



EDUCACIÓN

SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO
NACIONAL DE MÉXICO

Tecnológico Nacional de México

**Centro Nacional de Investigación
y Desarrollo Tecnológico**

Tesis de Maestría

**Reconstrucción geométrica multivista
del cuerpo humano en 3D**

presentada por

Lic. Alexandra Ehrlich López

como requisito para la obtención del grado de
Maestra en Ciencias de la Computación

Director de tesis
Dr. Raúl Pinto Elías

Cuernavaca, Morelos, México. Septiembre de 2025.



Cuernavaca, Mor., 25/Agosto/2025

OFICIO No. DCC/236/2025

Asunto: Aceptación de documento de tesis

CENIDET-AC-004-M14-OFICIO

CARLOS MANUEL ASTORGA ZARAGOZA
SUBDIRECTOR ACADÉMICO
PRESENTE

Por este conducto, los integrantes de Comité Tutorial de **Alexandra Ehrlich López** con número de control **M23CE063**, de la Maestría en Ciencias de la Computación, le informamos que hemos revisado el trabajo de tesis de grado titulado **"Reconstrucción Geométrica Multivista del Cuerpo Humano en 3D"** y hemos encontrado que se han atendido todas las observaciones que se le indicaron, por lo que hemos acordado aceptar el documento de tesis y le solicitamos la autorización de impresión definitiva.

ATENTAMENTE

Excelencia en Educación Tecnológica®
"Conocimiento y Tecnología al Servicio de México"


Dr. Raúl Pinto Elías
Director de Tesis


Dra. Andrea Magadán Salazar
Revisor 1


Dr. Noé Alejandro Castro Sánchez
Revisor 2

C.c.p. Depto. Servicios Escolares.
Expediente / Estudiante



2025
Año de
La Mujer
Indígena

Interior Internado Palmira S/N, Col. Palmira,
C. P. 62490, Cuernavaca, Morelos Tel. 01 (777) 3627770, ext. 3201,
e-mail: dcc_cenidet@tecnm.mx tecnm.mx | cenidet.tecnm.mx

cenidet
Centro Nacional de Investigación
y Desarrollo Tecnológico





Educación
Secretaría de Educación Pública



TECNOLÓGICO
NACIONAL DE MÉXICO



Centro Nacional de Investigación y Desarrollo tecnológico
Subdirección Académica

Cuernavaca Mor, 25/agosto/2025

Oficio No. SAC/209/2025

Asunto: Autorización de impresión de tesis

ALEXANDRA EHRLICH LOPEZ
CANDIDATA AL GRADO DE MAESTRA
EN CIENCIAS DE LA COMPUTACIÓN
P R E S E N T E

Por este conducto, tengo el agrado de comunicarle que el Comité Tutorial asignado a su trabajo de tesis titulado **"Reconstrucción Geométrica Multivista del Cuerpo Humano en 3D"**, ha informado a esta Subdirección Académica, que están de acuerdo con el trabajo presentado. Por lo anterior, se le autoriza a que proceda con la impresión definitiva de su trabajo de tesis.

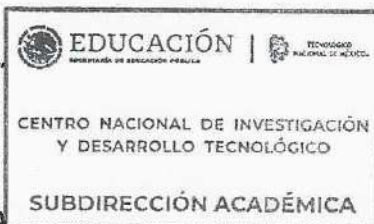
Esperando que el logro del mismo sea acorde con sus aspiraciones profesionales, reciba un cordial saludo.

ATENTAMENTE

Excelencia en Educación Tecnológica®

"Conocimiento y Tecnología al Servicio de México"

CARLOS MANUEL ASTORGA ZARAGOZA
SUBDIRECTOR ACADÉMICO



c.c.p. Departamento de Ciencias Computacionales
Departamento de Servicios Escolares

CMAZ/lmz



2025
Año de
La Mujer
Indígena

Interior Internado Palmira S/N, Col. Palmira,
C. P. 62490, Cuernavaca, Morelos Tel. 01 (777) 3627770, ext. 4104,
e-mail: acad_cenidet@tecnm.mx tecnm.mx | cenidet.tecnm.mx

cenidet
Centro Nacional de Investigación y Desarrollo Tecnológico



*“El éxito está en los detalles, en el trabajo constante y en la valentía de enfrentarte
a lo desconocido.”*

ADA LOVELACE

Agradecimientos

A la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI), por el valioso apoyo brindado a través de la beca otorgada, la cual hizo posible el desarrollo de este trabajo de investigación.

Al Tecnológico Nacional de México (TecNM), a través del Centro Nacional de Investigación y Desarrollo Tecnológico (CENIDET), por abrirme sus puertas y brindarme las condiciones académicas, el acompañamiento y los recursos necesarios para la culminación de este proyecto.

A mi director de tesis, el Dr. Raúl Pinto, por su invaluable orientación, dedicación y confianza. Su experiencia, guía y constante apoyo fueron determinantes para mi formación académica y para la conclusión de este trabajo.

A la Dra. Andrea y al Dr. Noé, por el tiempo y esfuerzo dedicados a la revisión de esta tesis, así como por sus valiosos comentarios y sugerencias, que enriquecieron significativamente su contenido. Asimismo, agradezco al Dr. Jorge por todo el apoyo brindado a lo largo de este proceso.

A mi papá, por su apoyo, sus enseñanzas y por inculcarme desde siempre la importancia de la responsabilidad y el esfuerzo. Su ejemplo ha sido, y seguirá siendo, una fuente de inspiración en mi vida.

A mi mamá, por su amor inagotable, su paciencia y sus palabras de aliento en los momentos más difíciles. Su fortaleza y apoyo constante me dan la motivación necesaria para seguir adelante y nunca rendirme.

A mi hermana, por ser mi mayor inspiración, mi ejemplo y mi apoyo incondicional. Gracias por todo tu amor, por creer en mí incluso cuando yo dudaba, por animarme en los momentos más difíciles y por recordarme siempre que soy capaz de lograr lo que me proponga.

Finalmente, a quienes, sin estar en los reflectores o ser explícitamente mencionados en estas líneas, fueron parte esencial de este proceso y de mi crecimiento como persona y profesionista.

Y a todas aquellas personas que, de una u otra forma, han estado presentes en mi vida durante este proceso: gracias por sus palabras de aliento, su cariño, sus gestos de apoyo y por acompañarme, aunque fuera en silencio, en los momentos más complejos. Cada consejo, cada conversación y cada muestra de afecto hicieron que este camino fuera más llevadero y significativo.

Este logro representa no solo el fruto de mi esfuerzo, sino también la suma de las contribuciones, grandes y pequeñas, de todas aquellas personas que creyeron en mí y me tendieron la mano en el momento indicado.

Gracias infinitas por caminar conmigo este trayecto. Los quiero mucho.

Resumen

Esta tesis presenta un enfoque para la reconstrucción tridimensional del cuerpo humano a partir de estimaciones de postura en 2D, utilizando técnicas de visión por computadora. El trabajo es motivado por el creciente interés en el estudio del cuerpo humano en aplicaciones que abarcan desde la terapia física y el diagnóstico médico hasta el reconocimiento y seguimiento de personas en entornos reales y virtuales.

Se propone un método de reconstrucción tridimensional del cuerpo humano basado en información multivista y principios geométricos. El sistema captura videos sincronizados desde múltiples cámaras RGB para generar modelos de pose en 2D, que luego se fusionan mediante triangulación, obteniendo modelos paramétricos en 3D basados en el modelo *SMPL*. Esta propuesta busca superar las limitaciones de métodos existentes, que suelen depender de una sola vista y grandes volúmenes de datos, reduciendo además la carga computacional. Para la validación, se desarrolló un conjunto de datos propio en un entorno controlado y se utilizaron métricas de error angular y longitudinal.

Los resultados demuestran que el enfoque geométrico multivista es viable para generar modelos de pose en 3D con un nivel constante de precisión, logrando reconstrucciones tridimensionales del cuerpo humano con ciertas limitaciones en el control fino de las rotaciones articulares. En conjunto, esta propuesta representa una alternativa eficiente y reutilizable para tareas relacionadas con la visión artificial del cuerpo humano.

Palabras clave: Reconstrucción 3D multivista, estimación de pose del cuerpo humano, modelado geométrico, calibración de cámaras, triangulación.

Abstract

This thesis presents an approach for the three-dimensional reconstruction of the human body from 2D pose estimations using computer vision techniques, motivated by the growing interest in human body analysis across various applications—from physical therapy and medical diagnostics to person recognition, and tracking in both real and virtual environments.

The proposed method leverages multiview information and geometric principles to perform 3D reconstruction. The system captures synchronized videos from multiple RGB cameras to generate 2D pose models, which are then fused through triangulation to obtain 3D parametric models based on the SMPL model. This approach aims to overcome the limitations of current methods, which often rely on single-view input and large data volumes, while also reducing computational load. A custom dataset was developed in a controlled environment, and angular and longitudinal error metrics were used for evaluation.

The results show that the multiview geometric approach is a viable method for generating 3D pose models with an acceptable level of accuracy. This method enables three-dimensional reconstruction of the human body, albeit with some limitations in fine-grained joint rotation control. Overall, the proposed method provides an efficient and reusable alternative for tasks related to artificial vision of the human body.

Keywords: *Multiview 3D reconstruction, human body pose estimation, geometric modeling, camera calibration, triangulation.*

Contenido

1. Introducción	1
1.1. Planteamiento del problema	2
1.1.1. Delimitación del problema	3
1.1.2. Complejidad del problema	3
1.2. Planteamiento general de la solución	4
1.2.1. Justificación	4
1.2.2. Objetivo	5
1.2.3. Alcances y limitaciones	6
1.2.4. Antecedentes del proyecto	7
1.2.5. Organización del Documento	8
2. Análisis del Estado del Arte y Marco Teórico	10
2.1. Marco Teórico	10
2.1.1. Calibración	10
2.1.2. Triangulación	11
2.1.3. Modelo Paramétrico	12
2.1.4. Proporciones antropométricas del cuerpo humano	14
2.1.5. Axis-Angles	14
2.2. Estado del Arte	15
2.2.1. Reconstrucción 3D del cuerpo humano	15
2.2.2. Modelos de estimación de pose 2D	19
2.2.3. Estimación de pose 3D	19
2.2.4. Conclusiones del estado del arte	20

3. Especificación y modelado de la solución	22
3.1. Especificación de la solución	22
3.1.1. Modelado de la solución	22
3.2. Implementación de la solución	23
3.2.1. Preparación y captura de video	23
3.2.2. Procesamiento	26
3.2.3. Generación del modelo 3D	32
4. Experimentación y resultados	33
4.1. Metodología experimental	33
4.1.1. Herramientas y tecnologías	33
4.1.2. Métricas para el análisis de resultados	34
4.2. Desarrollo experimental	35
4.2.1. Resumen experimental	35
4.2.2. Generación del <i>Dataset</i>	36
4.2.3. Generación de modelos de pose 2D	38
4.2.4. Fusión de modelos de pose 2D a modelos de pose 3D	40
4.2.5. Generación de modelos del cuerpo humano en 3D	44
4.2.6. Generación del <i>dataset</i> 3D	55
4.3. Discusión de Resultados	63
5. Conclusiones	66
5.1. Conclusiones generales	66
5.2. Trabajo futuro	68
5.3. Productos generados	68
Referencias	69
A. Documentos Generados	75
B. Tablas	78
C. Imágenes	85
D. Productos Generados	86

Figuras

2.1. Ejemplo gráfico del proceso de calibración [1]	11
2.2. Ejemplo gráfico del proceso de triangulación (modificada de [2]) . . .	12
2.3. Modelo paramétrico SMPL (adaptada de [3])	13
2.4. Representación gráfica de un <i>axis-angle</i> (Figura adaptada de [4]) . . .	14
3.1. Diagrama de solución diseñado	23
3.2. Distribución de cámaras en el espacio de grabación	24
3.3. Ejemplos gráficos de los ejercicios	25
3.4. Puntos detectados por <i>mediapipe</i> [5]	26
3.5. Cámaras empleadas para la triangulación	27
3.6. <i>T-pose</i>	28
3.7. Transformación de la postura	29
3.8. Ejemplo de medidas del modelo	30
3.9. Articulaciones que tienen valores de rotación	31
4.1. Escenario de captura	37
4.2. Ejemplo de captura mediante <i>OBS</i> [6]	37
4.3. Ejemplo de captura y distribución de cámaras entre equipos	38
4.4. Ejemplos estimación de pose	40
4.5. Patrones de calibración	41
4.6. Ejemplos de triangulación SVD y calibración manual	43
4.7. Resultados de triangulación para cada método	44
4.8. Modelo <i>mediapipe pose</i> y <i>SMPL</i>	45
4.9. Ejemplo de transformación con <i>axis-angle</i>	47
4.10. Ejemplo de cálculo de ángulos	48
4.11. Ejemplo de contenido vectores de pose	49

4.12. Jerarquía de movimiento del modelo <i>SMPL</i> (modificada de [7])	50
4.13. Ejemplo de modelos generados en <i>T-pose</i>	54
4.14. Ejemplos de modelos incorrectos	54
4.15. Ejemplos de modelos correctos	55
4.16. Distribución de error angular por segmento del ejercicio 1	57
4.17. Distribución de error angular por segmento del ejercicio 2	58
4.18. Distribución de error angular por segmento del ejercicio 3	58
4.19. Distribución de error angular por segmento del ejercicio 4	59
4.20. Distribución de error angular por segmento del ejercicio 5	59
4.21. Distribución de error angular por segmento del ejercicio 6	60
4.22. Distribución de error angular por segmento del ejercicio 7	60
4.23. Distribución de error angular por segmento del ejercicio 8	61
4.24. Distribución de error angular por segmento del ejercicio 9	61
4.25. Distribución de error angular por segmento del ejercicio 10	62
4.26. Distribución de error angular por segmento del ejercicio 11	62
4.27. Distribución de error angular por segmento del ejercicio 12	63
A.1. Ejemplo del contenido de un archivo <i>OBJ</i>	75
A.2. Ejemplo del contenido del archivo de medidas generado	76
A.3. Ejemplo del contenido del archivo de coordenadas generado	77
C.1. Personas grabadas para el <i>dataset</i>	85
D.1. Publicación en el CITCA	87
D.2. Publicación en revista REDDI+Tez	88
D.3. Retribución Social: Participación como STAFF en el CITCA	89

Tablas

2.1. Comparativa de los métodos del estado del arte en términos de <i>data-sets</i> y las arquitecturas utilizadas	18
4.1. Tabla resumen de experimentos	36
4.2. Comparativa de modelos de estimación de pose humana	39
4.3. Errores de reproyección	42
4.4. Comparativa de métodos de triangulación	42
4.5. Correspondencia generada entre los modelos <i>SMPL</i> y <i>MediaPipe Pose</i>	45
4.6. Diferencia angular promedio por segmento	48
4.7. Puntos de inicio y fin de <i>MediaPipe</i> de los segmentos corporales	51
4.8. Comparativa de medidas reales y estimadas	52
4.9. Relación de parámetros de forma <i>SMPL</i> y calculados	52
4.10. Resumen de resultados	56
B.1. Errores de la Persona 1	78
B.2. Errores de la Persona 2	79
B.3. Errores de la Persona 3	79
B.4. Errores de la Persona 4	80
B.5. Errores de la Persona 5	80
B.6. Errores de la Persona 6	81
B.7. Errores de la Persona 7	81
B.8. Errores de la Persona 8	82
B.9. Errores de la Persona 9	82
B.10. Errores de la Persona 10	83
B.11. Errores de la Persona 11	83
B.12. Errores de la Persona 12	84

Lista de Acrónimos

2D Dos dimensiones.

3D Tres dimensiones.

DeepMultiCap Deep Multi-view Human Capture: Captura multivista de cuerpo humano sin marcadores utilizando aprendizaje profundo.

FLEX Free muLti-view rEconstruXion: Reconstrucción 3D multivista de pose humana sin necesidad de parámetros de cámara mediante aprendizaje de rotaciones y longitudes óseas.

HMR Human Mesh Recovery: Recuperación de una malla 3D del cuerpo humano a partir de una imagen RGB mediante redes neuronales.

HMR 2.0 Human Mesh Recovery 2.0: Versión mejorada de HMR que estima mallas humanas más precisas y maneja múltiples vistas o secuencias.

HS-Nets Hierarchical Surface prediction Networks: Red neuronal jerárquica para predicción precisa de la superficie del cuerpo humano desde múltiples vistas.

HUMOR HUMAn mOtion model for Realistic animation: Modelo generativo de movimiento humano 3D para animaciones realistas.

MoCap Motion Capture: Técnica para registrar el movimiento corporal mediante sensores o marcadores.

MPJPE Mean Per Joint Position Error: Error medio por articulación entre la predicción y el valor real en coordenadas 3D.

PA-MPJPE Procrustes Aligned Mean Per Joint Position Error: MPJPE tras una alineación Procrustes sin escala entre la predicción y la verdad de terreno.

PARE Part Attention Regression: Estimación robusta de cuerpo humano 3D con atención a partes visibles.	SMPL Skinned Multi-Person Linear: Modelo paramétrico 3D de cuerpo humano basado en una malla deformable.
PyMAF Pyramidal Mesh Alignment Feedback: Estimación iterativa de mallas humanas 3D con realimentación piramidal de alineamiento.	SMPL-X SMPL eXpressive: Extensión del modelo SMPL que incluye manos.
PyMAF-X Pyramidal Mesh Alignment Feedback eXpressive: Extensión de PyMAF que utiliza el modelo SMPL-X para capturar cuerpo.	SMPLify SMPL fitting from 2D joints: Algoritmo que ajusta el modelo SMPL a puntos clave 2D para estimar pose y forma.
RGB Modelo de color basado en luz.	
RGB-D Imagen o sensor que combina color (RGB) con información de profundidad (Depth).	SVD Singular Value Decomposition: Descomposición de valores singulares de una matriz.

Capítulo 1

Introducción

El cuerpo humano ha sido y continua siendo objeto de extensos estudios. En el área de visión por computadora, se han desarrollado diversos modelos para llevar a cabo una variedad de tareas relacionadas con él. Entre sus aplicaciones se incluyen el reconocimiento de la postura humana, el rastreo y seguimiento de personas, la creación de entornos virtuales en 3D (Tres dimensiones), la identificación de individuos, el análisis de comportamientos, el diagnóstico médico y la terapia física, así como la localización automática de personas, entre otras [8–10]. La información utilizada para estas actividades suele recopilarse mediante dispositivos de uso común, lo que generalmente implica el uso de datos en dos dimensiones obtenidos por cámaras o teléfonos celulares, ya sean de alta o baja resolución. No obstante, mediante técnicas de visión por computadora, como el procesamiento de imágenes y el modelado en 3D, es posible transformar datos 2D (Dos dimensiones) en representaciones tridimensionales, lo que permite obtener una visión más completa del cuerpo humano.

La reconstrucción del cuerpo humano se refiere al proceso mediante el cual se recupera su forma y apariencia tridimensional a partir de imágenes, videos o información espacial. Esta tarea se ha beneficiado significativamente de los avances tecnológicos y del desarrollo de dispositivos innovadores, lo que ha permitido que sea más sencilla y precisa. Su objetivo principal es generar modelos computacionales en 3D a partir de una o varias imágenes. Llevar a cabo esta tarea ofrece múltiples ventajas y satisface diversas necesidades en distintos campos, ya que permite obtener representaciones visuales detalladas del cuerpo humano.

En los últimos años, el campo de investigación centrado en la reconstrucción del cuerpo humano ha experimentado un notable crecimiento. Como resultado, se han desarrollado modelos computacionales que permiten una reconstrucción tridimensional más precisa, generando modelos mallados que, gracias a los avances tecnológicos en *hardware* y *software*, ofrecen una representación más fiel y cercana a la realidad [11].

Según la exploración del estado del arte, los algoritmos más destacados para la reconstrucción 3D del cuerpo humano se basan, en su mayoría, en información proveniente de cámaras monoculares y de una sola vista. Sin embargo, estos requieren modelos de redes neuronales de gran tamaño y, además, cuando se presentan obstrucciones del cuerpo, deben estimar las superficies ocultas, lo que disminuye la precisión de los modelos mallados resultantes. En este contexto, contar con información proveniente de múltiples vistas del mismo cuerpo representa una ventaja significativa para la reconstrucción, ya que, aunque los datos sean 3D, la capacidad de establecer correspondencias entre imágenes permite obtener información más completa para el proceso.

La consideración del procesamiento de datos múltiples, ya sean videos o imágenes, plantea preocupaciones sobre la eficiencia de los equipos de cómputo necesarios para su análisis. En este sentido, es fundamental lograr un equilibrio entre la cantidad de datos a procesar y la carga computacional, de manera que se optimice el rendimiento sin comprometer la calidad de la reconstrucción.

1.1. Planteamiento del problema

La caracterización del cuerpo humano en términos de acciones, poses o comportamientos suele realizarse a partir de imágenes o secuencias de video en formato RGB. Esto implica trabajar con representaciones bidimensionales (2D) de un objeto tridimensional (3D) y flexible, cuyas características visuales varían según su posición en el espacio, el entorno, e incluso factores como las oclusiones o autoocclusiones.

En este contexto, el desempeño de los sistemas de reconocimiento depende en gran medida de la capacidad del modelo para aprender patrones asociados a situaciones específicas. Sin embargo, este aprendizaje es, por lo general, sensible al dominio; es decir, lo que el modelo aprende para un conjunto de datos o una aplicación específica no se transfiere fácilmente a otro dominio. Esto complica la reutilización del conocimiento entre distintas tareas relacionadas con el cuerpo humano.

Este fenómeno limita la escalabilidad de los sistemas actuales y genera una dependencia constante de nuevos datos y reentrenamientos cuando se desea aplicar el sistema en un contexto diferente.

1.1.1. Delimitación del problema

Se identificó que los modelos desarrollados para la caracterización del cuerpo humano tienden a ser parciales y están diseñados específicamente para una tarea o aplicación concreta. Como consecuencia, cada nueva aplicación requiere el desarrollo de un modelo distinto, así como la recopilación de un conjunto de datos específico.

1.1.2. Complejidad del problema

La caracterización del cuerpo humano regularmente se efectúa mediante la detección de puntos clave. Esta tarea se lleva a cabo mediante modelos computacionales diseñados específicamente para identificar las posiciones relevantes del cuerpo en imágenes 2D.

Dado que se utilizan datos bidimensionales para representar un objeto tridimensional, flexible y cambiante como el cuerpo humano, surgen limitaciones inherentes a esta representación. La pérdida de información espacial provoca que pequeñas imprecisiones en la detección de puntos clave dificulten el análisis posterior, esto debido a que no se conoce a qué profundidad se encuentran dichos puntos, lo cual se ve agravado en condiciones en las que se lleva ropa holgada, o se presentan oclusiones o poses inusuales, situaciones que introducen ambigüedad visual y reducen la capacidad del modelo para capturar adecuadamente la estructura completa del cuerpo.

1.2. Planteamiento general de la solución

La idea de generar un conjunto de datos específico y un modelo entrenado desde cero para cada tarea relacionada con el cuerpo humano parece innecesaria, considerando que todas comparten el mismo objeto de estudio: el cuerpo humano.

Un modelo 3D mallado del cuerpo humano que pueda ser reutilizado como base común para distintas aplicaciones permitiría reducir significativamente la necesidad de recopilar grandes volúmenes de datos para cada nuevo problema. No sería necesario aprender desde cero las reglas de cada tarea, sino simplemente mapear los datos de entrada a una representación conocida. Por lo tanto, si se cuenta con un modelo 3D reutilizable, el problema puede plantearse como una correspondencia entre lo observado y el conocimiento existente sobre modelos tridimensionales.

Este trabajo se centra en la generación automática de dicho modelo 3D. Se propone un enfoque geométrico que, a diferencia de los métodos basados en aprendizaje profundo, no requiere entrenamiento intensivo ni grandes cantidades de datos. Además, el uso de información multivista permite enriquecer la reconstrucción y plantea la posibilidad de adaptarse a condiciones complejas como oclusiones y auto-occlusiones.

1.2.1. Justificación

Las soluciones actuales para la reconstrucción tridimensional del cuerpo humano se basan, en su mayoría, en imágenes RGB capturadas desde una única vista [11]. Este enfoque presenta limitaciones importantes, ya que trabaja con representaciones bidimensionales de un objeto tridimensional y flexible, lo que afecta la precisión en la estimación de poses y movimientos, especialmente en presencia de oclusiones, auto-occlusiones o variaciones en el entorno [12].

Además, muchos de estos métodos se apoyan en modelos de aprendizaje profundo los cuales implican arquitecturas complejas que requieren grandes volúmenes de datos y un elevado consumo de recursos computacionales como se documenta en [13, 14]. Esta dependencia dificulta la adaptabilidad de los modelos cuando se pretende aplicar la solución en contextos distintos, lo que limita su reutilización y escalabilidad.

Frente a estas limitaciones, este trabajo propone un enfoque alternativo basado en el uso de información multivista y principios geométricos. Esta estrategia permite aprovechar la variedad de información proveniente de múltiples ángulos para reconstruir el cuerpo humano con mayor precisión, sin requerir de entrenamiento ni grandes cantidades de datos.

1.2.2. Objetivo

A continuación se presentan los objetivos de la presente investigación.

Objetivo General

Diseñar e implementar un modelo de sistema de visión artificial que permita la adquisición sincronizada de secuencias de imágenes (video) de varias cámaras en formato RGB para generar modelos 2D, así como la fusión de las vistas en un único modelo 3D.

Objetivos específicos

1. Generar un conjunto de datos de video sincronizado de múltiples vistas.
2. Implementar y comparar los algoritmos de generación de modelos 2D del cuerpo humano, para seleccionar el que mejor se acople a las necesidades.
3. Generar modelos 2D del cuerpo humano para cada una de las vistas manteniendo la sincronización de los videos.
4. Investigar arquitecturas que permitan la fusión de modelos 2D del cuerpo humano, estudiar las encontradas y seleccionar la que se adapte de mejor forma al objetivo.
5. Fusionar los modelos 2D generados para obtener un modelo 3D.
6. Evaluar la precisión del modelo 3D generado.
7. Generar un conjunto de datos 3D a partir de los videos capturados.

