`: Se asume que el participante es biomecánicamente simétrico entre el lado derecho e izquierdo. Se establece en false solo si hay razones específicas para creer lo contrario (ej., una extremidad protésica).
- `default_height = 1.7`: Un valor por defecto para la altura en metros, utilizado si el cálculo automático de la altura del participante falla.
- `remove_individual_scaling_setup = true`: Si es true, los archivos de configuración individuales generados por el proceso de escalado se eliminan para mantener el directorio de resultados organizado.
- `remove_individual_ik_setup = true`: Si es true, los archivos de configuración individuales generados por el proceso de cinemática inversa también se eliminan.
- `fastest_frames_to_remove_percent = 0.1`: Corresponde al 10% de los fotogramas con mayor velocidad que se eliminan para el escalado.
- `close_to_zero_speed_m = 0.2`: Umbral en metros por fotograma para identificar y eliminar fotogramas con velocidad cercana a cero.
- `large_hip_knee_angles = 45`: Define un umbral de 45 grados; los ángulos de cadera y rodilla por debajo de este valor se consideran potencialmente imprecisos, lo que puede influir en la lógica de eliminación de outliers durante el escalado.
- `trimmed_extrema_percent = 0.5`: Corresponde al 50% de los valores más extremos de la longitud de los segmentos que se eliminan antes de calcular la media para el escalado.

5.2.1.21.4 Entradas y Salidas del Módulo:

Las entradas para este proceso incluyen:

- El `config_dict`, que es un diccionario generado a partir del fichero de calibración `.toml` del proyecto, conteniendo todos los parámetros de configuración.
- El fichero `.trc` con las coordenadas 3D de los keypoints (filtrados y/o aumentados).

Las salidas principales de la fase de procesamiento cinemático son:

- Un modelo `.osim` escalado y personalizado para cada persona, que representa su esqueleto con las dimensiones ajustadas.
- Ficheros de datos de ángulos articulares en formato `.mot` para cada persona. Estos ficheros contienen la serie temporal de los ángulos de todas las articulaciones del modelo, listos para un análisis cuantitativo detallado.

Opcionalmente, los ficheros de configuración de escalado y cinemática inversa de OpenSim se guardan en el directorio `Kinematics/` para referencia.

Además, se generan registros (logs) de Pose2Sim y OpenSim que se guardan en archivos, útiles para la depuración y la verificación del proceso.

VISUALIZACIÓN Y EXPORTACIÓN DE RESULTADOS

La fase final del flujo de trabajo de Pose2Sim se centra en la visualización y exportación de los datos cinemáticos obtenidos. Este paso permite la validación cualitativa de los resultados, la depuración de posibles errores en etapas previas, y la preparación de los datos para un análisis cuantitativo posterior. Pose2Sim ofrece diversas alternativas para acceder y representar la información generada.

5.2.1.22 Formatos de Salida y su Propósito:

Pose2Sim genera varios tipos de ficheros que contienen la información de la pose reconstruida en 3D y los resultados cinemáticos:

- Ficheros .trc (TRaCking files): Estos archivos de texto contienen las coordenadas tridimensionales (X, Y, Z) de los keypoints (o marcadores aumentados) a lo largo del tiempo. Son ideales para la inspección directa de datos numéricos y para la realización de análisis personalizados mediante scripts en entornos como Python, permitiendo una comparación sencilla de coordenadas. No incluyen una representación gráfica del esqueleto, sino únicamente los puntos y las líneas que los unen.
- Ficheros .c3d: El formato .c3d es un estándar ampliamente reconocido en el campo de la captura de movimiento biomecánica. Pose2Sim puede generar estos archivos (si la opción `make_c3d = true` está activada), lo que facilita la compatibilidad y la importación de los datos en una amplia gama de softwares de análisis biomecánico y visualizadores.
- Ficheros .osim (OpenSim Model files): Estos archivos representan el modelo musculoesquelético escalado a las dimensiones del sujeto específico. Son la base para cualquier análisis biomecánico detallado en el entorno de OpenSim, ya que contienen la geometría, la masa, los segmentos corporales y las definiciones articulares del modelo.
- Ficheros .mot (Motion files): Asociados a los modelos .osim, los ficheros .mot contienen las series temporales de los ángulos articulares calculados mediante la cinemática inversa. Estos datos son la información fundamental para el análisis cinemático cuantitativo, permitiendo cuantificar la magnitud y el patrón del movimiento de cada articulación.

5.2.1.23 Herramientas de Visualización y Análisis:

Para la interpretación y el análisis de estos resultados, se pueden emplear diversas herramientas:

- Visualización Directa de Coordenadas (.trc, .c3d): Los ficheros .trc pueden ser abiertos con cualquier editor de texto o procesados con lenguajes de programación como Python para análisis numéricos personalizados. Para los ficheros .c3d, se requiere un visualizador específico de este formato.

- OpenSim: Es la plataforma principal para el análisis biomecánico detallado. Permite cargar el modelo .osim escalado y los ficheros .mot asociados para visualizar el movimiento del modelo musculoesquelético y analizar los ángulos, velocidades y aceleraciones articulares. OpenSim es la herramienta preferida para análisis cinemáticos y dinámicos complejos, como la estimación de fuerzas y momentos articulares.
- Blender con el Plugin de Pose2Sim: Esta es una herramienta potente para la validación cualitativa visual. El plugin de Pose2Sim para Blender permite importar el modelo .osim y sincronizarlo con los vídeos originales de la grabación. De esta forma, se puede visualizar el modelo esquelético superpuesto directamente sobre el movimiento real del sujeto. Es importante aclarar que, a diferencia del fichero .trc que solo contiene los keypoints y líneas que los unen, la visualización en Blender o OpenSim utiliza el modelo esquelético completo y escalado del fichero .osim.

A continuación, se presentan imágenes de las visualizaciones desarrolladas en Python plotly a partir de los ficheros .trc:

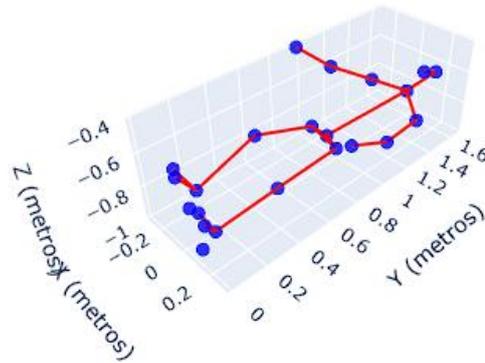
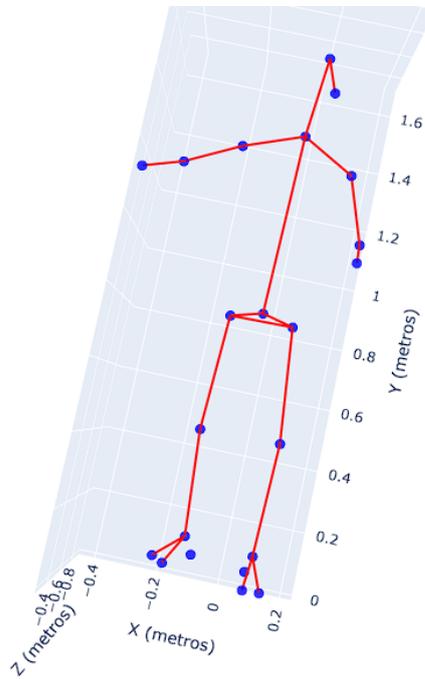


Ilustración 18 - Ejemplo 1, visualización de ficheros .trc



Frame: 95



Ilustración 19 - Ejemplo 2, visualización de ficheros .trc

A continuación, se presentan imágenes del modelo esquelético visualizado en OpenSim, que ilustran la representación final de la pose reconstruida en 3D.

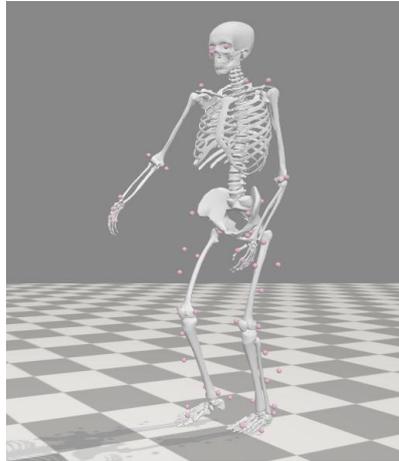


Ilustración 20 - Ejemplo 1, visualización con Pose2Sim

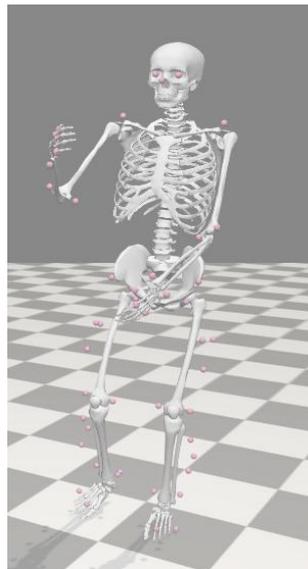


Ilustración 21 - Ejemplo 2, visualización con Pose2Sim

Es crucial señalar que las figuras presentadas en este documento corresponden a instantes estáticos o capturas de pantalla de las simulaciones. Es fundamental comprender que los ficheros de salida generados por Pose2Sim, como los modelos .osim (que representan el modelo biomecánico escalado del sujeto) y los datos .trc (que contienen las coordenadas tridimensionales de los keypoints a lo largo del tiempo), encapsulan la totalidad de los fotogramas de la grabación original. Las herramientas de visualización especializadas, como

OpenSim o el plugin de Pose2Sim para Blender, permiten una exploración dinámica y un análisis detallado del movimiento fotograma a fotograma. No obstante, debido a las limitaciones inherentes al formato de un documento estático (como Word), la presentación de los resultados se restringe a estas imágenes fijas para ejemplificar la pose y la configuración del modelo esquelético.

Para aquellos interesados en replicar o extender la visualización avanzada, los enlaces a los tutoriales oficiales para la descarga de OpenSim y la instalación del plugin de Pose2Sim se pueden encontrar en el repositorio github oficial de Pose2Sim. Son los siguientes:

- Descarga de OpenSim: <https://simtk.org/projects/opensim>
- Plugin de Pose2Sim para Blender:
https://github.com/davidpagnon/Pose2Sim_Blender

5.3 DISEÑO DEL EXPERIMENTO

El proceso de desarrollo y validación del sistema de estimación de pose sin marcadores en el presente Trabajo de Fin de Grado (TFG) se llevó a cabo a través de una serie de experimentos, diseñados para evaluar el rendimiento del algoritmo bajo diversas condiciones. Estos experimentos se clasificaron en distintas categorías según su propósito y el entorno de grabación, permitiendo una progresión desde pruebas exploratorias en entornos controlables hasta ensayos en un entorno de interés, más complejo y con restricciones.

Se realizaron diversos experimentos que podemos dividir en pruebas en entorno controlable o preliminares, pruebas en un entorno más controlado y realista y finalmente las pruebas en el entorno de interés.

EXPERIMENTOS EN ENTORNO CONTROLABLE

Esta primera fase consistió en un total de cuatro experimentos realizados en un entorno doméstico. De estos, tres resultaron fallidos inicialmente, sirviendo como una valiosa curva de aprendizaje para familiarizarse con la configuración de Pose2Sim, la correcta disposición de las cámaras, los parámetros de grabación y los desafíos inherentes a la estimación de pose. Estas pruebas preliminares fueron esenciales para comprender la importancia de la visibilidad completa del cuerpo del sujeto y la estabilidad de las cámaras.

