

Revisión sistemática de literatura del análisis de movimiento del tren superior del cuerpo humano mediante visión por computador

Daniel Esteban Quintero Unigarro

Juan Diego Rivera Caicedo

Universidad CESMAG

Facultad de Ingeniería

Programa Ingeniería Electrónica

San Juan de Pasto

2026

Revisión sistemática de literatura del análisis de movimiento del tren superior del cuerpo humano mediante visión por computador

Daniel Esteban Quintero Unigarro
Juan Diego Rivera Caicedo

Proyecto de Trabajo de grado presentado al Comité Curricular del Programa de Ingeniería
Electrónica
Modalidad: Monografía

Asesor
Mg. Mario Fernando Henao Rosero

Universidad CESMAG
Facultad de Ingeniería
Programa Ingeniería Electrónica
San Juan de Pasto
2026

Nota de aceptación

Aprobado por el Comité de
Investigación y Proyección Social en
Cumplimiento de los Requisitos
exigidos por la Universidad CESMAG
Para optar al título de Ingeniería
Electrónica.

Mag. Mario Fernando Henao Rosero



Asesor



Jurado



Jurado

San Juan de Pasto, abril 2026

Página de Nota de Exclusión de Responsabilidad Intelectual

“El pensamiento que se expresa en esta obra es exclusivamente responsabilidad de sus autores y no compromete la ideología de la Universidad CESMAG”

Agradecimientos

Queremos expresar nuestro más sincero agradecimiento a todas las personas que nos han acompañado a lo largo de este proceso y que hicieron posible la realización de este proyecto de grado.

A nuestras familias, por su apoyo incondicional, su amor y comprensión, que fueron fundamentales en cada etapa de este camino. Sin su respaldo constante, no habría sido posible culminar este logro.

A nuestros profesores, por su compromiso con nuestra formación académica y por compartir con nosotros sus conocimientos y experiencia. De manera especial, agradecemos al profesor Mag. Mario Fernando Henao por su orientación, dedicación y paciencia, así como por el tiempo y apoyo brindados durante el desarrollo de este trabajo.

A nuestros amigos y compañeros, quienes nos acompañaron durante este proceso académico, compartiendo momentos tanto difíciles como felices, y brindándonos ánimo y motivación para seguir adelante.

Finalmente, agradecemos a todas aquellas personas que, de una u otra forma, contribuyeron al desarrollo y culminación de este proyecto. Cada consejo, cada palabra y cada gesto de apoyo fueron valiosos en este camino. Este logro también es de ustedes.

Dedicatoria

Dedico este logro a la persona más importante en mi vida: mi madre, quien me ha brindado un apoyo incondicional y un amor infinito. Ella me enseñó que la educación es uno de los pilares fundamentales para crecer como persona. Sin su esfuerzo, confianza y acompañamiento, este sueño no habría sido posible. Espero seguir contando con su apoyo y que pueda verme alcanzar muchos más logros en mi vida.

A mis abuelos, quienes han estado a mi lado desde siempre. Sus consejos, llamados de atención y palabras de aliento han sido muestras de su amor y preocupación. Este logro también es de ustedes, pues han sido una parte esencial de mi formación personal.

A mis tíos, con quienes he compartido gran parte de mi vida, por sus enseñanzas, su cariño y su apoyo constante. Deseo que se sientan orgullosos de este momento y espero seguir contando con ustedes en los próximos pasos de mi camino.

Finalmente, a mis profesores, quienes me han brindado uno de los regalos más valiosos: el conocimiento. Gracias por sus enseñanzas, las cuales llevaré conmigo a lo largo de mi vida, con el compromiso de seguir aprendiendo y creciendo.

Daniel Esteban Quintero Unigarro

Dedicatoria

A mis hermanos, por su compañía permanente, por preocuparse por mí en cada etapa de este proceso y por estar siempre presentes con sus consejos, su ánimo y sus felicitaciones en cada pequeño y gran logro.

A mis abuelos, pilares fundamentales en mi vida. A mi abuelo, quien, aunque ya no está físicamente, sigue siendo una fuente de inspiración y fortaleza en cada paso que doy; su recuerdo vive en este logro. A mi abuela, por su amor, su apoyo y por acompañarme con su presencia y sus palabras en este camino, siendo un motor constante para seguir adelante.

A mis profesores, por ser guía y orientación a lo largo de estos años de formación. Su dedicación, conocimiento y compromiso han sido clave en mi crecimiento académico, personal y profesional.

A mis amigos, en especial a Nicolás, quien estuvo presente en mis momentos más difíciles y en mis mejores etapas durante la carrera, brindándome su apoyo sincero, sus consejos directos y su manera única de impulsarme a seguir adelante.

Finalmente, a mis compañeros de estudio y a todas las personas que formaron parte de este proceso académico, con quienes compartí experiencias, aprendizajes y momentos únicos e irrepetibles que quedarán para siempre.

Juan Diego Rivera Caicedo

Contenido

Introducción.....	14
1. El problema de investigación.....	16
1.1 Objeto o Tema de Investigación.....	16
1.2 Línea de Investigación.....	16
1.3 Sub Línea de Investigación.....	16
1.4 Planteamiento o Descripción del Problema.....	16
1.5 Formulación del Problema.....	17
1.6 Objetivos.....	17
1.6.1 Objetivo General.....	17
1.6.2 Objetivos Específicos	17
1.7 Justificación.....	18
1.8 Delimitación del problema	19
2. Tópicos del marco teórico	20
2.1 Antecedentes.....	20
2.1.1 Software Para La Estimación De Posición y Velocidad De La Prueba Tapping Test Mediante Visión Por Computador (2024)	20
2.1.2 A Review of the Evolution of Vision-Based Motion Analysis and the Integration of Advanced Computer Vision Methods Towards Developing a Markerless System. (2018)	20
2.1.3 An overview of Human Action Recognition in sports based on Computer Vision. (2022).....	21

2.1.4	Human Motion Analysis & Synthesis using Computer Vision and Graphics Techniques. State of Art and Applications. (2001)	22
2.1.5	Human Action Recognition Systems: A Review of the Trends and State-of-the-Art (2024)	22
2.2	Enunciados de los supuestos teóricos	23
2.2.1	La relación entre coordinación y desempeño motor	23
2.2.2	La relación entre desempeño motor y análisis de movimientos mediante video	23
2.2.3	Revisión sistemática de literatura (RSL)	24
3.	Metodología.....	25
3.1	Tipo de investigación	25
3.2	Protocolo de revisión.....	25
	Figura 1. Formulación de investigación.....	26
3.2.1	Planificación	26
3.2.2	Búsqueda.....	27
3.2.3	Selección de artículos	27
3.2.4	Evaluación	27
3.2.5	Extracción y síntesis	28
3.3	Técnicas de recolección de información	28
3.4	Instrumentos de recolección de la información.....	28
4.	Resultados.....	29

	10
4.1 Selección de artículos mediante la revisión sistemática de literatura.....	29
4.1.1 Planificación	29
4.1.2 Búsqueda.....	30
4.1.3 Selección de artículos	31
4.1.4 Evaluación de resultados	32
4.2 Análisis documental de la Revisión Sistemática de Literatura	34
4.2.1 Técnicas de visión por computador	34
4.2.2 Root Mean Square Error (RMSE)	42
4.2.3 Comparación de las técnicas según su nivel de error	42
4.2.4 Determinación de nivel según su técnica de visión por computador.....	44
4.2.5 Tipos de movimientos más estudiados del tren superior	45
4.2.6 Evaluación de las técnicas de visión por computador para cada tipo de movimiento	53
4.2.7 Tecnologías que se utilizaron para registrar el movimiento del tren superior del cuerpo.....	55
4.2.8 Oportunidades de investigación.....	61
5. Discusión de resultados	66
Conclusiones	70
Recomendaciones	71
Referencias	72

Lista de figuras

Figura 1. Formulación de investigación	26
Figura 2. Esquemático PRISMA	33
Figura 3. Clasificación de los diferentes niveles de estimación.....	45
Figura 4. Flexión y extensión del codo[52].....	48
Figura 5. Abducción, aducción y elevación del hombro[22]	49
Figura 6. Movimientos funcionales compuestos[22]	50
Figura 7. Posturas del tronco superior[33]	51
Figura 8. Muñeca/antebrazo[52]	52
Figura 9. Clasificación de niveles de exactitud.....	54
Figura 10. Error angular de cada tecnología	61

Lista de tablas

Tabla 1. Resumen del archivo de recolección de los metadatos (Excel).....	34
Tabla 2. Técnicas de visión por computador.....	35
Tabla 3. Cuadro comparativo de las técnicas.	43
Tabla 4. Movimientos más estudiados.	46
Tabla 5. Cuadro comparativo tipos de movimientos.....	53
Tabla 6. Tipos de tecnologías para registrar el movimiento del tren superior.	55
Tabla 7. Cuadro comparativo tipos de tecnologías.....	59
Tabla 8. Oportunidades de investigación.	62

Glosario de palabras

Abreviatura	Significado
CNN	Convolutional Neural Networks (Redes Neuronales Convolucionales)
Deep Learning	Aprendizaje profundo
GCN / GNN	Graph Convolutional Networks (Redes neuronales de grafos)
IMU	Inertial Measurement Unit (Unidad de Medición Inercial)
LSTM	Long Short-Term Memory (Memoria a Largo Corto Plazo)
Markerless	Estimación de pose sin marcadores
Pose Estimation	Estimación de pose
RMSE	Root Mean Square Error (Error Cuadrático Medio)
RNN	Recurrent Neural Networks (Redes neuronales recurrentes)
SVM	Support Vector Machine (Máquina de Vectores de Soporte)
RGB	Red Green Blue (Rojo Verde Azul)
RGB-D	Red Green Blue-Depth (Rojo Verde Azul profundidad)
HAR	Human Activity Recognition (Reconocimiento de Acciones Humanas)
FPS	Frames per second (Fotogramas por segundo)

Introducción

La visión es la capacidad de analizar el entorno a través de los ojos. Este proceso se lleva a cabo cuando la luz entra por la pupila y atraviesa el cristalino, proyectando una imagen en la retina. La visión por computador es una sub área de la inteligencia artificial que imita la visión humana para procesar datos obtenidos del mundo real y digitalizar la información. Esta tecnología se ha convertido en una herramienta fundamental en diversos campos de la industria, como la seguridad, la atención médica, la agricultura, el entretenimiento, así como en las industrias textil y automotriz (Ayma Quirita et al., 2022).