1.2.3. Alcances y limitaciones

A continuación, se plantean los alcances y las limitaciones del proyecto.

Alcances

- Se utilizan cámaras de captura de información RGB.
- Se considera el uso de cuatro cámaras para la obtención del conjunto de datos para ser usado en el trabajo.
- Los videos son capturados en un ambiente controlado, con condiciones de iluminación, tamaño del espacio y fondo.
- El modelo de visión artificial, es capaz de generar un modelo de malla parametrizado del cuerpo humano a partir de imágenes en formato RGB.
- El sistema genera el modelo tridimensional de una persona en escena.

Limitaciones

- La cantidad de información a procesar es dependiente de los sistemas de *hardware* a los que se tiene acceso.
- No se compromete el funcionamiento del modelo en tiempo real.
- La fusión de las vistas es únicamente a nivel volumen, obteniendo como resultado un modelo mallado; es decir, que no se tiene información en el modelo 3D de componentes como textura e iluminación.
- No se compromete la capacidad del sistema para generar modelos 3D de personas que vistan ropa holgada.
- El sistema no genera de múltiples modelos en escena.
- El modelo de visión artificial generado no realiza el reconocimiento de acciones ni identificación de personas.

1.2.4. Antecedentes del proyecto

En el Centro Nacional de Investigación y Desarrollo Tecnológico, en el Departamento de Ciencias Computacionales, en la línea de investigación de Inteligencia Artificial, se han desarrollado diversos trabajos que se centran en tareas de reconocimiento humano. El desarrollo de dichos trabajos, si bien no muestran directa relación con el enfoque, muestran que el cuerpo humano es ampliamente estudiado y que el desarrollo de esta investigación podría beneficiarlos de manera directa. A continuación, se enumeran algunos de los trabajos relevantes del estudio del cuerpo humano.

Caracterización de Imágenes en Movimiento: Correr y Caminar

(Saldaña García, 2007) [15]

Desarrolló un sistema automático de visión artificial para analizar características del movimiento corporal al caminar y correr. A diferencia de estudios previos, este trabajo incorpora el análisis de nuevas regiones, como la inclinación de la espalda y la posición de codos y manos. El sistema es no invasivo, permitiendo registrar el movimiento de forma natural.

Caracterización Visual de Movimientos “Sospechosos” de Personas en Estacionamientos

(Alcantar Juárez, 2013) [16]

Desarrolló un sistema de video-vigilancia inteligente capaz de detectar y caracterizar personas, así como identificar cinco movimientos sospechosos predefinidos: correr, levantar una o ambas manos, merodear, agacharse y acostarse. El sistema se estructura en tres módulos: el primero realiza la segmentación de objetos en movimiento y valida si se trata de una persona; el segundo caracteriza los movimientos; y el tercero detecta y genera alertas ante comportamientos sospechosos.

Seguimiento y caracterización automática de la trayectoria de una persona, en un entorno multicámaras de video-vigilancia

(Rosado Castellanos, 2019) [17]

Presenta un sistema de video-vigilancia multicámara capaz de realizar el seguimiento de un sujeto de interés a través de distintas cámaras, sin verse afectado por variacio-

nes de iluminación, escala o forma. Además, el sistema clasifica la acción realizada por la persona, identificando específicamente caminar en línea recta, girar a la izquierda, girar a la derecha y permanecer quieta.

Detección de Anomalías en el Comportamiento de una Persona

(Urquiza Flores, 2022) [18]

Desarrolló un modelo para el monitoreo de actividades diarias con el objetivo de reconocer patrones y detectar cambios en la rutina de una persona. El sistema identifica comportamientos normales, anormales e inhabituales mediante la detección de anomalías. Para ello, se recopilaban datos de una rutina habitual, los cuales sirvieron como referencia para detectar desviaciones en actividades futuras.

Estimación de la posición del cuerpo en secuencias de imágenes

(Arriaga Santos, 2025) [19]

Este trabajo propone un sistema para estimar y clasificar poses humanas mediante puntos clave del cuerpo, utilizando visión por computadora. Se compararon diversos modelos de estimación de pose y se evaluaron varios clasificadores en conjuntos de datos propios y de la literatura. Los resultados demostraron un buen desempeño, incluso ante oclusiones parciales, validando la efectividad del enfoque propuesto.

1.2.5. Organización del Documento

Este documento se encuentra estructurado en cinco capítulos, además de anexos que complementan el desarrollo de la investigación. A continuación, se describe brevemente el contenido de cada uno:

- Capítulo 1 – Introducción: Se presenta el planteamiento del problema, incluyendo su delimitación y complejidad. También se expone la propuesta general de solución, junto con la justificación, objetivos, alcances y antecedentes del proyecto.
- Capítulo 2 – Análisis del Estado del Arte y Marco Teórico: Se dividen los fundamentos en dos secciones. En la primera, se introducen los conceptos teóricos

clave necesarios para esta investigación como calibración, triangulación, modelos paramétricos, proporciones antropométricas y representaciones con *axis-angles*. En la segunda sección se realiza una revisión del estado del arte, abordando los enfoques existentes para la reconstrucción 3D del cuerpo humano, estimación de pose en 2D y 3D, y se concluye con un análisis comparativo entre los métodos revisados.

- Capítulo 3 – Especificación y modelado de la solución: Se detalla la propuesta planteada, incluyendo su especificación funcional y su modelado conceptual. Además, se describe el proceso de implementación, desde la captura de video, procesamiento y generación de los modelos 3D.
- Capítulo 4 – Experimentación y resultados: Se presenta la metodología experimental, las herramientas empleadas, las métricas utilizadas para la evaluación, y el desarrollo de las pruebas. En este capítulo se presentan y discuten los resultados obtenidos a lo largo de los distintos experimentos realizados.
- Capítulo 5 – Conclusiones: Se resumen las principales conclusiones del trabajo y se proponen posibles líneas de trabajo futuro.
- Finalmente, se incluyen los apartados de referencias bibliográficas y anexos, donde se recopilan documentos generados, tablas, imágenes y productos derivados del desarrollo de la solución propuesta.

Capítulo 2

Análisis del Estado del Arte y Marco Teórico

2.1. Marco Teórico

En esta sección se presentan los conceptos principales que fueron base para el desarrollo de la presente investigación. Su definición es necesaria para comprender adecuadamente el tema de estudio.

2.1.1. Calibración

Según los autores de [20], la calibración de cámaras se define como el proceso mediante el cual se determinan los parámetros internos y externos de una cámara, con el objetivo de establecer una relación precisa entre las coordenadas tridimensionales del mundo real y sus proyecciones bidimensionales en una imagen. El proceso incluye la estimación de los parámetros intrínsecos (como la distancia focal, el punto principal y los coeficientes de distorsión óptica) y los parámetros extrínsecos (como la posición y orientación de la cámara respecto al sistema de coordenadas del mundo real).

Este tipo de procedimiento requiere de un patrón de dimensiones conocidas (comúnmente de tipo ajedrez) capturado en el espacio. Mediante algoritmos de visión computacional, se detectan automáticamente los puntos de intersección del patrón.

Con esta información y las dimensiones conocidas, es posible calcular los parámetros de calibración utilizando métodos como el propuesto por Zhang [21] o funciones específicas de *OpenCV*.

En un proceso de autocalibración, como el utilizado en este proyecto, se requiere capturar, para cada cámara, varias imágenes del patrón de ajedrez en distintas posiciones y orientaciones (como se ilustra en la Figura 2.1). En cada una de ellas se detectan las esquinas del patrón y, para cada imagen, se calcula una matriz de homografía que relaciona las coordenadas del patrón 2D con sus proyecciones en la imagen. A partir de las matrices de homografía obtenidas de varias imágenes, se construye un sistema de ecuaciones que permite estimar la matriz de parámetros intrínsecos de la cámara [21].

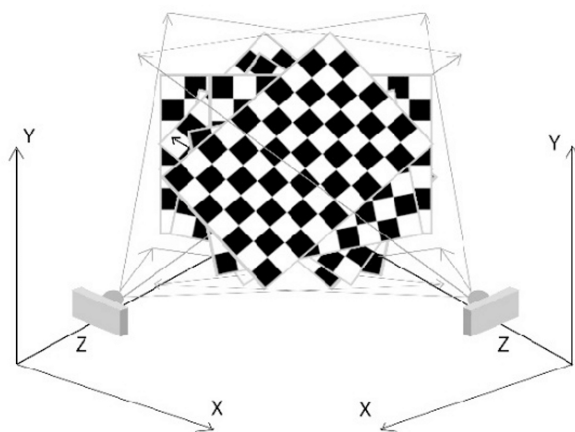


Figura 2.1: Ejemplo gráfico del proceso de calibración [1]

Una calibración adecuada permite obtener reconstrucciones tridimensionales más precisas, corrigiendo las distorsiones de la imagen, lo cual mejora significativamente el rendimiento de las aplicaciones de visión por computadora.

2.1.2. Triangulación

La triangulación es el método utilizado para estimar la posición tridimensional de un punto en el espacio, a partir de su proyección en dos o más imágenes capturadas desde distintos puntos de vista. Este proceso se basa en la geometría epipolar y requiere conocer los parámetros intrínsecos y extrínsecos de las cámaras involucradas [22].

Los mismos autores, definen formalmente la triangulación como la solución al problema geométrico:

Dado un conjunto de puntos $x_i \in R^2$ observados en i imágenes, y sus correspondientes matrices de proyección $P_i \in R^{3 \times 4}$, se busca el punto $X \in R^3$ tal que:

$$x_i \sim P_i X$$

para todo i .

En la práctica, debido a ruido y errores de correspondencia, este sistema rara vez se cumple exactamente, por lo que se estima X como el punto que minimiza el error de reproyección entre las observaciones x_i y las proyecciones modeladas

$$\vec{x}_i = P_i X$$

En la Figura 2.2 se muestra gráficamente la idea del proceso de triangulación.

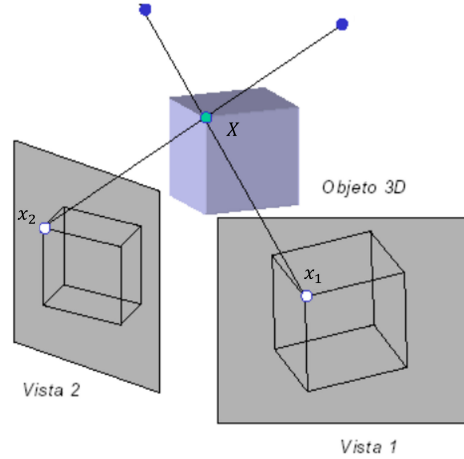


Figura 2.2: Ejemplo gráfico del proceso de triangulación (modificada de [2])

2.1.3. Modelo Paramétrico

Un modelo paramétrico del cuerpo humano es una representación matemática tridimensional del cuerpo en la que su forma y/o pose se controlan mediante un conjunto reducido de parámetros. Estos modelos se construyen a partir de datos

anatómicos reales (por ejemplo, escaneos 3D) y permiten generar nuevas instancias del cuerpo humano al modificar dichos parámetros [23].

Formalmente, un modelo paramétrico puede representarse como una función:

$$M(p) : R^n \rightarrow R^{3 \times V}$$

donde:

$p \in R^n$ denota el conjunto de parámetros del modelo.

$R^{3 \times V}$ corresponde al espacio tridimensional de los V vértices de una malla que representa el cuerpo.

El modelo paramétrico *SMPL* (*Skinned Multi-Person Linear*) [3] representa el cuerpo humano mediante una malla tridimensional compuesta por 6890 vértices, cuya forma y postura están controladas por dos vectores de parámetros:

- Forma: $\beta \in R^{n_\beta}$: codifica variaciones corporales como la altura y la complexión.
- Pose: $\theta \in R^{3K}$: codifica las rotaciones de las K articulaciones del cuerpo humano mediante representaciones en *axis-angle*.

En la Figura 2.3 se muestra un ejemplo del modelo paramétrico *SMPL*, en a) se presenta el modelo base, en b) se modifican los parámetros de forma β , mientras que en c) se modifican los parámetros de pose θ .



Figura 2.3: Modelo paramétrico *SMPL* (adaptada de [3])

2.1.4. Proporciones antropométricas del cuerpo humano

Las medidas antropométricas son descripciones cuantitativas de las dimensiones físicas del cuerpo humano.

Según [24], estas medidas se recopilan para definir la geometría de los distintos segmentos corporales (como el muslo, el torso o el antebrazo), así como su masa, centro de gravedad y momentos de inercia.

Las medidas típicas incluyen: longitudes segmentarias (por ejemplo, del brazo o la pierna), altura, peso corporal, circunferencias y diámetros (por ejemplo, de la cintura, la cadera o la muñeca), entre otros.

2.1.5. Axis-Angles

Una rotación en un espacio tridimensional puede representarse mediante un vector *axis-angle*. Este modelo describe una rotación como una combinación de un eje y un ángulo, donde el eje define la dirección en torno a la cual se rota, y el ángulo representa la magnitud de dicha rotación [25]. Esto se ilustra en la Figura 2.4.

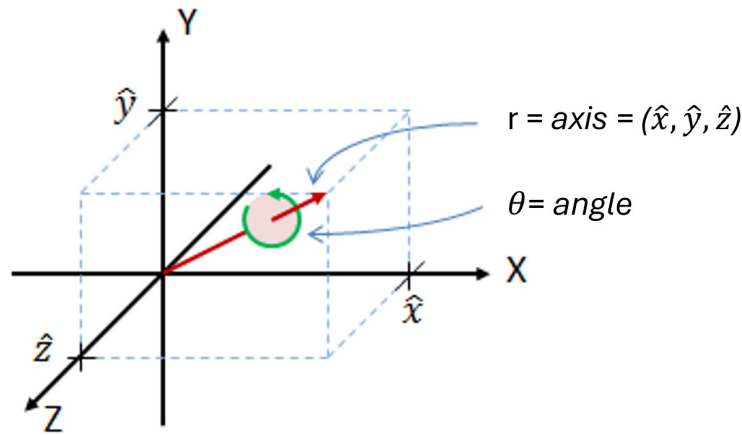


Figura 2.4: Representación gráfica de un *axis-angle* (Figura adaptada de [4])

Formalmente, una rotación *axis-angle* se representa con un vector $w \in R^3$, donde:

$$w = \theta r$$

Dónde:

$\theta \in R$ es el ángulo de rotaciones (en radianes).

$r \in R^3$ es un vector unitario que indica el eje de rotación.

2.2. Estado del Arte

Las tareas de reconocimiento, estimación de la pose y reconstrucción del cuerpo humano son un campo ampliamente estudiado. Su uso tiene diversos enfoques, como la medicina [8], realidad aumentada [9], realidad virtual [26], fisioterapia [27] con ayuda de sensores, y aplicaciones 3D como estimación de pose [10] y detección de actividades humanas [28], por mencionar algunas.

A continuación, se presenta el estado del arte organizado de la siguiente manera: se aborda el tópico de reconstrucción 3D del estado del cuerpo humano, describiendo diversos enfoques y métodos para esta tarea. Posteriormente, se analizan los modelos de estimación de pose 2D que son utilizados actualmente, y por último se introducen los modelos desarrollados para la estimación de pose 3D.

2.2.1. Reconstrucción 3D del cuerpo humano

La reconstrucción 3D consiste en recuperar la estructura geométrica de un objeto a partir de una o más imágenes bidimensionales. Este proceso permite estimar la forma, posición y orientación en el espacio tridimensional utilizando información visual en 2D, ya sea proveniente de vistas múltiples, sensores especializados o mediante técnicas de visión por computadora [29]. En el caso del cuerpo humano, la reconstrucción 3D puede definirse como la recuperación de su forma tridimensional a partir de información bidimensional, generando así una representación digital detallada.

En el campo de la visión artificial, actualmente existen diversos modelos desarrollados para la reconstrucción 3D del cuerpo humano mediante distintas técnicas y dispositivos.

Principalmente, existen métodos que emplean dispositivos de captura con sensores para obtener posiciones exactas en el espacio tridimensional. Por ejemplo, el trabajo de [30] utiliza una combinación de trajes MoCap con marcadores, cámaras de captura, sensores de profundidad estéreo y micrófonos para generar un conjunto

de datos detallado. En [31], se emplean sensores blandos colocados en una playera como alternativa para capturar la forma del torso del cuerpo humano. Finalmente, el estudio presentado en [32] se basa en el uso de una cámara RGB-D para realizar una reconstrucción completa y en tiempo real del cuerpo humano.

A pesar de las ventajas que ofrecen estos sensores, al proporcionar datos confiables sobre la posición y forma en el espacio real, su uso se ve limitado por su carácter invasivo y por su falta de adecuación a diversos entornos. Por esta razón, los modelos basados únicamente en técnicas computacionales han ganado relevancia, ya que permiten realizar la reconstrucción a partir del uso exclusivo de imágenes o video. Con base en la revisión del estado del arte, se observa que los enfoques computacionales pueden agruparse en dos categorías, según el tipo de resultados que generan: “modelos parametrizados” y “modelos no parametrizados”.

En términos generales, los modelos no parametrizados se caracterizan por no utilizar representaciones como vectores para describir la geometría del cuerpo. Es decir, a partir de la entrada (imágenes o videos), estos modelos generan directamente una malla tridimensional. Aunque este enfoque no es el más común, existen diversos métodos que lo utilizan, generando modelos del cuerpo mediante redes neuronales [33, 34], transformadores [35], campos de distancia implícitos [36] y voxelización [37].

Por otro lado, un modelo paramétrico se basa en la representación mediante un conjunto de parámetros, los cuales describen la malla tridimensional generada del cuerpo humano. En este enfoque, se ha observado que gran parte de los desarrollos utilizan una única cámara como entrada. El primer modelo desarrollado en esta línea fue *SMPL* [3], el cual tiene como base un modelo 3D del cuerpo humano promedio, que es modificado por medio de parámetros obtenidos de una imagen 2D de entrada. Estos parámetros se dividen en dos: parámetros de forma, que se encargan de controlar aspectos como altura, peso y circunferencia del cuerpo; y parámetros de pose, los cuales permiten representar distintas posturas del cuerpo.

Este modelo ha servido como base para modelos de reconstrucción del cuerpo humano más recientes, como *SMPL-X* [38], el cual utiliza una mayor cantidad de

parámetros para una mejor representación, enfocándose en la cara y las manos; y SMPLify [39], el cual se desarrolló como un modelo que simplifica y optimiza las estimaciones de pose y forma del cuerpo humano de *SMPL*.

A partir de estos avances, se propusieron modelos como HMR [40], que introdujo un enfoque de inferencia directa desde una imagen hacia los parámetros de *SMPL* mediante redes neuronales profundas, sin necesidad de etapas de optimización explícita. Este modelo permitió obtener resultados en tiempo real a partir de imágenes individuales. Más recientemente, HMR 2.0 [41] amplió esta idea incorporando supervisión multivista, mayor robustez ante variaciones en la pose, y mejoras en la calidad de la reconstrucción, lo que lo convierte en una de las propuestas más precisas y eficientes actualmente.

También, se han propuesto nuevos modelos que mejoran la robustez y precisión de las estimaciones. Uno de ellos es PARE [42], que introduce un enfoque basado en atención para refinar la predicción de la pose, incluso en casos donde el cuerpo se encuentra parcialmente ocluido. Por su parte, PyMAF [43] propone un método de ajuste iterativo basado en regresores multinivel que permiten una estimación más precisa y coherente de la forma y pose del cuerpo. Este enfoque fue ampliado en PyMAF-X [44], el cual extiende sus capacidades para capturar no solo el cuerpo, sino también los detalles de rostro, manos y expresiones faciales, integrando así una representación más completa del cuerpo humano. Finalmente, HUMOR [45] se enfoca en la estimación de la dinámica del cuerpo en secuencias de video, permitiendo modelar el movimiento humano de forma realista mediante la incorporación de información temporal en su optimización.

Además se han desarrollado enfoques multivista que aprovechan imágenes capturadas desde múltiples ángulos para mejorar la precisión en la reconstrucción del cuerpo humano. Estos métodos ofrecen resultados más detallados y robustos, especialmente en contextos con oclusiones, interacciones entre múltiples personas o movimientos complejos. Por ejemplo, HS-Nets [46] estima la forma del cuerpo humano a partir de siluetas extraídas de vistas múltiples utilizando redes convolucionales. Por otro lado, DeepMultiCap [47] se enfoca en la captura de múltiples personas en movimiento utilizando un conjunto reducido de cámaras, logrando reconstrucciones

precisas en escenas dinámicas. De igual manera, [48] combina información de múltiples vistas para mejorar tanto la pose como la forma estimada de cada individuo. Finalmente, hay modelos [49–51], que proponen estrategias que no dependen de calibración precisa entre cámaras, facilitando la implementación práctica y robusta de estos sistemas en escenarios no controlados.

En la Tabla 2.1 se presenta una comparativa de los métodos descritos anteriormente, considerando los *datasets* y las arquitecturas utilizadas. De dicha comparación, se puede resumir que los *datasets* comúnmente empleados son datos propios, *Human3.6M*, *Fit3D* y *MPI-INF-3DHP*. Asimismo, se observa que la mayoría de las arquitecturas están basadas en modelos de redes neuronales.

Tabla 2.1: Comparativa de los métodos del estado del arte en términos de *datasets* y las arquitecturas utilizadas

Modelo	Dataset	Arquitectura
[3]	Escaneos propios	Modelo paramétrico estadístico
[38]	Escaneos propios	Optimización de SMPL
[39]	LSP	Optimización de SMPL
[40]	Human3.6M, MPI-INF-3DHP, LSP	CNN + Regresor
[41]	3DPW, Human3.6M, MPI-INF-3DHP, COCO	Transformers (ViT + Decoder)
[42]	COCO, LSP, Human3.6M, 3DPW	CNN (ResNet50) + Atención por Partes
[43]	MPI-INF-3DHP, LSP-Extended	CNN + Retroalimentación de Alineación de Malla Piramidal
[44]	No detallado	CNN + PyMAF + Integración Adaptativa
[45]	AMASS, Human3.6M, 3DPW	Autoencoder + Dinámica Temporal
[46]	HumanEva, Human3.6M	CNN (Basado en siluetas)
[47]	No detallado	CNN + Atención Espacial + Fusión Temporal
[48]	No detallado	Optimización Grueso a fino + SMPL
[49]	No detallado	CNN + Optimización
[50]	No detallado	MLP + Transformer Decoder
[51]	No detallado	CNN + Fusión de Vistas

2.2.2. Modelos de estimación de pose 2D

La estimación de la pose humana es el proceso de detectar y localizar automáticamente partes clave del cuerpo (como cabeza, hombros, codos, muñecas, caderas, rodillas, entre otros) en imágenes o secuencias de video, con el fin de describir la postura corporal en 2D o 3D. Esta tarea resulta fundamental en el campo de la visión por computadora y posee aplicaciones en análisis de movimiento, interacción humano-computadora, realidad aumentada, vigilancia y rehabilitación médica [52].

En general, los modelos que estiman la pose en 3D pueden presentar sensibilidad ante condiciones ambientales, tales como iluminación y resolución, motivo por el cual gran parte de la literatura se centra en la estimación de pose en 2D, que resulta más robusta y ampliamente empleada. Entre los modelos más reconocidos se encuentran:

- *OpenPose* [52]: uno de los primeros y más populares modelos para la estimación de pose en 2D, capaz de detectar múltiples personas y sus articulaciones en tiempo real mediante redes neuronales convolucionales.
- *MediaPipe Pose* [53]: desarrollado por *Google*, ofrece una solución eficiente y ligera para la estimación de pose, optimizada para dispositivos móviles y aplicaciones en tiempo real.
- *AlphaPose* [54]: conocido por su alta precisión en la estimación de poses múltiples, utiliza técnicas avanzadas para mejorar la detección en escenarios con oclusiones y posturas complejas.
- *PoseNet* [55]: modelo ligero y rápido diseñado para operar en navegadores *web*, permitiendo la estimación de pose en tiempo real con dispositivos estándar.
- *MoveNet* [56]: modelo reciente que ofrece alta precisión y velocidad, ideal para aplicaciones interactivas en tiempo real, con arquitectura optimizada para el rendimiento en dispositivos móviles.

2.2.3. Estimación de pose 3D

La estimación de pose 3D se refiere al proceso de determinar la posición y orientación espacial de un objeto o sujeto en un espacio tridimensional, generalmente a

partir de datos visuales obtenidos mediante cámaras o sensores [57]. Esta tarea es fundamental en campos como la visión por computadora, robótica, realidad aumentada y análisis de movimiento, debido a su capacidad para representar la estructura y dinámica del entorno de manera precisa.

En la actualidad gracias a las redes neuronales, se han desarrollado métodos capaces de estimar la pose 3D a partir de imágenes monoculares. En [58] propusieron un modelo sencillo basado en redes neuronales que mapea coordenadas 2D a 3D con alta precisión; en contraste, en [59] se introduce un enfoque volumétrico que estima mapas de probabilidad 3D para cada articulación, logrando un mejor desempeño en poses complejas. Además, arquitecturas como *Hourglass Networks* [60] han sido ampliamente adoptadas para capturar información contextual a múltiples escalas, mejorando la estimación tanto en 2D como en 3D.

Sin embargo históricamente, los métodos clásicos para la estimación de pose 3D se basaron en técnicas geométricas y modelado explícito, utilizando sistemas multicámara para reconstruir la pose a partir de correspondencias de 2D a 3D [61], con las limitaciones de dependencia de múltiples vistas y su sensibilidad.

2.2.4. Conclusiones del estado del arte

Del análisis del estado del arte se concluye que la estimación y reconstrucción 3D del cuerpo humano es un área en constante evolución, dominada por enfoques basados en redes neuronales profundas.

Estos métodos han permitido superar varias de limitaciones de reconstrucción, especialmente en la capacidad de estimar poses tridimensionales a partir de imágenes monoculares, oclusiones y trabajo en tiempo real. Sin embargo, a pesar del éxito de las mismas, se puede destacar que dependen en gran medida de grandes conjuntos de datos para su entrenamiento y validación. Entre los *datasets* más utilizados en el área destacan *Human3.6M* [62], *Fit3D* [63] y *MPI-INF-3DHP* [64].