El cuarto experimento en este entorno, aunque no óptimo, arrojó resultados aceptables, permitiendo obtener datos coherentes y realizar las primeras visualizaciones satisfactorias del modelo esquelético. Durante estos experimentos iniciales, la colocación de las cámaras se realizó con métodos caseros y adaptados a la situación, utilizando objetos cotidianos como libros u otros soportes improvisados.

EXPERIMENTO EN ENTORNO REALISTA (GRABACIÓN SATISFACTORIA EN ICAI)

Este grupo comprende un único experimento, el cual resultó ser satisfactorio y fue la base para el análisis cuantitativo detallado del TFG. Se llevó a cabo en un aula de la Universidad Pontificia Comillas (ICAI), un entorno que ofrecía condiciones considerablemente mejores que las pruebas caseras. Para esta grabación, se utilizaron cuatro dispositivos móviles (un iPad, un iPhone 13, un iPhone 15 y un teléfono Android), posicionados estratégicamente para maximizar la cobertura visual del sujeto.

Las condiciones fueron óptimas gracias a la amplitud del aula, la disponibilidad de un patrón de calibración de mayor tamaño y calidad (a diferencia del utilizado en casa), y la posibilidad de ubicar las cámaras en ángulos que favorecían una visión completa y sin oclusiones del sujeto. En el siguiente apartado entraremos en más detalle de la posición de las cámaras y de los requisitos necesarios.

EXPERIMENTOS EN ENTORNO DE INTERÉS (GRABACIONES FALLIDAS EN MOBIOS)

Finalmente, se realizaron dos experimentos de grabación en el laboratorio MOBIOS, considerados como el entorno de interés debido a su relevancia para aplicaciones biomédicas. Sin embargo, ambas pruebas resultaron fallidas para los propósitos de este TFG, principalmente debido a las limitaciones y restricciones inherentes al propio laboratorio.

Las cámaras en MOBIOS estaban fijas en posiciones preestablecidas, con ángulos que dificultaban en gran medida la obtención de una calibración correcta y una visión completa del sujeto. Estas ubicaciones provocaron oclusiones significativas, como la visualización de solo la mitad superior del cuerpo, o el ocultamiento de las piernas por un asiento específico de los experimentos de MOBIOS. Además de las restricciones de posicionamiento de las cámaras, la dificultad en la organización de horarios y la limitada disponibilidad de recursos para adaptar el setup a nuestras condiciones de grabación (por ejemplo, el asiento principal utilizaba pirotecnia de un solo uso y costosa, lo que impedía modificarlo para evitar oclusiones) impidieron obtener datos útiles para el análisis cinemático deseado.

En esta ocasión, se emplearon trípodes y sistemas de soporte más estables para asegurar la inmovilidad de las cámaras tras la calibración.

CONFIGURACIÓN GENERAL DE LA ADQUISICIÓN DE DATOS

A lo largo de todos los experimentos, independientemente del entorno, se siguieron una serie de directrices generales para la adquisición de datos:

Calibración Previa: En todos los casos, se realizó una calibración intrínseca y extrínseca de las cámaras previo a cada sesión de grabación. Este paso es fundamental para establecer la relación entre el espacio 2D de la imagen y el espacio 3D real, y para corregir las distorsiones de la lente.

Visibilidad del Cuerpo Entero: Se priorizó siempre la grabación desde ángulos que permitieran una visión completa del cuerpo del sujeto. La experiencia demostró que las

oclusiones o la visualización parcial del cuerpo afectan drásticamente la capacidad del algoritmo para realizar una estimación de pose precisa y una triangulación 3D coherente.

Estabilidad de las Cámaras: Se hizo hincapié en asegurar la inmovilidad de las cámaras una vez realizada la calibración extrínseca. Inicialmente, esto se logró con métodos caseros y, progresivamente, se emplearon trípodes y sistemas de soporte más robustos.

Capítulo 6. ANÁLISIS DE RESULTADOS

6.1 INTRODUCCIÓN AL ANÁLISIS DE RESULTADOS

El presente capítulo tiene como objetivo principal la evaluación del rendimiento, la precisión y la robustez del sistema de estimación de pose humana sin marcadores basado en Pose2Sim. Para ello, se realizará un análisis crítico de los datos obtenidos de diversas configuraciones experimentales. La investigación se centrará en dos vertientes: por un lado, un análisis cuantitativo detallado de una grabación exitosa realizada en condiciones controladas en la universidad; por otro, una investigación cualitativa de las pruebas fallidas llevadas a cabo en el laboratorio MOBIOS. Esta dualidad en el análisis permitirá no solo cuantificar las capacidades del algoritmo en escenarios óptimos, sino también comprender las limitaciones y los factores críticos que pueden comprometer la obtención de resultados fiables, extrayendo así conclusiones significativas para el futuro del sistema en entornos aplicados.

CONDICIONES EXPERIMENTALES Y CONFIGURACIÓN DE LA GRABACIÓN EXITOSA EN LA UNIVERSIDAD

La grabación seleccionada para el análisis cuantitativo principal de este estudio se realizó en un aula de la universidad, un entorno que ofreció condiciones buenas para la adquisición de datos de alta calidad. Este escenario permitió la obtención de resultados coherentes, fundamentales para evaluar de manera fiable el rendimiento de Pose2Sim.

Para la captura de los vídeos se utilizaron cuatro dispositivos móviles: un iPad, un iPhone 13, un iPhone 15 y un teléfono Android proporcionado por el tutor del TFG. La disposición de las cámaras se planificó para maximizar la visibilidad del sujeto y cubrir el espacio de grabación de manera eficiente.

A continuación, se presenta un diagrama ilustrativo de la colocación de las cámaras durante esta sesión de grabación:

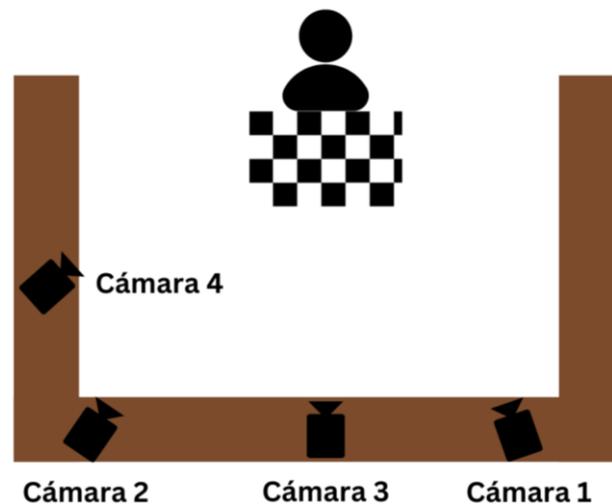


Ilustración 22 - Ubicación de las Cámaras grabación ICAI

Las condiciones del aula fueron ventajosas debido a su amplitud, lo que proporcionó un espacio suficiente para la configuración de las cámaras y el movimiento del sujeto sin obstrucciones significativas. Además, se contó con un patrón de calibración de tamaño adecuado, lo que no fue posible en pruebas previas realizadas en un entorno doméstico y contribuyó significativamente a la precisión de la calibración de las cámaras. La ubicación de las cuatro cámaras se seleccionó cuidadosamente para ofrecer diversas perspectivas, lo que es esencial para una triangulación 3D robusta. Durante la grabación, el sujeto realizó movimientos simples, diseñados para ser representativos de actividades cotidianas y para facilitar la detección de keypoints por el algoritmo.

Para ilustrar el entorno y la calidad de las grabaciones, a continuación, se muestran algunas imágenes de una de las cámaras utilizadas en esta sesión:



Ilustración 23 - Imagen Calibración Extrínseca Cámara 1

6.2 METODOLOGÍA DEL ANÁLISIS

HERRAMIENTAS Y SELECCIÓN DE DATOS PARA EL ANÁLISIS CUANTITATIVO

Para llevar a cabo el análisis cuantitativo del rendimiento de Pose2Sim, se realizó una selección estratégica de los datos de grabación. La elección recayó en la sesión de grabación llevada a cabo en el aula de la universidad, dado que fue la realizada bajo las condiciones más favorables y arrojó resultados satisfactorios. Esta grabación se benefició de un mayor número de cámaras (cuatro dispositivos móviles: un iPad, un iPhone 13, un iPhone 15 y un teléfono Android proporcionado por el tutor del TFG) y un patrón de calibración de alta calidad, elementos que no estuvieron presentes en las pruebas preliminares realizadas en un entorno doméstico. Aunque las grabaciones en casa también produjeron resultados visuales interesantes, la calidad subóptima del tablero de calibración utilizado comprometió la precisión de la calibración, lo que las hizo menos idóneas para un análisis de consistencia riguroso.

Es fundamental contrastar este éxito con las grabaciones intentadas en el laboratorio MOBIOS, las cuales, a pesar de utilizar las mismas cámaras y configuraciones, no resultaron

funcionales para el análisis cuantitativo debido a problemas significativos que se detallarán en el apartado 6.4.

El análisis de la información de los datos 3D se realizó mediante un script de Python desarrollado específicamente para este proyecto. Este script, ejecutado en un entorno Jupyter Notebook, se encarga de extraer la información de todos los puntos (keypoints) contenida en los ficheros .trc generados por Pose2Sim. Su funcionalidad permite seleccionar huesos específicos para la comparación (por ejemplo, el húmero derecho y la tibia derecha, como se definirá en el apartado 0), y calcular las medidas estadísticas que se explicarán en el siguiente apartado (0 Métricas de Evaluación Cuantitativa). Este enfoque programático garantiza la automatización y la consistencia en el procesamiento de los datos para el análisis cuantitativo.

MÉTRICAS DE EVALUACIÓN CUANTITATIVA

Para evaluar de manera objetiva la consistencia y fiabilidad de las longitudes de los segmentos óseos reconstruidos en 3D por Pose2Sim, se emplearon varias métricas estadísticas clave que permiten cuantificar su valor central y su dispersión. Estas métricas se estudian a lo largo de todos los fotogramas del vídeo.

La Media (Promedio) de la longitud de un hueso a lo largo de todos los fotogramas de una grabación proporciona una estimación del tamaño promedio de dicho hueso tal como lo calcula el algoritmo, siendo fundamental para determinar si la longitud estimada se mantiene cercana a lo esperado anatómicamente.

La Desviación Típica es una medida de dispersión que cuantifica la variabilidad absoluta de las longitudes de los huesos; un valor bajo indica que las mediciones son consistentes y estables a lo largo del tiempo.

La Varianza, siendo el cuadrado de la desviación típica, también mide la dispersión de los datos respecto a la media, proporcionando una cuantificación de la magnitud total de la variabilidad, aunque su unidad la hace menos intuitiva para la interpretación directa.

Finalmente, el Coeficiente de Variación (CV) es una métrica crucial para la comparación de la consistencia relativa entre diferentes huesos o bajo distintas condiciones de grabación, ya que normaliza la dispersión por la magnitud de la media. Un CV más bajo indica una mayor consistencia relativa en la longitud estimada del hueso, permitiendo evaluar la estabilidad de forma comparativa.

CONFIGURACIONES EXPERIMENTALES A COMPARAR

Para el análisis cuantitativo del rendimiento del sistema Pose2Sim en la grabación realizada en el entorno universitario, se establecieron y compararon las siguientes configuraciones experimentales:

Configuraciones de Cámaras: Se evaluó el impacto del número de cámaras en la estimación de pose 3D mediante simulaciones con:

- 4 cámaras.
- 3 cámaras.
- 2 cámaras.