El reconocimiento es una función esencial y compleja dentro de la visión por computador. A través de esta capacidad, el sistema puede aprender a identificar formas y clasificarlas adecuadamente. Además, el seguimiento de objetos permite analizar y observar el comportamiento y las trayectorias de los objetos en estudio. Para obtener datos correctamente, es necesario dotar al ordenador de la capacidad de percibir y comprender una imagen, tratando de imitar el proceso que realizan los seres humanos. Aunque, la arquitectura y las etapas de un sistema de visión dependen de su aplicación (David et al., n.d.).

En este estudio, se desarrolló una monografía utilizando la metodología de revisión sistemática de la literatura propuesta por la Universidad de Keele (Carrizo & Moller, 2018). Se realiza una búsqueda exhaustiva de artículos en bases de datos como Scopus, IEEE y Web of Science, siguiendo un proceso que incluye las etapas de planificación, búsqueda, selección, evaluación, extracción y síntesis de la información. El propósito de esta revisión fue analizar las técnicas de visión por computador utilizadas en el seguimiento del tren superior del cuerpo humano, examinar los tipos de movimientos que se han estudiado y comprender las razones detrás de estos análisis. Asimismo, se busca identificar las oportunidades de investigación en esta área.

La investigación actual en el análisis del movimiento humano evidencia una clara transición hacia sistemas de visión por computador sin marcadores (markerless) y arquitecturas de aprendizaje profundo como redes convolucionales, transformers y redes de grafos, las cuales permiten evaluaciones biomecánicas precisas y de bajo costo en entornos clínicos, deportivos y ergonómicos. Diferentes estudios han demostrado que estos enfoques pueden alcanzar niveles de exactitud comparables a los sistemas ópticos profesionales considerados como estándar de

referencia, empleando cámaras monoculares o RGB-D. Entre los hallazgos más relevantes se destaca la eficacia de la fusión visual-inercial apoyada en marcadores artificiales para simplificar la calibración, así como la importancia de la ubicación de la cámara para lograr un análisis equilibrado del tren superior. Las futuras líneas de investigación apuntan al desarrollo de sistemas de fusión multimodal más robustos y modelos optimizados para computación en el borde, con el objetivo de habilitar soluciones portátiles y monitoreo biomecánico en tiempo real fuera de entornos de laboratorio.

1. El problema de investigación

1.1 Objeto o Tema de Investigación

Análisis de movimiento del tren superior del cuerpo humano mediante visión por computador.

1.2 Línea de Investigación

Sistemas de automatización y control: La línea de sistemas de automatización y control de la Universidad CESMAG desarrolla procesos investigativos orientados al modelamiento, simulación, diseño, desarrollo y evaluación de algoritmos de control, sistemas de control, sistemas inteligentes, control de procesos industriales, sistemas embebidos, acondicionamiento y procesamiento de señales, robótica, domótica e inteligencia artificial (Electrónica, 2015).

1.3 Sub Línea de Investigación

Biometría: La línea de investigación en biometría hace referencia a la identificación medición y análisis de las características fisiológicas (morfología de las manos, huellas digitales, iris en los ojos) y de comportamiento (forma de caminar, forma de teclear) de un individuo a través del uso combinado de sistemas hardware y software. (Electrónica, 2015).

1.4 Planteamiento o Descripción del Problema

La visión por computador, como área de la inteligencia artificial, permite interpretar escenas a través de imágenes tomadas por cámaras, convirtiéndose en una herramienta fundamental para el análisis del movimiento humano en áreas como la medicina, rehabilitación, deporte y en la ergonomía. El análisis del movimiento del tren superior del cuerpo es de suma importancia ya que forma una parte esencial en la evaluación del desarrollo motor infantil, la rehabilitación musculoesquelética, la valoración del desempeño deportivo y el monitoreo funcional en poblaciones adultas mayores. Sin embargo, aunque existen numerosos trabajos dedicados al análisis general del movimiento humano, las investigaciones más recientes demuestran una escasez de estudios centrados exclusivamente en el tren superior. Esta limitación se observa tanto en las metodologías empleadas como en la descripción detallada de la exactitud de las mediciones y su aplicación real en ámbitos clínicos, deportivos o ergonómicos. La ausencia de información dificulta el avance de las tecnologías creadas para monitorear el tren superior del cuerpo, generando complicaciones de aplicabilidad en entornos reales. Se necesitan sistemas eficaces,

económicos, adaptables y de fácil implementación. Por lo tanto, es necesario analizar la documentación existente para evaluar la investigación realizada sobre el tren superior mediante visión por computador, los métodos empleados, su eficiencia y sus campos de aplicación con el propósito de responder a las siguientes preguntas de investigación:

- ¿Qué técnicas de visión por computador han sido utilizadas para realizar el seguimiento del tren superior del cuerpo?
- ¿Qué tipo de movimientos del tren superior del cuerpo humano han sido estudiados y por qué razón?
- ¿Qué tecnologías se usan para registrar el movimiento del tren superior del cuerpo?
- ¿Qué oportunidades de investigación existen en el área?

1.5 Formulación del Problema

¿Qué análisis basados en visión por computador han sido usados en el estudio de movimientos del tren superior del cuerpo humano?

1.6 Objetivos

1.6.1 Objetivo General

Determinar a nivel bibliográfico el tipo de análisis basados en visión por computador que han sido empleados para caracterizar los movimientos del tren superior del cuerpo humano.

1.6.2 Objetivos Específicos

- Seleccionar documentos científicos relevantes en relación con el análisis del movimiento del tren superior del cuerpo a través de visión por computador.
- Analizar sistemáticamente los documentos seleccionados de acuerdo con las preguntas que orientan la investigación.
- Sintetizar la información obtenida para dar respuesta a las preguntas que orientan la investigación.

1.7 Justificación

Este proyecto se desarrolló en el marco del programa de Ingeniería Electrónica, en la línea de Automatización, Control y Biometría, con el objetivo de interpretar y analizar las diferentes técnicas utilizadas en la visión por computador para el estudio del movimiento del tren superior del cuerpo humano. La visión por computador ha sido implementada progresivamente en aplicaciones médicas, deportivas y ergonómicas, debido a la necesidad de reducir la subjetividad que presentan las evaluaciones realizadas únicamente por expertos humanos. Esta tecnología permite obtener mediciones más objetivas, precisas y reproducibles, facilitando el análisis sistemático de actividades relacionadas con la movilidad, la fuerza y la coordinación.

A pesar de este crecimiento, durante la revisión preliminar se evidenció que existen muy pocos estudios centrados específicamente en el tren superior del cuerpo humano dentro de contextos como el deporte, el ámbito laboral y la rehabilitación. Aunque se han desarrollado diversas técnicas con distintos niveles de rendimiento (dependiendo del tipo de movimiento, el entorno de captura y la finalidad del análisis) no se encontró una fuente que recopile y organice de forma sistemática estos métodos aplicados exclusivamente al tren superior. Esta ausencia de información hace necesario realizar un proceso de búsqueda exhaustivo que permita identificar, comparar y clasificar las técnicas existentes dando lugar a esta revisión sistemática de literatura.

Además, la importancia de analizar el desempeño del tren superior del cuerpo se basa en que se utiliza en diversas actividades físicas y manuales que requieren coordinación, precisión, fuerza. Evaluar su rendimiento es indispensable para lograr optimizar la ergonomía en el diseño de instrumentos, herramientas, incrementar el rendimiento en movimientos repetitivos, así como evitar el riesgo de lesiones. Además, un tren superior funcional lleva a una buena distribución de cargas en actividades biomecánicas, que mejora el rendimiento físico en aplicaciones del día a día, así como de deportes (Tanineh & Halaweh, 2023).

1.8 Delimitación del problema

El objetivo de este proyecto fue realizar una revisión sistemática de literatura sobre artículos relacionados con la caracterización de movimientos del tren superior del cuerpo humano mediante visión por computador. Se han establecido criterios de inclusión y exclusión para la selección de artículos, los cuales deben ser de carácter científico y estar en español o inglés. Se excluirán documentos con más de 5 años de antigüedad, así como libros y revisiones de literatura, porque no constituyen fuentes primarias de datos experimentales ni estudios originales. La síntesis de la información recolectada responde a las siguientes preguntas orientadoras:

- ¿Qué técnicas de visión por computador se utilizan para analizar el movimiento del cuerpo?
- ¿Qué tipos de movimientos del tren superior del cuerpo humano han sido investigados y con qué propósito?
- ¿Qué tecnologías se usan para registrar el movimiento del tren superior del cuerpo?
- ¿Qué oportunidades de investigación existen en el área?