Las métricas utilizadas en la literatura para la comparación del desempeño de modelos de estimación de pose y reconstrucción 3D son *Mean Per Joint Position Error (MPJPE)* y su variante *Procrustes Aligned Mean Per Joint Position Error (PA-MPJPE)*.

En resumen, el campo presenta un gran desempeño en enfoques basados en redes neuronales gracias a su poder de generalización, manejo de variabilidad, precisión y control, sin embargo estos modelos requieren de grandes cantidades de datos y equipo computacional de alto rendimiento por lo que métodos geométricos multivista, podrían destacar especialmente en escenarios donde la disponibilidad de datos y de equipo es limitada.

Capítulo 3

Especificación y modelado de la solución

3.1. Especificación de la solución

A partir del análisis del estado del arte y la investigación realizada, se ha definido un enfoque de solución que prioriza dos aspectos fundamentales.

En primer lugar, se adopta un enfoque multivista que permite acceder a información que podría permanecer oculta o ser poco visible desde una única perspectiva.

En segundo lugar, se presenta como una alternativa al uso de redes neuronales. En lugar de recurrir a técnicas basadas en aprendizaje profundo, se opta por un enfoque que se fundamenta únicamente en principios de geometría básica, lo cual permite mantener la solución más liviana en términos de requerimientos computacionales.

3.1.1. Modelado de la solución

En la Figura 3.1 se muestra el modelo de solución propuesto, el cual se compone de 3 bloques principales:

- Preparación y captura de datos
- Procesamiento de los datos

■ Generación y evaluación del modelo 3D

donde las flechas rojas indican los datos generados en cada bloque y las flechas verdes los datos necesarios para el procesamiento realizado en cada bloque.

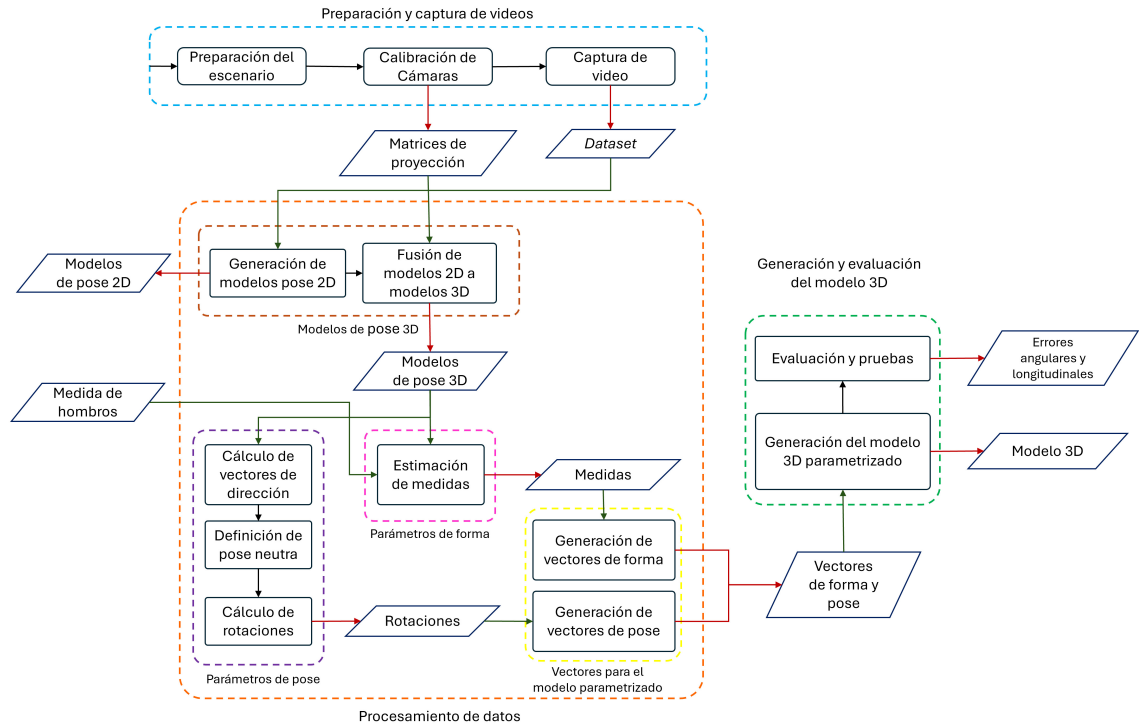


Figura 3.1: Diagrama de solución diseñado

Cada uno de los bloques se componen por diversos módulos los cuales en conjunto permiten la generación de los modelos 3D. Cada uno de los módulos se describen en las siguientes secciones.

3.2. Implementación de la solución

3.2.1. Preparación y captura de video

El bloque de preparación y captura de video consiste en la configuración física del entorno de grabación, incluyendo la disposición de las cámaras. Según lo localizado en el estado del arte, se optó por mantener la configuración empleada en los *datasets*

más utilizados en el área, como *Human3.6M* y *Fit3D*, donde las cámaras capturan a los actores según la configuración mostrada en la Figura 3.2 a, mientras realizan diversos ejercicios. Adicionalmente, se incorporaron cuatro cámaras que graban el cuerpo humano con la configuración de la Figura 3.2 b, para generar un conjunto de datos más completo y reutilizable.

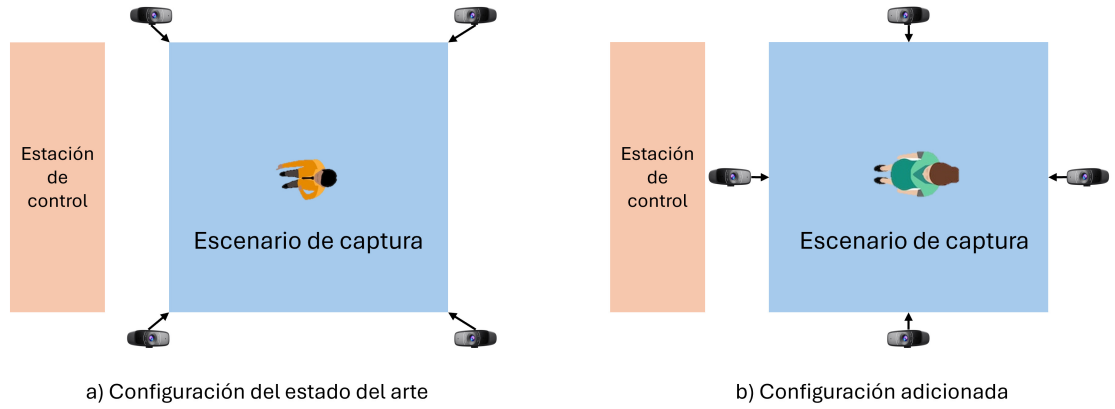


Figura 3.2: Distribución de cámaras en el espacio de grabación

Para este trabajo fue esencial realizar un calibrado de las cámaras, con el fin de conocer los parámetros intrínsecos de cada una. Para ello, se utilizó un patrón plano de tipo ajedrez, guardando para cada cámara su respectiva matriz de proyección.

La captura de datos se realizó mediante *YoloV8*, utilizando un módulo de grabación simultánea. Se emplearon dos equipos de cómputo, cada uno grabando 4 de las 8 cámaras disponibles. En uno de los equipos se concentraron las cámaras ubicadas en las esquinas de las habitaciones, y en el otro las cámaras restantes, de modo que una posible desincronización entre ambos equipos no impacte directamente en los resultados de la presente investigación.

Se capturaron videos de 12 personas, de forma individual, mientras realizaban cada uno de los diecisiete ejercicios considerados, esto hace un total de $12 \times 17 = 204$ videos para cada una de las 8 cámaras, generando un *dataset* de 16632 videos, de 30 segundos cada uno grabados a 30 *fps*. Los ejercicios abarcan movimientos de brazos,

piernas y cuerpo completo. A continuación se listan los ejercicios capturados y en la Figura 3.3 se ilustran los movimientos.

1. Movimientos de brazos (Figura 3.3 a-g): movimientos circulares, subir y bajar al frente, subir y bajar a los laterales, estirarse, nadar, simular empujar algo y golpear al frente.
2. Movimientos de piernas (Figura 3.3 h-m): desplantes laterales, pasos laterales, desplantes frontales, marchar, levantar rodillas al pecho y llevar el pie al glúteo.
3. Movimientos de cuerpo completo (Figura 3.3 n-q): agacharse, simular sentarse, saltar y realizar movimientos de robot.

Además, como medida complementaria, se registraron la altura y el ancho de hombros de cada una de las personas participantes.

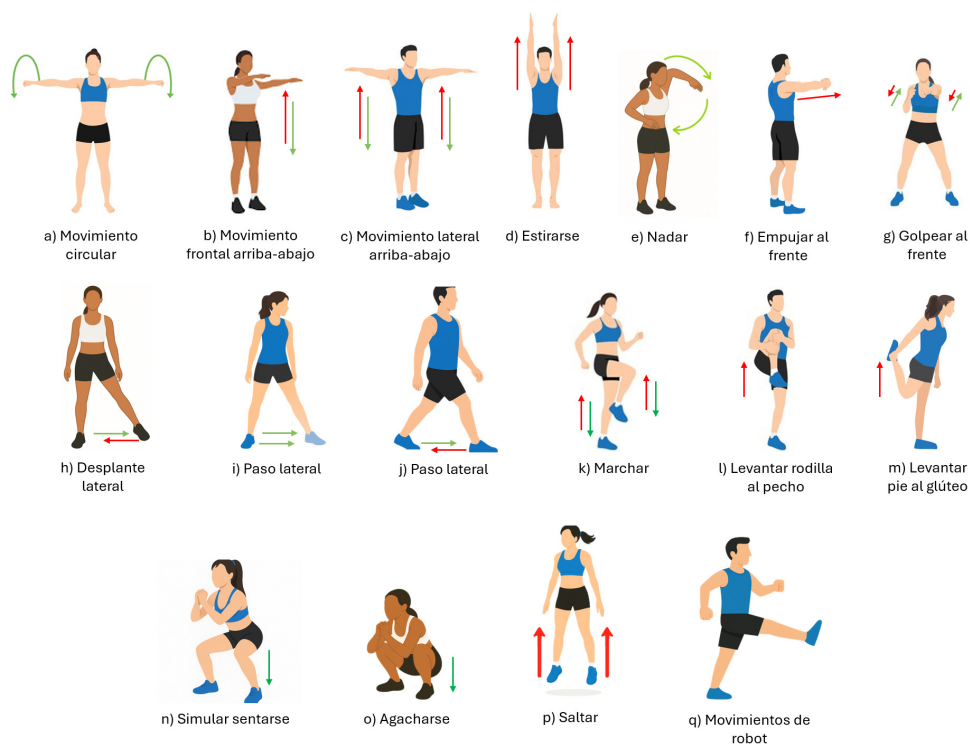


Figura 3.3: Ejemplos gráficos de los ejercicios

3.2.2. Procesamiento

La etapa de procesamiento se divide en varias fases, las cuales se describen a continuación. Los experimentos que justifican las decisiones tomadas en esta configuración pueden consultarse en la sección 4 (Experimentación y Resultados).

Modelo de pose 3D

Dado que el modelado de los videos generados es suave en transición se puede hacer un muestreo tomando una menor cantidad *frames* para realizar el modelado de una manera más simple y liviana. Un procesamiento inteligente seleccionaría los *frames* que presentan una mayor variabilidad en la pose; sin embargo, esto requiere de una carga computacional adicional, por esta razón se realiza de forma empírica muestreando cada $10 * n$ con n un entero a partir de 1.

Para cada uno de los videos, se realiza la detección de pose en dos dimensiones mediante el modelo *MediaPipe Pose*, obteniendo para cada vista y *frame* del video un total de 34 puntos por pose, los cuales se muestran en la Figura 3.4, con coordenadas (x, y, z) para cada uno de los puntos estimados, donde z corresponde a una estimación de profundidad. Para cada una de las cuatro vistas utilizadas en este proceso, se generaron archivos correspondientes a cada cámara, donde cada línea contiene los valores (x, y) de las poses estimadas, respetando el orden definido por *MediaPipe Pose* y omitiendo el valor z debido a su baja confiabilidad.

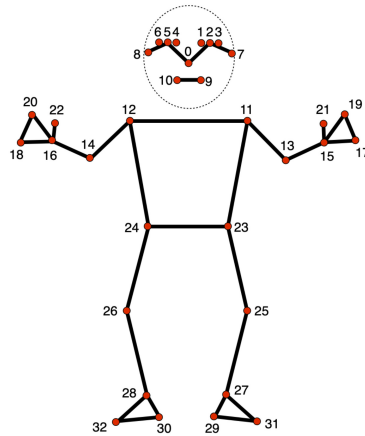


Figura 3.4: Puntos detectados por *mediapipe* [5]

Una vez realizada la estimación de pose, se lleva a cabo un proceso de triangulación entre puntos (de cada *frame*), utilizando las matrices de proyección y los puntos obtenidos con *MediaPipe Pose*. Para esta implementación en particular, se optó por una triangulación iterativa empleando únicamente las dos cámaras frontales ubicadas en las diagonales, las cuales se señalan en la Figura 3.5, ya que mostraron el mejor desempeño.

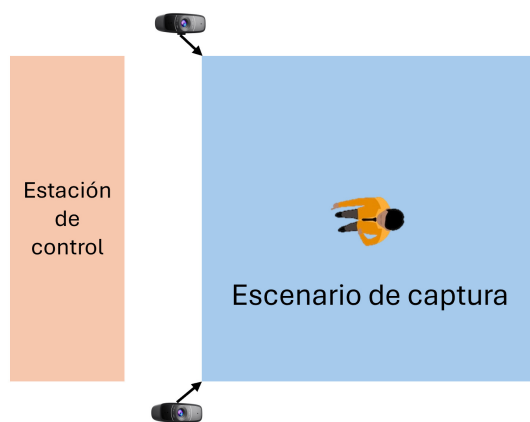


Figura 3.5: Cámaras empleadas para la triangulación

Como resultado, se obtiene un modelo de pose en 3D, donde los valores de profundidad son estimados mediante una relación entre el espacio real y lo observado por las cámaras. En este punto, se guarda en un archivo de texto las coordenadas del modelo tridimensional calculado, donde cada línea corresponde a la pose de un *frame*, respetando el orden de los puntos de *MediaPipe Pose*, con la diferencia de que cada coordenada se representa ahora con x , y y z . Para ver un ejemplo del archivo generado ver Anexo A Figura A.3.

Parámetros de pose

Para obtener los parámetros de pose, cada uno de los videos comienza con una pose que es fundamental para el modelado realizado. Por ello, se solicitó a los actores iniciar el video en una *T-pose*, la cual consiste en que la persona se coloque de pie con

los brazos completamente extendidos hacia los costados, a la altura de los hombros (como se muestra en la Figura 3.6). Esta pose se considera como la pose neutra.

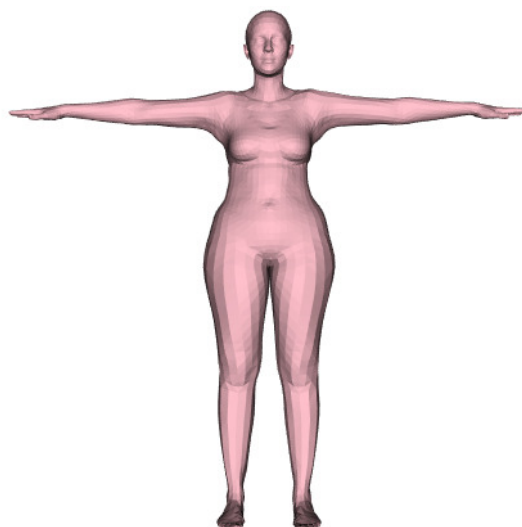


Figura 3.6: *T-pose*

Para cada una de las poses, se realiza el cálculo de vectores de dirección entre nodos (o articulaciones), los cuales están definidos como la dirección desde un nodo padre hacia un nodo hijo (por ejemplo: del hombro al codo), utilizando una correspondencia entre los nodos usados en *MediaPipe Pose* y los del modelo 3D a emplear, *SMPL*.

Una vez obtenidos los vectores de dirección, es necesario realizar una comparación entre el vector de dirección de la *T-pose*, que siempre servirá como referencia, y el vector de dirección correspondiente a la pose en un *frame* arbitrario x mediante operaciones vectoriales simples. Esta comparación da como resultado una matriz de rotación, la cual puede ser convertida a un vector del tipo *axis-angle* mediante operaciones matemáticas.

Para ejemplificar los vectores generados, se toma un vector en la postura T y un vector en una pose arbitraria, el resultado de la rotación según el *axis-angle* generado se puede ver visualmente en la Figura 3.7 dónde la posición y rotación del brazo es:

En *T-pose* es:

$$[0,158207, -0,361927, 0,004388]$$

En el *frame* arbitrario es:

$$[0,090904, -0,0187249, -0,0336379]$$

El *axis-angle* generado es:

$$[1,287970, 0,537176, 0,318162]$$

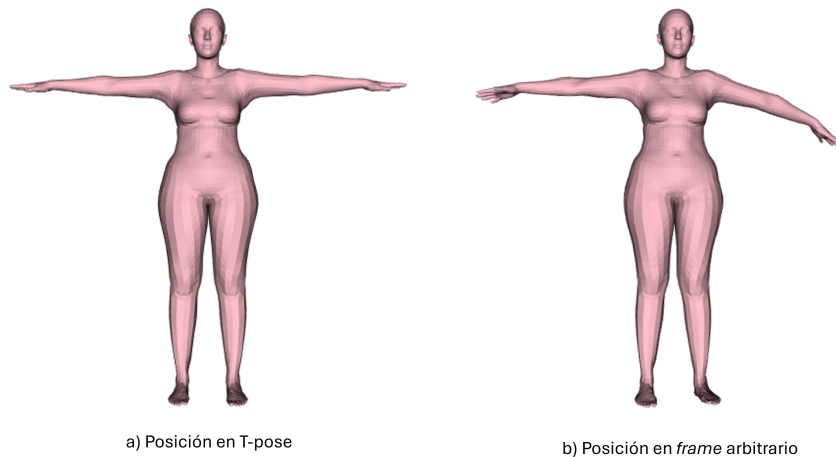


Figura 3.7: Transformación de la postura

Parámetros de forma

Dado que se cuenta con información de calibración, es posible realizar una estimación de medidas a partir del modelo de pose 3D generado. Sin embargo, debido a los errores asociados al equipo de captura y al error de calibración obtenido, se opta por utilizar la distancia real de los hombros del actor como referencia inicial. Esto permite calcular otras medidas antropométricas durante la triangulación. Las medidas obtenidas en este punto son:

- Ancho de caderas
- Largo de brazos y piernas

- Altura desde el talón hasta el hombro

las cuales se ejemplifican en la Figura 3.8.

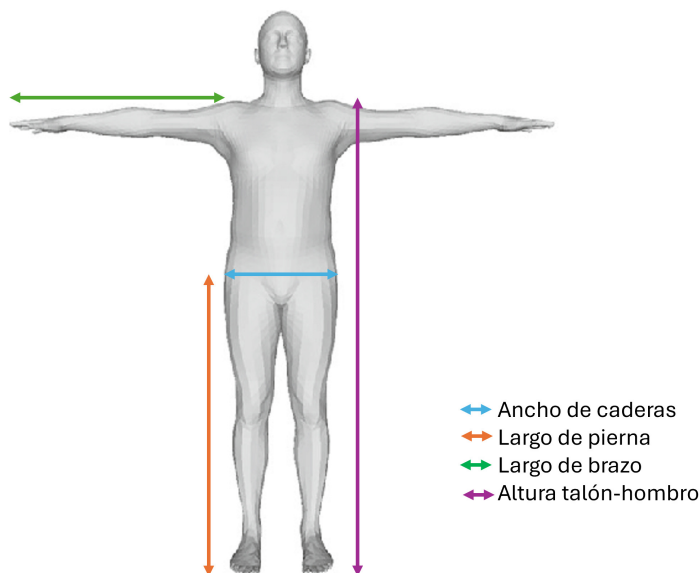


Figura 3.8: Ejemplo de medidas del modelo

No obstante, dado que el modelo *SMPL* requiere ciertas medidas antropométricas específicas para la generación del modelo paramétrico, la medida de altura total del cuerpo humano es indispensable para lograr una reconstrucción más precisa. Para estimarla, se recurrió al estudio anatómico de Leonardo da Vinci en su obra más reconocida, *El Hombre de Vitruvio* [65], donde se menciona que “la cabeza, desde la barbilla hasta la coronilla, mide una octava parte de todo el cuerpo”.

Por lo tanto, el cálculo de la altura total se realizó considerando que la distancia desde el talón hasta el hombro representa $\frac{7}{8}$ de la altura total de la persona en escena.

Vectores para el modelo *SMPL*

El modelo *SMPL* recibe, para su modificación, dos vectores de parámetros. En su versión más simple, que ha servido como base para diversos trabajos posteriores en el estado del arte (como son *SMPLify*, *SMPL-X*, *HMR*, entre otros), requiere de un

vector de pose $\theta \in R^{72}$, donde los primeros tres valores representan la rotación de la pelvis, y los 69 restantes representan mediante valores (x, y, z) las rotaciones de 23 articulaciones. Por otro lado, recibe un vector $\beta \in R^{10}$ que representa las variaciones morfológicas del cuerpo, como altura, peso, complexión, entre otros.

Como parte de la correspondencia que se establece entre los puntos obtenidos por *MediaPipe Pose* y el modelo *SMPL*, se realiza una adaptación en la cual únicamente 16 de las 24 articulaciones contienen valores de rotación (las cuales se pueden identificar en la Figura 3.9 rodeadas en color rojo); el resto se mantienen con valores de cero, evitando así modificaciones no conocidas.

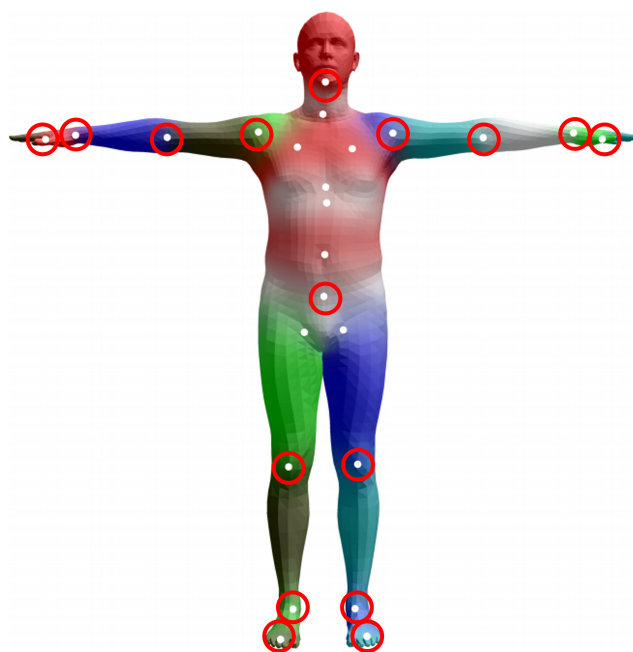


Figura 3.9: Articulaciones que tienen valores de rotación

En el caso de los parámetros antropométricos, se cuenta con 6 de los 10 valores requeridos por el modelo. Dado que *SMPL* es según sus autores un modelo pre-entrenado y optimizado, y el envío de ceros en los valores faltantes permite que el modelo realice una estimación razonable de dichos parámetros basándose en los datos disponibles.

3.2.3. Generación del modelo 3D

La generación del modelo 3D se realiza de manera directa, modificando el modelo *SMPL* en su forma base según los vectores calculados en las etapas anteriores. Se pasan al modelo *SMPL* los vectores de parámetros θ y β , y se aplican las modificaciones correspondientes, cuando se realiza este proceso el modelo *SMPL* nos devuelve el valor del ángulo que rota cada una de las articulaciones, este proceso se repite para cada uno de los *frames* previamente muestreados.

Para cada *frame*, se guarda un modelo generado en formato *.obj* (en el Anexo A Figura A.1) junto con una imagen del mismo, con el objetivo de facilitar la visualización final. Además, en un archivo adicional se almacenan los datos de error de ángulo de rotación y de longitud, con el fin de ser analizados posteriormente.

Dado que se emplea un enfoque geométrico, no es posible utilizar métricas como *MPJPE* o *PA-MPJPE* usadas en modelos de aprendizaje profundo, por lo que la evaluación del modelo se realiza utilizando métricas estadísticas tradicionales, como la media aritmética y la desviación estándar de los errores obtenidos para cada uno de los videos procesados.

Capítulo 4

Experimentación y resultados

4.1. Metodología experimental

4.1.1. Herramientas y tecnologías

Para la realización de este proyecto se consideraron dos etapas: la primera consistió en la adquisición de datos, y la segunda en el desarrollo del sistema, que incluyó el procesamiento de los videos y la evaluación de los resultados.

Para la adquisición de datos se utilizaron ocho cámaras *ATVIO*, con una resolución de video de 1280×720 pixeles y una velocidad de captura de 30 *fps*, así como dos computadoras con las siguientes características:

- La primera, empleada exclusivamente para la adquisición: sistema operativo *Windows 10 Home 64 bits*, procesador *Intel Core i7* (6ª generación), 8 GB de memoria *RAM* y tarjeta gráfica *NVIDIA GeForce GTX 960*.
- La segunda, utilizada tanto para la captura como para la etapa de desarrollo del sistema: sistema operativo *Windows 10 Pro 64 bits*, procesador *Intel Core i5* (13ª generación), 32 GB de memoria *RAM* y tarjeta gráfica *NVIDIA GeForce RTX 4060 Ti*.

El procesamiento de los datos se realizó utilizando Anaconda, en un entorno virtual con *Python 3.8.20*, empleando las siguientes librerías principales: *OpenCV*, *NumPy*, *MediaPipe*, *SMPLX*, *Torch*, *Scipy* y *Trimesh*.