Efecto del Filtrado: Para la configuración con 4 cámaras, se realizó una comparación específica para determinar la influencia del filtrado de datos:

- Datos con filtro Butterworth aplicado.
- Datos sin filtro aplicado.

Segmentos Óseos Analizados: Los huesos seleccionados para el análisis de longitud y consistencia fueron el Húmero derecho y la Tibia derecha. Estos segmentos fueron elegidos estratégicamente por su representatividad en el cuerpo humano y por la visibilidad relativa

de sus keypoints asociados (hombro-codo y rodilla-tobillo, respectivamente) en las condiciones de grabación, permitiendo una evaluación robusta del comportamiento del algoritmo en segmentos de diferente longitud y función.

6.3 RESULTADOS CUANTITATIVOS DE LA GRABACIÓN EN LA UNIVERSIDAD

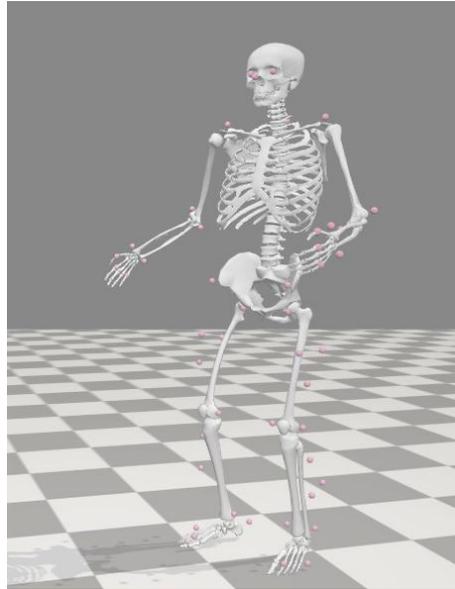
Esta sección presenta los hallazgos cuantitativos derivados del análisis de la grabación realizada en el aula de la universidad, seleccionada por sus condiciones óptimas de captura. Se detallarán las métricas estadísticas clave obtenidas para las longitudes de segmentos óseos específicos, explorando el impacto del número de cámaras y la aplicación de filtrado, con el fin de determinar las configuraciones que arrojan los resultados más consistentes.

ANÁLISIS DE LONGITUDES ÓSEAS POR CONFIGURACIÓN DE CÁMARAS

Este subapartado se enfoca en la evaluación de la consistencia de las longitudes de los segmentos óseos reconstruidos en 3D en función del número de cámaras utilizadas para la triangulación. Se presentarán los resultados estadísticos (media, desviación estándar y varianza) para el húmero derecho y la tibia derecha en configuraciones de 4, 3 y 2 cámaras, con el objetivo de dilucidar la influencia de la redundancia de datos en la precisión de las mediciones del algoritmo. Es relevante añadir, que dado que Pose2Sim dispone de las funciones de filtrado y de Marker Augmentation, utilizaremos ambas funciones para el análisis de 2, 3 y 4 cámaras, y estudiaremos a parte el resultado sin filtrado para el caso de 4 cámaras.

6.3.1.1 Análisis con 4 Cámaras

6.3.1.1.1 Análisis a simple vista



La observación cualitativa de las simulaciones y los vídeos generados con el modelo esquelético superpuesto, utilizando la configuración de 4 cámaras, revela una notable coherencia y fluidez en el movimiento reconstruido. A simple vista, el modelo sigue de forma precisa la pose del sujeto a lo largo de toda la grabación, mostrando una progresión natural de las articulaciones y los segmentos corporales. No se aprecian desperfectos significativos, saltos abruptos, oclusiones prolongadas del modelo o artefactos importantes que comprometan la interpretación visual del movimiento. Esta evaluación preliminar sugiere una estimación de pose 3D de alta calidad, proporcionando una base sólida para el análisis cuantitativo subsiguiente.

6.3.1.1.2 Análisis estadístico de los ficheros .trc

Para complementar el análisis cualitativo, se realizó un estudio estadístico de las longitudes del húmero derecho y la tibia derecha, extraídas directamente de los ficheros .trc generados por Pose2Sim con la configuración de 4 cámaras. Este análisis tiene como objetivo cuantificar la consistencia y la variabilidad de estas mediciones a lo largo de la secuencia de movimiento.

Los resultados obtenidos para la media, desviación típica, varianza y el Coeficiente de Variación (CV) son los siguientes:

Hueso	Media (cm)	Desviación Típica (cm)	Varianza (cm^2)	Coefficiente de variación (%)
Húmero derecho	27.34	2.66	7.084	9.74
Tibia Derecha	37.96	2.78	7.70	7.31

Tabla 3 - Resultados ensayo con 4 cámaras

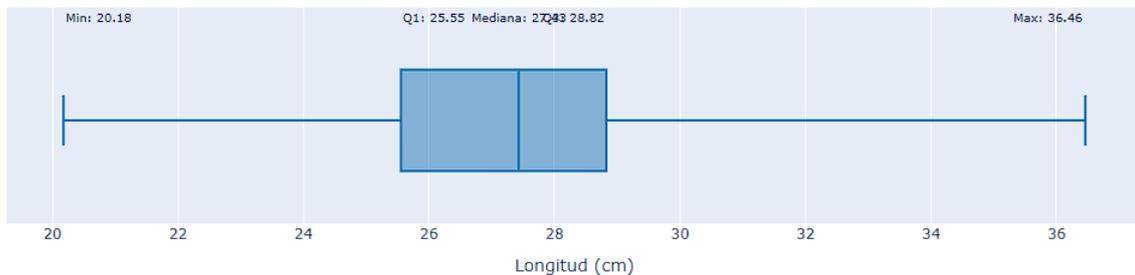


Ilustración 24 - Distribución de Longitud del Húmero derecho (4 Cámaras)

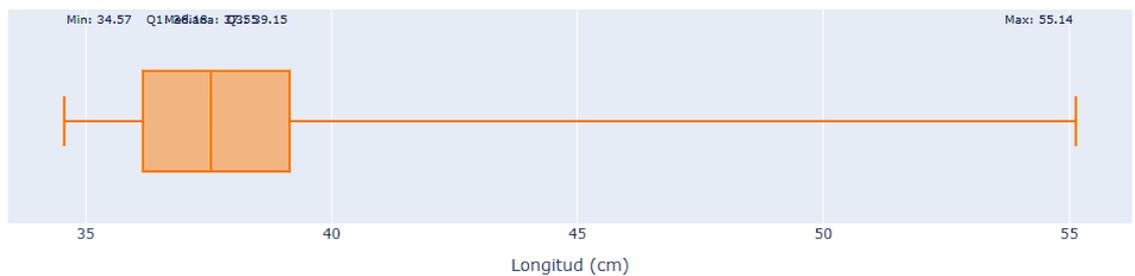


Ilustración 25 - Distribución de longitud de la Tibia derecha (4 Cámaras)

Interpretación de los Resultados:

- Media de las Longitudes: La media de la longitud del húmero derecho se sitúa en 27.34 cm y la de la tibia derecha en 37.96 cm. Estos valores representan la estimación promedio de la longitud de estos segmentos óseos a lo largo de toda la secuencia de movimiento analizada. La razonabilidad de estas medias se evaluará en un apartado posterior mediante la comparación con mediciones antropométricas de referencia del sujeto.

- **Dispersión (Desviación Típica y Varianza):** Los valores de desviación típica y varianza cuantifican la dispersión de las mediciones de longitud del hueso alrededor de su media. Un valor menor en estas métricas sugiere una mayor estabilidad y consistencia en la estimación de la longitud del hueso a lo largo de los fotogramas.
 - Para el húmero derecho, la desviación típica de 2.66 cm y la varianza de 7.0846 cm^2 indican una cierta fluctuación en su longitud estimada.
 - En contraste, la tibia derecha presenta una desviación típica de 2.78 cm y una varianza de 7.7021 cm^2 .
- **Coefficiente de Variación (CV):** El CV permite evaluar la dispersión relativa de los datos, siendo una métrica valiosa para comparar la consistencia entre segmentos de diferentes longitudes o bajo distintas condiciones de manera normalizada.
 - El húmero derecho muestra un CV del 9.74%.
 - La tibia derecha, por su parte, presenta un CV del 7.31%.

En conclusión, el análisis estadístico de los datos obtenidos con 4 cámaras, sin aplicar el marker augmentation, revela que, si bien el algoritmo proporciona una estimación de las longitudes óseas, la consistencia en estas mediciones varía entre segmentos. La tibia derecha demuestra una mayor estabilidad relativa en su longitud estimada a lo largo del tiempo, con un CV ligeramente inferior, en comparación con el húmero derecho. Estos hallazgos serán fundamentales para comprender las limitaciones y fortalezas del sistema, y para futuras comparaciones con otras configuraciones de cámaras y la aplicación de filtrado.

6.3.1.2 Análisis con 3 cámaras

6.3.1.2.1 Análisis a simple vista

La inspección visual de las simulaciones y los vídeos generados con el modelo esquelético utilizando una configuración de 3 cámaras, muestra resultados similares en coherencia y fluidez a los observados con 4 cámaras. El seguimiento de la pose del sujeto se mantiene en un nivel satisfactorio, y el movimiento reconstruido es, en general, fluido y libre de grandes desperfectos. Si bien podría esperarse una ligera degradación de la calidad debido a la menor redundancia de información, esta no es evidente a una simple observación visual, lo que

sugiere que 3 cámaras aún proporcionan suficientes datos para una estimación de pose 3D cualitativamente buena en las condiciones de grabación controladas.

6.3.1.2.2 Análisis estadístico de los ficheros .trc

Para el análisis cuantitativo de la configuración con 3 cámaras, se siguió la misma metodología que para la configuración de 4 cámaras, extrayendo las longitudes del húmero derecho y la tibia derecha de los ficheros .trc. Este análisis también se realizó **sin la aplicación del módulo de *Marker Augmentation***.

Hueso	Media (cm)	Desviación Típica (cm)	Varianza (cm^2)	Coefficiente de variación (%)
Húmero derecho	25.5	3.71	13.77	14.55
Tibia Derecha	37.10	2.24	5.03	6.05

Tabla 4 - Resultados del ensayo con 3 cámaras

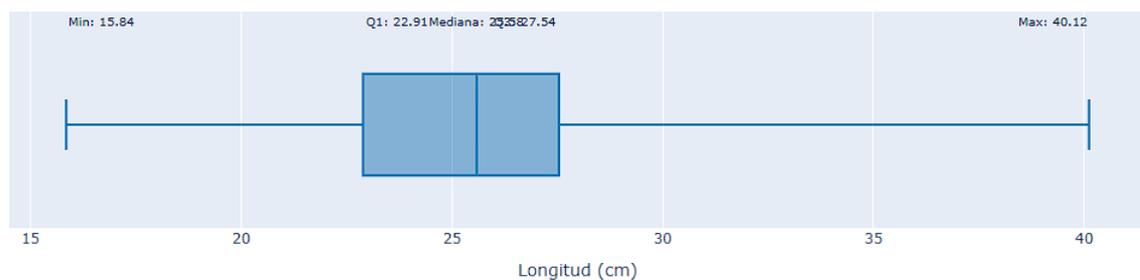


Ilustración 26 - Distribución de Longitud del Húmero derecho (3 Cámaras)

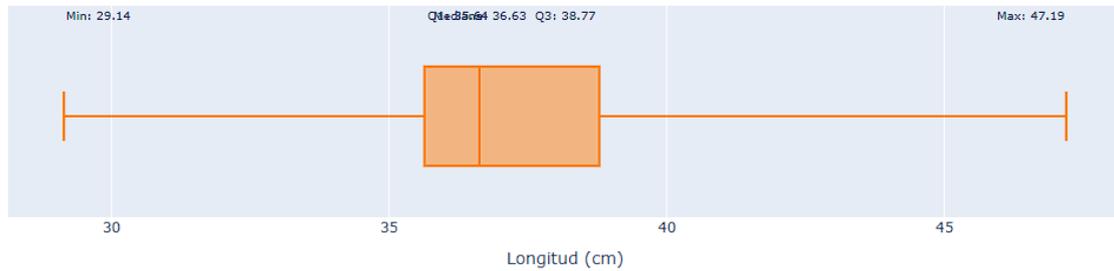


Ilustración 27 - Distribución de longitud de la tibia derecha (3 Cámaras)

Interpretación de los Resultados:

Media de las Longitudes: Las medias de longitud estimadas para el húmero derecho (25.5 cm) y la tibia derecha (37.10 cm) son consistentes con las observadas en la configuración de 4 cámaras, lo que indica que la reducción de una cámara no ha alterado significativamente la estimación promedio de la magnitud de estos segmentos óseos.