2. Tópicos del marco teórico

2.1 Antecedentes

2.1.1 Software Para La Estimación De Posición y Velocidad De La Prueba Tapping Test Mediante Visión Por Computador (2024)

El trabajo (Alexander & Urbina, 2024) se presenta como un antecedente fundamental para esta revisión sistemática de literatura debido a su enfoque en la automatización y objetividad de las pruebas de coordinación motora. Mediante el uso de herramientas de visión artificial (OpenCV y Mediapipe) y la captura de video de alta velocidad a 240 FPS (Frames per second), los autores proponen un sistema capaz de mitigar la subjetividad y el error humano asociados a la evaluación manual tradicional del desempeño motor. El rigor metodológico del documento queda evidenciado en su validación estadística, donde alcanzó una exactitud del 83.91% demostrando una alta fiabilidad en la medición de tiempos y posiciones. Se considera un antecedente fundamental para el análisis de movimiento porque aporta evidencia empírica directa sobre la viabilidad de los sistemas markerless (OpenCV y Mediapipe) para mitigar la subjetividad y el error del evaluador humano, problemas centrales identificados en la RSL. El estudio presenta la transición hacia mediciones objetivas al reportar un nivel de exactitud del 83.91%. Mientras la revisión sistemática categoriza globalmente las tecnologías aplicadas a hombro, codo y tronco, este trabajo profundiza en la precisión cinemática de una de esas tecnologías (estimación de pose) aplicada a una evaluación motora concreta, validando así la eficacia de las arquitecturas de aprendizaje profundo que la RSL busca clasificar.

2.1.2 A Review of the Evolution of Vision-Based Motion Analysis and the Integration of Advanced Computer Vision Methods Towards Developing a Markerless System. (2018)

El trabajo (Colyer et al., 2018) se centra en el análisis de movimiento en general, incluyendo movimientos en deportes, biomecánica y rehabilitación, desde métodos manuales y basados en marcadores físicos, hasta las técnicas más recientes que utilizan visión por computador sin marcadores. Aunque estos nuevos sistemas, impulsados por la industria del entretenimiento, ofrecen la posibilidad de capturar el movimiento sin interrumpir el entorno natural del usuario, aún enfrentan desafíos en cuanto a su precisión y robustez para ser útiles en biomecánica y rehabilitación. Los autores destacan la necesidad de una mayor colaboración entre expertos en

visión por computador y biomecánicos para desarrollar sistemas que puedan aplicarse de manera práctica en entornos de entrenamiento y rehabilitación cotidianos. El documento se centra en el análisis de movimiento en general, incluyendo movimientos en deportes, biomecánica y rehabilitación. El trabajo que se propone a diferencia del ya existente se centró únicamente en el análisis del movimiento del tren superior del cuerpo mediante las técnicas de visión por computador existentes.

2.1.3 An overview of Human Action Recognition in sports based on Computer Vision. (2022)

El trabajo (Host & Ivašić-Kos, 2022) se centra en el Reconocimiento de Acciones Humanas (HAR) en deportes utilizando visión por computador. En este estudio, se buscó identificar, rastrear y clasificar las acciones de los jugadores en deportes como voleibol, fútbol, baloncesto y tenis, empleando HAR para monitorear el rendimiento de los atletas durante entrenamientos, competencias o partidos. Los autores se enfocaron tanto en las acciones individuales como en las interacciones entre jugadores, y resaltaron la complejidad de las acciones deportivas, que van desde movimientos simples, como correr, hasta acciones más complejas que involucran múltiples movimientos o interacciones con objetos, como golpear un balón. Este estudio también destacó el uso de métodos avanzados de aprendizaje automático y aprendizaje profundo, tales como redes neuronales convolucionales (CNN) y redes recurrentes (RNN), para mejorar la precisión del HAR. Sin embargo, este trabajo se centra principalmente en el análisis del cuerpo completo de los jugadores y en el uso de HAR en deportes específicos. La presente revisión se diferencia de esta investigación en su enfoque especializado. A diferencia de los estudios previos que consideran el cuerpo completo, la revisión propuesta se centra únicamente en el análisis del movimiento del tren superior del cuerpo humano mediante visión por computador. Con un enfoque más sistemático y detallado, esta revisión permitió profundizar en las técnicas y algoritmos que resultan particularmente útiles para el seguimiento y análisis del tren superior en actividades deportivas, identificando también oportunidades y vacíos de investigación en el área. Este enfoque específico aportó conocimientos valiosos para optimizar y mejorar las aplicaciones de visión por computador en el análisis de movimientos complejos y específicos del tren superior del cuerpo humano.

2.1.4 Human Motion Analysis & Synthesis using Computer Vision and Graphics Techniques. State of Art and Applications. (2001)

Uno de los trabajos más destacados relacionados con el análisis del movimiento del tren superior mediante visión por computador es (Perales, 2001). Un sistema que trata sobre el análisis y la síntesis del movimiento humano utilizando técnicas de visión por computador y gráficos. Se enfoca en la captura y el análisis del movimiento humano para diversas aplicaciones, desde la animación por computador y los videojuegos, hasta el diagnóstico médico y los sistemas de vigilancia. En esta investigación exploraron los métodos para capturar y reconstruir movimientos humanos, utilizando modelos gráficos y técnicas de visión por computador para crear representaciones tridimensionales precisas. También analizaron el seguimiento del movimiento humano sin usar sensores invasivos, con un enfoque en sistemas que permitan la captura en tiempo real en entornos controlados, y aunque los autores señalan que muchos de los sistemas y técnicas propuestos están en una etapa de desarrollo, lograron avances significativos, especialmente en la comparación de tecnologías de captura de movimiento y en la exploración de métodos no invasivos basados en visión por computador. Este antecedente es relevante para la presente monografía, ya que ambos estudios se centran en la utilización de visión por computador para el análisis del movimiento. Sin embargo, mientras el trabajo mencionado se enfoca más en aplicaciones como videojuegos, animación por computador, sistemas de vigilancia etc., el presente estudio busca explorar las técnicas que mejor podrían acoplarse al análisis del movimiento del tren superior del cuerpo, proporcionando un enfoque más integral y diversificado sobre las aplicaciones de estas técnicas en campos como la medicina, el deporte y la rehabilitación.

2.1.5 Human Action Recognition Systems: A Review of the Trends and State-of-the-Art (2024)

El avance de los sistemas que identifican el movimiento humano por medio de la visión por computador también está vinculado con esta labor. En este contexto, el trabajo realizado por (Karim et al., 2024) propone una visión extensa de las tendencias y la situación actual en los procedimientos para identificar acciones humanas, enfatizando la importancia de tecnologías como las redes neuronales profundas, los métodos que emplean esqueletos corporales y la estimación de poses. Su investigación se centra en la transición hacia sistemas que son menos invasivos, más precisos y eficientes. Esto va de acuerdo con el objetivo de esta investigación, que es examinar

meticulosamente el movimiento del tren superior del cuerpo y emplear técnicas de aprendizaje profundo para su análisis. Mientras que la investigación citada discute tendencias generales y algoritmos vinculados a la categorización de actos y dinámicas corporales en su totalidad, este trabajo se centra en los esquemas biomecánicos, las uniones del torso y tronco, además de las tecnologías y los métodos empleados para evaluar clínicamente su desempeño mediante el uso de métodos de visión por computador.

2.2 Enunciados de los supuestos teóricos

2.2.1 La relación entre coordinación y desempeño motor

La coordinación es fundamental en el desempeño motor, ya que se refiere a la capacidad de sincronizar y ajustar los movimientos de distintas partes del cuerpo para realizar tareas motoras de manera efectiva. La literatura indica que una coordinación adecuada mejora la precisión y la eficacia en la ejecución de movimientos, lo que es crucial en actividades deportivas y en la vida diaria. Además, se ha observado que el entrenamiento enfocado en el desarrollo de la coordinación puede traducirse en mejoras significativas en el rendimiento motor (Morillo-Baro et al., 2015). Este tipo de entrenamiento no solo fortalece la capacidad de ejecutar movimientos coordinados, sino que también fomenta el aprendizaje motor, facilitando la adquisición de nuevas habilidades y la adaptación a diferentes demandas físicas.

2.2.2 La relación entre desempeño motor y análisis de movimientos mediante video

El análisis de movimientos mediante video se ha consolidado como una herramienta crucial en la evaluación del desempeño motor. Esta metodología permite la observación detallada y la descomposición de habilidades motrices, facilitando la identificación de errores técnicos y patrones de movimiento. La investigación ha demostrado que el uso del análisis de video contribuye significativamente a la optimización del rendimiento motor, proporcionando retroalimentación visual que ayuda a los atletas a realizar ajustes precisos en su técnica (Ramzan & Dawn, 2019). Además, el análisis de movimientos mediante video fomenta una mayor conciencia corporal, lo que es vital para el desarrollo de habilidades motoras en diversas disciplinas deportivas. Por lo tanto, la implementación de esta técnica en entornos de entrenamiento se presenta como una herramienta fundamental para maximizar el desempeño motor y mejorar la ejecución de habilidades.