4.1.2. Métricas para el análisis de resultados

Para evaluar cuantitativamente el desempeño del método geométrico multivista aplicado para la reconstrucción 3D con el modelo *SMPL*, se utilizaron las siguientes métricas, que permiten medir la exactitud de la reconstrucción espacial y angular: error de reproyección, error de triangulación, error angular y error longitudinal.

Error de Reproyección

Se define como la distancia euclidiana promedio, en píxeles, entre las proyecciones de los puntos reconstruidos en las vistas 2D y sus posiciones correspondientes en las imágenes originales. Formalmente, si p_i es la posición observada en píxeles y $\hat{p}_i$ la reproyección del punto 3D reconstruido, el error se calcula como:

$$E_{reproyeccion} = \frac{1}{N} \sum_{i=1}^N \|p_i - \hat{p}_i\|$$

donde N es el número total de puntos considerados.

Error de triangulación

Corresponde a la distancia euclidiana en el espacio 3D entre el punto reconstruido $\hat{X}_i$ y el punto de referencia X_i . Este error mide la precisión espacial de la reconstrucción 3D y se calcula como:

$$E_{triangulacion} = \frac{1}{N} \sum_{i=1}^N \|X_i - \hat{X}_i\|$$

Error Angular

Representa la diferencia en grados entre los vectores estimados por el modelo *SMPL* $\hat{v}_i$ y los de referencia v_i . Este error se calcula mediante:

$$\theta_i = \arccos \left(\frac{\hat{v}_i \cdot v_i}{\|\hat{v}_i\| \|v_i\|} \right)$$

Error longitudinal

Se calcula como la diferencia en centímetros entre las longitudes de segmentos generadas por el modelo *SMPL* $\hat{L}_i$ y las longitudes reales L_i para todos los segmentos considerados. Se calcula mediante:

$$E_{longitud} = \frac{1}{N} \sum_{i=1}^N |\hat{L}_i - L_i|$$

Para las métricas de error angular y longitudinal, se calcularon el valor promedio (media aritmética) y la desviación estándar con el fin de obtener una medida representativa de la tendencia central y la dispersión de los errores observados.

Donde:

- Promedio: Indica el valor central o de la métrica a lo largo del video. Un promedio bajo en cualquiera de las métricas implica que, en general, la reconstrucción es precisa y cercana a los valores reales.
- Desviación estándar: Mide la variabilidad o consistencia de los errores alrededor del promedio. Una desviación estándar baja señala que los errores son similares y la reconstrucción es estable y confiable, mientras que una desviación alta indica variabilidad significativa, lo que puede sugerir inconsistencias.

4.2. Desarrollo experimental

4.2.1. Resumen experimental

La Tabla 4.1 muestra cómo se alinean los experimentos realizados con los objetivos del proyecto. Los objetivos se corresponden con la siguiente numeración:

1. Generar un conjunto de datos de video sincronizado de múltiples vistas y personas.
2. Implementar y comparar los algoritmos de generación de modelos 2D del cuerpo humano, para seleccionar el que mejor se acople a las necesidades.
3. Generar modelos 2D del cuerpo humano para cada una de las vistas manteniendo la sincronización de los videos.

4. Investigar arquitecturas que permitan la fusión de modelos 2D del cuerpo humano, estudiar las encontradas y seleccionar la que se adapte de mejor forma al objetivo.
5. Fusionar los modelos 2D generados para obtener un modelo 3D.
6. Generar un conjunto de datos 3D de los videos capturados.

Tabla 4.1: Tabla resumen de experimentos

Experimento	Objetivos					
	1	2	3	4	5	6
1	x					
2		x	x			
3				x		
4				x		
5				x		
6				x		
7					x	
8						x
9						x
10						x

4.2.2. Generación del *Dataset*

Experimento 1: Captura de datos con múltiples cámaras

La generación del *dataset* requirió de la instalación de cámaras en el escenario disponible para la captura de datos. Estas se colocaron en un área de 2.7×3.0 m, a una altura de 2.10 metros, distribuidas como se señala en la Figura 4.1, adicionalmente se agrega una numeración de las mismas, para referencias futuras.

Inicialmente, se contempló realizar la captura mediante una programación multi-hilo en *Python*. Sin embargo, la implementación solo permitió la recepción de video en tiempo real desde máximo tres cámaras.

Por esta razón, se optó por utilizar un software externo que permitiera la captura sincronizada desde las ocho cámaras. Esta solución generaba una imagen combinada de toda la escena, como se ilustra en la Figura 4.2. No obstante, requería un

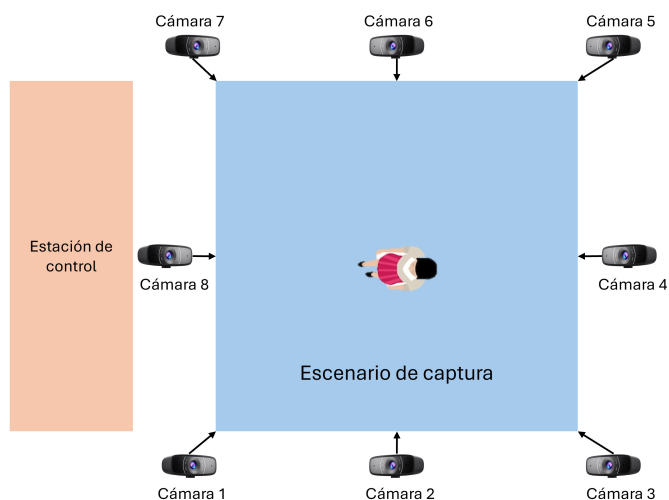


Figura 4.1: Escenario de captura

procesamiento adicional para dividir el video, lo que causaba pérdida de calidad y generaba dependencia de software externo. Debido a estas limitaciones, esta opción fue descartada.



Figura 4.2: Ejemplo de captura mediante *OBS* [6]

Finalmente, se optó por utilizar un módulo de grabación simultánea disponible en *YOLOv8*. Sin embargo, esta implementación tampoco logró recibir video de las ocho cámaras al mismo tiempo. Para superar esta limitación, la tarea se dividió entre dos equipos de cómputo: uno recibió la señal de las cuatro cámaras ubicadas en

las esquinas de la habitación, y el otro procesó las cámaras situadas en el centro de cada pared. Aunque esta solución resultó funcional, comprometió la sincronización precisa entre todas las cámaras. La Figura 4.3 muestra un ejemplo de captura y la distribución de las cámaras entre los dos equipos.



Figura 4.3: Ejemplo de captura y distribución de cámaras entre equipos

4.2.3. Generación de modelos de pose 2D

Experimento 2: Evaluación comparativa de modelos de estimación de pose 2D

Para la generación de modelos de estimación de pose 2D, se realizó un análisis comparativo entre los modelos más reconocidos en esta tarea. La Tabla 4.2 muestra la comparación entre ellos en términos de partes del cuerpo detectadas, número de puntos clave detectados, compatibilidad con plataformas y lenguajes, y disponibilidad de documentación.

Uno de los principales criterios de evaluación fue la capacidad de los modelos para identificar un mayor número de puntos clave del cuerpo humano, incluyendo escenarios con poses inusuales, oclusiones y fondos complejos. Para las pruebas se utilizó un conjunto de 14 imágenes recolectadas de internet, que incluían movimientos acrobáticos, posturas deportivas y actividades en interiores y exteriores. Adicionalmente, se evaluó el rendimiento en tiempo real mediante una *webcam* a 30 *fps*.

Tabla 4.2: Comparativa de modelos de estimación de pose humana

Modelo	Partes detectadas	Número de puntos	Plataformas y lenguajes	Documentación
AlphaPose	Cuerpo, manos, cara	136	Linux, Ubuntu, Python	Poca disponible
OpenPose	Cuerpo, pies, manos, cara	135	Windows, Ubuntu, MacOS, Python, C++	Mucha disponible
PoseNet	Cuerpo	17	Windows, Ubuntu, MacOS, Web, Python, Java	Poca disponible
MoveNet	Cuerpo	17	Windows, Ubuntu, MacOS, Web, Python, Java	Disponible
MediaPipe	Cuerpo, pies, manos, cara	543	Windows, Ubuntu, MacOS, Python, C++, Web	Mucha disponible

El modelo *AlphaPose* fue considerado inicialmente, dado que sus autores afirman mejoras respecto a *OpenPose* en cuanto a precisión. Sin embargo, en la versión más reciente (*v0.6.0*, julio de 2022) dependía de paquetes desactualizados e incompatibles con el entorno actual, lo cual dificultó su implementación y motivó su descarte.

OpenPose mostró un rendimiento aceptable en imágenes estáticas con poses básicas (véase Figura 4.4 a), pero presentó fallos en la detección de puntos clave en escenarios complejos (Figura 4.4 b), omitiendo algunas articulaciones o incluso todas. Adicionalmente en las pruebas de funcionamiento en tiempo real, el modelo evidenció una tasa de detección inconsistente en movimientos rápidos.

En contraste, *MediaPipe Pose* mostró ser el modelo más robusto y versátil. En las pruebas con imágenes estáticas, logró detectar correctamente las articulaciones del cuerpo en aproximadamente el 70 % de las imágenes; en el 30 % restante, localizó parcialmente las articulaciones (véase 4.4 c). Incluso en los casos más complejos, aunque algunos puntos fueron mal posicionados, el modelo realizó una detección. Por otro lado, en tiempo real, ofreció una experiencia fluida y estable, adecuada para aplicaciones que requieren procesamiento en vivo.



Figura 4.4: Ejemplos estimación de pose

En conclusión, debido a su amplia cobertura de puntos del cuerpo, facilidad de integración multiplataforma, buena documentación y rendimiento superior en condiciones reales de uso, se seleccionó el modelo *MediaPipe Pose* para la implementación final de este proyecto.

4.2.4. Fusión de modelos de pose 2D a modelos de pose 3D

Para transformar los modelos 2D a modelos 3D, inicialmente se evaluó el método *Free multi-view rEconstruXion (FLEX)* [66], el cual no requiere calibración previa al no depender de los parámetros intrínsecos de la cámara. No obstante, fue descartado por su dependencia de redes neuronales y por dificultades técnicas en su implementación. Por lo anterior y con el objetivo de mantener un enfoque geométrico, se decidió emplear calibración de cámaras para permitir la reconstrucción tridimensional a partir de correspondencias 2D.

Calibración

Se compararon dos técnicas de calibración mediante cuatro experimentos:

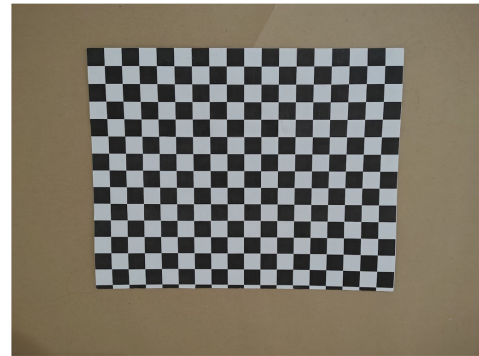
Experimento 3: Calibración manual con patrón 3D

Se utilizó un patrón tridimensional (ver Figura 4.5 a) compuesto por dos hojas perpendiculares de 40×30 cm, cada una con 20 cuadros de ancho por 15 cuadros de alto, de 2 cm cada uno. El procedimiento realizado fue:

- Captura de una fotografía del patrón por cámara.
- Selección manual de puntos correspondientes a intersecciones del patrón.
- Generación de correspondencias 3D (X_i) y 2D (x_i).
- Estimación de la matriz de proyección con *RANSAC* a partir de subconjuntos aleatorios de seis puntos.



a) Patrón de calibración 3D



b) Patrón de calibración 2D

Figura 4.5: Patrones de calibración

Aunque los errores de reproyección para este experimento fueron bajos (ver Tabla 4.3 columna “calibración manual”), la triangulación presentó errores considerables (los detalles se presentan en el experimento de triangulación).

Experimentos 4–6: Autocalibración con patrones 2D

Para estos experimentos se empleó un patrón de ajedrez con un tamaño de 72×56 cm, conformado por 18×14 cuadros de 4 cm (ver Figura 4.5 b). Cada experimento usó un conjunto distinto de imágenes:

- **Experimento 4:** 30 imágenes por cámara a nivel del suelo.
- **Experimento 5:** 90 imágenes con el patrón en distintas alturas, inclinaciones y posiciones.
- **Experimento 6:** 65 imágenes con el patrón ubicado en la zona que normalmente ocuparía una persona.

La Tabla 4.3 presenta los errores de reproyección obtenidos para los experimentos realizados.

Tabla 4.3: Errores de reproyección

Cámara	Calibración Manual	Autocalibración 1	Autocalibración 2	Autocalibración 3
1	2.9058	0.2443	0.2792	0.0957
3	0.5365	0.2784	0.2114	0.1292
5	1.6362	0.3477	0.3406	0.1024
7	1.7109	0.3185	0.2301	0.0925
Promedio	1.6973	0.2977	0.2653	0.1044

Triangulación

La triangulación permite generar modelos de pose 3D a partir de coordenadas 2D y matrices de proyección. Se compararon tres métodos de triangulación considerando la linealidad del método, en enfoque de resolución, la precisión y su velocidad (ver Tabla 4.4).

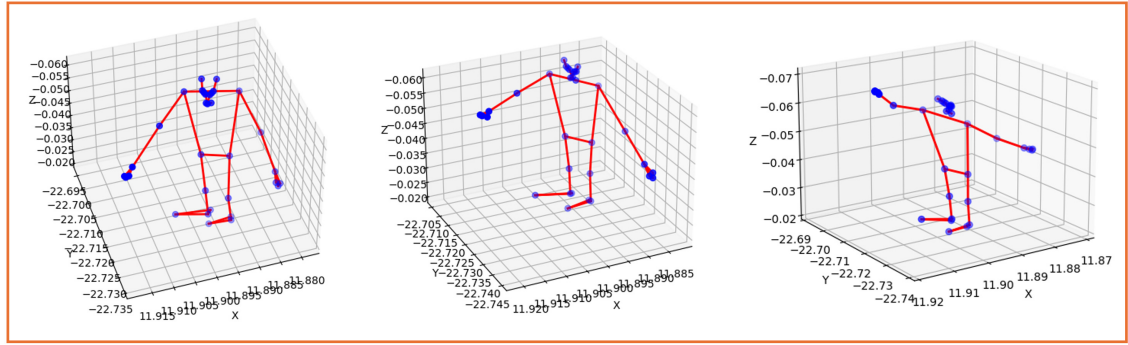
Tabla 4.4: Comparativa de métodos de triangulación

Método	Tipo	Enfoque	Precisión	Velocidad
SVD (Lineal)	Lineal	Sistema $AX = 0$	Media	Alta
Iterativa	No lineal	Punto 3D (error de reproyección)	Alta	Media
Bundle Adjustment	No lineal	Cámaras + puntos 3D	Muy alta	Baja

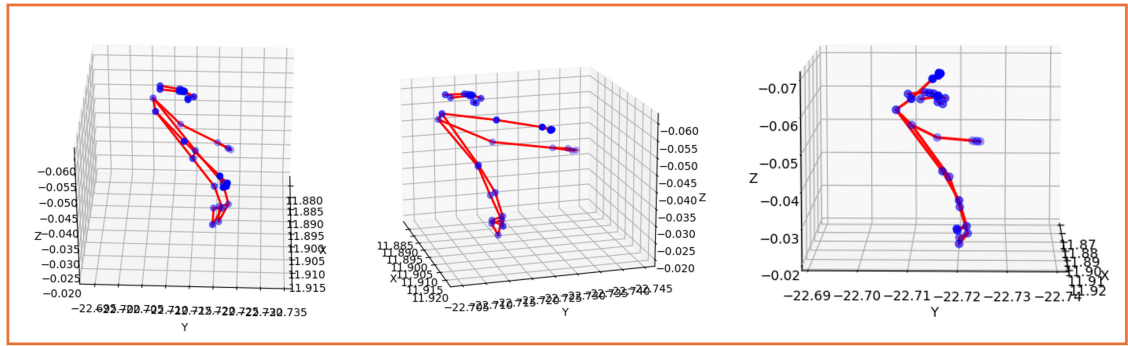
Experimento 7: Comparación de métodos de triangulación

Inicialmente se realizaron pruebas de triangulación *Singular Value Decomposition* (SVD) con las matrices de proyección resultantes de la calibración manual. Aunque la reconstrucción frontal fue aceptable como se muestra en la Figura 4.6 a, la estimación de profundidad fue incorrecta (ver Figura 4.6 b), por lo cual este enfoque fue descartado y se evaluó el resto de métodos de triangulación y calibración.

Para los experimentos posteriores se seleccionó el conjunto de matrices de proyección generadas mediante el método de autocalibración perteneciente al experimento 6. Se aplicaron los tres métodos de triangulación, los errores obtenidos se listan a continuación.



a) Vistas frontales de la triangulación



b) Vistas laterales de la triangulación

Figura 4.6: Ejemplos de triangulación SVD y calibración manual

- SVD: 1424.25 pixeles
- Iterativo: 424.23 pixeles
- *Bundle Adjustment*: 1012.25 pixeles

Debido a los errores elevados, se realizó como medida de control una prueba con el *dataset Fit3D*, con el método de triangulación iterativo, ya que fue el que mostró el menor error en el *dataset* propio, como resultado de esta prueba se obtuvo un error de 1257.54 pixeles.

Adicional a los resultados cuantitativos, se realizó un análisis cualitativo de los resultados de la triangulación, en la Figura 4.7, incisos a), b) y c), se presenta lo obtenido (para el mismo *frame*) para cada uno de los tres métodos empleados (*SVD*, Iterativo y *Bundle Adjustment*) con una vista frontal (parte superior de la imagen),

mientras que en el inciso d) se muestra un ejemplo de triangulación iterativa para el dataset *Fit3D*.

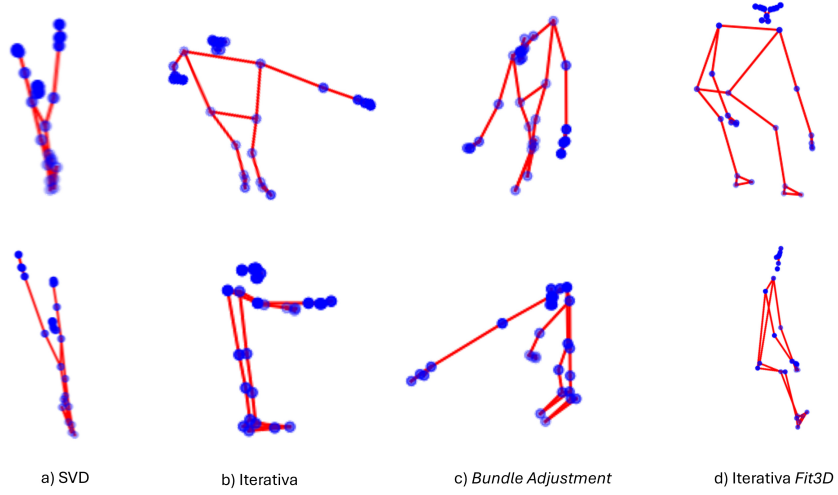


Figura 4.7: Resultados de triangulación para cada método

Como conclusión el método iterativo mantiene un balance entre precisión y velocidad mientras ofrece el mejor rendimiento tanto cuantitativo como cualitativo. Su principal desventaja es que solo permite triangulación entre dos cámaras, por lo cual se seleccionaron las cámaras 1 y 7 como base para generar el modelo 3D.

4.2.5. Generación de modelos del cuerpo humano en 3D

Experimento 8: Construcción y validación del modelo del cuerpo *SMPL*

La generación del modelo 3D del cuerpo humano se llevó a cabo mediante uno de los modelos paramétricos más utilizados en el estado del arte: el modelo *SMPL*. Este modelo requiere dos vectores de parámetros: uno de pose y otro de forma.

Parámetros de pose:

El vector de parámetros de pose está compuesto por 24 articulaciones, cada una representada mediante una rotación en formato *axis-angle* (3 valores). Por tanto, el vector de pose θ es de la forma: $\theta = [\theta_0, \theta_1, \dots, \theta_{71}]$ donde cada tripleta $(\theta_0, \theta_1, \theta_2)$

representa la rotación de una articulación en el espacio 3D.

Dado que el modelo de articulaciones de *SMPL* y el de *MediaPipe Pose* difieren, fue necesario establecer una correspondencia entre ambos. La Tabla 4.5 muestra las equivalencias entre los puntos clave. Además se pueden visualizar dichas correspondencias en la Figura 4.8.

Tabla 4.5: Correspondencia generada entre los modelos *SMPL* y *MediaPipe Pose*

SMPL	MediaPipe	SMPL	MediaPipe	SMPL	MediaPipe	SMPL	MediaPipe
0	-	6	-	12	-	18	13
1	23	7	27	13	-	19	14
2	24	8	28	14	-	20	15
3	-	9	-	15	-	21	16
4	25	10	31	16	11	22	19
5	26	11	32	17	12	23	20

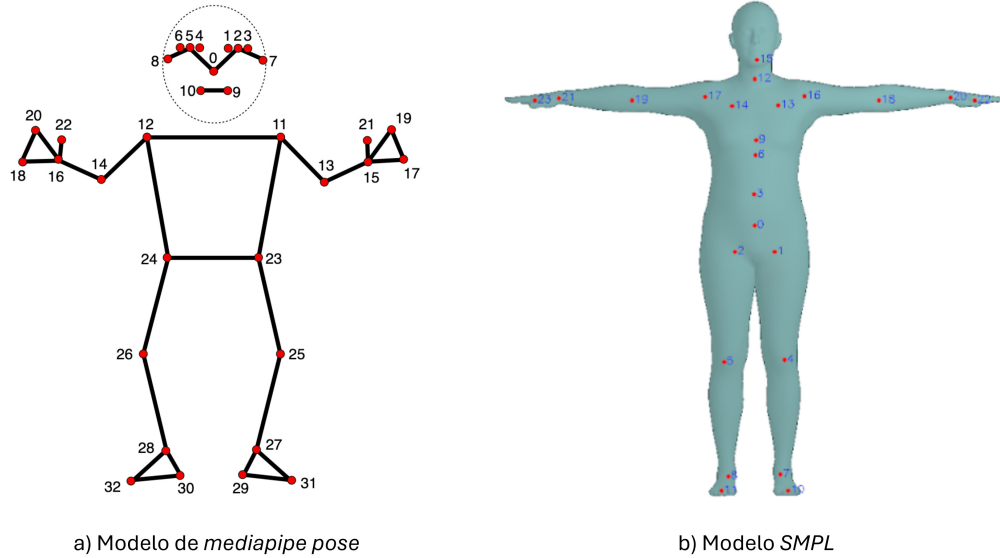


Figura 4.8: Modelo *mediapipe pose* y *SMPL*

Cálculo de vectores de dirección y rotaciones

El vector de parámetros de forma requiere las rotaciones en formato *axis-angle* para cada una de las articulaciones a mover, por lo que es necesario un proceso para

obtener dichos ángulos. El proceso realizado para obtenerlos se describe a continuación.

Primero, se generan los vectores de dirección para cada articulación. Un vector unitario (de tres componentes) que indica una dirección en el espacio 3D, independientemente de su posición o longitud. Por ejemplo, el vector que va de la cadera a la rodilla indica la dirección del muslo.

Para la generación de un vector de dirección se realiza el siguiente procedimiento, dados dos puntos $A = (x_1, y_1, z_1)$ y $B = (x_2, y_2, z_2)$ el vector de dirección $\vec{v}$ se calcula como:

$$\vec{v} = \frac{(x_2 - x_1, y_2 - y_1, z_2 - z_1)}{\|x_2 - x_1, y_2 - y_1, z_2 - z_1\|}$$

Luego, se calcula:

El eje de rotación:

$$\vec{axis} = \vec{v}_{pose_{neutral}} \times \vec{v}_{pose_{arbitraria}}$$

El ángulo de rotación:

$$\theta = \cos^{-1}(\vec{v}_{pose_{neutral}} \cdot \vec{v}_{pose_{arbitraria}})$$

Y el *axis-angle*:

$$\vec{v} = \vec{axis} \times \theta$$

El resultado de realizar estos cálculos permite efectuar las rotaciones de las articulaciones del modelo *SMPL*. En la Figura 4.9 se muestra un ejemplo de cómo los vectores en formato *axis-angle* permiten rotar dichas articulaciones. Se considera el ejemplo descrito en la sección 3, en el cual el antebrazo izquierdo, en la postura *T-pose*, tiene el valor $[0,158207, -0,361927, 0,004388]$, mientras que en un *frame* arbitrario toma el valor $[0,090904, -0,0187249, -0,0336379]$. El *axis-angle* generado es $[1,287970, 0,537176, 0,318162]$, lo que permite transformar el antebrazo en el modelo 3D de la posición en a), a la posición en b).

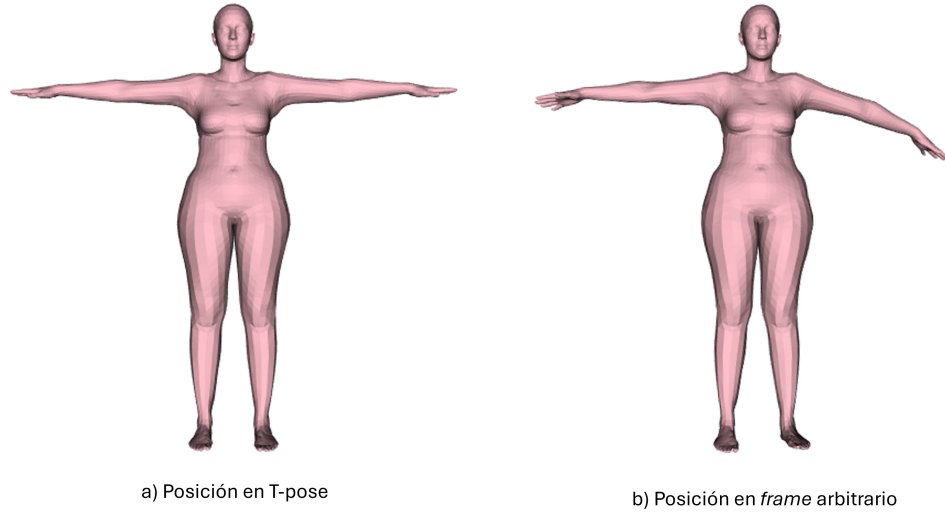


Figura 4.9: Ejemplo de transformación con *axis-angle*

Para este procedimiento se realizaron pruebas de dos formas. La primera consistió en la implementación de la función *align_vectors* de la librería *scipy*; sin embargo, este método devolvía valores de rotación desproporcionadamente grandes. Por esta razón, se optó por realizar el proceso manualmente, utilizando los cálculos mencionados anteriormente. En la Figura 4.10, se muestran los cálculos del ángulo de rotación obtenidos tanto manualmente como mediante la función *align_vectors*, en los segmentos corporales de tres *frames* distintos.