Al comparar estos coeficientes de variación con los obtenidos en la configuración de 4 cámaras (9.74% para el húmero y 7.31% para la tibia, según los datos proporcionados para la sección 6.3.1.1.2), se observa una tendencia interesante. Para el húmero, el CV ha aumentado de 9.74% a 14.55%, lo que sugiere una mayor variabilidad relativa en la estimación de su longitud al reducir una cámara. Esto refuerza la hipótesis de que los segmentos más cortos o aquellos con keypoints más susceptibles a errores (como los del brazo/hombro) son más sensibles a la disminución de la redundancia de cámaras.

En contraste, el CV de la tibia derecha ha disminuido de 7.31% a 6.05% con la configuración de 3 cámaras. Esta aparente mejora en la consistencia podría deberse a una optimización no lineal en la triangulación o a que la eliminación de una cámara específica en la configuración de 4 cámaras (que pudo haber contribuido a una mayor dispersión para la tibia) ha resultado en un conjunto de 3 cámaras más efectivo para este segmento en particular. Sin embargo, esta observación contraria a la intuición general de que más cámaras conducen a mayor precisión, subraya la complejidad de la triangulación y la interacción entre la calidad de las detecciones 2D y la geometría de las cámaras.

En resumen, la reducción de 4 a 3 cámaras parece tener un impacto diferenciado en la consistencia de la estimación de longitud de los segmentos óseos. El húmero muestra una mayor sensibilidad a esta reducción, incrementando su variabilidad, mientras que la tibia, inesperadamente, muestra una ligera mejora en su consistencia. Este análisis resalta la importancia de evaluar cada segmento individualmente y considerar la geometría específica de las cámaras al optimizar la configuración de un sistema de estimación de pose.

6.3.1.3 Análisis con 2 cámaras

6.3.1.3.1 Análisis a simple vista

Sorprendentemente, la observación cualitativa de las simulaciones y los vídeos generados con el modelo esquelético utilizando una configuración de tan solo 2 cámaras sigue revelando un muy buen resultado a simple vista. El movimiento reconstruido se mantiene notablemente suave y la estimación de pose presenta una coherencia considerable. Este desempeño inesperado probablemente se deba a dos factores principales: la excelente posición de las cámaras durante la grabación, que permitió una buena visualización de los keypoints cruciales a pesar de la menor redundancia, y la efectividad del filtrado posterior, que contribuye significativamente a la suavidad percibida del movimiento y a la reducción de artefactos. Esta observación resalta la capacidad del sistema para generar visualizaciones cualitativamente aceptables incluso con un número limitado de cámaras, siempre y cuando las condiciones de captura sean favorables.

6.3.1.3.2 Análisis estadístico de los ficheros .trc

Para el análisis cuantitativo de la configuración con 2 cámaras, se siguió la misma metodología que en los casos anteriores, extrayendo las longitudes del húmero derecho y la tibia derecha de los ficheros .trc. Este análisis también se realizó sin la aplicación del módulo de Marker Augmentation.

Los resultados obtenidos para la media, desviación típica, varianza y el Coeficiente de Variación (CV) son los siguientes:

Hueso	Media (cm)	Desviación Típica (cm)	Varianza (cm^2)	Coeficiente de variación (%)
Húmero derecho	25.32	2.93	8.59	11.58
Tibia Derecha	37.03	2.07	4.30	5.6

Tabla 5 - Resultados del ensayo con 2 cámaras

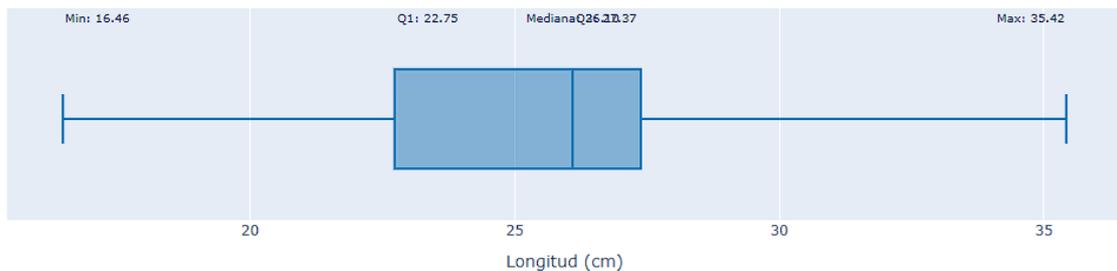


Ilustración 28 - Distribución de Longitud del Húmero derecho (2 Cámaras)

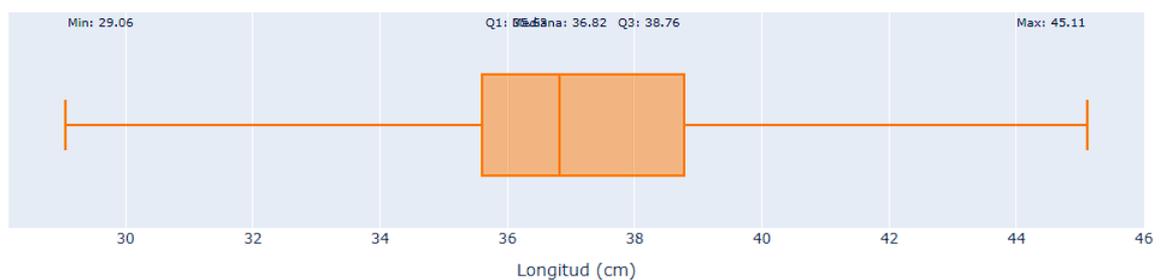


Ilustración 29 - Distribución de Longitud de la tibia derecha (2 Cámaras)

Interpretación de los resultados

Media de las Longitudes: Las medias de longitud estimadas para el húmero derecho (25.32 cm) y la tibia derecha (37.03 cm) se mantienen consistentes con las observadas en las configuraciones de 4 y 3 cámaras. Esto indica que, en términos de la estimación promedio

de la longitud del segmento, el sistema de 2 cámaras aún es capaz de producir valores comparables.

La desviación típica del húmero derecho es de 2.93 cm, con una varianza de 8.59 cm^2 . Al comparar con la configuración de 4 cámaras (2.66 cm y 7.08 cm^2) y 3 cámaras (3.71 cm y 13.77 cm^2), se observa que la dispersión del húmero con 2 cámaras es mayor que con 4 cámaras, pero notablemente menor que con 3 cámaras. Esto sugiere una compleja interacción entre el número de cámaras y la geometría específica que puede afectar la estabilidad de la triangulación para segmentos cortos como el húmero.

Para la tibia derecha, la desviación típica es de 2.07 cm y la varianza es de 4.30 cm^2 . Comparando con las configuraciones de 4 cámaras (2.78 cm y 7.70 cm^2) y 3 cámaras (2.24 cm y 5.03 cm^2), se evidencia que la configuración de 2 cámaras arroja la menor dispersión absoluta para este segmento. Esto refuerza la observación de que, para la tibia en las condiciones de grabación dadas, dos cámaras bien posicionadas pueden ser altamente efectivas.

El húmero derecho presenta un CV del 11.58%. Al comparar este valor con el de 4 cámaras (9.74%) y 3 cámaras (14.55%), se observa que el CV del húmero con 2 cámaras se sitúa en un punto intermedio, ligeramente superior al de 4 cámaras, pero inferior al de 3 cámaras.

La tibia derecha muestra un CV del 5.60%. Este es el valor de CV más bajo observado para la tibia en todas las configuraciones de cámaras analizadas hasta ahora (4 cámaras: 7.31%; 3 cámaras: 6.05%). Esto indica que, para este segmento, la configuración de 2 cámaras ha resultado en la estimación de longitud más consistente y menos variable en términos relativos. Este resultado es particularmente notable, ya que desafía la intuición de que una mayor redundancia de cámaras siempre conduce a una mayor precisión y consistencia.

En resumen, el análisis con 2 cámaras revela un rendimiento sorprendentemente robusto. Si bien se esperaría una mayor variabilidad debido a la menor redundancia, la calidad de la posición de las cámaras y la efectividad del filtrado parecen mitigar en gran medida este efecto. La tibia derecha, en particular, demuestra una consistencia excepcional con solo dos

cámaras, lo que sugiere que para ciertos segmentos y bajo condiciones de grabación óptimas, un sistema de dos cámaras puede ser suficiente para obtener mediciones de longitud fiables. Esto demuestra que, a veces, sucede lo que ocurre en las regresiones: más datos no siempre implican mayor precisión, sino que pueden llevar a una menor si se tienen demasiados datos o si no se gestionan adecuadamente. Estos hallazgos son importantes para la optimización de recursos y la simplificación de setups en futuras aplicaciones. Como siguiente paso, aunque no se estudiará en este TFG, sería crucial evaluar con una herramienta como Vicon, de la que ya se conoce su alta precisión, si la ubicación estimada por este algoritmo es la real.

IMPACTO DEL FILTRADO BUTTERWORTH (CASO 4 CÁMARAS)

Este apartado se centrará en analizar el efecto de la aplicación del filtro Butterworth sobre la calidad de los datos de pose 3D, utilizando como base la configuración de 4 cámaras, que ofrece la mayor redundancia de información de entrada.

6.3.1.4 Analisis a simple vista

La observación cualitativa de las simulaciones y los vídeos generados con el modelo esquelético, cuando no se aplica ningún tipo de filtrado a los datos triangulados, revela una notable pérdida de suavidad en el movimiento reconstruido. A simple vista, aunque la estimación de la pose y la posición del sujeto puede ser globalmente correcta, los movimientos individuales son abruptos y la trayectoria de los keypoints presenta una gran cantidad de ruido. Esto se manifiesta en la aparición de falsas o incorrectas estimaciones, con "saltos" y fluctuaciones rápidas que no se corresponden con un movimiento biológico real. La consistencia visual de la estimación se ve comprometida, lo que dificulta la interpretación cualitativa y la extracción de información significativa del movimiento, evidenciando la necesidad de un proceso de suavizado posterior a la triangulación.

La mejor forma de representar en imágenes el efecto del filtrado, es mediante las gráficas de filtrado que Pose2Sim nos entrega. A continuación, se presentan varias gráficas de varios

puntos clave comparando los valores sin filtrar (en azul) y filtrados (en naranja) en donde se puede apreciar el ruido existente, sobre todo en los valores de Y. Este ruido, es el que genera que, a simple vista, los resultados no sean idóneos.