2.2.3 *Revisión sistemática de literatura (RSL)*

La revisión sistemática de literatura es un enfoque estructurado y metódico que busca compilar, evaluar y sintetizar la evidencia disponible sobre un tema específico. Este proceso no solo proporciona una visión integral del estado actual del conocimiento, sino que también ayuda a identificar lagunas en la investigación y a formular nuevas preguntas. Según Keele et al (Kitchenham, B., Charters, S., Budgen, D., Brereton, P., Turner, M., Linkman, S., Jørgensen, M., Mendes, E., & Visaggio, G. (2007). Guidelines for performing systematic literature reviews in software engineering. EBSE Technical Report EBSE-2007-01, Keele U, 2007), la revisión sistemática incluye varias etapas clave: la formulación de preguntas de investigación claras, la definición de criterios precisos de inclusión y exclusión, y la realización de búsquedas exhaustivas en fuentes relevantes. Este enfoque riguroso asegura que los hallazgos sean confiables y reproducibles, lo que lo convierte en una herramienta invaluable para investigadores y profesionales que buscan basar sus decisiones en evidencia sólida.

3. Metodología

3.1 Tipo de investigación

Esta investigación es de tipo descriptiva, ya que tiene como propósito analizar estudios relacionados con el movimiento del tren superior del cuerpo humano mediante técnicas de visión por computador. Para ello se realizó una revisión sistemática de literatura utilizando artículos científicos provenientes de las bases de datos Scopus, IEEE y Web of Science. A partir de la selección y evaluación de estos documentos, se identificó y organizó la información más relevante para dar cumplimiento al objetivo propuesto de investigación.

3.2 Protocolo de revisión

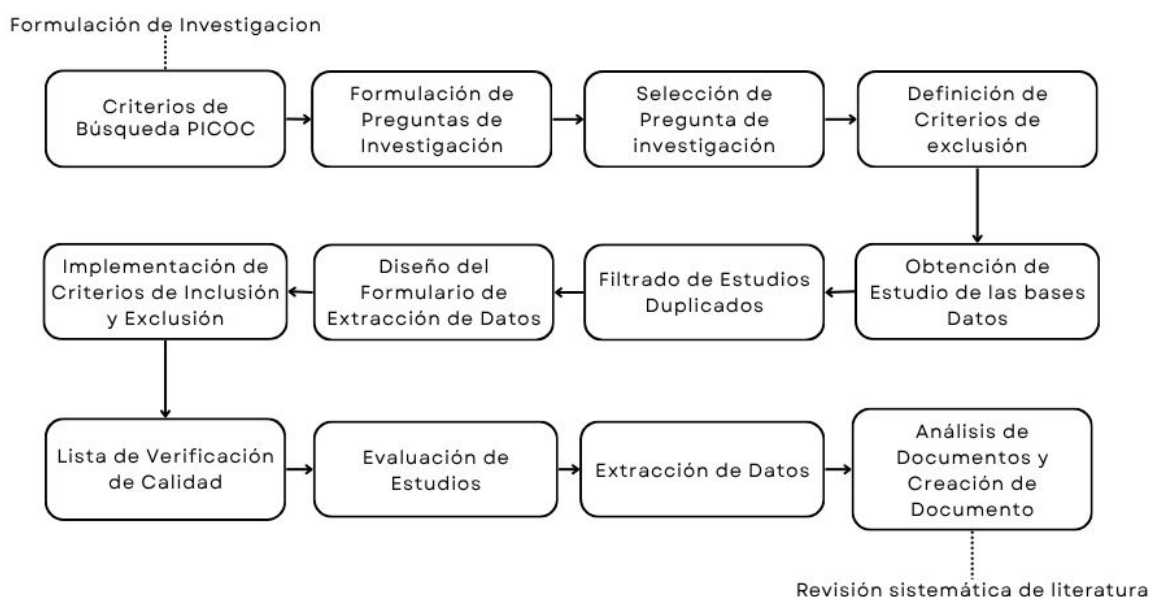
El diseño se basa en la Revisión Sistemática de Literatura de la Universidad de Keele (Kitchenham, B., Charters, S., Budgen, D., Brereton, P., Turner, M., Linkman, S., Jørgensen, M., Mendes, E., & Visaggio, G. (2007). Guidelines for performing systematic literature reviews in software engineering. EBSE Technical Report EBSE-2007-01, Keele U, 2007), la cual es conocida por su capacidad de presentar una estructura rigurosa, clara e iterativa para construir revisiones sistemáticas de literatura. Este protocolo fue elegido porque permitió realizar una investigación ordenada, transparente y compatible con los estándares académicos de ingeniería.

El protocolo establece tres fases principales: planificación, ejecución y documentación. En la fase de planificación, se formulan los objetivos, se definen las preguntas de investigación y se establecen los criterios que orientarán la selección de estudios relevantes. Esta etapa también se observa la definición preliminar de palabras clave y la identificación de fuentes potenciales de información. La fase de ejecución comprende la construcción de estrategias de búsqueda basadas en palabras clave relacionadas con el campo de estudio, la consulta de bases de datos científicas reconocidas y la aplicación de los criterios de inclusión y exclusión previamente establecidos. En esta fase también se realiza la evaluación crítica de los estudios con el fin de determinar su pertinencia y calidad metodológica, siguiendo las recomendaciones del protocolo. Finalmente, en la fase de documentación, el protocolo orienta la organización, síntesis e integración de la

información obtenida. Aquí se clasifican los estudios según categorías conceptuales o técnicas, se identifican patrones, vacíos de investigación y tendencias, y se definen los elementos necesarios para construir la base teórica y analítica del estudio.

En esta ocasión se utilizaron las herramientas: Excel, Parsifal y Zotero (de uso libre) para almacenar y organizar todo tipo de datos recopilados de los estudios seleccionados como se muestra en la figura 1.

Figura 1. Formulación de investigación



3.2.1 Planificación

En primera instancia fue necesario desarrollar una cadena de búsqueda que incluyera términos relacionados con la detección de movimiento del tren superior y visión por computador. Esta cadena debe contener palabras clave que formen un patrón de búsqueda adecuado para las bases de datos seleccionadas. Los términos de búsqueda deben enfocarse principalmente en la detección de movimiento del tren superior del cuerpo mediante visión por computador. Además, la búsqueda se realiza en inglés, ya que la mayoría de los artículos publicados están en este idioma con palabras de búsqueda como ("body") AND ("motion" OR "movement") AND ("Computer Vision").

3.2.2 Búsqueda

Para realizar la búsqueda de estudios relevantes, se utilizaron las bases de datos IEEE Xplore, Scopus y Web of Science, ya que estas plataformas ofrecen acceso a una amplia gama de investigaciones sobre el análisis de movimiento del tren superior del cuerpo humano mediante visión por computador. Se seleccionaron estas bases debido a su reputación en el ámbito académico y su extenso contenido relacionado con la biomecánica y la visión por computador. Se emplearon palabras clave como "análisis de movimiento del tren superior", "visión por computador", "captura de movimiento humano" y "biomecánica", con el objetivo de identificar estudios relacionados. Además, se delimitó la búsqueda a artículos publicados en los últimos cinco años para asegurar que la información utilizada fuera actual y relevante. Por otra parte, para filtrar los resultados, se aplicaron criterios que incluyeron solo artículos revisados por pares, disponibles en texto completo y en los idiomas inglés y español. Posteriormente, se recopiló los resultados y se eliminó los duplicados para quedarse únicamente con los estudios más relevantes para esta revisión sistemática de literatura.

3.2.3 Selección de artículos

Para la selección de los artículos se establecieron criterios de búsqueda, inclusión y exclusión dirigidos exclusivamente a documentos científicos. Inicialmente, se descartaron estudios a partir del análisis del título y del resumen, eliminando aquellos que no se relacionaban con el análisis del movimiento del cuerpo humano o con la visión por computador. También se excluyeron artículos cuya antigüedad comprometiera su vigencia (2019-2025), publicaciones que no pertenecían a revistas científicas reconocidas y documentos duplicados. La estrategia de búsqueda se construyó empleando términos clave como "body", "motion", "movement" y "computer vision", con el fin de identificar literatura pertinente al campo de estudio.

3.2.4 Evaluación

Se llevó a cabo una lectura cuidadosa de los documentos y artículos que se ha seleccionado previamente, con el objetivo de identificar los más relevantes para esta investigación. Este proceso es fundamental, ya que ayudó a responder de manera clara y precisa a las preguntas formuladas en los objetivos, que incluyen tanto la pregunta de investigación principal como el problema que se busca abordar. Al enfocarse en estos textos clave, se logró extraer información valiosa y sintetizar

los hallazgos, lo que permitió construir una base sólida para el desarrollo de la monografía. Este enfoque no solo enriquece el análisis, sino que también se acerca a una comprensión más profunda del tema en cuestión.

3.2.5 Extracción y síntesis

Para la extracción y síntesis, se usó el sitio web parsif.al como plataforma para recopilar las respuestas a las preguntas de investigación, a partir de los artículos seleccionados. Una vez recopiladas, estas respuestas se clasificaron y agruparon de acuerdo con su contenido, lo que permitió organizarlas de manera lógica. Este proceso de agrupamiento facilitó la identificación clara de las respuestas a cada una de las preguntas de investigación formuladas previamente. De este modo, se logró sistematizar la información obtenida, asegurando una mejor comprensión y estructura en los resultados del estudio.

3.3 Técnicas de recolección de información

Las técnicas de recopilación de información se basan principalmente en la consulta de bases de datos académicas de alto impacto como Scopus, IEEE Xplore y Web of Science. Estas plataformas fueron seleccionadas debido a su relevancia científica y a las características particulares que ofrecen para una investigación en el campo de la ingeniería y la visión por computador.