Para realizar una comparación completa, se seleccionó un video y se calculó la diferencia angular promedio de las rotaciones calculadas por ambos métodos, por cada segmento del cuerpo. En la Tabla 4.6 se presentan los resultados obtenidos.

Finalmente, se construyó el vector completo de pose respetando el orden definido por *SMPL*, dadas las correspondencias establecidas anteriormente. Aquellas articulaciones sin correspondencia conocida se mantienen con valores en cero.

Como ejemplo, en la Figura 4.11 se muestra el contenido de tres vectores de pose. Para una interpretación más sencilla, se segmenta en el formato pose $i : [x, y, z]$, el cual representa la rotación del ángulo i .

<pre> ===== FRAME 3 ===== --- Ángulos (manuales) --- muslo_izquierdo: 6.44° muslo_derecho: 2.03° torso: 2.02° pantorrilla_derecha: 2.52° pie_izquierdo: 6.57° pie_derecho: 2.37° brazo_izquierdo: 12.08° brazo_derecho: 36.78° antebrazo_izquierdo: 15.91° antebrazo_derecho: 49.09° mano_izquierda: 15.60° mano_derecha: 22.35° --- Ángulos (align_vectors) --- muslo_izquierdo: 89.17° muslo_derecho: 7.42° torso: 167.28° pantorrilla_izquierda: 172.56° pantorrilla_derecha: 108.42° pie_izquierdo: 140.65° pie_derecho: 6.16° brazo_izquierdo: 19.85° brazo_derecho: 43.78° antebrazo_izquierdo: 56.99° antebrazo_derecho: 165.34° mano_izquierda: 24.52° mano_derecha: 76.65° </pre>	<pre> ===== FRAME 25 ===== --- Ángulos (manuales) --- muslo_izquierdo: 5.47° muslo_derecho: 3.94° torso: 5.43° pantorrilla_derecha: 3.10° pie_izquierdo: 6.33° pie_derecho: 2.78° brazo_izquierdo: 13.17° brazo_derecho: 38.21° antebrazo_izquierdo: 15.00° antebrazo_derecho: 45.91° mano_izquierda: 16.94° mano_derecha: 41.17° --- Ángulos (align_vectors) --- muslo_izquierdo: 90.64° muslo_derecho: 5.30° torso: 38.35° pantorrilla_izquierda: 172.68° pantorrilla_derecha: 90.11° pie_izquierdo: 179.43° pie_derecho: 2.90° brazo_izquierdo: 38.62° brazo_derecho: 44.45° antebrazo_izquierdo: 61.56° antebrazo_derecho: 47.84° mano_izquierda: 153.34° mano_derecha: 56.53° </pre>	<pre> ===== FRAME 68 ===== --- Ángulos (manuales) --- muslo_izquierdo: 4.12° muslo_derecho: 6.33° torso: 6.94° pantorrilla_izquierda: 1.07° pantorrilla_derecha: 4.52° pie_izquierdo: 4.10° pie_derecho: 2.94° brazo_izquierdo: 49.06° brazo_derecho: 45.84° antebrazo_izquierdo: 29.44° antebrazo_derecho: 32.60° mano_izquierda: 30.67° mano_derecha: 23.83° --- Ángulos (align_vectors) --- muslo_izquierdo: 89.39° muslo_derecho: 6.33° torso: 62.13° pantorrilla_izquierda: 90.21° pantorrilla_derecha: 179.85° pie_izquierdo: 8.53° pie_derecho: 176.95° brazo_izquierdo: 169.14° brazo_derecho: 62.48° antebrazo_izquierdo: 35.74° antebrazo_derecho: 52.07° mano_izquierda: 52.96° mano_derecha: 41.61° </pre>
--	---	--

Figura 4.10: Ejemplo de cálculo de ángulos

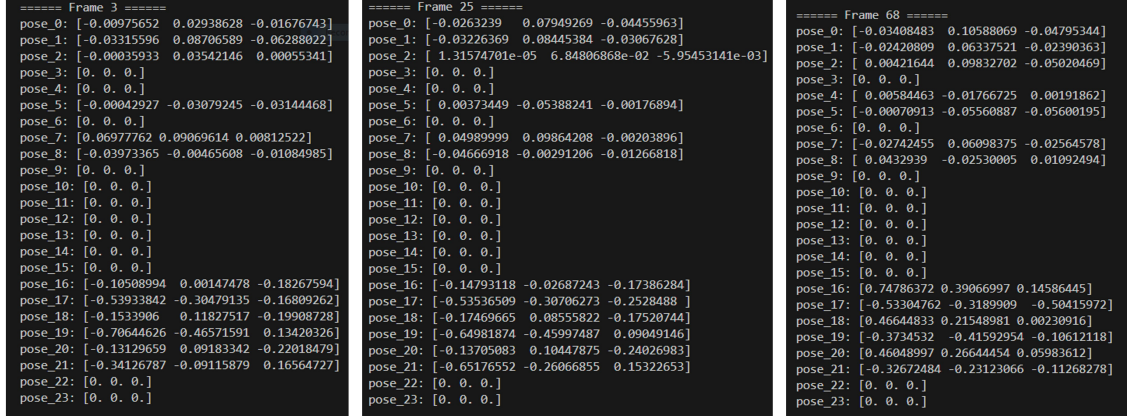
Tabla 4.6: Diferencia angular promedio por segmento

Segmento	Diferencia Promedio	
	Lado izquierdo	Lado derecho
Mano	58.91°	59.54°
Antebrazo	46.85°	54.27°
Brazo	66.29°	55.56°
Muslo	85.60°	75.34°
Pantorrilla	89.00°	81.49°
Pie	87.01°	92.07°
Torso	76.67°	

Es importante destacar que el modelo **SMPL** utiliza un sistema de referencia distinto con la correspondencia:

$$(x, y, z) \rightarrow (x, z, -y)$$

lo cual es importante tener en cuenta al momento de construir los vectores de rotación del modelo.



```

===== Frame 3 =====
pose_0: [-0.00975652  0.02938628 -0.01676743]
pose_1: [-0.03315596  0.08706589 -0.06288022]
pose_2: [-0.00035933  0.03542146  0.00055341]
pose_3: [0. 0. 0.]
pose_4: [0. 0. 0.]
pose_5: [-0.00042927 -0.03079245 -0.03144468]
pose_6: [0. 0. 0.]
pose_7: [0.06977762  0.09069614  0.00812522]
pose_8: [-0.03973365 -0.00465608 -0.01084985]
pose_9: [0. 0. 0.]
pose_10: [0. 0. 0.]
pose_11: [0. 0. 0.]
pose_12: [0. 0. 0.]
pose_13: [0. 0. 0.]
pose_14: [0. 0. 0.]
pose_15: [0. 0. 0.]
pose_16: [-0.10508994  0.00147478 -0.18267594]
pose_17: [-0.53933842 -0.30479135 -0.16809262]
pose_18: [-0.1533906  0.11827517 -0.19908728]
pose_19: [-0.70644626 -0.46571591  0.13420326]
pose_20: [-0.13129659  0.09183342 -0.22018479]
pose_21: [-0.34126787 -0.09115879  0.16564727]
pose_22: [0. 0. 0.]
pose_23: [0. 0. 0.]

===== Frame 25 =====
pose_0: [-0.0263239  0.07949269 -0.04455963]
pose_1: [-0.03226369  0.08445384 -0.03067628]
pose_2: [ 1.31574701e-05  6.84806868e-02 -5.95453141e-03]
pose_3: [0. 0. 0.]
pose_4: [0. 0. 0.]
pose_5: [ 0.00373449 -0.05388241 -0.00176894]
pose_6: [0. 0. 0.]
pose_7: [ 0.04989999  0.09864208 -0.00203896]
pose_8: [-0.04666918 -0.00291206 -0.01266818]
pose_9: [0. 0. 0.]
pose_10: [0. 0. 0.]
pose_11: [0. 0. 0.]
pose_12: [0. 0. 0.]
pose_13: [0. 0. 0.]
pose_14: [0. 0. 0.]
pose_15: [0. 0. 0.]
pose_16: [-0.14793118 -0.02687243 -0.17386284]
pose_17: [-0.53536509 -0.30706273 -0.2528488 ]
pose_18: [-0.17469665  0.08555822 -0.17520744]
pose_19: [-0.64981874 -0.45997487  0.09049146]
pose_20: [-0.13705083  0.10447875 -0.24026983]
pose_21: [-0.65176552 -0.26066855  0.15322653]
pose_22: [0. 0. 0.]
pose_23: [0. 0. 0.]

===== Frame 68 =====
pose_0: [-0.03408483  0.10588069 -0.04795344]
pose_1: [-0.02420809  0.06337521 -0.02390363]
pose_2: [ 0.00421644  0.09832702 -0.05020469]
pose_3: [0. 0. 0.]
pose_4: [ 0.00584463 -0.01766725  0.00191862]
pose_5: [-0.00070913 -0.05560887 -0.05600195]
pose_6: [0. 0. 0.]
pose_7: [-0.02742455  0.06098375 -0.02564578]
pose_8: [ 0.0432939  -0.02530005  0.01092494]
pose_9: [0. 0. 0.]
pose_10: [0. 0. 0.]
pose_11: [0. 0. 0.]
pose_12: [0. 0. 0.]
pose_13: [0. 0. 0.]
pose_14: [0. 0. 0.]
pose_15: [0. 0. 0.]
pose_16: [0.74786372  0.39066997  0.14586445]
pose_17: [-0.53304762 -0.3189909 -0.50415972]
pose_18: [0.46644833  0.21548981  0.00230916]
pose_19: [-0.3734532 -0.41592954 -0.10612118]
pose_20: [0.46048997  0.26644454  0.05983612]
pose_21: [-0.32672484 -0.23123066 -0.11268278]
pose_22: [0. 0. 0.]
pose_23: [0. 0. 0.]

```

Figura 4.11: Ejemplo de contenido vectores de pose

Adicionalmente el modelo usa un sistema jerárquico para la rotación de las articulaciones (este tipo de sistemas es comúnmente utilizado en robótica para el movimiento de robots articulados), en esta jerarquía, o cadena cinemática, cada articulación hija hereda las transformaciones de su articulación padre, por lo que el cálculo de rotaciones debe realizarse en el sistema de referencia local correspondiente. En la Figura 4.12 se muestra cómo funciona el sistema jerárquico, se pueden observar líneas de distintos colores que indican la jerarquía de articulaciones, siguiendo la numeración de las mismas.

Para entenderlo de una forma más simple esta jerarquía se explica a continuación cómo se mueve el tobillo izquierdo de una persona. Para mover el tobillo izquierdo (articulación 7), es necesario considerar toda la jerarquía de articulaciones que le antecede: la articulación 4 (rodilla izquierda), la cual depende de la articulación 1 (cadera izquierda), la cual a su vez depende de la articulación 0 (la pelvis), suponiendo que las rotaciones locales para cada articulación son las siguientes:

- Articulación 0 (pelvis): $[10^\circ, 0^\circ, 0^\circ]$
- Articulación 1 (cadera izquierda): $[5^\circ, 2^\circ, 0^\circ]$

- Articulación 4 (rodilla izquierda): $[3^\circ, 1^\circ, -1^\circ]$
- Articulación 7 (tobillo izquierdo): $[2^\circ, 0^\circ, -2^\circ]$

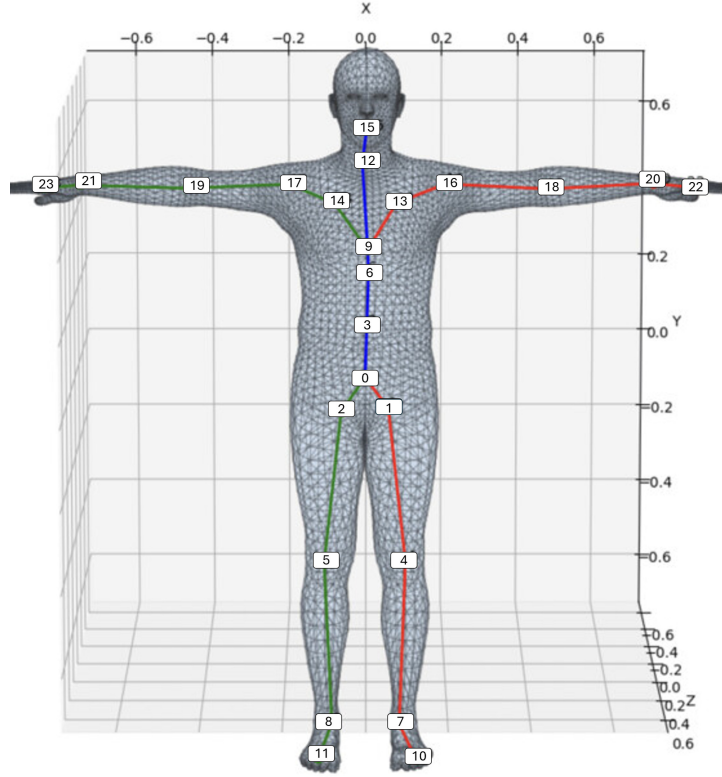


Figura 4.12: Jerarquía de movimiento del modelo *SMPL*(modificada de [7])

Para obtener la rotación global del tobillo izquierdo, se deben componer estas rotaciones acumulativamente a lo largo de la jerarquía. Si se asume una composición aditiva simple para ilustración (aunque en la práctica se usan multiplicaciones de matrices, es decir antes de la conversión a *axis-angle*), el resultado sería:

$$[10 + 5 + 3 + 2, 0 + 2 + 1 + 0, 0 + 0 - 1 - 2] = [20^\circ, 3^\circ, -3^\circ]$$

Este vector representa la rotación global aproximada del tobillo izquierdo con respecto al sistema de referencia raíz. Por lo que, respetar esta jerarquía fue necesario al momento de generar los vectores de rotación de la pose en la que se encuentra la persona.

Parámetros de forma

El vector de parámetros de forma contiene medidas del cuerpo como altura, largo de extremidades, ancho de hombros y caderas, entre otras. Dado que el enfoque busca ser poco invasivo, no se recurre a la toma directa de las medidas, por esto, se utiliza solo una medida real: la distancia entre hombros, la cual se emplea como referencia para encontrar la relación entre el modelo y la medida real. Las medidas se calculan como sigue:

1. Se leen las coordenadas 3D resultantes de la triangulación.
2. Para el primer *frame* se calcula la distancia euclidiana entre los puntos relevantes los cuales se listan en la Tabla 4.7.
3. Se obtiene una correspondencia de medidas a partir de la distancia obtenida entre los hombros en el modelo de coordenadas 3D generadas y la medida real.
4. Con la correspondencia obtenida, se escalan los segmentos necesarios.

Tabla 4.7: Puntos de inicio y fin de *MediaPipe* de los segmentos corporales

Segmento	Punto de inicio	Punto final
Largo de pierna	23, 24	29, 30
Largo de brazo	11, 12	19, 20
Ancho de hombros	12	11
Ancho de caderas	24	23
Altura talón-hombro	29, 30	11, 12

En la Tabla 4.8, se muestra una comparativa entre las medidas reales y las estimadas con el sistema, así como la diferencia entre las mismas, se puede observar que no se tiene una diferencia constante por lo que realizar un ajuste para la disminución del error no es viable. Dado que el modelo de *MediaPipe Pose* no cuenta con puntos de referencia en la cabeza, por lo que la altura total requirió de una estimación, para esto se utilizó la proporción anatómica clásica (de Leonardo da Vinci), donde la altura corresponde a 8 veces el tamaño de la cabeza, para interpolar la altura se asumió que, del talón al hombro hay 7 cabezas.

Tabla 4.8: Comparativa de medidas reales y estimadas

Hombros			Altura		
Real	Estimado	Diferencia	Real	Estimado	Diferencia
34 cm	40.2 cm	6.2 cm	167 cm	74.3 cm	92.7 cm
41 cm	61.2 cm	20.2 cm	176 cm	124 cm	52 cm
48 cm	70.5 cm	22.5 cm	172 cm	122.7 cm	49.3 cm
39 cm	42.8 cm	3.8 cm	178 cm	85.4 cm	92.6 cm
36 cm	45.5 cm	9.5 cm	160 cm	87.5 cm	72.5 cm
37 cm	51.8 cm	14.8 cm	164 cm	91.8 cm	72.2 cm
37 cm	47.6 cm	10.6 cm	167 cm	90.6 cm	76.4 cm
37 cm	30.1 cm	6.9 cm	162 cm	48.7 cm	113.3 cm
40 cm	41.9 cm	1.9 cm	168 cm	75.8 cm	92.2 cm
40 cm	48.2 cm	8.2 cm	166 cm	77.5 cm	88.5 cm
36 cm	41.1 cm	5.1 cm	160 cm	76.2 cm	83.8 cm
36 cm	41.4 cm	5.4 cm	148 cm	70 cm	78 cm

Para el vector de forma se requiere también de la correspondencia entre los parámetros de forma calculados y el vector para el modelo *SMPL*, dado que no todos los valores fueron calculados, en la Tabla 4.9 se muestra detalladamente qué controla cada uno de los 10 parámetros y cuáles de ellos fueron calculados. Según los creadores del modelo sí se usan más de 10 valores (hasta 300 en el modelo más extenso), se obtienen más detalles corporales; sin embargo, los 10 mencionados ya explican el 95 % de la variabilidad de la forma del cuerpo humano.

Tabla 4.9: Relación de parámetros de forma *SMPL* y calculados

Parámetro	Controla ...	Se calcula el valor:
β_0	Altura general	Si
β_1	Grosor del pecho	No
β_2	Ancho de hombros	Si
β_3	Longitud de piernas	Si
β_4	Ancho de cadera	Si
β_5	Tamaño de pecho	No
β_6	Tamaño de cabeza	Si
β_7	Longitud de brazos	Si
β_8	Forma de las piernas	No
β_9	Distribución de grasa	No

El procedimiento descrito anteriormente permite generar vectores de parámetros de pose y forma compatibles con el modelo *SMPL* a partir de datos obtenidos con *MediaPipe Pose*, mediante un enfoque geométrico.

Generación del modelo *SMPL*

Experimento 9: Comparación entre generación directa y optimizada del modelo *SMPL*

La generación del modelo *SMPL* se realiza utilizando los dos vectores descritos previamente: el de pose y el de forma. El modelo *SMPL* puede generarse mediante un proceso de optimización (el cual genera modelos a los que se hace referencia como *SMPL_O*) o directamente al utilizar los vectores calculados (denotándose el modelo generado como *SMPL_V*).

Para este proyecto, se optó por el enfoque *SMPL_O*, en línea con el objetivo de emplear un modelo geométrico, sin embargo se realizaron pruebas exploratorias utilizando ambos enfoques para realizar una comparación cualitativa de los modelos generados con ambos enfoques.

En la Figura 4.13 se muestra una comparación visual del *frame* original y los modelos generados con ambos enfoques: *SMPL_O* (b) y *SMPL_V* (c), utilizando el mismo *frame* y vectores de parámetros en ambos casos de una persona en *T-pose*.

- En a), se presenta la imagen original
- En b), se observa el modelo *SMPL_O*, el cual muestra deformaciones las cuales son atribuibles al proceso de ajuste automático.
- En c), se muestra el modelo *SMPL_V*, el cual genera como resultado una estructura corporal más coherente y visualmente precisa (Por motivos prácticos la visualización de las imágenes de los modelos 3D generados se muestran completamente de frente).

Se puede apreciar que el método directo (*SMPL_V*) produce resultados más estables en la *T-pose*, sin las distorsiones introducidas por la optimización automática. Por este motivo, se adoptó como enfoque final.

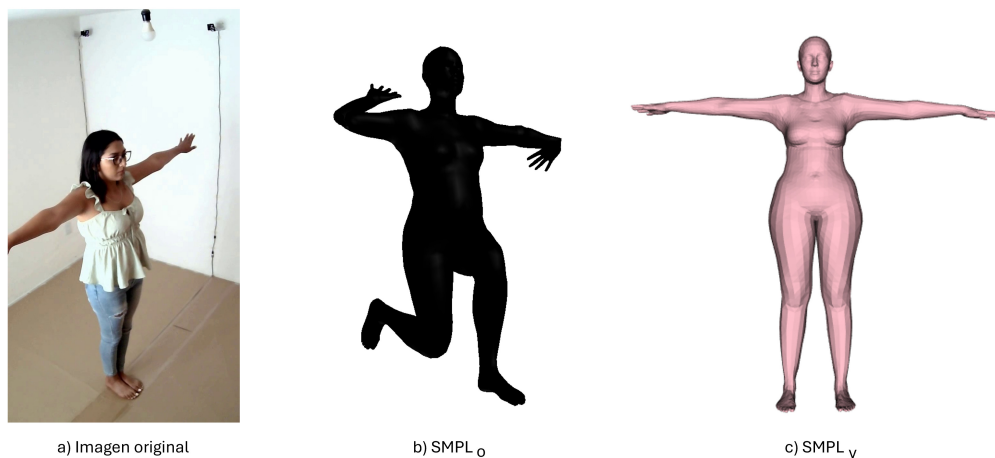


Figura 4.13: Ejemplo de modelos generados en *T-pose*

Es importante agregar que aún con el método directo se observan demasiadas distorsiones en las poses, las cuales son dependientes de la estructura y codificación del modelo *SMPL* lo cual se ve reflejado en el error angular presentado anteriormente y visualmente en la mayoría de los modelos generados como los que se muestran en la Figura 4.14.

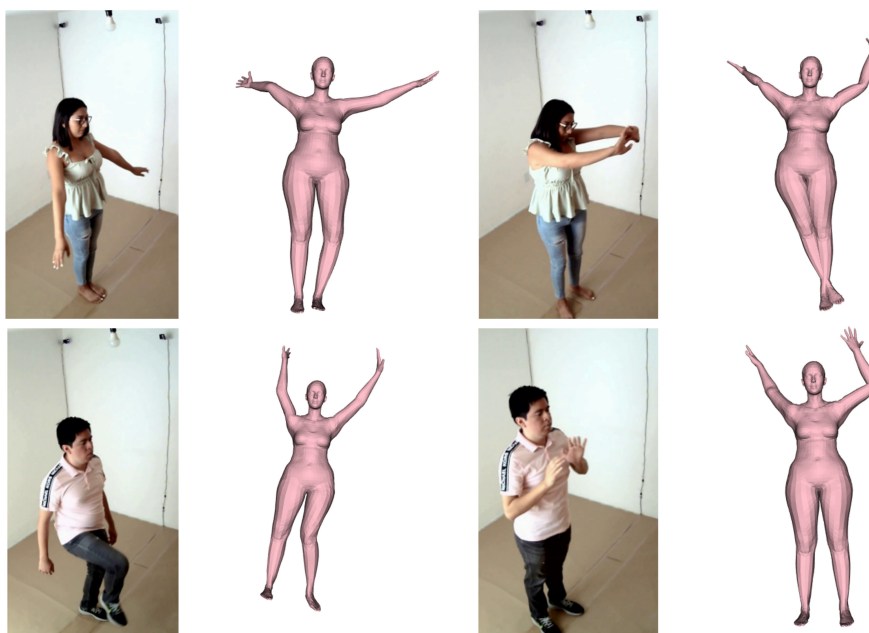


Figura 4.14: Ejemplos de modelos incorrectos

Mientras que en la Figura 4.15 se presentan algunos de los ejemplos relativamente correctos de la generación de modelos de tipo $SMPL_V$. Para fines de esta investigación, dado que el objetivo no se centró en el reconocimiento, se utilizó el modelo $SMPL$ neutro, tanto para hombres como de mujeres, lo cual genera una estructura corporal similar para ambos géneros.

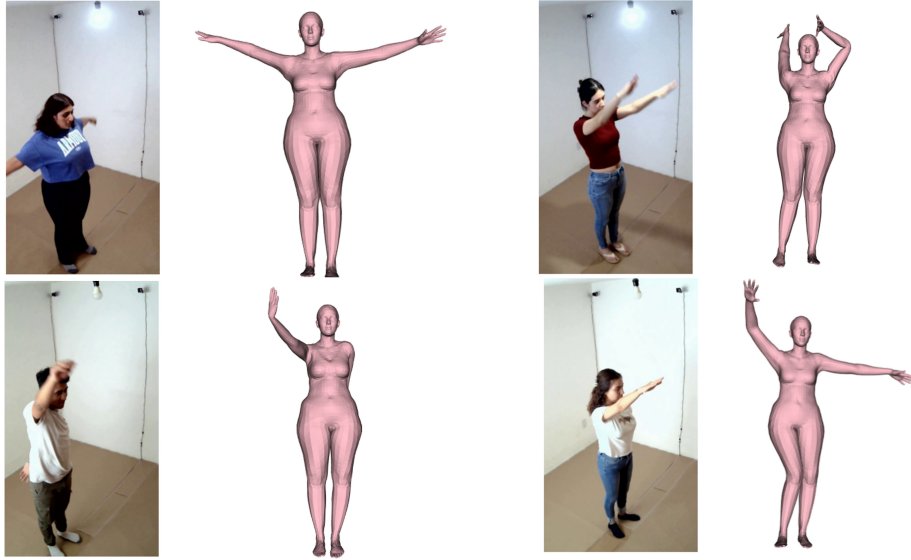


Figura 4.15: Ejemplos de modelos correctos

4.2.6. Generación del *dataset* 3D

Experimento 10: Construcción del *dataset* 3D y evaluación cuantitativa

El *dataset* 3D generado contiene múltiples archivos por cada acción realizada por cada persona. Para cada caso, se produce la siguiente salida:

- Un archivo por vista de coordenadas 2D, detectadas mediante *MediaPipe Pose*, con un muestreo en *frames* $10 \times n$, con n entero positivo.
- Un archivo de coordenadas 3D, calculadas mediante triangulación de los datos de los archivos 2D.