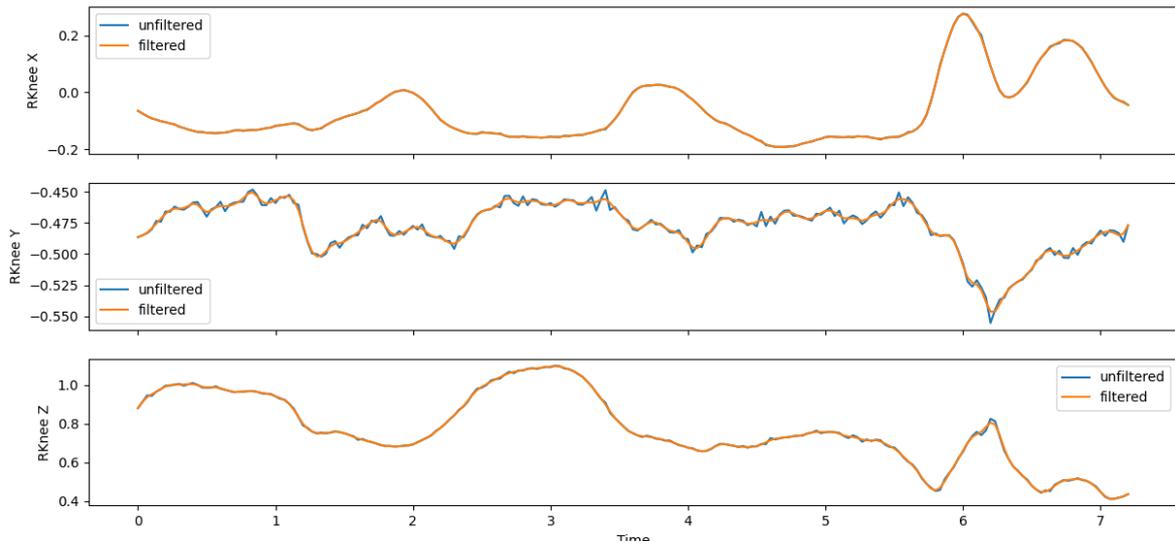


Ilustración 30 - Comparación de valores filtrados RKnee

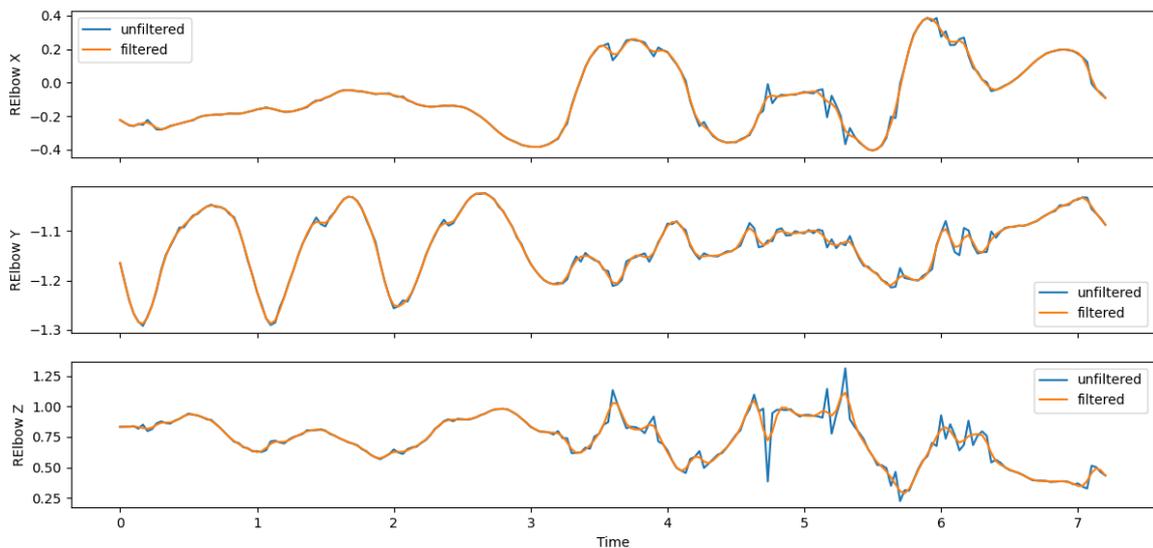


Ilustración 31 - Comparación de valores filtrados RElbow

6.3.1.5 Análisis estadístico de los ficheros .trc

Para cuantificar el impacto del filtrado, se presenta a continuación el análisis estadístico de las longitudes del húmero derecho y la tibia derecha, obtenidos de los ficheros .trc generados con 4 cámaras sin la aplicación del filtro Butterworth. Posteriormente, se compararán estos resultados con los obtenidos con el filtro aplicado (6.3.1.1 Análisis con 4 Cámaras)

Hueso	Media (cm)	Desviación Típica (cm)	Varianza (cm^2)	Coefficiente de variación (%)
Húmero derecho	27.51	3.03	9.17	11.01
Tibia Derecha	37.98	3.18	10.13	8.38

Tabla 6 - - Resultados del ensayo con 4 cámaras sin filtrado

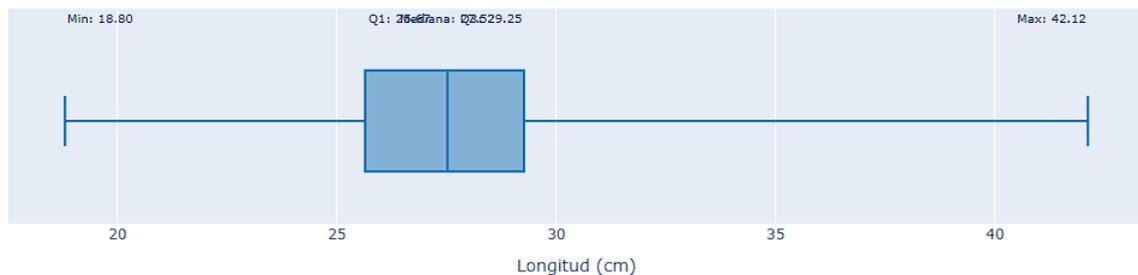


Ilustración 32 - Distribución de Longitud del Húmero derecho (4 Cámaras sin filtro)

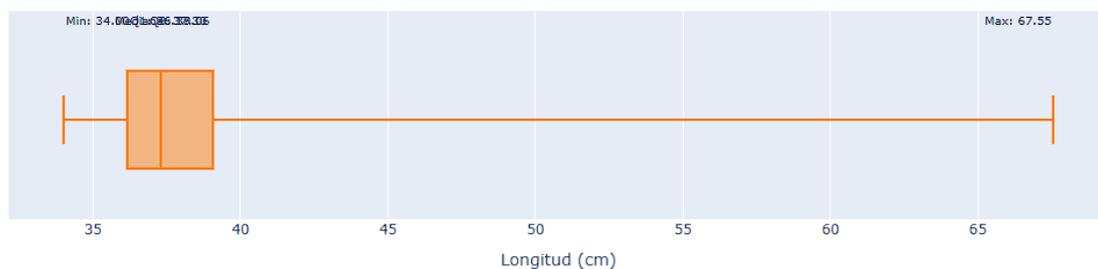


Ilustración 33 - Distribución de longitud de la Tibia derecha (4 Cámaras sin filtro)

Media de las Longitudes: Las medias de longitud para el húmero derecho (27.51 cm) y la tibia derecha (37.98 cm) en los datos sin filtro son muy similares a las obtenidas con el filtro aplicado (27.34 cm y 37.96 cm, respectivamente). Esto era de esperar, ya que la aplicación de un filtro de suavizado como Butterworth no busca alterar la magnitud promedio de la señal, sino reducir el ruido y la variabilidad.

Dispersión (Desviación Típica y Varianza):

- Para el húmero derecho, la desviación típica es de 3.03 cm y la varianza de 9.17 cm^2 sin filtro. Al comparar con los datos filtrados (desviación típica de 2.66 cm y varianza de 7.08 cm^2), se observa un aumento en la dispersión sin la aplicación del filtro. Esto confirma que el filtro Butterworth reduce la variabilidad absoluta de las longitudes estimadas del húmero.
- Para la tibia derecha, la desviación típica es de 3.18 cm y la varianza de 10.13 cm^2 sin filtro. Al comparar con los datos filtrados (desviación típica de 2.78 cm y varianza de 7.70 cm^2), también se aprecia un aumento en la dispersión sin el filtro. Esto demuestra que el filtro es efectivo en suavizar las mediciones de la tibia, reduciendo su variabilidad.

Coefficiente de Variación (CV):

- El húmero derecho presenta un CV del 11.01% sin filtro. Al comparar con el CV del húmero filtrado (9.74%), se observa que el filtro ha logrado una reducción de la variabilidad relativa de este segmento. Un CV más bajo indica una mayor consistencia de las mediciones de longitud.
- La tibia derecha muestra un CV del 8.38% sin filtro. Comparando con el CV de la tibia filtrada (7.31%), también se evidencia una reducción en la variabilidad relativa gracias al filtro Butterworth.

En conclusión, el análisis estadístico valida la observación cualitativa: la aplicación del filtro Butterworth tiene un impacto positivo y significativo en la consistencia de las longitudes de los segmentos óseos. Si bien las medias de longitud se mantienen estables, la desviación

típica, la varianza y, crucialmente, el Coeficiente de Variación, se reducen notablemente con el filtro aplicado. Esto indica que el filtro logra suavizar el ruido inherente a la triangulación 3D, resultando en trayectorias de keypoints más estables y, por ende, en mediciones de longitud de segmentos óseos más consistentes, lo cual es fundamental para cualquier análisis biomecánico preciso.

COMPARACIÓN CON MEDIDAS ANTROPOMÉTRICAS DE REFERENCIA

Finalmente, se ha decidido no realizar un análisis detallado en este apartado. Para poder obtener una precisión relevante en la comparación con medidas antropométricas de referencia, sería indispensable contar con datos obtenidos mediante técnicas de alta fiabilidad, como radiografías, resonancias magnéticas o mediciones realizadas con instrumentación especializada. Las mediciones "a ojo", aunque útiles para una primera aproximación cualitativa, resultan insuficientes para validar la precisión del algoritmo a un nivel adecuado, dado que la variabilidad intrínseca del estudio se sitúa en el rango de los centímetros. La baja precisión inherente a la medición visual haría que cualquier comparación en este contexto resultara inconsistente y carente de validez científica para los objetivos de este trabajo.

6.4 ANÁLISIS DE LAS PRUEBAS FALLIDAS EN MOBIOS

DESCRIPCIÓN DE LAS CONDICIONES Y DESAFÍOS EN EL LABORATORIO MOBIOS

Las pruebas realizadas en el laboratorio MOBIOS de la Universidad Pontificia Comillas, un centro de biomecánica de impacto que se especializa en simulaciones de seguridad vial, presentaron desafíos significativos que impidieron la obtención de datos útiles para el análisis cinemático mediante Pose2Sim. Este laboratorio está equipado con una máquina de ensayo de impacto que utiliza un pistón para acelerar una plataforma instantáneamente a la velocidad deseada, normalmente unos 40 km/h, simulando la experiencia de un cuerpo humano en el momento de un accidente, y frenando la plataforma casi al instante. Los ensayos habituales se realizan con cadáveres reales, lo que impone estrictas implicaciones

legales, de higiene, limpieza y respeto al cadáver, limitando drásticamente el acceso y la manipulación del entorno.