Scopus fue escogida por su amplia cobertura en múltiples áreas y por integrar artículos provenientes de distintas bases de datos y editoriales, lo que permite acceder a una cantidad más completa y diversa de estudios científicos. Web of Science destaca por la facilidad de obtención de artículos y por su riguroso proceso de selección, lo que garantiza la calidad y coherencia de los documentos incluidos. Finalmente, IEEE Xplore se seleccionó por ser una base de datos especializada en ingeniería, electrónica, computación y tecnologías afines, lo que la convierte en un medio indispensable para estudios relacionados con visión por computador y análisis de movimiento.

3.4 Instrumentos de recolección de la información

Este trabajo se basa en la revisión sistemática de artículos científicos y académicos relacionados con el análisis de movimiento del tren superior del cuerpo humano mediante visión

por computador. La importancia de Parsifal radica en su capacidad para estructurar el protocolo de investigación de manera detallada. Permite definir con precisión el alcance mediante el marco PICOC (Población, Intervención, Comparación, Resultados y Contexto), establecer las preguntas de investigación y diseñar estrategias de búsqueda eficaces mediante palabras clave y sinónimos. (José & Peñalvo, 2017) Esta organización previa es crucial para evitar revisiones deficientes que podrían invalidar el resto de la monografía.

4. Resultados

4.1 Selección de artículos mediante la revisión sistemática de literatura

Para la selección de artículos se adoptó el modelo metodológico de revisión sistemática propuesto por Keele (Carrizo & Moller, 2018). Este modelo propone una estructura organizada que permite la planificación, ejecución y documentación de la búsqueda de literatura de manera clara y repetible. Su modelo garantiza que el proceso de recopilación y análisis de información desarrolle unos resultados de calidad, respondiendo a criterios previamente definidos, lo que incrementa la calidad y veracidad de los hallazgos.

En este trabajo, el modelo se utilizó como marco general para la definición de palabras claves, la elección de bases de datos, la definición de criterios de inclusión y exclusión, la síntesis de la información obtenida con el fin de responder las preguntas de investigación. El uso de esta metodología permitió construir una base de artículos pertinentes y alineados con los objetivos de investigación planteados previamente.

4.1.1 Planificación

La fase de planificación consistió en la definición de la estrategia de búsqueda y de los criterios que guiarían la revisión. Se estableció la siguiente cadena de búsqueda: ("**upper body**" OR "**upper limbs**") AND ("**Action recognition**" OR "**Pose estimation**" OR "**Computer Vision**" OR "**movement analysis**") AND (2019:2025).

Con esta cadena de búsqueda se identificó y ajustó de manera progresiva las palabras clave a partir de una revisión preliminar de la literatura, utilizando un método estructurado que combina el esquema PICOC con el proceso iterativo de Parsifal para construir la cadena de búsqueda. Durante este proceso, se identificaron tres elementos esenciales: el segmento corporal que se quiere

analizar (por ejemplo, "upper limb", "upper body"), la perspectiva tecnológica "computer vision" y el tipo de análisis del movimiento "pose estimation", "motion analysis" y otros). La elección de un número acotado de sinónimos responde a la necesidad de preservar un balance entre precisión y amplitud, con el fin de evitar obtener una cantidad excesiva de resultados irrelevantes. Además, los términos empleados no se presentan como sinónimos directos, sino más bien como ideas complementarias. Sin embargo, debido a que el método es iterativo, se puede extender la cadena con términos relacionados adicionales para aumentar la cobertura y disminuir el peligro de dejar fuera estudios importantes.

Se obtuvo artículos científicos publicados en los últimos cinco años (principalmente en inglés), disponibles en texto completo, y relacionados de manera directa con el análisis de movimiento del tren superior del cuerpo humano. Esta etapa facilitó la búsqueda de documentos pertinentes y la identificación de información específica para la investigación, asegurando que el proceso se desarrollara de forma ordenada y bajo parámetros claros, lo que fortaleció la validez del estudio.

La utilización de la cadena de búsqueda en conjunto con bases de datos reconocidas como IEEE Xplore, Web of Science y Scopus facilitó tanto la localización de artículos clave, como también aseguró la consistencia y la agilidad del proceso de desarrollo de la revisión sistemática.

4.1.2 Búsqueda

En la fase de búsqueda se emplearon cadenas de búsqueda específicas para cada base de datos, con el fin de encontrar estudios relacionados con el análisis del movimiento del tren superior mediante visión por computador. Estas cadenas se construyeron combinando términos clave en inglés y operadores booleanos, adaptándose a la sintaxis propia de cada plataforma.

Para IEEE Xplore, se utilizó la siguiente cadena: (**“upper body” OR “upper limb” OR “upper extremity”**) AND (**“motion analysis” OR “movement analysis”**) AND (**“computer vision” OR “pose estimation”**).

Para Scopus, cuya estructura contiene una amplia cantidad de documentos de distintas bases de datos, se empleó la siguiente cadena: **TITLE-ABS-KEY (“upper body” OR “upper limb” OR “upper extremity”) AND TITLE-ABS-KEY (“movement” OR “motion”) AND TITLE-ABS-KEY (“computer vision” OR “pose estimation” OR “human motion analysis”)**.

Para Web of Science, debido a su sistema de indexación por temas y categorías, se aplicó la cadena: **TS = (“upper limb” OR “upper body”) AND TS = (“motion analysis” OR “movement analysis”) AND TS = (“computer vision” OR “pose estimation”)**.

4.1.3 Selección de artículos

Para la selección de artículos se aplicaron criterios de exclusión que permitieron dividir la cantidad inicial de documentos. Estos criterios permitieron la eliminación de duplicados, publicaciones con más de cinco años de antigüedad, libros, artículos de conferencias, documentos de acceso restringido, revisiones sistemáticas y aquellos con escasa relevancia para los objetivos de la investigación, esto es, solo se mantuvieron los estudios directamente relacionados con el análisis de movimiento del tren superior mediante visión por computador.

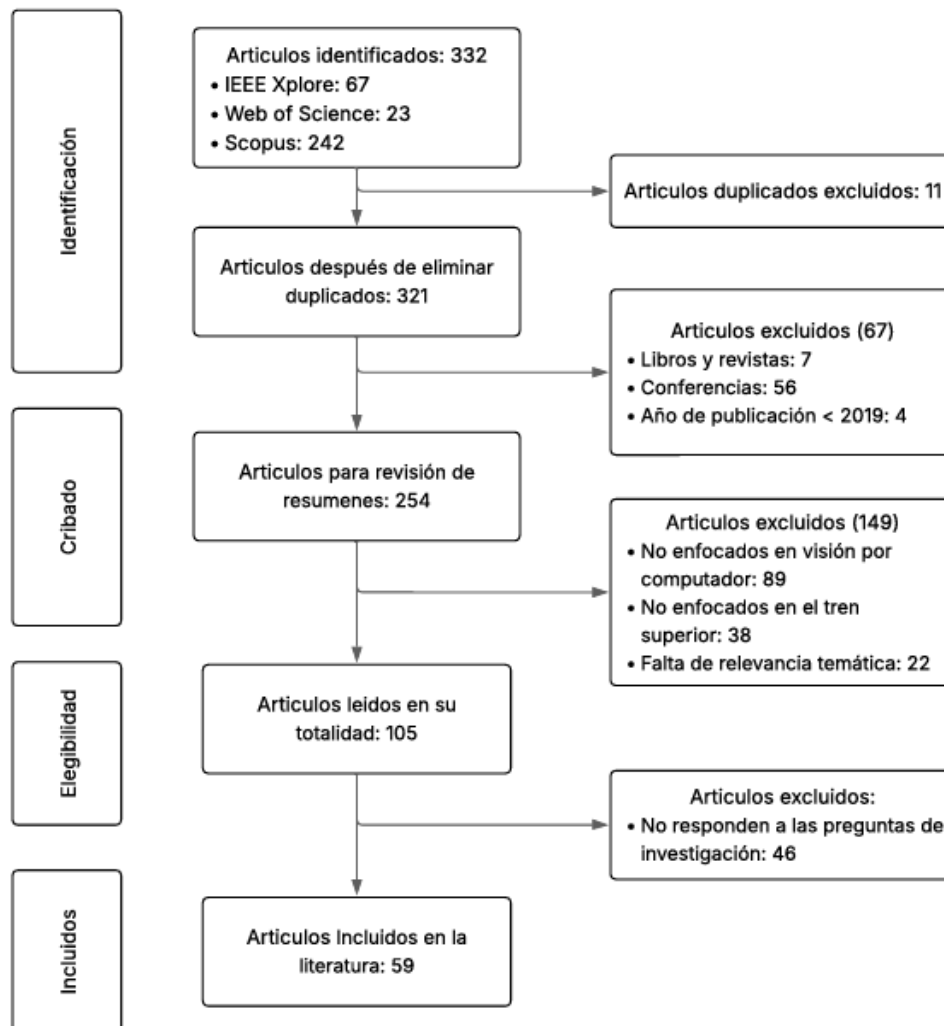
La cantidad inicial de documentos por bases de datos fue la siguiente: Web of Science entregó 23 documentos, IEEE Xplore 67 y Scopus 242, sumando un total de 332 publicaciones. Tras la aplicación de los criterios de exclusión, se descartaron 11 documentos por duplicados, quedando 321. Posteriormente, se eliminaron publicaciones por antigüedad, libros, conferencias, artículos que no pudieron ser obtenidos y revisiones sistemáticas, reduciéndose a 254. Se descartaron 149 artículos por no tener relación directa con el objeto de estudio. En su mayoría, estos documentos analizaban el movimiento humano de forma general o se enfocaban exclusivamente en el tren inferior sin visión por computador, la marcha, el cuerpo completo o en aplicaciones no vinculadas al análisis del tren superior mediante visión por computador. Tras esta depuración temática, quedaron 105 artículos que sí abordaban de manera pertinente el estudio del movimiento del tren superior. Finalmente, con los 105 documentos, se realizó la respectiva descarga y lectura. Al realizar las actividades anteriores se descartó un total de 46 documentos ya que no aclaran que técnicas de visión por computador emplean, o qué movimientos del cuerpo humano analizan, o simplemente no responde a ninguna pregunta de investigación, dejando así una total de 59 artículos que contienen información, que al menos responden una pregunta de investigación (ver Tabla 1).