- Una carpeta con modelos 3D en formato `.obj`, con un archivo para triangulación realizada.
- Una carpeta de imágenes en formato `.png`, con una correspondencia 1-a-1 con los modelos `.obj` generados.
- Un video generado a partir de las imágenes `.png`.
- Un archivo de texto con las medidas corporales estimadas por persona. (Para ver un ejemplo del contenido del archivo ver Anexo A Figura A.2).

Evaluación de resultados

Las métricas utilizadas para la evaluación de los resultados son: error de triangulación, error angular y error de longitud. En la Tabla 4.10 se presentan el reporte general de resultados.

Tabla 4.10: Resumen de resultados

Persona	Promedio			Desviación Estándar	
	Error Angular (grados)	Error Longitud (cm)	Error Triangulación (píxeles)	Error Angular (grados)	Error Longitud (cm)
1	12.8991	0.0955	424.2822	21.7677	0.0669
2	11.0026	0.0841	424.2996	17.7546	0.0687
3	10.7971	0.0839	424.2881	17.3646	0.0656
4	9.1821	0.0708	424.2738	14.8844	0.0476
5	9.2899	0.0537	424.2433	13.7571	0.0455
6	10.9040	0.0856	424.3087	17.8653	0.0645
7	9.5095	0.0845	424.2617	15.2552	0.0654
8	9.8658	0.0601	424.2731	14.8252	0.0483
9	9.4230	0.0777	424.2825	15.5477	0.0587
10	10.2873	0.0597	424.2551	16.1284	0.0552
11	10.4055	0.0851	424.2522	17.0184	0.0620
12	11.2899	0.0681	424.2335	18.4949	0.0625
Promedio	10.4047	0.0757	424.2712	16.7220	0.0593

Para cada acción realizada por persona, se calculo la media y la desviación estándar del error angular y del error longitudinal. Posteriormente, para el reporte general de resultados, se obtuvo el promedio de dichos errores por persona. Para un mayor detalle de los errores por acción, se recomienda consultar los informes detallados en el Anexo B de la Tabla B.1 a la Tabla B.12.

Adicionalmente, en las Figuras 4.16 a la 4.27 se presenta un diagramas de tipo *boxplot* que muestra la distribución de errores por segmento y por ejercicio realizado.

De todos los gráficos presentados se puede determinar lo siguiente:

- Los segmentos que presentan errores angulares cercanos a cero en la mayoría de los ejercicios son: muslo derecho, muslo izquierdo, torso, brazo izquierdo y brazo derecho, lo que sugiere que son los segmentos que mantienen una menor movilidad en los ejercicios realizados.
- Los segmentos pantorrilla izquierda, pantorrilla derecha, pie derecho y pie izquierdo muestran errores en un nivel intermedio con una mayor dispersión entre los ejercicios (especialmente en ejercicios como el 10 y 12), lo cual puede estar causado por una mayor complejidad en las rotaciones y por la precisión limitada en la captura de rotaciones en tobillos y rodillas.
- Los segmentos antebrazo izquierdo, antebrazo derecho, mano izquierda y mano derecha presentan los errores angulares más elevados y dispersos. Las cajas de los diagramas muestran alta variabilidad y valores extremos (*outliers*) frecuentes, esto se debe probablemente a la alta movilidad de las articulaciones de codo y muñeca, o a errores en la captura.