Para la adquisición de datos de vídeo, MOBIOs cuenta con su propio sistema de cámaras. Se intentó utilizar las tres cámaras de alta velocidad del laboratorio, configuradas a 120 FPS y posteriormente reframing a 60 FPS para compatibilidad. Aunque el laboratorio dispone de cuatro cámaras principales, una de ellas está específicamente posicionada para capturar el cinturón de seguridad y no ofrecía una vista útil para la estimación de pose del cuerpo, por lo que no pudo ser empleada. Las otras dos cámaras estaban colocadas enfrentadas para capturar el asiento desde una perspectiva lateral, y una cuarta cámara estaba ubicada directamente encima del asiento. Todas estas cámaras estaban fijadas en trípodes permanentes, inamovibles debido a la configuración especializada del laboratorio para sus propios experimentos. A continuación, un diagrama con la posición de las cámaras fijas de MOBIOs.

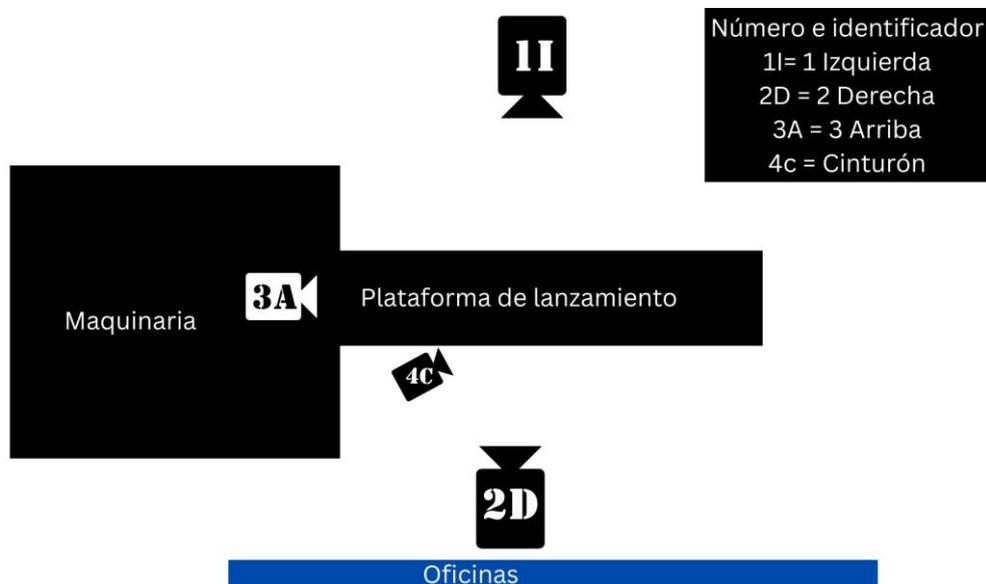


Ilustración 34 - Diagrama de Posición de las cámaras de MOBIOs

La fijación de las cámaras en posiciones preestablecidas representó un obstáculo crítico para la calibración con el tablero de ajedrez. La geometría de las cámaras (dos laterales enfrentadas y una cenital) hacía extremadamente difícil encontrar una posición óptima del

tablero donde todas las cámaras pudieran verlo de manera aceptable simultáneamente. Esto resultó en calibraciones con altos errores de reproyección o, en algunos casos, directamente inviábiles, comprometiendo la precisión de la reconstrucción 3D.

Adicionalmente, la imposibilidad de realizar el test con un cadáver, debido a las restricciones de acceso y manipulación, nos obligó a utilizar un sujeto vivo. Las posiciones que podíamos adoptar sobre la plataforma eran muy limitadas, y en el último intento de grabación, el asiento del simulador tapaba una parte significativa del cuerpo, ocasionando una oclusión severa de la mitad inferior del mismo.

Las restricciones logísticas y de horarios también contribuyeron al fracaso de las pruebas. El personal de MOBIOs trabaja a tiempo completo en sus propios proyectos de investigación, lo que dificultaba coordinar los tiempos para realizar nuestros ensayos sin interferir con sus actividades diarias. Esto implicaba que no podíamos acceder al laboratorio en los días o momentos en que se realizaban sus ensayos principales, y conseguir cuadrar una visita era complejo. Un ejemplo claro de las limitaciones de recursos fue la imposibilidad de adaptar el asiento del simulador: para simular los accidentes, se utiliza un respaldo en el asiento que es eyectado con un equipo pirotécnico justo antes de la aceleración de la plataforma. Este sistema de pirotecnia es costoso y de un solo uso por ensayo, lo que impidió la instalación de un respaldo adaptable para nuestros propósitos y contribuyó a la oclusión de las piernas del sujeto.

En resumen, la combinación de la disposición inamovible de las cámaras con ángulos desfavorables, las severas oclusiones causadas por el equipamiento del laboratorio, y las estrictas restricciones logísticas y de recursos, hicieron que las grabaciones en MOBIOs resultaran inviábiles para el análisis cinemático deseado con Pose2Sim.

DESAFÍOS EN LA CALIBRACIÓN

Un factor determinante en el fracaso de las pruebas en MOBIOs, independientemente de la oclusión directa del sujeto, fueron los desafíos inherentes a la calibración de las cámaras. Debido a la posición fija y a menudo comprometida de las cámaras del laboratorio, resultaba

extremadamente difícil conseguir que todas las cámaras a la vez viesen con un ángulo suficientemente amplio y claro el tablero de calibración.

La disposición de las cámaras (dos enfrentadas lateralmente al asiento y una cenital) no facilitaba la visualización simultánea del tablero desde múltiples perspectivas óptimas. Esto condujo a la necesidad de colocar el tablero en posiciones extrañas y complicadas dentro de la plataforma de impacto, comprometiendo la calidad de las imágenes de calibración. A pesar de los esfuerzos, no existía una posición sencilla en la plataforma que permitiera a todas las cámaras capturar el tablero de manera idónea para una calibración robusta.

Aunque finalmente se consiguió realizar la calibración, es muy posible que la calidad de esta calibración fuese deficiente debido a estos motivos. Una calibración imprecisa se traduce directamente en errores en la reconstrucción 3D de los keypoints, independientemente de la calidad de la detección 2D. Este fue otro de los motivos cruciales por los que se decidió añadir más cámaras a la configuración del segundo experimento (en la universidad), buscando asegurar una calibración de mayor calidad que sería fundamental para obtener resultados fiables.

PROPAGACIÓN DE ERRORES POR OCLUSIÓN EN LOS DATOS 3D

Los desafíos encontrados en el laboratorio MOBIO, particularmente las severas oclusiones, tuvieron un impacto directo y determinante en la capacidad de Pose2Sim para generar datos 3D coherentes y utilizables.

En las dos primeras pruebas, el sujeto vivo (yo) se encontraba sentado en la plataforma de impacto. En esta configuración, una parte intrínseca de la plataforma ocultaba significativamente la ubicación de los pies del sujeto. Esto sucedía en la cámara cenital, que a pesar de su posición elevada, tenía dificultades para obtener una visión completa del cuerpo.

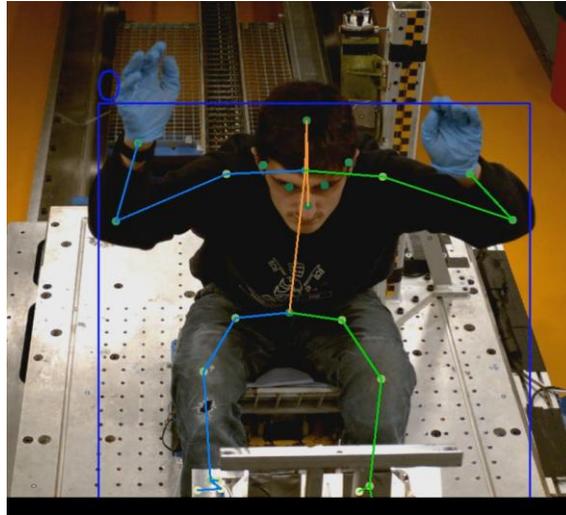


Ilustración 35 - Visión de la cámara cenital

En el caso de las grabaciones laterales, las cámaras solo ofrecían una vista lateral parcial del sujeto, careciendo de la otra mitad del cuerpo. Esto significaba que no se disponía de una visión **completa** del sujeto desde múltiples ángulos, lo que descubrimos es esencial para una triangulación 3D precisa.

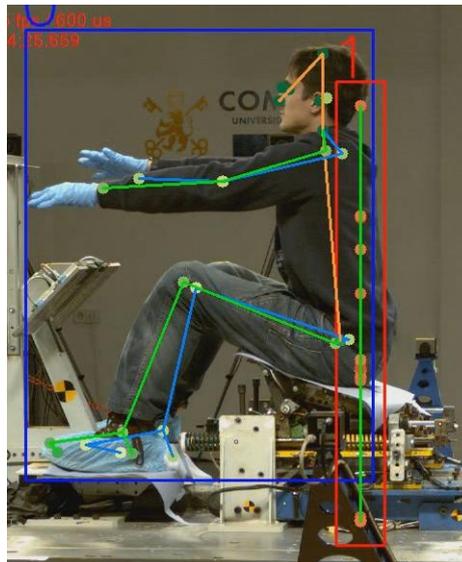


Ilustración 36 - Vista de una cámara lateral

Finalmente, en la última prueba, se intentó una configuración en la que el sujeto se encontraba de cuclillas sobre la plataforma pero con más cámaras. Sin embargo, incluso en

esta postura adaptada, el asiento de la plataforma continuó tapando las piernas del sujeto de forma persistente.

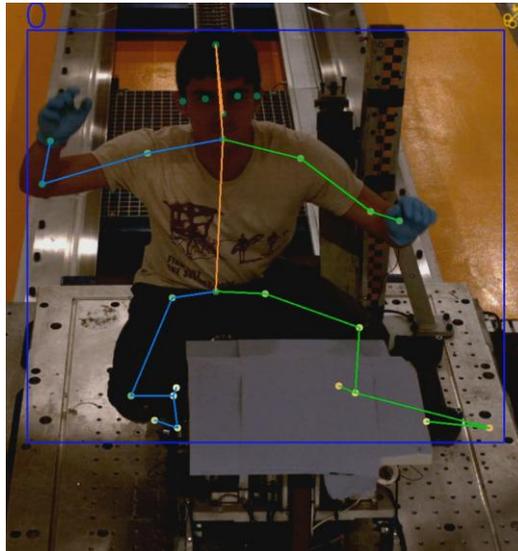


Ilustración 37 - Imagen grabada de la última prueba

Esta constante presencia de oclusiones en todas las pruebas en MOBIOS implicó que, en la fase de Estimación de Pose 2D, el algoritmo de detección tuviera una información visual incompleta o errónea. Como resultado, los keypoints correspondientes a las partes ocluidas del cuerpo (principalmente los pies y las piernas, pero también otras articulaciones que quedaban ocultas lateralmente) eran detectados con una confianza extremadamente baja, se ubicaban en posiciones incoherentes (a menudo "inventadas" por el algoritmo) o, directamente, no se detectaban.

Estos keypoints erráticos o ausentes se propagaban a la fase de Triangulación 3D. Al carecer de proyecciones 2D consistentes y fiables desde múltiples cámaras para los mismos keypoints, el proceso de triangulación se volvía inestable. Los valores de muchos keypoints resultaban incoherentes entre sí al ser totalmente erráticos a causa de la oclusión, lo que impedía la reconstrucción 3D precisa de las longitudes de los segmentos óseos y del movimiento articular. En consecuencia, la salida de Pose2Sim carecía de la calidad necesaria para cualquier análisis biomecánico, justificando el fracaso de estas pruebas a pesar de los esfuerzos realizados.