A partir de este conjunto fue posible, identificar patrones generales: técnicas aplicadas,

partes del tren superior analizadas, herramientas empleadas y oportunidades de investigación. Se evidenció una recurrencia en el uso de métodos de estimación de pose y movimientos del tren superior.

4.1.4 Evaluación de resultados

De los 332 documentos encontrados en las bases de datos Scopus, IEEE Xplore y Web of Science, se seleccionaron 105 estudios para su revisión preliminar, los cuales fueron descargados y analizados. Durante este proceso se evaluó la relación de cada documento con base en el análisis del movimiento del tren superior mediante visión por computador, descartando aquellos que, a pesar de estar vinculados al movimiento humano, no abordaban técnicas, segmentos corporales o aplicaciones coherentes con los objetivos de la investigación. Además de la relación temática, se consideró la calidad metodológica de los estudios, analizando elementos como claridad en el diseño experimental, consistencia en la presentación de resultados y precisión en la descripción de las técnicas empleadas. A partir de este proceso, 46 documentos fueron descartados por no cumplir con los criterios de relevancia e interés previamente establecidos, formando así un conjunto total de 59 artículos que aportan información válida y suficiente para responder las preguntas de investigación planteadas. La figura 2 muestra el esquemático PRISMA del proceso empleado para seleccionar y filtrar artículos científicos con el objetivo de llegar a un conjunto final para su análisis. Este procedimiento asegura que los artículos incluidos sean los más relevantes para el estudio, permitiendo un análisis profundo y centrado en la literatura científica más pertinente.

Figura 2. Esquemático PRISMA

La tabla 1 evidencia que la mayoría de los documentos incluyen técnicas avanzadas de análisis basadas en aprendizaje profundo y estimación de pose, lo que indica un proceso de integración y estructura. También, se identificó que gran parte de los estudios se concentran en zonas específicas del tren superior, principalmente hombros y manos, mientras que otras regiones son menos exploradas, lo que revela vacíos relevantes en la literatura. Del mismo modo, las herramientas empleadas, entre las que destacan librerías de código abierto y frameworks de uso extendido, reflejan los estándares comunes que fortalecen la reproducibilidad y accesibilidad de las investigaciones. Finalmente, una proporción importante de los artículos analizados señalan limitaciones y proponen futuras oportunidades de investigación, orientadas a mejorar la precisión clínica, integrar dispositivos portátiles y validar los modelos en entornos no controlados.

Tabla 1. Resumen del archivo de recolección de los metadatos (Excel).

Pregunta	Artículos que mencionan una respuesta en forma directa
¿Qué técnicas de visión por computador han sido utilizadas para realizar el seguimiento del tren superior del cuerpo?	59
¿Qué tipo de movimientos del tren superior del cuerpo humano han sido estudiados y por qué razón?	47
¿Qué tipo de tecnologías se ocupan para lograr registrar el movimiento del tren superior del cuerpo?	53
¿Cuáles son las oportunidades de investigación en el área?	39

4.2 Análisis documental de la Revisión Sistemática de Literatura

4.2.1 Técnicas de visión por computador

Se identifica una marcada prevalencia de dos técnicas: las Convolutional Neural Networks (CNN), empleadas en 37 estudios, y los métodos de Pose Estimation Markerless en 2D y 3D, presentes en 40 estudios.

La Tabla 2 muestra las diferentes técnicas de visión por computador identificadas. Se encontraron nueve técnicas distintas, cada una apoyada por múltiples fuentes. La relación de estas técnicas de visión por computador permite entender cuáles son los métodos más utilizados. La tabla relaciona la técnica con la fuente, es decir con la referencia bibliográfica del documento que la utiliza.

Tabla 2. Técnicas de visión por computador.

N	Técnicas de visión por computador	Fuente
1	Pose Estimation Markerless (2D/3D)	<p>(Moreira et al., 2024)(Li & Dong, 2023)(Koul & Novembre, 2025)(Scott et al., 2022)(Adolf et al., 2025)(Kim et al., 2025)(Boudreault-Morales et al., 2025)(Ceglia et al., 2025)(Lee & Tan, 2025)(Jia et al., 2025)(Edriss et al., 2024)(Jin et al., 2025)(Kwon, 2025)</p> <p>(Harrison et al., 2025)(Behera et al., 2020)(Sun et al., 2025)(Q. Gao et al., 2025)(Chang et al., 2021)(Zhou et al., 2025)(Suo et al., 2025)(Jackson et al., 2023)(Abromavičius et al., 2025)(A. Zhao et al., 2025)(Paiva et al., 2025)(Xu & Xu, 2025)(Xu & Xu, 2024)</p> <p>(Lovanshi et al., 2024)(Zaher et al., 2025)(Cao et al., 2019)(Ryu et al., 2025)(Tanthuwapathom et al., 2025)(Esmail et al., 2025)(Y. Gao et al., 2025)(Doong, 2025)(Lin et al., 2025)(Kryeem et al., 2025)(Dellepiane et al., 2025)(Li & Yu, 2023)(Wu et al., 2025)</p> <p>(Shahid et al., 2025)</p>
2	Convolutional Neural Networks (CNN)	<p>(Murillo et al., 2025)(Shi et al., 2024)(W. Zhao et al., 2024)(Q. Gao et al., 2025)(Shahid et al., 2025)(Sun et al., 2025)(Yan et al., 2025)(Zaher et al., 2025)(Ryu et al., 2025)(Cao et al., 2019)(Moreira et al., 2024)(Koul & Novembre, 2025)(Esmail et al., 2025)</p> <p>(Edriss et al., 2024)(Giulietti et al., 2025)(Li & Dong, 2023)(Boudreault-Morales et al., 2025)(Adolf et al., 2025)(Abromavičius et al., 2025)(A. Zhao et al., 2025)(Scott et al., 2022)(Jackson et al., 2023)(Suo et al., 2025)(Kryeem et al., 2025)(Behera et al.,</p>

- 2020)(Harrison et al., 2025)
 (Doong, 2025)(Jin et al., 2025)(Dellepiane et al.,
 2025)(Tanthuwapathom et al., 2025)(Y. Gao et al.,
 2025)(Xu & Xu, 2025)(Paiva et al., 2025)(Xu & Xu,
 2024)(Lovanshi et al., 2024)(Kuang et al.,
 2025)(Wang et al., 2025)
- 3 Transformers** (Kryeem et al., 2025)(Behera et al., 2020)(Suo et al.,
 2025)(Wang et al., 2025)(Lovanshi et al., 2024)(Xu &
 Xu, 2024)(Kuang et al., 2025)(Y. Gao et al.,
 2025)(Cao et al., 2019)(Doong, 2025)(Paiva et al.,
 2025)(Xu & Xu, 2025)(Sun et al., 2025)
 (Zaher et al., 2025)(Wu et al., 2025)
- 4 RNN / LSTM** (Zaher et al., 2025)(Sun et al., 2025)(Kwon,
 2025)(Wang et al., 2025)(Abrar Ashraf et al.,
 2025)(Giulietti et al., 2025)(Cao et al., 2019)(Behera
 et al., 2020)(Kryeem et al., 2025)(Wu et al., 2025)(A.
 Zhao et al., 2025)(Doong, 2025)
- 5 Reconstrucción/Estimación 3D** (Lin et al., 2025)(Ceglia et al., 2025)(A. Zhao et al.,
 2025)(Li & Dong, 2023)(Li & Yu, 2023)(Kim et al.,
 2025)(Sun et al., 2025)(Jiang et al.,
 2025)(Abromavičius et al., 2025)(Jia et al., 2025)(Y.
 Gao et al., 2025)
- 6 Graph Convolutional Networks** (Lovanshi et al., 2024)(Wu et al., 2025)(Xu & Xu,
 (GCN/GNN) 2024)(Paiva et al., 2025)(Kuang et al., 2025)(Xu &
 Xu, 2025)
- 7 Filtros de Kalman / Seguimiento** (Chang et al., 2021)(Li & Yu, 2023)(Li & Dong,
 estadístico 2023)(Giulietti et al., 2025)(Abromavičius et al.,
 2025)(Esmail et al., 2025)
- 8 SVM / Métodos clásicos** (Kamal et al., 2025)(Doong, 2025)(Jin et al., 2025)(Jia
 et al., 2025)(Harrison et al., 2025)
-

Las técnicas de visión por computador utilizadas por los documentos revisados, se explican a continuación:

4.2.1.1 Pose Estimation Markerless (2D/3D)

La estimación de pose markerless se refiere al uso de cámaras RGB o RGB-D y algoritmos de visión por computador para identificar y seguir posturas de las articulaciones en 2D o 3D sin la necesidad de colocar marcadores físicos en el cuerpo. Esta técnica ha tenido relevancia en rehabilitación y análisis del movimiento porque no es tan invasiva, es más económica y permite aplicaciones en entornos clínicos y cotidianos. Al eliminar los marcadores, los sistemas se vuelven más accesibles para pacientes y profesionales, aunque presentan desafíos como la calibración de cámaras y la precisión en presencia de obstrucciones.