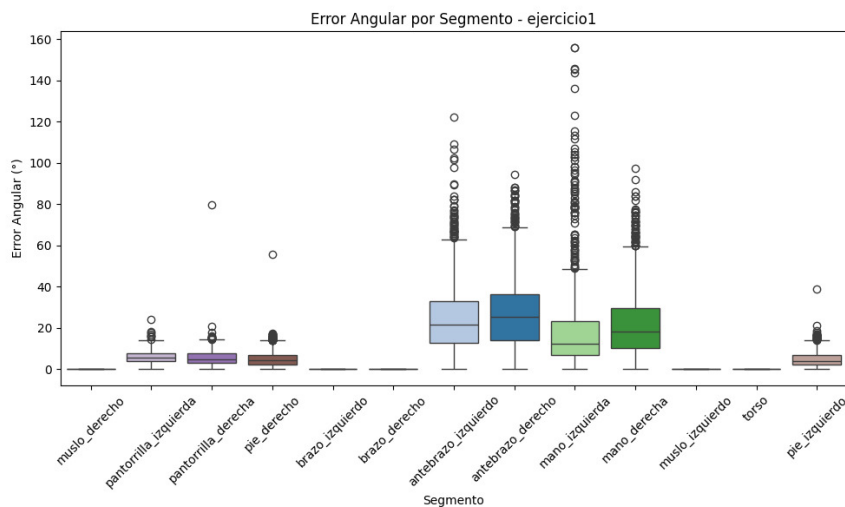


Figura 4.16: Distribución de error angular por segmento del ejercicio 1

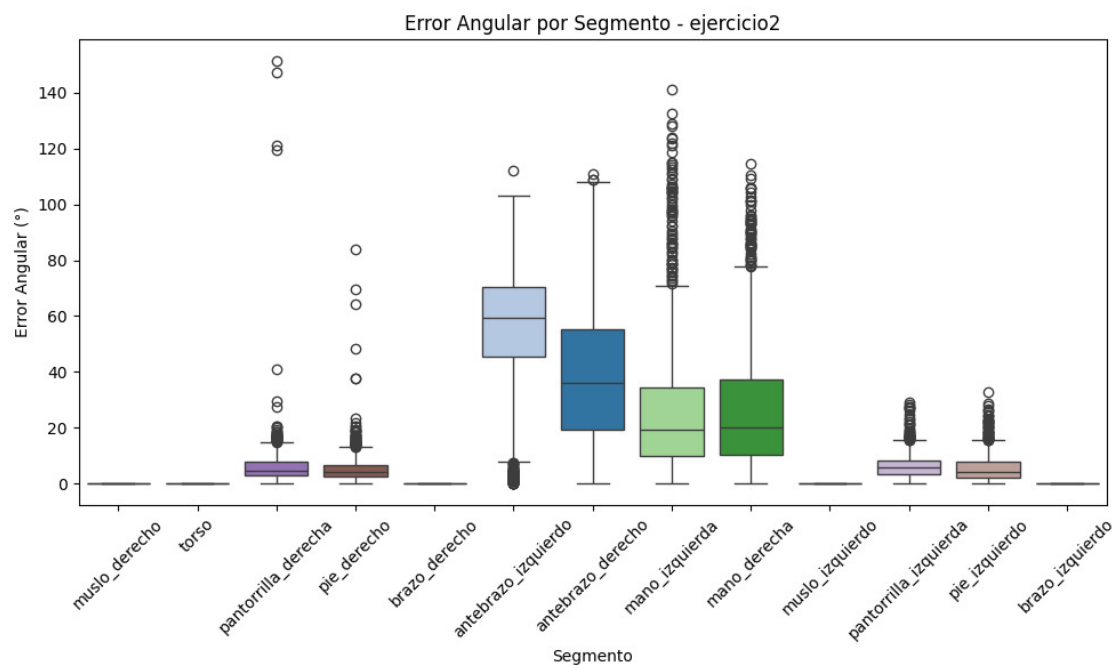


Figura 4.17: Distribución de error angular por segmento del ejercicio 2

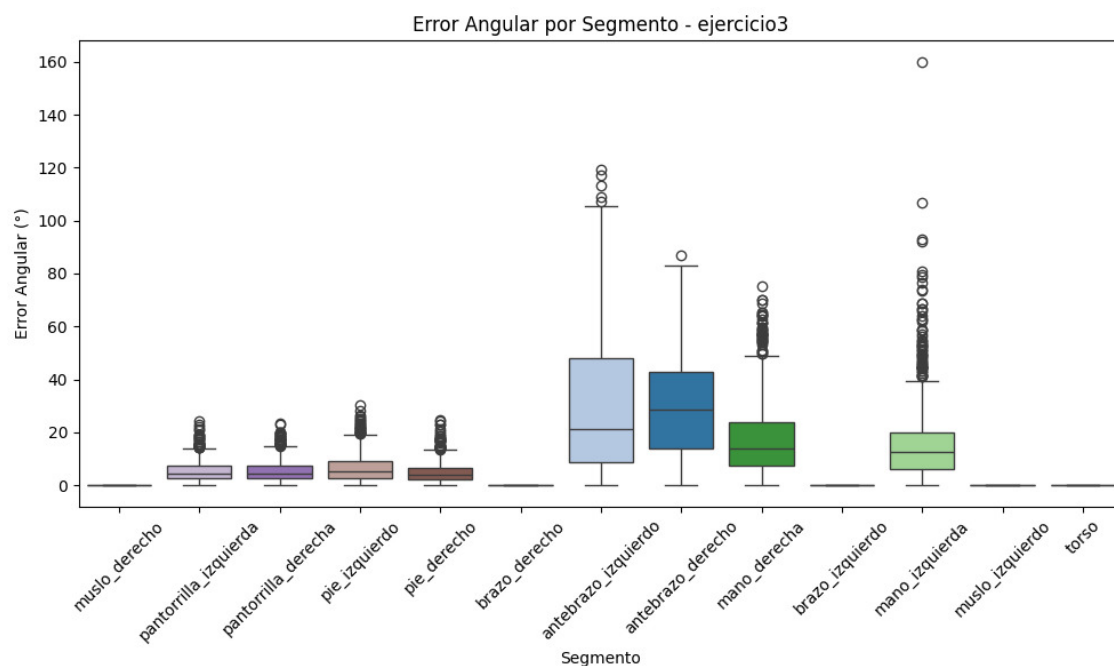


Figura 4.18: Distribución de error angular por segmento del ejercicio 3

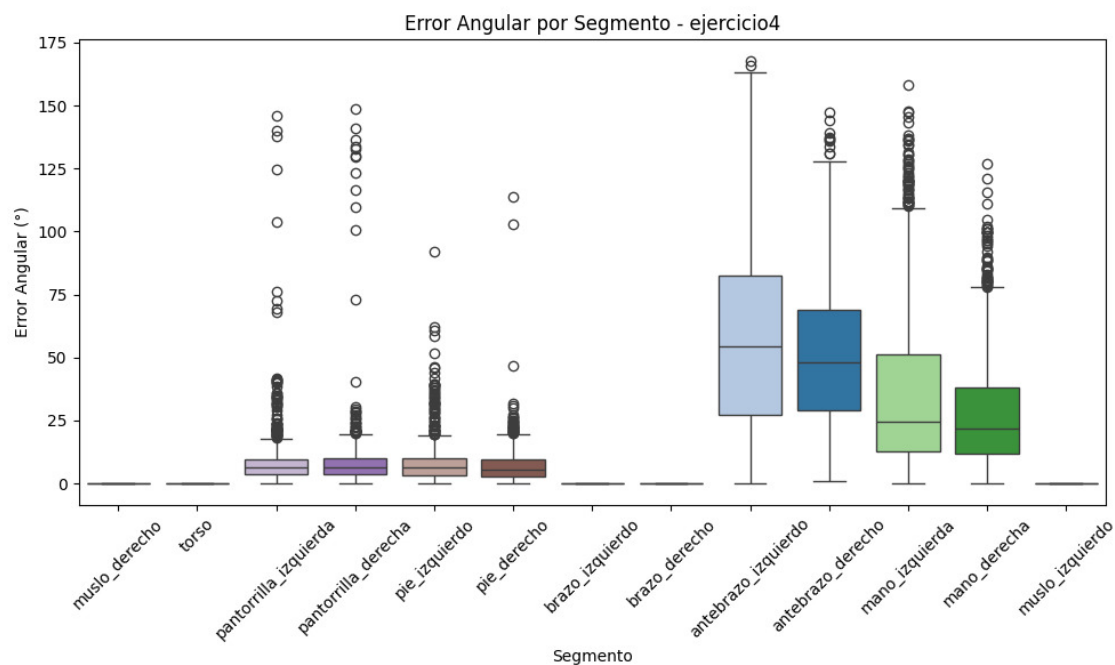


Figura 4.19: Distribución de error angular por segmento del ejercicio 4

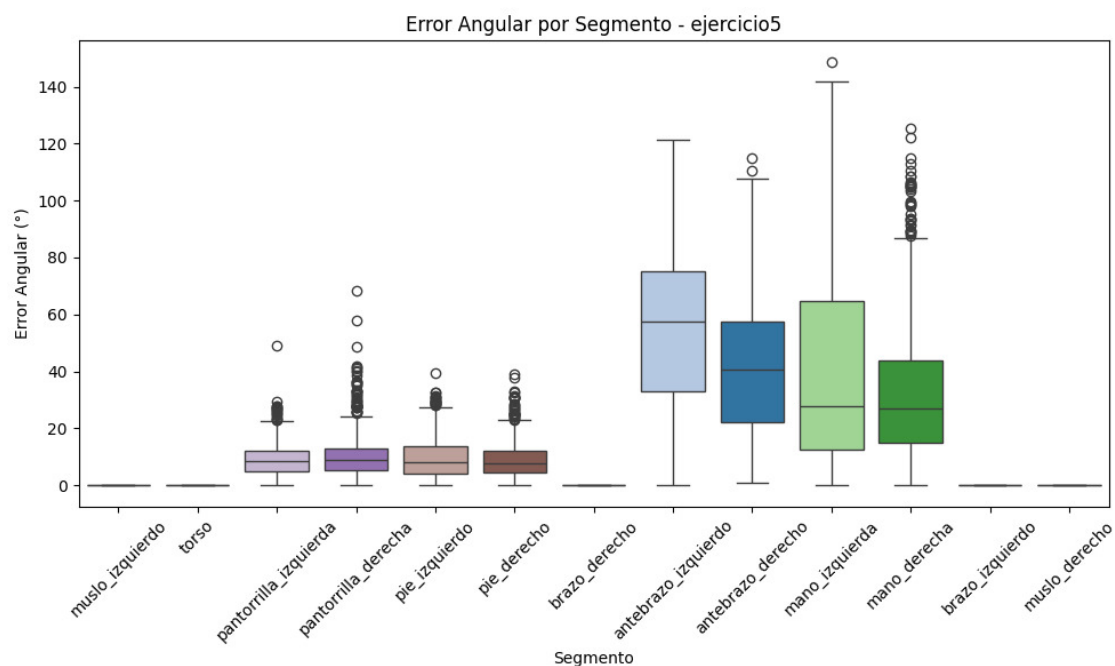


Figura 4.20: Distribución de error angular por segmento del ejercicio 5

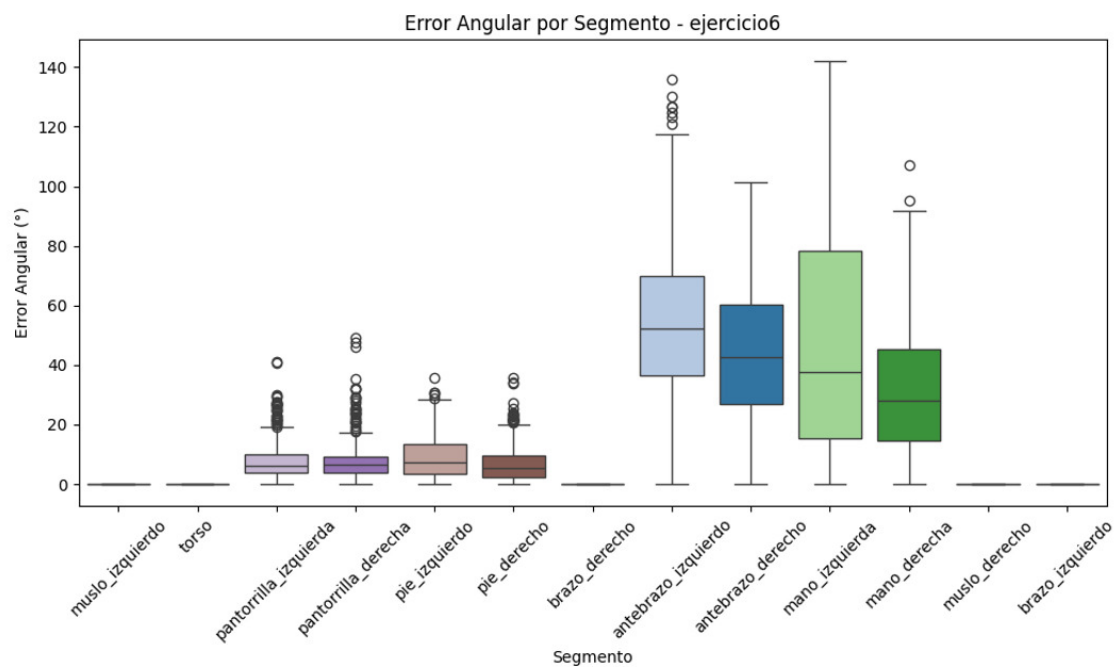


Figura 4.21: Distribución de error angular por segmento del ejercicio 6

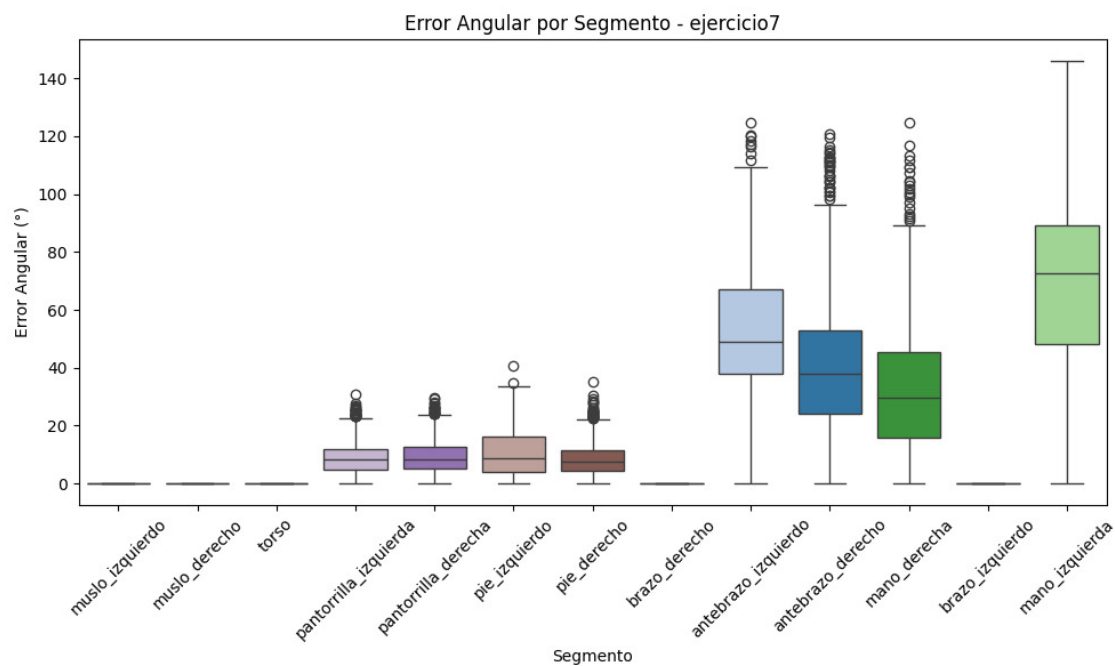


Figura 4.22: Distribución de error angular por segmento del ejercicio 7

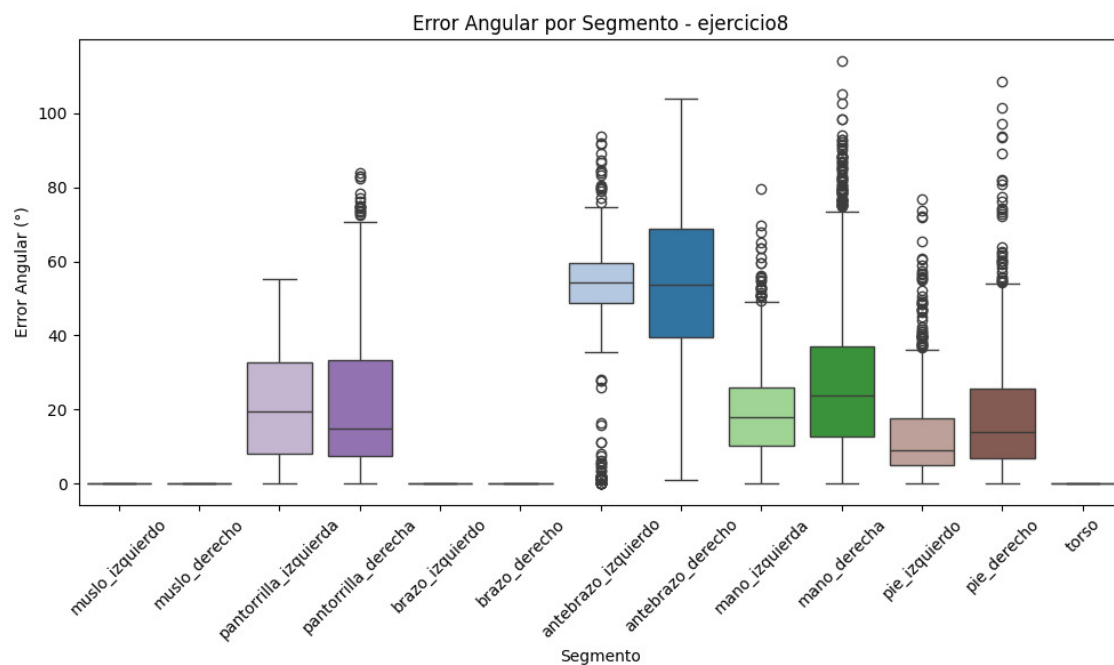


Figura 4.23: Distribución de error angular por segmento del ejercicio 8

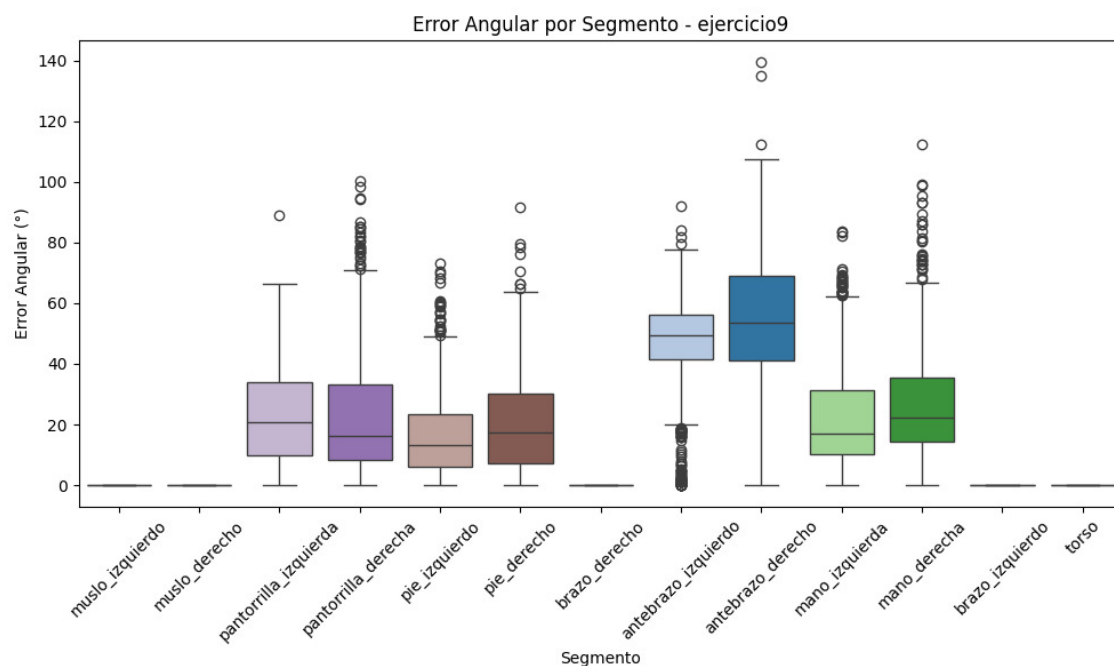


Figura 4.24: Distribución de error angular por segmento del ejercicio 9

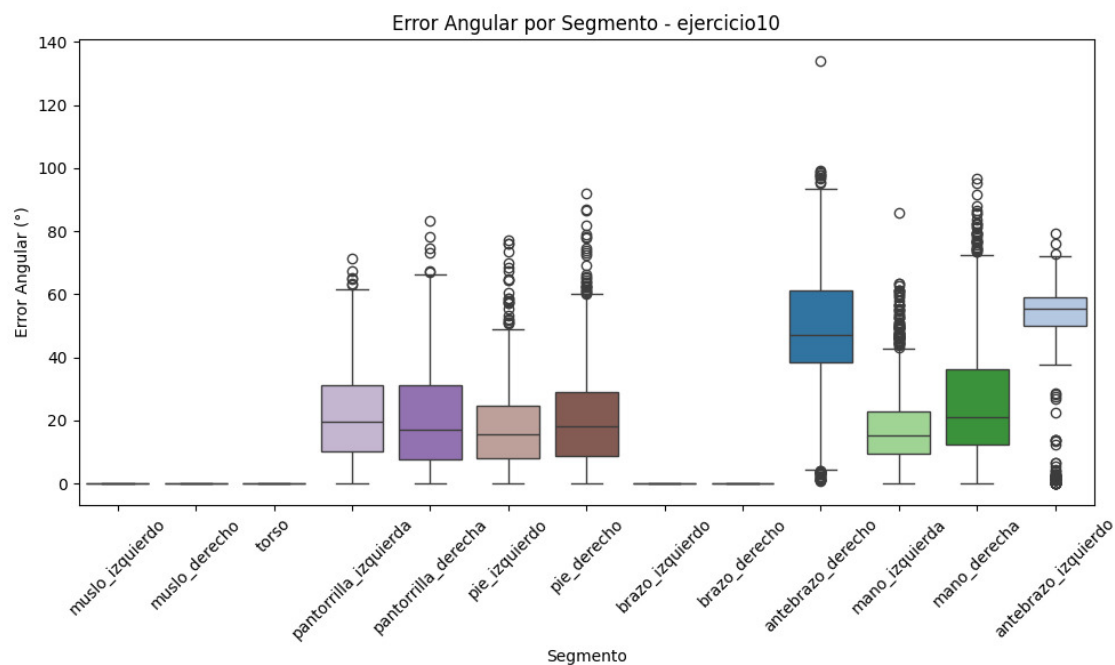


Figura 4.25: Distribución de error angular por segmento del ejercicio 10

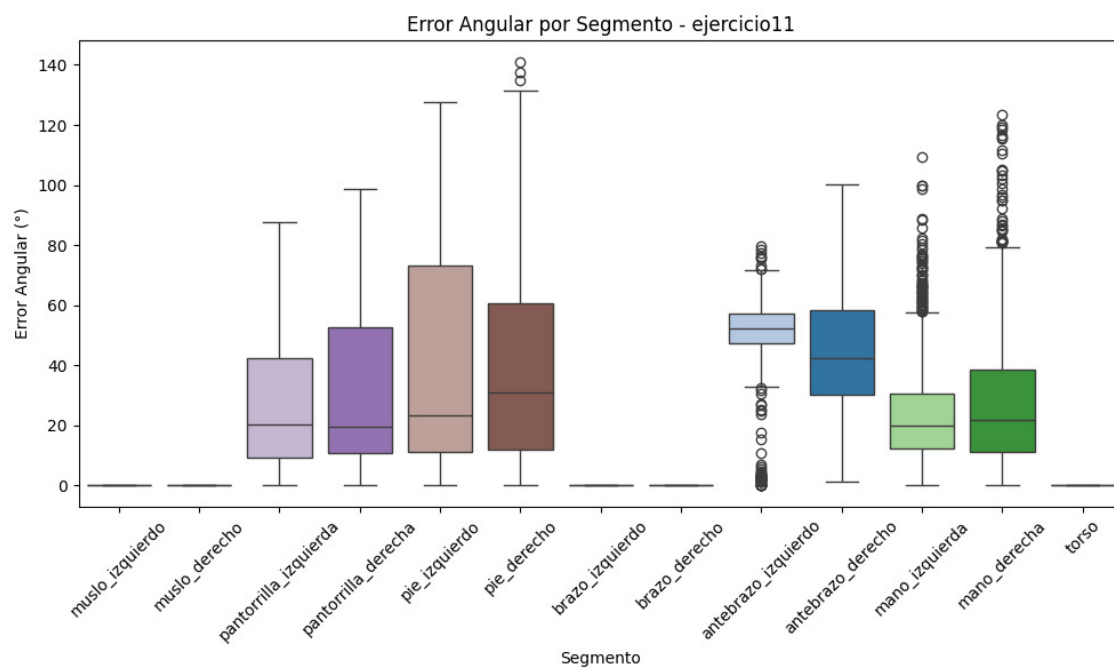


Figura 4.26: Distribución de error angular por segmento del ejercicio 11

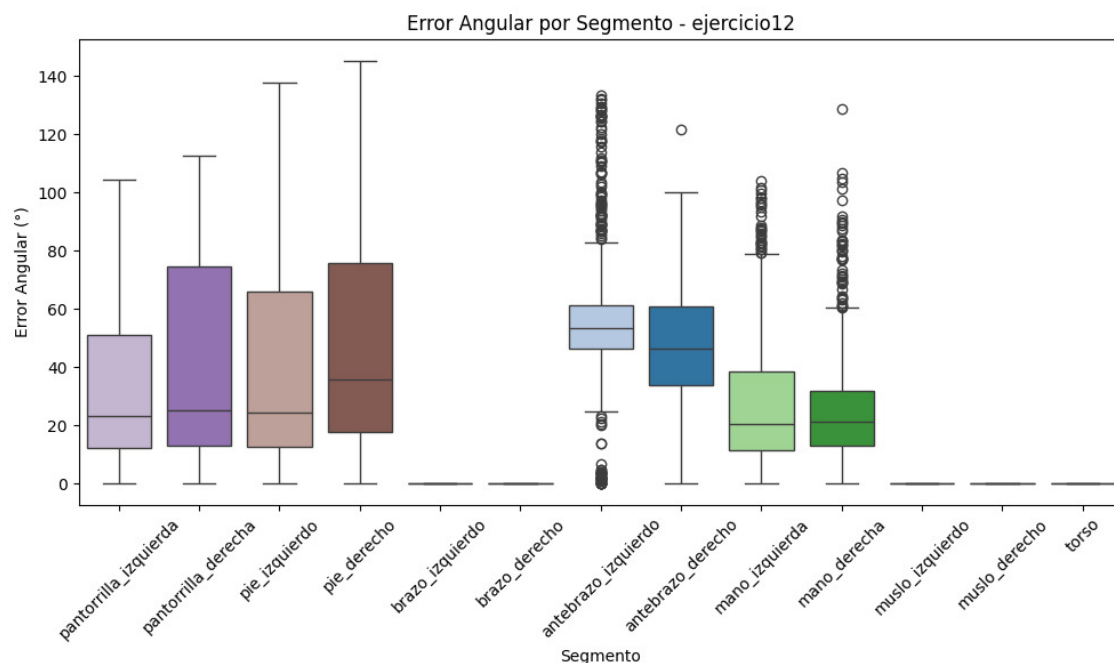


Figura 4.27: Distribución de error angular por segmento del ejercicio 12

4.3. Discusión de Resultados

El modelo que resultó más adecuado para la estimación de pose en este trabajo fue *Media Pipe Pose*, gracias a que mostró amplia cobertura de puntos clave y precisión en la detección de articulaciones, incluso en escenarios complejos.

Dado que los métodos de triangulación dependen de las coordenadas 2D proyectadas y de la geometría estimada de cámara, se considera que su desempeño es sensible tanto a la calidad de los datos como a la precisión en la calibración de las cámaras, por lo que la selección del método de triangulación se vio influenciado por estos datos.

Aunque el sistema mostró un comportamiento consistente, la mejora de la calibración mediante una mayor variabilidad en los puntos registrados podría incrementar significativamente la precisión de la triangulación dado que se contaría con una mejor cobertura espacial lo cual disminuiría el error. Sin embargo, reducir este error en un espacio amplio como el usado en este estudio, implicaría el uso de equipos especializados, como sensores de profundidad o cámaras de alta calidad, pues la precisión

depende de la calidad de las detecciones 2D y este tipo de equipos permite obtener estimaciones más precisas.

El modelo *SMPL* es una de las mejores opciones cuando se desea obtener un modelo paramétrico del cuerpo humano con nivel de detalle medio (bajo en cara y manos, para mayor detalle en estas zonas, se recomienda explorar versiones más recientes del modelo base). Sin embargo, actualmente se considera que existe un control limitado sobre su uso, lo que provoca que los modelos generados difieran significativamente de las poses reales en diversas ocasiones, difieran por completo de las poses reales, dado que existen errores relevantes en extremidades superiores e inferiores distales (manos, antebrazos, pies), especialmente en ejercicios complejos.

El análisis cuantitativo de errores mostró que:

- El error angular promedio global fue de 10.4° , con una desviación estándar de 16.7° , lo que refleja variabilidad en las rotaciones realizadas por el modelo *SMPL*.
- El error de longitud promedio fue de 0.076 cm , con una baja dispersión (0.059 cm), lo que sugiere consistencia en la escala generada por el modelo *SMPL*.
- El error de triangulación promedio fue de 424.27 píxeles y se mantuvo estable entre todos los participantes lo que indica un sistema de captura coherente.

Cabe destacar que los participantes fueron 6 mujeres y 6 hombres jóvenes. La vestimenta fue libre, incluyendo el caso de una participante (número 12) que usó ropa holgada y de color similar al suelo. A pesar de esto, el sistema generó los respectivos modelos 3D del cuerpo humano (En el Anexo C, Figura C.1 se muestran a los participantes). Lo que a su vez genera un *dataset* variado en ejercicios y participantes así como en condiciones de iluminación.

Aunque la variabilidad angular es evidente, se ha comprobado que está influenciada principalmente por las limitaciones del modelo *SMPL* y la documentación disponible. Se sugiere un mayor acceso al código fuente y especificaciones del modelo para permitir ajustes detallados y correctos.

Este enfoque ha mostrado ser útil, especialmente en la generación de modelos de pose 3D, para aplicaciones con un enfoque geométrico que permite reducir el consumo de recursos computacionales. Es especialmente viable en entornos controlados, donde se puede asegurar una calibración adecuada y múltiples vistas. Para escenarios no controlados o con un solo video, se recomiendan otras soluciones del estado del arte que requieren menos restricciones. Adicionalmente, aunque se observó que este enfoque tiene potencial para manejar casos con oclusión gracias al uso de múltiples cámaras, se requiere investigación adicional para validar esta teoría.

Capítulo 5

Conclusiones

5.1. Conclusiones generales

Este trabajo presenta un enfoque completo para la generación de modelos tridimensionales del cuerpo humano a partir de estimaciones de pose 2D, utilizando técnicas de calibrado de cámaras, triangulación y modelado paramétrico. A continuación, se resumen las principales conclusiones alineadas con los objetivos específicos planteados:

En primer lugar, se logró cumplir el objetivo de generar un conjunto de datos de video sincronizado a partir de múltiples vistas. La sincronización temporal entre cámaras fue clave para asegurar la coherencia en los modelos generados posteriormente. A partir de este conjunto de datos, se implementaron distintos algoritmos de estimación de pose 2D con el fin de identificar el que mejor se adecuara a las necesidades del proyecto, cumpliendo así el segundo objetivo. En este proceso, se concluyó que el modelo *MediaPipe Pose* ofrecía un desempeño eficiente, estable y suficiente para las condiciones del experimento, sin requerir grandes recursos computacionales.

Posteriormente, se procedió a la generación de modelos de pose 2D para cada vista capturada, manteniendo la sincronización entre secuencias. Esto permitió establecer una base para la triangulación, con estimaciones 2D que se corresponden entre cámaras. Al investigar arquitecturas para la fusión de vistas, se optó por un enfoque geométrico multivista mediante triangulación, dada su simplicidad y su ade-

cuación al objetivo planteado. Esta decisión responde al cuarto objetivo específico, que implicaba analizar distintas alternativas de fusión de modelos y seleccionar la más apropiada.

En cuanto a la fusión de modelos 2D para la obtención de un modelo 3D, se implementó un proceso de triangulación basado en la calibración precisa de las cámaras. Se comprobó que la calidad del modelo tridimensional dependía en gran medida de la exactitud de dicho proceso de calibración. Aunque se observaron errores de reconstrucción, estos se mantuvieron relativamente estables durante todo el procedimiento. Adicionalmente, se integró el modelo paramétrico *SMPL* para representar la estructura del cuerpo humano, permitiendo una reconstrucción completa del cuerpo que se encuentra limitada en cuanto a la manipulación directa de las rotaciones articulares.

Finalmente, se generó un conjunto de datos 3D estimado a partir de los videos capturados, cumpliendo con el último objetivo específico. Sin embargo, requiere ser trabajado debido a los errores encontrados por el limitado control en la manipulación del modelo *SMPL*, este conjunto proporciona una base para futuras investigaciones y abre la posibilidad de explorar mejoras en ambientes no controlados, donde se identificaron algunas limitaciones relacionadas con la sensibilidad del sistema a la iluminación, el fondo y las condiciones de calibración.

Se observa potencial en aplicaciones prácticas en áreas donde la reconstrucción precisa de la pose y la estructura corporal resulta útil. Por ejemplo, en el ámbito de la rehabilitación física, permitiría un seguimiento no invasivo del progreso de pacientes. Asimismo, podría aplicarse en entornos industriales y de seguridad para analizar patrones de movimiento y detectar situaciones de riesgo. Adicionalmente, el uso de modelos 3D abre la posibilidad de abordar tareas de clasificación o detección de patrones con mayor eficiencia, ya que contar con una representación tridimensional reduce la dependencia de grandes conjuntos de datos.

Este trabajo muestra que es posible construir un sistema geométrico multivista para reconstrucción 3D del cuerpo humano, con resultados aceptables en condiciones específicas.

5.2. Trabajo futuro

A partir del trabajo desarrollado y de sus limitaciones, se identifican las siguientes posibles áreas de trabajo:

- Se recomienda utilizar cámaras de mayor resolución o integrar sensores con la intención reducir el error de calibración, lo cual mejoraría la precisión general del sistema.
- Trabajar con detalle los errores de rotación del modelo *SMPL*, profundizando en el código del mismo para eliminar los mismos.
- Una vez trabajados los errores de las rotaciones se sugiere implementar modelos *SMPL* más detallados (como *SMPLify* y *SMPL-X*) para generar una representación con mayor detalle del cuerpo humano.
- Es posible adaptar el sistema a ambientes no controlados eliminando la necesidad de la calibración para generar el modelo de pose 3D, se recomienda evaluar el desempeño del sistema en este tipo de ambientes no controlados.

5.3. Productos generados

Como resultado del proyecto se generaron los siguientes productos:

- *Dataset* de videos 2D, que contiene 1624 videos.
- *Dataset* de modelos 3D, que contiene 18000 elemento `obj`.
- Código fuente del sistema
- Tesis
- Divulgación en la revista ReDDI+TEZ (Ver Anexo D Figura D.2)
- Artículo publicado en el CITCA (Ver Anexo D Figura D.1)

Referencias

- [1] Olga Krutikova, Aleksandrs Sisojevs, and Mihails Kovalovs. Creation of a depth map from stereo images of faces for 3d model reconstruction. *Procedia Computer Science*, 104:452–459, 12 2017. doi: 10.1016/j.procs.2017.01.159.
- [2] M. I. Ibáñez Durán. Fundamentos teóricos de un sistema de visión artificial (capítulo 2). Documento PDF, s.f. Recuperado el 5 de agosto de 2025, de <https://biblus.us.es/bibing/proyectos/use/abreproy/10622/fichero/CAPITULO+2%252FFUND-TE%C3%93RICOS.pdf>.
- [3] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. Smpl: A skinned multi-person linear model. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)*, 34(6):248:1–248:16, October 2015.
- [4] D. Rose. Rotation quaternions, and how to use them, May 2015. URL <https://danceswithcode.net/engineeringnotes/quaternions/quaternions.html>. Accessed: 2025-06-04.
- [5] Pose landmarker – mediapipe solutions, 2025. URL https://ai.google.dev/edge/mediapipe/solutions/vision/pose_landmarker?hl=es-419. consultada en 2025.
- [6] Likeotset.click. Product details, 2025. URL https://likeotset.click/product_details/20252753.html. Accedido 2025.
- [7] Unite.AI. Un nuevo sistema para la difusión estable y consistente de personajes de vídeo, n.d. URL <https://www.unite.ai/es/Un-nuevo-sistema-para-la-difusi%C3%B3n-estable-y-consistente-de-personajes-de-v%C3%ADdeo/>. 2025.
- [8] A. Mendez, S. Shapoval, Z. García, and I. Mugueta-Aguinaga. Biofeedback applied to interactive serious games to monitor frailty in an elderly population. *Applied Sciences*, 11(8), 2021.
- [9] J. Byun, H. Ro, Y. Park, and T. Han. Display methods of projection augmented reality based on deep learning pose estimation. In *ACM SIGGRAPH 2019 Posters*, SIGGRAPH '19, New York, NY, USA, 2019. Association for Computing Machinery.

- [10] Z. Huang, Z. Muhammad, and R. Khan. A review of 3d human body pose estimation and mesh recovery. *Digital Signal Processing*, 128:103628, 2022.
- [11] Helena A. Correia and José Henrique Brito. 3d reconstruction of human bodies from single-view and multi-view images: A systematic review. *Computer Methods and Programs in Biomedicine*, 239:107620, 2023. ISSN 0169-2607. doi: <https://doi.org/10.1016/j.cmpb.2023.107620>. URL <https://www.sciencedirect.com/science/article/pii/S0169260723002857>.
- [12] Maryem Belkaid, El Arbi Abdellaoui Alaoui, Achraf Berrajaa, Nabil El Akkad, and Mostafa Merras. Deep learning-based solution for 3d reconstruction from single rgb images. In *2024 International Conference on Circuit, Systems and Communication (ICCSC)*, pages 1–6, 2024. doi: 10.1109/ICCSC62074.2024.10617404.
- [13] Yuxin Tian, Hao Zhang, and Yebin Liu. Deep learning for 3d human pose estimation and mesh recovery: A survey. *arXiv preprint arXiv:2402.18844*, 2024. URL <https://arxiv.org/abs/2402.18844>.
- [14] Ramesh B Neupane, Kun Li, and Tsegaye F Boka. A survey on deep 3d human pose estimation. *Artificial Intelligence Review*, 2024. doi: 10.1007/s10462-024-10767-w. URL <https://doi.org/10.1007/s10462-024-10767-w>.
- [15] J. Saldaña. Caracterización de imágenes en movimiento: Correr y caminar. Tesis de maestría, CENIDET, 2007.
- [16] R. Juárez. Caracterización visual de movimientos “sospechosos” de personas en estacionamientos. Tesis de maestría, CENIDET, 2013.
- [17] D. Rosado. Seguimiento y caracterización automática de la trayectoria de una persona, en un entorno multicámaras de videovigilancia. Tesis de maestría, CENIDET, 2019.
- [18] G. Urquiza. Detección de anomalías en el comportamiento de una persona. Tesis de maestría, CENIDET, 2022.
- [19] L. Arriaga. Estimación de la posición del cuerpo en secuencias de imágenes. Tesis de maestría, CENIDET, 2024.
- [20] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.
- [21] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.
- [22] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2nd edition, 2004.

- [23] W.-C. Ma and Adrian Hilton. Modelling human motion using volumetric surface tracking. *Computer Vision and Image Understanding*, 108(2–3):163–181, 2007.
- [24] R. Drillis and R. Contini. Body segment parameters. Technical report, New York University, Institute of Environmental Medicine, New York, USA, 1966.
- [25] Richard M. Murray, Zexiang Li, and S. Shankar Sastry. *A Mathematical Introduction to Robotic Manipulation*. CRC Press, 1994.
- [26] S. Pan, J. Peng, X. Liu, X. Feng, and X. Zhao. Skeleton tracking based on kinect camera and the application in virtual reality system. In *Proceedings of the 4th International Conference on Virtual Reality (ICVR 2018)*, pages 21–25, New York, NY, USA, 2018. Association for Computing Machinery.
- [27] S. Ganesan and L. Anthony. Using the kinect to encourage older adults to exercise: A prototype. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems (CHI EA '12)*, pages 2297–2302, New York, NY, USA, 2012. Association for Computing Machinery.
- [28] Q. Ke, H. Rahmani, M. Bennamoun, Z. Sun, J. Liu, and G. Wang. Human action recognition from various data modalities: A review. *CoRR*, abs/2012.11866, 2020. URL <https://arxiv.org/abs/2012.11866>.
- [29] Richard Szeliski. *Computer Vision: Algorithms and Applications*. Springer, London, U.K., 2010.
- [30] Anargyros Chatzitofis, Leonidas Saroglou, Prodromos Boutis, Petros Drakoulis, Nikolaos Zioulis, Shishir Subramanyam, Bart Kevelham, Caecilia Charbonnier, Pablo Cesar, Dimitrios Zarpalas, Stefanos Kollias, and Petros Daras. Human4d: A human-centric multimodal dataset for motions and immersive media. *IEEE Access*, 8:176241–176262, 2020.
- [31] Zhiyong Chen, Ronghui Wu, Shihui Guo, Xiangyang Liu, Hongbo Fu, Xiaogang Jin, and Minghong Liao. 3d upper body reconstruction with sparse soft sensors. *Soft Robotics*, 8(2): 226–239, 2021.
- [32] Yang Lu, Han Yu, Wei Ni, and Liang Song. 3d real-time human reconstruction with a single rgb-d camera. *Applied Intelligence*, 53:8735–8745, 2022.
- [33] Zhihao Li, Jianzhuang Liu, Zhensong Zhang, Songcen Xu, and Youliang Yan. Cliff: Carrying location information in full frames into human pose and shape estimation, 2022. URL <https://arxiv.org/abs/2208.00571>.
- [34] Haibo Qiu, Chunyu Wang, Jingdong Wang, Naiyan Wang, and Wenjun Zeng. Cross view fusion for 3d human pose estimation, 2019. URL <https://arxiv.org/abs/1909.01203>.

- [35] Junhyeong Cho, Kim Youwang, and Tae-Hyun Oh. Cross-attention of disentangled modalities for 3d human mesh recovery with transformers, 2022. URL <https://arxiv.org/abs/2207.13820>.
- [36] Jianchuan Chen, Wentao Yi, Tiantian Wang, Xing Li, Liqian Ma, Yangyu Fan, and Huchuan Lu. Pixel2isdf: Implicit signed distance fields based human body model from multi-view and multi-pose images, 2022. URL <https://arxiv.org/abs/2212.02765>.
- [37] Zhongguo Li, Magnus Oskarsson, and Anders Heyden. Detailed 3d human body reconstruction from multi-view images combining voxel super-resolution and learned implicit representation, 2020. URL <https://arxiv.org/abs/2012.06178>.
- [38] Georgios Pavlakos, Vasileios Choutas, Nima Ghorbani, Timo Bolkart, Ahmed A. A. Osman, Dimitrios Tzionas, and Michael J. Black. Expressive body capture: 3D hands, face, and body from a single image. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 10975–10985, 2019.
- [39] Federica Bogo, Angjoo Kanazawa, Christoph Lassner, Peter Gehler, Javier Romero, and Michael J. Black. Keep it SMPL: Automatic estimation of 3d human pose and shape from a single image. In *Computer Vision – ECCV 2016*, Lecture Notes in Computer Science. Springer International Publishing, October 2016.
- [40] Angjoo Kanazawa, Michael J. Black, David W. Jacobs, and Jitendra Malik. End-to-end recovery of human shape and pose. In *Computer Vision and Pattern Recognition (CVPR)*, 2018.
- [41] Shubham Goel, Georgios Pavlakos, Jathushan Rajasegaran, Angjoo Kanazawa, and Jitendra Malik. Humans in 4d: Reconstructing and tracking humans with transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023.
- [42] Muhammed Kocabas, Chun-Hao P. Huang, Otmar Hilliges, and Michael J. Black. Pare: Part attention regressor for 3d human body estimation, 2021. URL <https://arxiv.org/abs/2104.08527>.
- [43] Hongwen Zhang, Yating Tian, Xinchu Zhou, Wanli Ouyang, Yebin Liu, Limin Wang, and Zhenan Sun. Pymaf: 3d human pose and shape regression with pyramidal mesh alignment feedback loop, 2021. URL <https://arxiv.org/abs/2103.16507>.
- [44] Hongwen Zhang, Yating Tian, Yuxiang Zhang, Mengcheng Li, Liang An, Zhenan Sun, and Yebin Liu. Pymaf-x: Towards well-aligned full-body model regression from monocular images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(10):12287–12303, October 2023. ISSN 1939-3539. doi: 10.1109/tpami.2023.3271691. URL <http://dx.doi.org/10.1109/TPAMI.2023.3271691>.

- [45] Davis Rempe, Tolga Birdal, Aaron Hertzmann, Jimei Yang, Srinath Sridhar, and Leonidas J. Guibas. Humor: 3d human motion model for robust pose estimation, 2021. URL <https://arxiv.org/abs/2105.04668>.
- [46] Endri Dibra, Himanshu Jain, Cengiz Öztireli, Remo Ziegler, and Markus Gross. Hs-nets: Estimating human body shape from silhouettes with convolutional neural networks. In *2016 Fourth International Conference on 3D Vision (3DV)*, pages 108–117, 2016. doi: 10.1109/3DV.2016.19.
- [47] Yang Zheng, Ruizhi Shao, Yuxiang Zhang, Tao Yu, Zerong Zheng, Qionghai Dai, and Yebin Liu. Deepmulticap: Performance capture of multiple characters using sparse multiview cameras, 2021. URL <https://arxiv.org/abs/2105.00261>.
- [48] Zijian Dong, Jie Song, Xu Chen, Chen Guo, and Otmar Hilliges. Shape-aware multi-person pose estimation from multi-view images, 2021. URL <https://arxiv.org/abs/2110.02330>.
- [49] Zhixuan Yu, Linguang Zhang, Yuanlu Xu, Chengcheng Tang, LUAN TRAN, Cem Keskin, and Hyun Soo Park. Multiview human body reconstruction from uncalibrated cameras. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 7879–7891. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/33610fba262d7b6fed0810b89f55e147-Paper-Conference.pdf.
- [50] Xiaoben Li, Mancheng Meng, Ziyan Wu, Terrence Chen, Fan Yang, and Dinggang Shen. Human mesh recovery from arbitrary multi-view images, 2024. URL <https://arxiv.org/abs/2403.12434>.
- [51] Yitao Zhu, Sheng Wang, Mengjie Xu, Zixu Zhuang, Zhixin Wang, Kaidong Wang, Han Zhang, and Qian Wang. Muc: Mixture of uncalibrated cameras for robust 3d human body reconstruction, 2024. URL <https://arxiv.org/abs/2403.05055>.
- [52] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(1):172–186, 2021.
- [53] Google. MediaPipe Pose. <https://google.github.io/mediapipe/solutions/pose.html>, 2021. Accessed: 2021-12-28.
- [54] Hao-Shu Fang, Jiefeng Li, Hongyang Tang, Chao Xu, Haoyi Zhu, Yuliang Xiu, Yong-Lu Li, and Cewu Lu. Alphapose: Whole-body regional multi-person pose estimation and tracking in real-time. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

- [55] Matthew Grimes Alexander Kendall and Roberto Cipolla. Posenet: A convolutional network for real-time 6-dof camera relocalization. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2015.
- [56] Google. Movenet: Ultra fast and accurate pose detection model. <https://blog.tensorflow.org/2020/11/movenet-ultrafast-and-accurate-pose-detection-model.html>, 2020. Accessed: 2025-05-29.
- [57] Xingyi Zhou and Vladlen Koltun. Objects as points. *arXiv preprint arXiv:1904.07850*, 2019.
- [58] Julieta Martinez, Rayat Hossain, Javier Romero, and James J Little. A simple yet effective baseline for 3d human pose estimation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2640–2649, 2017.
- [59] Georgios Pavlakos, Xiaowei Zhu, Xiaowei Zhou, and Kostas Daniilidis. Coarse-to-fine volumetric prediction for single-image 3d human pose. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7025–7034, 2017.
- [60] Alejandro Newell, Kai Yang, and Jia Deng. Stacked hourglass networks for human pose estimation. In *European Conference on Computer Vision*, pages 483–499. Springer, 2016.
- [61] Ijaz Akhter and Michael J Black. Pose-conditioned joint angle limits for 3d human pose reconstruction. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1446–1455, 2015.
- [62] Catalin Ionescu, Dragos Papava, Vlad Olaru, and Cristian Sminchisescu. Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(7):1325–1339, jul 2014.
- [63] Mihai Fieraru, Mihai Zanfir, Silviu-Cristian Pirlea, Vlad Olaru, and Cristian Sminchisescu. Ai-fit: Automatic 3d human-interpretable feedback models for fitness training. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2021.
- [64] Dushyant Mehta, Helge Rhodin, Dan Casas, Pascal Fua, Oleksandr Sotnychenko, Weipeng Xu, and Christian Theobalt. Monocular 3d human pose estimation in the wild using improved cnn supervision, 2017. URL <https://arxiv.org/abs/1611.09813>.
- [65] Leonardo da Vinci. Studio di proporzioni del corpo umano, noto come uomo vitruviano, 1490. URL <https://www.gallerieaccademia.it/studio-di-proporzioni-del-corpo-noto-come-uomo-vitruviano>. Dibujo, ca. 1490.
- [66] Brian Gordon, Sigal Raab, Guy Azov, Raja Giryes, and Daniel Cohen-Or. Flex: Extrinsic parameters-free multi-view 3d human motion reconstruction, 2022. URL <https://arxiv.org/abs/2105.01937>.