6.5 DISCUSIÓN GENERAL Y CONCLUSIONES DEL ANÁLISIS

Este apartado sintetiza los hallazgos más relevantes obtenidos del análisis cuantitativo de las longitudes de los segmentos óseos, explorando cómo la variación en el número de cámaras y la aplicación de filtrado impactan en la consistencia de las estimaciones de Pose2Sim.

IMPACTO DEL NÚMERO DE CÁMARAS

Al reflexionar sobre los datos de las longitudes del húmero derecho y la tibia derecha, la influencia del número de cámaras en la precisión y consistencia del algoritmo es un tema que invita a una discusión más profunda. Lo que hemos visto, de hecho, puede resultar bastante contraintuitivo y nos lleva a cuestionar la idea de que "más es siempre mejor".

Para facilitar esta discusión, presentamos a continuación una tabla consolidada con las métricas estadísticas clave para cada hueso y configuración de cámara, basadas en los resultados detallados en apartados previos (con filtro aplicado) basadas en los resultados detallados en los apartados 6.3.1.1, 6.3.1.2, 6.3.1.3:

Hueso	Configuración	Media (cm)	Desviación típica (cm)	Varianza (cm^2)	Coefficiente de Variación (%)
Húmero derecho	4 Cámaras	27.34	2.66	7.084	9.74
	3 Cámaras	25.5	3.71	13.77	14.55
	2 Cámaras	25.32	2.93	8.59	11.58
Tibia Derecha	4 Cámaras	37.96	2.78	7.70	7.31
	3 Cámaras	37.10	2.24	5.03	6.05
	2 Cámaras	37.03	2.07	4.30	5.6

Tabla 7 - Tabla comparativa de valores estadísticos en función del número de cámaras

Observando la tabla, para el húmero derecho, la variabilidad (como lo indican la desviación típica y el coeficiente de variación) sorprendentemente aumenta al pasar de 4 a 3 cámaras (de un CV del 9.74% a 14.55%). Sin embargo, al reducirnos a 2 cámaras, la dispersión disminuye notablemente (a un CV del 11.58%), aunque sigue siendo superior a la de 4

cámaras. Esta tendencia no lineal sugiere que la eliminación de una cámara específica entre el setup de 4 y 3 cámaras pudo haber desestabilizado la triangulación de ese segmento en particular, quizás eliminando una vista crucial que compensaba los errores de otras. Sin embargo, con solo dos cámaras, si su posicionamiento relativo fue particularmente óptimo para capturar los keypoints del húmero desde ángulos que permitían una buena triangulación, el algoritmo podría haber logrado una estimación más estable. Claramente, no es una relación lineal; parece que la geometría y la calidad de la vista de cada cámara son tan, o incluso más, importantes que la cantidad bruta. Es como si añadir una tercera cámara, en este caso, introdujera más "ruido" que "señal" útil para ese hueso en concreto.

La tibia derecha nos ofrece una perspectiva aún más interesante. Aquí, la tendencia fue de una mejora constante en la consistencia relativa (CV) a medida que reducíamos el número de cámaras, pasando de un CV del 7.31% con 4 cámaras, a 6.05% con 3 cámaras, y finalmente a 5.60% con 2 cámaras. Este es un resultado sorprendente y contraintuitivo. Uno esperaría que con más vistas, la redundancia de datos permitiría al algoritmo "promediar" y mitigar mejor los errores de detección en 2D, llevando a una mayor estabilidad en 3D. Sin embargo, en nuestro experimento, para la tibia, parece que la configuración con menos cámaras resultó ser la más "limpia" o la mejor optimizada en cuanto a los ángulos de triangulación. Podría ser que las cámaras "adicionales" (la tercera y cuarta) no aportaran información de valor significativo para la tibia, o incluso que introdujeran pequeñas inconsistencias o ambigüedades que el algoritmo, al intentar integrar, acababa propagando como variabilidad.

En definitiva, lo que estos resultados nos muestran es que la relación entre el número de cámaras y la precisión/consistencia no es tan simple como "a más cámaras, más preciso". Hay una compleja interacción entre la cantidad de cámaras, su disposición geométrica, la calidad intrínseca de las detecciones 2D en cada vista y la efectividad de los algoritmos de triangulación. Este fenómeno, donde más datos no siempre implican más precisión, se observa también en otros campos como las regresiones estadísticas, donde un exceso de variables redundantes o ruidosas puede llevar a un sobreajuste o a una menor generalización del modelo. Nuestros hallazgos sugieren que, con un posicionamiento de cámaras cuidadoso

y condiciones de grabación favorables, es posible lograr estimaciones de pose 3D cualitativa y cuantitativamente consistentes incluso con un setup minimalista.

EFFECTIVIDAD DEL FILTRADO

La aplicación de filtros a los datos cinemáticos es una práctica común en biomecánica para suavizar el ruido inherente a la adquisición de movimiento y obtener trayectorias más coherentes y fisiológicamente plausibles. En este estudio, se evaluó la efectividad del filtro Butterworth, comparando los resultados de longitud de los segmentos óseos en la configuración de 4 cámaras con y sin la aplicación de dicho filtro.

Para una mejor comprensión del impacto del filtro, la siguiente tabla muestra las métricas estadísticas para el húmero derecho y la tibia derecha con y sin filtro Butterworth (configuración de 4 cámaras):

Hueso	Configuración	Media (cm)	Desviación típica (cm)	Varianza (cm^2)	Coefficiente de Variación (%)
Húmero Derecho	Con Filtro	27.34	2.66	7.084	9.74
	Sin Filtro	27.51	3.03	9.17	11.01
Tibia Derecha	Con Filtro	37.96	2.78	7.70	7.31
	Sin Filtro	37.98	3.18	10.13	8.38

Tabla 8 - Tabla comparativa del efecto del filtrado.

La observación cualitativa de las simulaciones sin filtrar ya nos alertaba de una diferencia abismal: los movimientos se percibían abruptos, espasmódicos y llenos de ruido, lo que se traducía en una experiencia visual poco convincente y, lo que es más importante, en datos cinemáticos difíciles de interpretar. Al cuantificarlo en la tabla anterior, los números respaldan esta percepción: para ambos huesos, tanto la desviación típica como la varianza y el coeficiente de variación se reducen de forma tangible con el filtro aplicado. Esto significa que el filtro Butterworth cumple su función esencial: suavizar las fluctuaciones inherentes a la estimación de pose y la triangulación. Las mediciones de longitud de los segmentos óseos

se vuelven más estables y consistentes a lo largo del tiempo. Es el tipo de efecto que se esperaría de un buen filtro; no cambia la "verdadera" longitud del hueso (la media se mantiene prácticamente igual), sino que elimina el "temblor" y la inconsistencia del ruido.

Este punto es vital. La efectividad del filtrado no solo mejora la calidad visual de las reconstrucciones, sino que, de hecho, hace que los datos cuantitativos sean más fiables y representativos de un movimiento biológico real. Sin este paso, la utilidad de los datos para aplicaciones de ingeniería o médicas sería muy limitada. El filtro Butterworth actúa como un "purificador" de la señal, permitiendo que las tendencias reales del movimiento emerjan por encima del ruido de la adquisición, lo cual es indispensable para cualquier análisis preciso y significativo.

IMPORTANCIA DE LAS CONDICIONES DE GRABACIÓN Y VISIBILIDAD

El rendimiento de un sistema de estimación de pose sin marcadores, como Pose2Sim, no solo depende de la sofisticación de sus algoritmos o del número de cámaras. Un factor fundamental que hemos podido constatar a lo largo de este TFG es la importancia crítica de las condiciones de grabación y la visibilidad del sujeto. Las experiencias en el laboratorio MOBIOs contrastan de manera brutal con los resultados obtenidos en el aula de la universidad, ofreciendo una lección invaluable sobre los límites y las sensibilidades de esta tecnología.

Mientras que en el entorno controlado de la universidad, con una buena disposición de cámaras y visibilidad completa, pudimos obtener resultados consistentes e incluso sorprendentes con solo dos cámaras, en MOBIOs la situación fue diametralmente opuesta. Allí, a pesar de contar con cámaras de alta velocidad y un entorno profesional, las restricciones inherentes al laboratorio (cámaras fijas en posiciones desfavorables, oclusiones severas causadas por el equipamiento de simulación, y desafíos en la calibración) hicieron que la obtención de datos útiles fuera prácticamente imposible. No se trata de la potencia del hardware, sino de cómo se utiliza y las limitaciones del entorno.

Cuando un segmento corporal se ocluye, el algoritmo de estimación de pose 2D simplemente pierde la información visual necesaria. Esto no solo afecta al keypoint directamente ocluido, sino que sus efectos se propagan a la triangulación 3D. Un keypoint mal detectado o ausente en una o varias cámaras deprime la confianza del algoritmo en su reconstrucción 3D. El resultado son puntos 3D erráticos, "saltos" en las trayectorias de los huesos, y una pérdida total de la suavidad y consistencia en el movimiento reconstruido. De nada sirve un filtro o un algoritmo de triangulación muy avanzado si los datos de entrada están fundamentalmente comprometidos. Es el clásico "garbage in, garbage out": si la información original es defectuosa, el resultado final lo será también.

Por tanto, este estudio subraya que, para garantizar la fiabilidad de las mediciones y la coherencia del movimiento reconstruido, es imprescindible asegurar la máxima visibilidad del sujeto desde múltiples ángulos. Esto implica una planificación cuidadosa del setup de cámaras, la eliminación de obstáculos y la elección de entornos que permitan una calibración precisa. Si estas condiciones básicas no se cumplen, incluso la tecnología más prometedora se verá limitada, y la variabilidad en las estimaciones será tan grande que el análisis cuantitativo carecerá de valor. La calidad de la materia prima (las imágenes y la visibilidad de los keypoints) es el cimiento sobre el que se construye cualquier análisis de pose 3D fiable.

LIMITACIONES DEL ESTUDIO Y FUTURAS LÍNEAS

Este estudio, aunque ha ofrecido valiosos insights sobre el rendimiento de Pose2Sim bajo diferentes condiciones, presenta ciertas limitaciones inherentes que abren la puerta a futuras líneas de investigación.

Una limitación fundamental ha sido la ausencia de un "estándar de oro" de referencia concurrente para la validación absoluta de la precisión de las estimaciones 3D. Aunque se pretendía comparar con los sistemas Vicon de MOBIOS, las dificultades inherentes al laboratorio impidieron la adquisición de datos paralelos. Esto significa que, si bien hemos podido evaluar la consistencia y estabilidad de las longitudes óseas, no hemos podido determinar su precisión absoluta con respecto a valores reales o a la de un sistema de

referencia reconocido. La imposibilidad de obtener mediciones antropométricas precisas (como radiografías) de los huesos del sujeto mediante métodos fiables también limita la validación de la exactitud de las longitudes estimadas.

Además, el estudio se centró en un número limitado de movimientos y un único sujeto. La variabilidad de la pose y el movimiento en una población más diversa (diferentes rangos de edad, constituciones físicas, o condiciones patológicas) o la ejecución de movimientos más complejos y dinámicos podrían revelar otros desafíos y oportunidades para el algoritmo.