En los artículos analizados se ha identificado que los modelos de estimación de pose basados en deep learning han permitido cuantificar con detalle movimientos en rehabilitación, aportando ventajas frente a los sistemas de captura con marcadores por su menor costo y mayor facilidad de uso (Boudreault-Morales et al., 2025). Por otra parte, también se ha demostrado que la ubicación de la cámara es de gran relevancia en la precisión clínica de los resultados, lo que indica la necesidad de optimizar la geometría de captura en sistemas de este tipo (Kim et al., 2025).

4.2.1.2 Convolutional Neural Networks (CNN)

Las Redes Neuronales Convolucionales (CNN) son arquitecturas de aprendizaje profundo (Deep Learning) diseñadas para procesar datos visuales como imágenes o secuencias de video. Se diferencian por el uso de filtros convolucionales capaces de identificar patrones locales (como bordes o texturas) y combinarlos jerárquicamente para obtener resultados de mayor nivel. Esta estructura permite reconocer y modelar movimientos humanos de manera más precisa y efectiva.

En la literatura revisada se resalta que las CNN permiten el análisis de datos complejos del tren superior, al extraer automáticamente características relevantes de la ejecución de movimientos durante entrenamientos de rehabilitación, permitiendo detectar la eficacia de los ejercicios y diferenciar patrones de movimiento en pacientes con alteraciones funcionales (Shi et al., 2024). Por otro lado, otros artículos han mostrado cómo estas arquitecturas pueden emplearse para modelar y realizar movimientos del miembro superior en sistemas artificiales, evidenciando su potencial en ámbitos como la robótica y la biomecánica (Murillo et al., 2025).

La importancia de las CNN en los diferentes documentos se basa en su robustez frente a grandes volúmenes de datos y en su capacidad de adaptación a distintos escenarios, desde la rehabilitación hasta la robótica y la ergonomía. Por lo tanto, constituyen una técnica fundamental en la visión por computador aplicada al análisis del movimiento humano.

4.2.1.3 Transformers

Los Transformers constituyen una arquitectura de aprendizaje profundo inicialmente diseñada para el procesamiento de lenguaje natural, pero que en los últimos años ha sido adaptada de manera exitosa a tareas de visión por computador. La técnica se basa en el mecanismo de atención, que permite identificar y analizar de manera dinámica las relaciones más relevantes entre diferentes partes de una secuencia de datos. En el caso del análisis de movimiento del cuerpo humano, esta capacidad se identifica como una ventaja apreciativa para modelar dependencias espaciales y temporales de las diferentes partes del cuerpo.

A diferencia de las redes convolucionales, en los Transformers no es necesario un procesamiento estrictamente secuencial, optimizando la captura de información a gran escala y favoreciendo el entrenamiento en paralelo. Esto los convierte en una clara alternativa para abordar escenarios donde el movimiento humano debe ser entendido en su contexto completo, como la detección de micro gestos, la predicción de trayectorias articulares y la evaluación simultánea de múltiples articulaciones en tiempo real (Xu & Xu, 2024).

En la revisión realizada, se identificó que esta técnica aprovecha el reconocimiento de emociones a partir de esqueletos 3D, mostrando ventajas en la incorporación de información temporal y espacial que otras arquitecturas no logran capturar con la misma precisión (Paiva et al., 2025). También, se destacan sus aportes en sectores como la rehabilitación y la ergonomía, donde los Transformers han logrado facilitar la detección de posturas de riesgo y la evaluación del desarrollo en ejercicios clínicos, generando modelos más robustos y generalizables (Xu & Xu, 2025).

En conclusión, los Transformers se consolidan como una de las técnicas emergentes más prometedoras en el análisis de movimiento, ya que al integrar de manera eficiente la dimensión temporal y espacial. Su capacidad para adaptarse a distintas configuraciones de datos (videos, secuencias esqueléticas, sensores híbridos) amplía las posibilidades para investigaciones futuras, como aplicaciones clínicas y contextos de interacción humano-máquina.

4.2.1.4 RNN / LSTM

Una Red Neuronal Recurrente (RNN) es una arquitectura diseñada para el procesamiento secuencial de datos mediante conexiones que retroalimentan información de estados anteriores hacia el presente. Esto les permite modelar dependencias temporales en series de tiempo, lo cual es útil para el análisis del movimiento humano. Sin embargo, las RNN tradicionales presentan limitaciones a la hora de enfrentarse a secuencias largas, debido al problema del desvanecimiento del gradiente.

Para resolver esta problemática, se han desarrollado variantes como las Long Short-Term Memory (LSTM), que incorporan puertas de entrada, olvido y salida para gestionar de manera más eficiente la memoria a largo plazo. Gracias a esta estructura, las LSTM logran capturar patrones temporales complejos y mantener información relevante durante periodos más prolongados, lo que las hace ideales para el seguimiento del movimiento del tren superior en tareas de predicción y clasificación.

En los estudios revisados se identificaron distintas aplicaciones de esta técnica. En contextos de simuladores dinámicos en tiempo real, las LSTM se han utilizado para mejorar la precisión en la medición de la pose humana y la integración humano-máquina (Giulietti et al., 2025). En el ámbito deportivo, han demostrado su efectividad al ser usadas en redes espaciotemporales para la estimación 3D de la pose durante entrenamientos de artes marciales, capturando de manera eficiente tanto las dependencias espaciales como las temporales (Sun et al., 2025). Finalmente, en el campo de la rehabilitación, la combinación de RNN y CNN permitió clasificar con alta precisión secuencias de ejercicios del tren superior, mejorando la detección de patrones motores atípicos y el seguimiento del progreso clínico (Zaher et al., 2025).

4.2.1.5 Reconstrucción/Estimación 3D

La reconstrucción o estimación tridimensional (3D) es relevante en visión por computador aplicada al análisis de movimiento humano. Esta técnica busca generar representaciones volumétricas o modelos esqueléticos 3D a partir de imágenes 2D capturadas por cámaras Red Green Blue (RGB) o Red Green Blue - Depth (RGB-D). Su mayor atractivo facilita la captura más exacta del movimiento de las articulaciones, lo que resulta esencial para entender la biomecánica de la parte superior del cuerpo en contextos clínicos, deportivos o ergonómicos

En oposición a las metodologías planas, la evaluación en tres dimensiones minimiza las

confusiones que provienen del punto de vista y ofrece datos más detallados sobre los ángulos de las articulaciones, el rango de movimiento y la sinergia entre las partes del cuerpo. Esto es especialmente útil en la rehabilitación, donde se necesita un examen minucioso del avance en los ejercicios, así como en ámbitos industriales y deportivos que requieren un monitoreo exacto de la postura y la forma de realizar los movimientos.

En los artículos analizados, se pueden ver distintas metodologías para lograr este fin. Algunos estudios emplearon redes neuronales que predicen de manera directa la ubicación tridimensional de las articulaciones a partir de secuencias de video, resaltando mejoras en la precisión clínica y en su resistencia frente a obstrucciones (Lin et al., 2025). También se encontraron investigaciones que combinan técnicas multimodales, como la integración de sensores de profundidad y datos visuales, con el objetivo de mejorar la estimación de las articulaciones del hombro y el codo en tareas complejas (Jiang et al., 2025).

4.2.1.6 Graph Convolutional Networks (GCN/GNN)

Las Graph Convolutional Networks (GCN) y, en general, las Graph Neural Networks (GNN), son estructuras que han ganado importancia en la visión por computador porque permiten representar datos en forma de grafo. Esto es clave para el análisis del movimiento humano, ya que el esqueleto puede modelarse como un conjunto de articulaciones conectados por segmentos corporales.

A diferencia de las CNN, que analizan imágenes organizadas en una cuadrícula de píxeles, las GCN se centran directamente las relaciones entre articulaciones. De esta manera, cada nodo articula información tanto de sí mismo como de sus vecinos, logrando una representación más real de la biomecánica del tren superior.

En los artículos inspeccionados, estas técnicas se aplicaron en diferentes contextos. Por un lado, se usaron para reconocimiento de acciones y posturas, aprovechando la estructura natural del esqueleto humano (Xu & Xu, 2024). También se aplicaron en predicción de movimiento, al integrar información espacial y temporal para anticipar trayectorias articulares con mayor precisión (Kuang et al., 2025). Asimismo, en escenarios clínicos, las GCN fueron útiles para la evaluación de ejercicios de rehabilitación, ya que modelan las dependencias entre hombro, codo y muñeca de una manera más robusta que otros enfoques (Paiva et al., 2025).

El valor principal de estas técnicas radica en que no solo analizan articulaciones de forma aislada, sino también las interacciones entre ellas. Esto permite obtener modelos más fiables para

tareas como la rehabilitación física, la ergonomía ocupacional y el análisis deportivo, abriendo nuevas oportunidades de investigación que priorizan la relación entre segmentos corporales y no únicamente el movimiento individual de cada articulación.

4.2.1.7 Filtros de Kalman / Seguimiento estadístico

Un filtro de Kalman es una técnica que aplica un método estadístico que permite una estimación continua, facilitando la predicción y ajuste del estado de un sistema dinámico utilizando información imprecisa. Su función trata de unir un modelo de predicción con mediciones observadas, ajustando ambos en función de la falta de precisión presente. En el estudio del movimiento humano a través de la visión por computador, se emplean los filtros de Kalman y sus versiones (Extended Kalman Filter, Unscented Kalman Filter) para suavizar las trayectorias de las articulaciones, corregir fallos en la detección y aumentar la consistencia de los datos de la postura.