Anexo A

Documentos Generados

En la Figura A.1 se muestra un ejemplo de como se compone un archivo OBJ, dónde las líneas que comienzan con “v” representan un vértice en el espacio 3D mediante tres valores (x, y, z) , mientras que las líneas que comienzan con “f” contiene el número de 3 vértices los cuales se unen para formar una cara (en forma de triángulo). Los datos totales en el archivo describen un objeto tridimensional.

```
#Wavefront OBJ file
v 0.0 0.0 0.0
v 1.0 0.0 0.0
v 2.0 0.0 0.0
v 0.0 0.0 1.0
v 1.0 0.0 1.0
v 2.0 0.0 1.0
v 0.0 0.0 2.0
v 1.0 0.0 2.0
v 2.0 0.0 2.0
f 1 2 5
f 1 5 4
f 2 3 6
f 2 6 5
f 4 5 8
f 4 8 7
f 5 6 9
f 5 9 8
```

Figura A.1: Ejemplo del contenido de un archivo OBJ

En la Figura A.2 se muestra un ejemplo de contenido de un archivo de medidas.

```
Largo pierna,  
41.814387576510285  
  
Largo brazo,  
72.44231855958769  
  
Ancho hombros,  
51.889099548306  
  
Ancho caderas,  
28.142448346785084  
  
Altura talón-hombro,  
79.62110715315048  
  
Altura total,  
91.87050825363515  
  
Tamaño cabeza  
12.249401100484683|
```

Figura A.2: Ejemplo del contenido del archivo de medidas generado

En la Figura A.3 se muestra un ejemplo de contenido de un archivo de coordenadas 3D generado.

Figura A.3: Ejemplo del contenido del archivo de coordenadas generado

Figura A.3: Ejemplo del contenido del archivo de coordenadas generado

Anexo B

Tablas

En esta sección se concentran las tablas de error, por persona las cuales se concentran en la tabla 4.10 para el análisis de resultados.

Tabla B.1: Errores de la Persona 1

	Persona 1				
	Promedio General		Desviación Estándar		Error de triangulación (píxeles)
	Error Angular (grados)	Error Longitud (cm)	Error Angular (grados)	Error Longitud (cm)	
1	13.273516	0.072907	22.395966	0.051986	424.2734
2	16.940386	0.123696	27.291855	0.072883	424.2985
3	8.483634	0.090125	15.615567	0.075903	424.2748
4	18.560227	0.102749	30.035774	0.067356	424.2979
5	18.878718	0.102764	27.069948	0.061337	424.2926
6	19.983974	0.095693	29.265567	0.064283	424.2979
7	17.958511	0.106811	29.335122	0.068326	424.2955
8	18.898498	0.103118	22.884709	0.072277	424.2662
9	17.148325	0.116222	20.059713	0.08566	424.2572
10	16.979249	0.139991	19.56011	0.096034	424.2932
11	25.916056	0.144031	32.111638	0.114213	424.2619
12	26.69671	0.152515	33.633163	0.118031	424.2675
13	22.857369	0.143632	31.060526	0.114791	424.2751
14	36.501847	0.225426	46.250022	0.200806	424.1767
15	32.227585	0.148792	37.196267	0.111503	424.2418
16	21.875984	0.112265	29.87922	0.067653	424.2367
17	19.332428	0.094882	27.729471	0.06909	424.2762

Tabla B.2: Errores de la Persona 2

	Persona 2				
	Promedio General		Desviación Estándar		Error de triangulación (píxeles)
	Error Angular (grados)	Error Longitud (cm)	Error Angular (grados)	Error Longitud (cm)	
1	10.27694	0.09082	17.579754	0.075358	424.309
2	14.578397	0.079504	23.346016	0.060773	424.2978
3	8.152585	0.08213	12.338069	0.070236	424.2922
4	14.846313	0.083476	24.640936	0.058789	424.3297
5	16.850775	0.094822	23.004104	0.059931	424.2979
6	15.985118	0.116783	24.417808	0.078511	424.2953
7	17.184143	0.07358	23.512105	0.049037	424.2897
8	14.351071	0.106733	17.883482	0.063628	424.2617
9	14.4303	0.101429	17.760732	0.073123	424.2653
10	17.331962	0.111979	21.94154	0.06297	424.3041
11	18.958882	0.110373	26.909258	0.085588	424.2997
12	21.815109	0.138799	29.926811	0.106145	424.2832
13	21.007817	0.115026	28.172877	0.09266	424.2849
14	33.157122	0.266504	44.040573	0.253687	424.2043
15	23.978758	0.123145	30.940669	0.099297	424.2198
16	19.147141	0.084683	25.770549	0.092339	424.2317
17	20.209219	0.114659	25.624495	0.08283	424.2797

Tabla B.3: Errores de la Persona 3

	Persona 3				
	Promedio General		Desviación Estándar		Error de triangulación (píxeles)
	Error Angular (grados)	Error Longitud (cm)	Error Angular (grados)	Error Longitud (cm)	
1	8.720408	0.082092	15.032589	0.067112	424.2962
2	14.914159	0.101401	21.414311	0.061886	424.2924
3	8.756862	0.068358	15.647065	0.068073	424.2758
4	14.574652	0.104588	24.018933	0.084229	424.3196
5	17.882672	0.131079	25.469685	0.094984	424.2368
6	14.391011	0.074424	24.21524	0.052323	424.2777
7	17.59007	0.087334	24.260186	0.060285	424.2673
8	18.741534	0.125061	23.371658	0.080038	424.2325
9	18.56826	0.116949	22.573371	0.08443	424.2317
10	20.742043	0.130426	24.888828	0.084388	424.2602
11	19.894462	0.144208	23.247384	0.092975	424.2585
12	22.280009	0.155729	27.511463	0.109157	424.2772
13	21.349581	0.147214	25.985147	0.110954	424.2743
14	29.213426	0.172862	36.858851	0.151386	424.2298
15	13.434236	0.031799	21.967259	0.034939	424.2864
16	19.4174	0.112278	26.894112	0.077536	424.2668
17	18.084035	0.096817	25.265053	0.067244	424.2918

Tabla B.4: Errores de la Persona 4

	Persona 4				
	Promedio General		Desviación Estándar		Error de triangulación (píxeles)
	Error Angular (grados)	Error Longitud (cm)	Error Angular (grados)	Error Longitud (cm)	
1	6.755279	0.059469	8.789127	0.049591	424.2877
2	12.375062	0.095956	21.076152	0.056088	424.2711
3	8.416067	0.05698	14.788044	0.037309	424.2626
4	15.030796	0.065834	24.432222	0.057585	424.272
5	16.803069	0.076486	25.741461	0.053467	424.2794
6	15.270905	0.077566	22.109774	0.049935	424.2806
7	16.306819	0.088366	21.508505	0.052544	424.2892
8	13.204861	0.099771	18.333288	0.06371	424.242
9	14.966316	0.09	19.522963	0.099465	424.218
10	18.073107	0.092828	22.425655	0.071832	424.2526
11	22.632442	0.098151	25.860638	0.083887	424.2727
12	22.115281	0.15781	27.795904	0.115556	424.2881
13	23.582381	0.149218	29.778046	0.137151	424.2869
14	28.297636	0.185852	34.17606	0.180436	424.1804
15	25.366332	0.171712	30.80738	0.149051	424.225
16	21.539647	0.119318	26.855733	0.085271	424.2594
17	14.966853	0.113547	19.752464	0.073458	424.2474

Tabla B.5: Errores de la Persona 5

	Persona 5				
	Promedio General		Desviación Estándar		Error de triangulación (píxeles)
	Error Angular (grados)	Error Longitud (cm)	Error Angular (grados)	Error Longitud (cm)	
1	8.212721	0.063718	11.517579	0.047771	424.2461
2	6.741113	0.051455	11.196491	0.044161	424.2353
3	12.916156	0.045998	18.557349	0.044682	424.2486
4	16.63119	0.069844	28.418653	0.066679	424.2716
5	17.529474	0.071509	28.09334	0.050027	424.2863
6	12.394258	0.075894	19.192058	0.046598	424.2261
7	18.797781	0.070561	25.75826	0.047587	424.2362
8	14.260515	0.112908	17.512976	0.079694	424.2335
9	15.293057	0.106418	18.956115	0.070219	424.2253
10	15.393003	0.11362	18.54728	0.063217	424.2179
11	21.407906	0.151355	27.767216	0.1351	424.2328
12	19.348682	0.142696	25.061178	0.122673	424.2329
13	17.83964	0.099078	23.190449	0.086878	424.2197
14	NaN	NaN	NaN	NaN	NaN
15	24.091401	0.148616	29.488422	0.113233	424.2045
16	15.988092	0.104955	20.523622	0.067754	424.2714
17	18.124703	0.108387	24.147569	0.082113	424.2787

Tabla B.6: Errores de la Persona 6

	Persona 6				
	Promedio General		Desviación Estándar		Error de triangulación (píxeles)
	Error Angular (grados)	Error Longitud (cm)	Error Angular (grados)	Error Longitud (cm)	
1	7.884217	0.069309	11.578743	0.056792	424.2968
2	14.19453	0.111977	25.239771	0.073694	424.3133
3	10.633333	0.075597	16.77767	0.06329	424.3162
4	16.039882	0.078966	29.531927	0.078703	424.3266
5	14.844518	0.104221	23.145608	0.092142	424.3172
6	17.458426	0.108522	24.377012	0.07171	424.3209
7	20.599115	0.106331	29.799963	0.067913	424.3226
8	18.214808	0.102532	25.805756	0.079396	424.2729
9	17.582443	0.097257	23.607778	0.076857	424.266
10	20.08317	0.161526	24.771721	0.116182	424.2687
11	20.875523	0.124686	25.000957	0.096421	424.2771
12	22.751642	0.127612	28.014499	0.113961	424.2748
13	21.475105	0.128327	27.817434	0.111652	424.2811
14	31.476336	0.150846	36.540841	0.149397	424.2336
15	27.290312	0.103279	29.78964	0.084403	424.2363
16	21.00188	0.121568	27.199584	0.095323	424.2643
17	15.331847	0.096054	23.647477	0.063298	424.2995

Tabla B.7: Errores de la Persona 7

	Persona 7				
	Promedio General		Desviación Estándar		Error de triangulación (píxeles)
	Error Angular (grados)	Error Longitud (cm)	Error Angular (grados)	Error Longitud (cm)	
1	7.93549	0.064974	12.042078	0.053703	424.2664
2	13.172543	0.113125	21.094218	0.069418	424.2591
3	7.420691	0.07547	12.629553	0.073327	424.2597
4	19.641343	0.095146	32.599178	0.079676	424.2879
5	15.041846	0.097752	22.01184	0.062268	424.2566
6	14.532836	0.098815	20.811024	0.067223	424.2764
7	15.206092	0.094775	22.45325	0.064679	424.2882
8	15.760879	0.104221	18.124489	0.073466	424.2516
9	16.135901	0.121171	20.467663	0.079845	424.2606
10	16.908519	0.111202	19.239624	0.080177	424.266
11	20.734202	0.125118	26.500268	0.107485	424.2571
12	28.113741	0.141117	32.352636	0.12183	424.2602
13	23.610692	0.13875	30.763502	0.131508	424.2426
14	31.897408	0.169651	36.591112	0.173884	424.1881
15	26.79402	0.119377	34.041624	0.103002	424.251
16	23.659947	0.156988	32.19868	0.138959	424.2143
17	16.771543	0.091186	22.812313	0.06532	424.2583

Tabla B.8: Errores de la Persona 8

	Persona 8				
	Promedio General		Desviación Estándar		Error de triangulación (píxeles)
	Error Angular (grados)	Error Longitud (cm)	Error Angular (grados)	Error Longitud (cm)	
1	9.093475	0.044631	10.710855	0.039378	424.2722
2	10.019938	0.070652	14.848088	0.053795	424.2742
3	10.484252	0.065197	18.916889	0.051993	424.2731
4	10.186307	0.045801	17.676795	0.051264	424.2999
5	15.295456	0.090184	24.215379	0.06421	424.2604
6	12.941336	0.057066	21.094467	0.046318	424.2862
7	15.496748	0.056845	22.245961	0.039228	424.2845
8	18.553775	0.083177	19.572076	0.063034	424.2383
9	21.0325	0.083867	26.135133	0.079984	424.2169
10	17.54711	0.142698	19.769737	0.103172	424.2273
11	18.974171	0.111361	21.938576	0.087215	424.2595
12	26.854064	0.125555	30.667096	0.124119	424.2635
13	22.60087	0.089547	26.566205	0.084869	424.2369
14	35.497766	0.187725	39.260498	0.205571	424.1852
15	31.375809	0.141948	34.772679	0.140266	424.2109
16	20.360524	0.091012	25.232961	0.070375	424.2424
17	16.291388	0.069262	22.95386	0.051353	424.2577

Tabla B.9: Errores de la Persona 9

	Persona 9				
	Promedio General		Desviación Estándar		Error de triangulación (triangulación)
	Error Angular (grados)	Error Longitud (cm)	Error Angular (grados)	Error Longitud (cm)	
1	8.074801	0.070265	13.547716	0.059874	424.2866
2	12.796593	0.107952	20.455053	0.06548	424.2846
3	7.397697	0.054959	12.64048	0.050749	424.2764
4	15.384159	15.384159	27.779666	0.062131	424.3114
5	15.151364	0.076198	22.408363	0.05212	424.2946
6	16.672235	0.08189	27.506378	0.05627	424.2987
7	16.589375	0.082465	25.80704	0.057968	424.2967
8	15.403194	0.104752	18.469303	0.073789	424.2687
9	16.573088	0.095251	19.768323	0.07537	424.2551
10	15.814595	0.105014	19.05101	0.078134	424.271
11	21.014673	0.129477	25.670414	0.115411	424.2523
12	23.717504	0.148917	27.658531	0.135524	424.2589
13	16.005317	0.122229	21.370905	0.095985	424.2572
14	33.750305	0.22866	41.804861	0.240197	424.2041
15	25.522007	0.132433	28.808289	0.1067	424.224
16	18.356062	0.099263	23.424612	0.078123	424.2465
17	20.726696	0.098617	27.089853	0.07194	424.2616

Tabla B.10: Errores de la Persona 10

	Persona 10				
	Promedio General		Desviación Estándar		Error de triangulación (píxeles)
	Error Angular (grados)	Error Longitud (cm)	Error Angular (grados)	Error Longitud (cm)	
1	9.385044	0.068823	15.575959	0.06468	424.2648
2	12.815557	0.052404	20.764859	0.053936	424.2573
3	8.661365	0.058119	12.044386	0.047031	424.2434
4	17.765608	0.085232	31.21283	0.059451	424.29
5	12.333693	0.096437	19.542225	0.067829	424.241
6	17.499657	0.104376	29.676943	0.072565	424.2767
7	19.025252	0.090148	26.053408	0.057587	424.2747
8	23.642342	0.110867	27.91738	0.087051	424.2293
9	22.319349	0.101536	25.439519	0.078253	424.2309
10	19.58404	0.136372	21.93825	0.0927	424.2287
11	25.374913	0.144488	30.910801	0.125189	424.2598
12	25.527901	0.168534	32.122752	0.137936	424.2412
13	13.921913	0.0586	20.806847	0.064372	424.2697
14	28.808302	0.202435	34.555223	0.199253	424.1845
15	24.291328	0.156259	30.507757	0.156055	424.2008
16	30.893209	0.180174	35.844425	0.168469	424.2046
17	18.918456	0.122905	23.759265	0.084333	424.2592

Tabla B.11: Errores de la Persona 11

	Persona 11				
	Promedio General		Desviación Estándar		Error de triangulación (píxeles)
	Error Angular (grados)	Error Longitud (cm)	Error Angular (grados)	Error Longitud (cm)	
1	8.854559	0.061129	15.782701	0.050622	424.2573
2	14.08405	0.110036	22.986746	0.070527	424.2593
3	8.278117	0.084378	12.285932	0.064854	424.2402
4	16.981975	0.08555	21.504376	0.080457	424.2553
5	14.308702	0.069321	21.137669	0.052444	424.2479
6	14.227771	0.091209	19.696986	0.057379	424.2393
7	17.252258	0.079259	28.536591	0.049717	424.2488
8	17.957258	0.115536	22.531678	0.075568	424.2088
9	17.510505	0.100566	17.824481	0.071796	424.2203
10	16.532981	0.107487	19.935021	0.070578	424.2283
11	21.376499	0.127088	24.707118	0.103447	424.2263
12	20.297661	0.126652	24.278865	0.109882	424.2392
13	18.451019	0.130141	21.702282	0.09112	424.2034
14	27.965167	0.186441	34.359651	0.161422	424.1705
15	21.167842	0.137485	22.960169	0.097526	424.192
16	15.335222	0.120061	18.043286	0.088111	424.2115
17	20.745053	0.093119	29.786408	0.06268	424.2327

Tabla B.12: Errores de la Persona 12

	Persona 12				
	Promedio General		Desviación Estándar		Error de triangulación (píxeles)
	Error Angular (grados)	Error Longitud (cm)	Error Angular (grados)	Error Longitud (cm)	
1	8.804523	0.054019	14.271724	0.051935	424.2283
2	17.004405	0.093623	28.147831	0.068658	424.241
3	8.060906	0.056946	13.065159	0.066917	424.2312
4	16.485359	0.095129	26.06535	0.059719	424.2621
5	15.784257	0.101195	26.064187	0.082797	424.2579
6	20.473301	0.114194	29.713098	0.074619	424.2578
7	23.202047	0.091409	34.949809	0.060619	424.2527
8	21.096094	0.113191	26.898848	0.072838	424.2061
9	21.781104	0.08993	26.227223	0.065733	424.2021
10	11.872283	0.08874	16.702488	0.079353	424.2181
11	25.458605	0.136364	28.557561	0.104371	424.2236
12	29.331008	0.156412	34.294647	0.115081	424.2325
13	23.093023	0.127088	26.801525	0.111726	424.1989
14	NaN	NaN	NaN	NaN	424.2177
15	20.625275	0.16018	26.368577	0.106748	424.1683
16	15.685424	0.12745	20.729421	0.073149	424.1987
17	17.645026	0.110532	25.264701	0.078667	424.2251

Anexo C

Imágenes

En la Figura C.1 se muestran a los participantes grabados para la creación del *dataset*.



Figura C.1: Personas grabadas para el *dataset*

Anexo D

Productos Generados

En esta sección se presentan los productos generados durante el desarrollo de la Tesis.

- Una publicación en el Congreso Internacional de Tecnología y Ciencia Aplicada (Figura D.1)

Así como dos retribuciones sociales:

- Una publicación de divulgación en la revista REDDI+Tez (Figura D.2).
- Participación como parte del *STAFF* en el CITCA 2024-2 (D.3).



Educación
Secretaría de Educación Pública



TECNOLÓGICO
NACIONAL DE MÉXICO

EL TECNOLÓGICO NACIONAL DE MÉXICO A TRAVÉS DEL CENTRO NACIONAL DE INVESTIGACIÓN Y DESARROLLO TECNOLÓGICO

OTORGA EL PRESENTE

RECONOCIMIENTO

A

**Alexandra Ehrlich López, Raúl Pinto Elías, Andrea
Magadan Salazar, Jorge A. Fuentes Pacheco**

Por su destacada participación con la presentación del Artículo
en el área Ciencias de la Computación, eje estratégico Otros titulado:
Reconstrucción 3D del esqueleto humano

En el marco del 3er Congreso Internacional de Tecnología y Ciencias Aplicadas (CITCA), celebrado en el
Tecnológico Nacional de México, campus Centro Nacional de Investigación y Desarrollo Tecnológico
(TecNM/CENIDET), del 19 al 22 de noviembre de 2024

Cuernavaca, Morelos, 22 de noviembre de 2024



CITCA54424
<http://constancias.cenidet.tecnm.mx>

LORENZO O. HERNÁNDEZ
DIRECTOR DEL CENTRO NACIONAL DE
INVESTIGACIÓN Y DESARROLLO TECNOLÓGICO

Sello Digital:

c4v4aA51B7QiFCkIIyYf4THnbIVDpgxiatDHH2qwlmmqyRfQ0vhsQ3pH58ym5yNGyNI7DR71GyUv96HLZ41fTL
jU6cvTGyXzFPw6CRQ89+wd1U6SITAIINy3/2bqTW4uyeKObag+2/GyDFU4ZQPWdrW3001XM4s3eL06JppD61H/
lNyDItRsXoguK6LBjw8FLSoAmwxyE9/FpgoIv4WxPOx58oKOY/zLiwpFpfT2MMjUjbbzAecPvL4WgxigFF5yA
BXNNiLZ3XnxYtFUnCk89UeY94jzKS7Kii1BdCSq1OKxBbOMmRAL6MEPsA9gOl7ETMX3Ahuli2S1dbSZ8M/6A==

ALICIA MARTÍNEZ REBOLLAR
COORDINADORA GENERAL
III CITCA

cenidet
Centro Nacional de Investigación
y Desarrollo Tecnológico

CITCA
3er. CONGRESO INTERNACIONAL
DE TECNOLOGÍA Y CIENCIA APLICADA



Figura D.1: Publicación en el CITCA

Estudios y Ensayos

UNA REVISIÓN DE MÉTODOS DE RECONSTRUCCIÓN 3D DEL CUERPO HUMANO

AUTOR 1.

ALEXANDRA EHRLICH LÓPEZ

CENTRO NACIONAL DE INVESTIGACIÓN Y DESARROLLO TECNOLÓGICO

NO. ORCID: 0000-0001-6880-8426

AUTOR 2.

RAÚL PINTO ELÍAS

CENTRO NACIONAL DE INVESTIGACIÓN Y DESARROLLO TECNOLÓGICO

NO. ORCID: 0000-0001-9141-058X

AUTOR 3.

NIMROD GONZÁLEZ FRANCO

CENTRO NACIONAL DE INVESTIGACIÓN Y DESARROLLO TECNOLÓGICO

NO. ORCID: 0000-0002-1051-1379

AUTOR 4.

ANDREA MAGADAN

CENTRO NACIONAL DE INVESTIGACIÓN Y DESARROLLO TECNOLÓGICO

NO. ORCID: 0000-0001-5474-4150

KEYWORDS

Human body, 3D model, reconstruction

PALABRAS CLAVE

Cuerpo humano, modelo 3D, reconstrucción.

RESUMEN

Actualmente el cuerpo humano es ampliamente estudiado en diversas aplicaciones, normalmente dichas aplicaciones son realizadas por medio de información 2D obtenida de dispositivos comunes; gracias a los avances tecnológicos tanto en hardware como en software, es posible realizar

las tareas relacionadas al cuerpo humano en tres dimensiones, lo cual provee de información extra para la realización de tareas como clasificación y reconocimiento, por medio de la reconstrucción en modelo 3D del cuerpo humano. Esta área de investigación lleva ya varios años en desarrollo por lo que se pueden encontrar múltiples algoritmos para generación de los modelos 3D mencionados anteriormente, los cuales pueden o no utilizar sensores adicionales a una cámara y distintas técnicas para la reconstrucción. En este artículo se presenta una síntesis de diversos modelos computacionales para la tarea de reconstrucción del cuerpo humano.

ABSTRACT

Currently, the human body is widely studied in various applications, usually through 2D information obtained from common devices. Thanks to technological advances in both hardware and software, it is possible to perform related tasks to the human body in three dimensions, which provides extra information for tasks such as classification and recognition through the 3D reconstruction of the human body. This area of research has been in development for several years, so multiple algorithms can be found for generating the aforementioned 3D models, which may or may not use additional sensors to a camera and different techniques for reconstruction. This article presents a synthesis of various computational models for the task of reconstructing the human body.

INTRODUCCIÓN

El cuerpo humano es objeto de extenso estudio en la actualidad y se han desarrollado diversos modelos de visión por computadora para llevar a cabo tareas variadas relacionadas con él. Estas tareas incluyen el reconocimiento

10

Figura D.2: Publicación en revista REDDI+Tez



Figura D.3: Retribución Social: Participación como STAFF en el CITCA