En cuanto a las futuras líneas de investigación, el camino a seguir es claro:

Validación con un Estándar de Oro: El paso más crucial sería la realización de un ensayo donde se capturen simultáneamente datos con Pose2Sim y con un sistema de captura de movimiento basado en marcadores de alta precisión, como Vicon. Esto permitiría:

- Homogeneizar los sistemas de coordenadas: Entender bien cómo alinear los ejes de coordenadas de ambos sistemas para que los datos sean directamente comparables.
- Comparación directa de keypoints y trayectorias: Evaluar la diferencia de ubicación entre los keypoints estimados por Pose2Sim y los marcadores de Vicon en las mismas condiciones.
- Análisis de la variabilidad vs. precisión: Determinar no solo la consistencia, sino también la exactitud de Pose2Sim en función de los datos de Vicon, entendiendo las desviaciones absolutas.

Exploración de Marker Augmentation: En este trabajo, no se profundizó en la opción de marker augmentation debido a la complejidad de los scripts de procesamiento y las limitaciones de tiempo. Sin embargo, su estudio futuro es ideal, ya que podría ofrecer una mayor densidad de marcadores y potencialmente una mayor estabilidad en el modelo biomecánico, especialmente en escenarios con menos cámaras.

Nuevas Pruebas en Entornos Realistas: Sería positivo realizar nuevas pruebas en laboratorios como MOBIOS, pero con un diseño experimental optimizado. Esto implicaría:

- Posicionamiento estratégico de cámaras: La instalación de cámaras en ubicaciones flexibles y con trípodes lo suficientemente altos para evitar oclusiones (incluso en la plataforma de impacto) o con objetivos de gran angular y asegurar una visión completa del sujeto desde múltiples ángulos.
- Sujetos adecuados: Contar con una persona que pueda colocarse en la posición correcta para el estudio, permitiendo la visibilidad de todos los segmentos de interés.

Optimización y Desarrollo de Algoritmos:

- Comparación exhaustiva de filtros: Explorar y comparar la efectividad de los diferentes tipos de filtros disponibles en Pose2Sim (Kalman, Gaussian, etc.) y ajustar sus parámetros para distintas condiciones de ruido y tipos de movimiento.
- Investigación de otros algoritmos de pose 2D/3D: Evaluar la integración y el rendimiento de los últimos avances en estimación de pose sin marcadores.
- Entrenamiento de un algoritmo personalizado: Si las necesidades de aplicación lo justifican, una línea de investigación muy avanzada sería entrenar un algoritmo de estimación de pose propio y específico para las condiciones y patrones de movimiento de las simulaciones de accidentes, aprovechando los avances más recientes para maximizar la precisión en ese nicho. Esto iría más allá del uso de modelos preentrenados y buscaría una solución a medida.

Estas futuras líneas de investigación permitirían no solo validar de manera más rigurosa el potencial de Pose2Sim, sino también empujar los límites de la estimación de pose sin marcadores hacia aplicaciones más exigentes y de gran impacto.

Capítulo 7. CONCLUSIONES Y TRABAJOS FUTUROS

El presente Trabajo Fin de Grado se propuso evaluar la precisión de los algoritmos de posicionamiento 3D implementados en Pose2Sim en el contexto de simulaciones de accidentes, en contraposición a los sistemas tradicionales basados en marcadores. Para ello, se buscó incrustar un modelo esquelético en las grabaciones de vídeo con el fin de visualizar el movimiento y analizar la viabilidad de que Pose2Sim pudiera complementar o incluso sustituir los métodos actuales empleados en el laboratorio MOBIOs.

A lo largo de este proyecto, se ha logrado una evaluación cuantitativa del rendimiento de Pose2Sim en un entorno controlado. Los análisis estadísticos de las longitudes de segmentos óseos específicos, como el húmero derecho y la tibia derecha, revelaron que, bajo condiciones de grabación óptimas, el algoritmo es capaz de producir estimaciones consistentemente estables. Se observó una relación no lineal y, en ciertos casos, contraintuitiva, entre el número de cámaras y la consistencia de las mediciones, sugiriendo que la disposición geométrica y la calidad visual de las cámaras son factores tan cruciales como su cantidad bruta. En particular, la tibia derecha demostró una notable consistencia incluso con un número reducido de cámaras, desafiando la premisa de que una mayor redundancia de datos siempre conduce a una mayor precisión.

Una aportación fundamental de este estudio ha sido la confirmación empírica de la efectividad del filtrado Butterworth. La aplicación de este filtro demostró una mejora significativa en la suavidad de las trayectorias de los keypoints y una notable reducción del ruido inherente al proceso de triangulación 3D. Esto no solo repercute en una mejor calidad visual de las reconstrucciones, sino que, lo que es más relevante, confiere una mayor fiabilidad a los datos cuantitativos para su interpretación biomecánica. El modelo esquelético incrustado en las grabaciones de vídeo, visualizado tanto en entornos de programación como en OpenSim, ha permitido una validación cualitativa del movimiento, facilitando la comprensión del impacto en las estructuras óseas y el seguimiento de la reacción corporal. Asimismo, la capacidad de Pose2Sim para inferir la posición de la cadera a partir de los

keypoints detectados se perfila como una ventaja significativa, dada la dificultad inherente para posicionar marcadores en esa región con los sistemas tradicionales en escenarios como los accidentes de tráfico.

No obstante, este TFG ha puesto de manifiesto la importancia crítica de las condiciones de grabación y la visibilidad del sujeto, evidenciando las limitaciones de la tecnología frente a entornos no controlados. Las experiencias en el laboratorio MOBIO, caracterizadas por cámaras fijas en posiciones subóptimas y oclusiones severas, impidieron la obtención de datos útiles para un análisis cinemático robusto. Esto resalta que, por muy sofisticados que sean los algoritmos, la calidad de la información de entrada es el pilar sobre el que se asienta la fiabilidad de cualquier reconstrucción 3D. El sistema, en su estado actual, ofrece un valor considerable para la creación de ilustraciones y visualizaciones de alta calidad de forma más sencilla y rápida que los métodos Vicon, lo cual ya justificaría su exploración por parte del laboratorio.

De cara al futuro, este trabajo abre diversas líneas de investigación. La más crucial es la validación absoluta de la precisión de Pose2Sim mediante una comparación directa con un "estándar de oro", como un sistema Vicon, en un entorno con sistemas de coordenadas homogeneizados. Además, sería valioso explorar la funcionalidad de marker augmentation en Pose2Sim, así como realizar nuevas pruebas en laboratorios realistas con un diseño experimental optimizado que garantice una visibilidad óptima y la flexibilidad en el posicionamiento de cámaras. Finalmente, la investigación en la optimización de los algoritmos de filtrado y la evaluación de otros modelos de estimación de pose, o incluso el entrenamiento de un algoritmo personalizado para las condiciones específicas de las simulaciones de accidentes, representan avenidas prometedoras para llevar esta tecnología a su máximo potencial en aplicaciones de ingeniería y biomecánica de impacto.

Capítulo 8. BIBLIOGRAFÍA

- Cao, Z. a.-E. (2019). Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE transactions on pattern analysis and machine intelligence*, 172--186.
- Ce Zheng, W. W. (2020). Computer Vision and Pattern Recognition. *arXiv:2012.13392*.
- Contributors, M. (s.f.). OpenMMLab Pose Estimation Toolbox and Benchmark. *OpenMMLab Pose Estimation Toolbox and Benchmark*.
- Federolf, P. A. (2013). A novel approach to solve the “missing marker problem” in marker-based motion analysis that exploits the segment coordination patterns in multi-limb motion data. *PloS one*, e78689.
- Glassdoor. (s.f.). https://www.glassdoor.es/Sueldos/ingeniero-junior-sueldo-SRCH_KO0,16.htm. Obtenido de Glassdoor: https://www.glassdoor.es/Sueldos/ingeniero-junior-sueldo-SRCH_KO0,16.htm
- Glassdoor. (s.f.). *Sueldo: Ingeniero Junior en España en Junio, 2023 | Glassdoor*. Recuperado el 11 de mayo de 2025, de https://www.glassdoor.es/Sueldos/ingeniero-junior-sueldo-SRCH_KO0,16.htm
- Mao, W. a. (2021). Tfpose: Direct human pose estimation with transformers. *arXiv preprint arXiv:2103.15320*.
- Martinez, J. a. (2017). A simple yet effective baseline for 3d human pose estimation. *Proceedings of the IEEE international conference on computer vision*, 2640--2649.
- Metzler, S. A. (2007). *Motion Capture Applications in Forensic Injury Accident Reconstruction*. SAE Technical Paper.
- Minh, V. T. (2020). Model Predictive Control for Modeling Human Gait Motions Assisted by Vicon Technology. *Journal Europeen des Systemes Automatises*.

- Moeslund, T. B. (2006). A survey of advances in vision-based human motion capture and analysis. *Computer vision and image understanding*, 90--126.
- Newell, A. a. (2016). Stacked hourglass networks for human pose estimation. *Computer Vision--ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VIII 14*, 483--499.
- Pagnon, D. a. (2022). Pose2Sim: an end-to-end workflow for 3D markerless sports kinematics—part 2: accuracy. *Sensors*, 2712.
- Pagnon, D. a. (2022). Pose2Sim: An open-source Python package for multiview markerless kinematics. *Journal of Open Source Software*.
- Pagnon, D. a. (6530). Pose2Sim: an end-to-end workflow for 3D markerless sports kinematics—part 1: robustness. *Sensors*, 2021.
- Pagnon, D. a. (s.f.). *Diagrama de funcionamiento de Pose2Sim* .
- Smirnova, V. a. (2022). The automatization of the GAIT analysis by the Vicon video system: A pilot study. *Sensors*, 7178.
- Tekin, B. a.-N. (2017). Learning to fuse 2d and 3d image cues for monocular body pose estimation. *Proceedings of the IEEE international conference on computer vision*, 3941--3950.
- Toshev, A. a. (2014). Deeppose: Human pose estimation via deep neural networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*.
- Yoganandan, N. a. (1998). Head-neck biomechanics in simulated rear impact. *Annual Proceedings/Association for the Advancement of Automotive Medicine*, 208.
- Zheng, C. a. (2023). Deep learning-based human pose estimation: A survey. *ACM Computing Surveys*, 1--37.

ANEXO I: ALINEACIÓN DEL PROYECTO CON LOS ODS

Este proyecto se alinea con los siguientes ODS:



Figure 1- OS3, Salud y Bienestar

- ODS 3: Salud y Bienestar: El análisis de la cinemática en accidentes laborales y de tráfico, tiene un impacto directo en la seguridad de las personas involucradas al permitir el desarrollo por parte de los fabricantes de mejores sistemas de protección tanto activos como pasivos. En este proyecto, trabajaremos para plantear un nuevo sistema de evaluación de estos eventos, lo que puede contribuir a desarrollar medidas de prevención y procedimientos de seguridad más efectivos, protegiendo así la salud y el bienestar de los trabajadores y de quienes usan la infraestructura de transporte.



Figure 2 - ODS 9 Industria, innovación e infraestructura

- ODS 9: Industria, Innovación e Infraestructura. La adopción de tecnología sin marcadores responde a la necesidad de innovación en el análisis de movimientos y en la seguridad industrial. Al evaluar Pose2Sim como una alternativa a los sistemas tradicionales, este proyecto promueve el desarrollo y la accesibilidad de tecnologías innovadoras, contribuyendo a la modernización de infraestructuras del laboratorio y de sus sistemas de análisis. Además, gracias al trabajo del laboratorio de MOBIO, y al uso de herramientas como esta, podrán proponerse nuevas tecnologías de infraestructuras de seguridad como guardarraíles, airbags u otras medidas de seguridad activa y pasivas.