En el conjunto de artículos revisados se encuentran investigaciones que utilizan esta metodología con diferentes objetivos. Por ejemplo, se utilizó un filtrado adaptable para mejorar la estimación de posiciones en videos, lo que aumentó la fiabilidad ante obstrucciones y cambios en la detección (Chang et al., 2021). Además, se implementaron filtros de Kalman en sistemas visuales e inerciales para la determinación del tren superior, logrando una mayor exactitud en los ángulos de las articulaciones gracias a la fusión de predicciones estadísticas con datos de sensores (Li & Yu, 2023).

4.2.1.8 SVM / Métodos clásicos

Las técnicas tradicionales de aprendizaje automático, como Support Vector Machines (SVM), los árboles de decisión o ensambles, han desempeñado un papel clave en el estudio del movimiento humano antes del surgimiento de las redes neuronales profundas. Estas metodologías se distinguen por utilizar características que han sido diseñadas de manera más manual (por ejemplo, ángulos de articulaciones, vectores de keypoints o métricas de similitud), lo que las convierte en opciones más ligeras en términos de computación y menos dependientes de grandes cantidades de datos etiquetados.

En la investigación analizada se identificaron herramientas que respaldan esta efectividad. Por ejemplo, se utilizaron SVMs para clasificar automáticamente las actividades humanas basándose en características híbridas obtenidas de videos, presentando una opción eficiente en

comparación con arquitecturas profundas que son más costosas a nivel de implementación de tecnologías como cámaras, sensores o personal experimentado para la manipulación de dichas tecnologías (Kamal et al., 2025). De manera análoga, en el ámbito de la ergonomía y la seguridad laboral, se han aplicado modelos tradicionales para anticipar los niveles de riesgo postural a partir de imágenes y datos esqueléticos, proporcionando herramientas de evaluación que son accesibles y de fácil adaptación (Doong, 2025). Por último, también se menciona el uso de técnicas que se basan en conjuntos, como Random Forest o métodos de aprendizaje colaborativo, para identificar posturas sedentarias y patrones de actividad en tiempo real, considerando su sencillez, los hace ideales para el uso en entornos clínicos o en hogares (Jin et al., 2025).

4.2.2 *Root Mean Square Error (RMSE)*

El error cuadrático medio (RMSE, por sus siglas en inglés) es una medida estadística que se utiliza para analizar el rendimiento de un modelo. Se define como la raíz cuadrada del promedio de los errores al cuadrado entre las cifras estimadas y los valores verdaderos. Cuando los errores del modelo no están sesgados y tienen una distribución normal, el RMSE es particularmente adecuado porque proporciona una representación precisa de la magnitud total del error y penaliza más severamente los errores grandes. Asimismo, el RMSE tiene características matemáticas de una métrica de distancia, lo cual lo hace apropiado para la evaluación cuantitativa del rendimiento de modelos. Esto incluye su uso en visión artificial, donde se contraponen estimaciones con valores de referencia (Chai, 2014).

4.2.3 *Comparación de las técnicas según su nivel de error*

La comparación de las técnicas de visión por computador revela una jerarquía de precisión donde las Redes Neuronales Convolucionales (CNN) optimizadas lideran con un RMSE de entre 3.21° y 4.25°. En un rango de alta a media exactitud se encuentran la reconstrucción 3D y los Transformers, cuyos valores de RMSE oscilan entre 2.0° y 9.0°, seguidos por la fusión visual-inercial que registra un RMSE de 5.13° a 7.86°. Por su parte, las redes de grafos (GCN/GNN) presentan un RMSE de entre 5.13° y 9.6°, mientras que las arquitecturas RNN/LSTM muestran errores ligeramente superiores, situándose entre 5.0° y 12.0°. Los sistemas de estimación de pose sin marcadores (markerless) más comunes reportan un RMSE de 7.7° para OpenPose y de 10.9° para MediaPipe, aunque este último puede alcanzar un RMSE de 27.3° en articulaciones complejas

como la muñeca. Finalmente, técnicas como los Filtros de Kalman y los métodos clásicos (SVM) se consideran de baja precisión para mediciones cinemáticas al superar habitualmente un RMSE de 15° , quedando distantes de los sistemas ópticos de referencia que mantienen errores menores a 1° .

Tabla 3. Cuadro comparativo de las técnicas.

Técnica	Valor de error reportado (RMSE°)	Nivel de exactitud	Fuentes
CNN (Redes Neuronales Convolucionales)	$3.21^\circ - 4.25^\circ$	Alto	(Yan et al., 2025)
Pose Estimation Markerless (2D/3D)	7.7° (OpenPose) / 10.9° (MediaPipe)	Medio	(Abrar Ashraf et al., 2025)(Maggi oni et al., 2025)
Transformers	$2.0^\circ - 9.0^\circ$ (Rango general en estimación 3D)	Alto / Medio	(Zhou et al., 2025)
RNN / LSTM	$5.0^\circ - 12.0^\circ$ (Estimado en fusión temporal)	Medio	(Dellepiane et al., 2025)
Reconstrucción / Estimación 3D	$2.0^\circ - 9.0^\circ$	Alto / Medio	(Zhou et al., 2025)
GCN / GNN	$5.13^\circ - 9.6^\circ$	Medio	(Adjel et al., 2025)(Li & Dong, 2023)
Filtros de Kalman	$5.13^\circ - 7.86^\circ$	Medio	(Maggioni et al., 2025)(Paiva et al., 2025)
SVM / Métodos clásicos	$> 15^\circ$ (Suelen usarse para clasificación, no para grados precisos)	Bajo	(Kuang et al., 2025)

4.2.4 Determinación de nivel según su técnica de visión por computador.

Para concluir si la exactitud de una técnica es Baja, Media o Alta, se encontró que los estudios utilizan criterios basados en la aplicación clínica y la desviación estadística, así:

1. Nivel Alto (Error < 5°):

- Es el "estándar de oro" para aplicaciones clínicas críticas (como el análisis de la marcha o diagnósticos médicos). Porque errores menores a 5° se consideran insignificantes para la estimación de tensiones musculoesqueléticas, representando menos del 10% de error total en simulaciones biomecánicas. Técnicas como las CNN optimizadas logran este rango (Zhou et al., 2025).

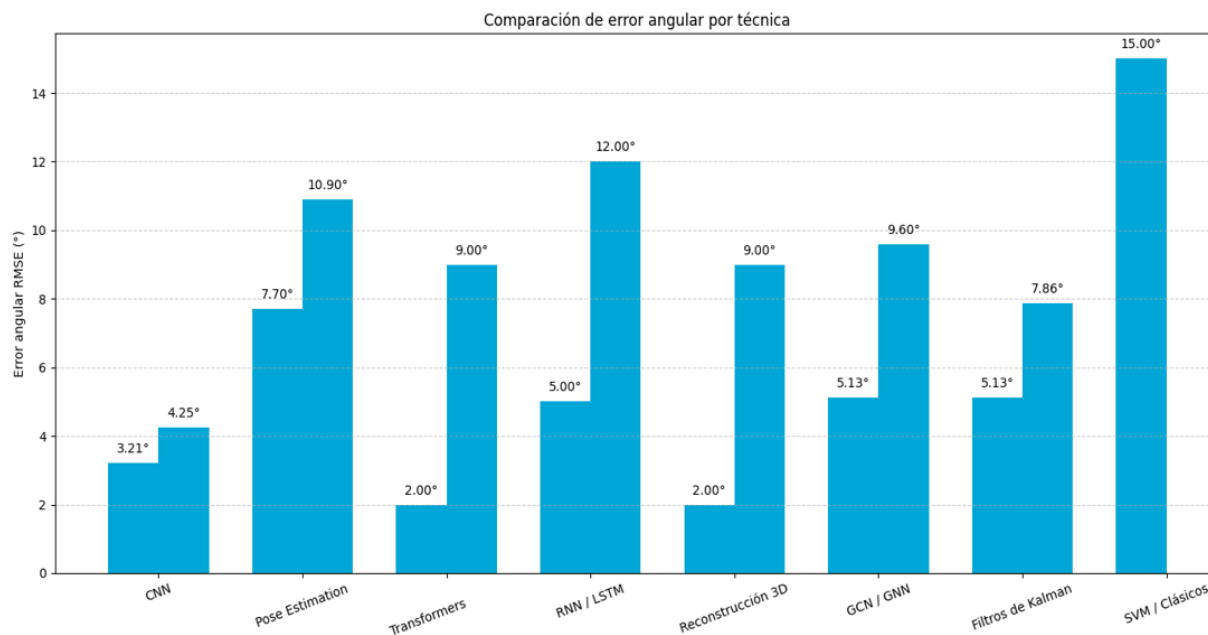
2. Nivel Medio (Error entre 5° y 15°):

- Se considera un rango aceptable para la tele-rehabilitación y el monitoreo de ejercicios en casa. Porque aunque existe una desviación, es suficiente para detectar tendencias de mejora, amplitud de movimiento y correctitud del ejercicio, La mayoría de los sistemas Markerless (MediaPipe/OpenPose) y los Filtros de Kalman caen en esta categoría debido al ruido inherente de los sensores económicos (Dellepiane et al., 2025).

3. Nivel Bajo (Error > 15° o 20°):

- Inadecuado para mediciones cinemáticas precisas; útil solo para reconocimiento general de actividades. Ya que valores superiores a 15° suelen deberse a oclusiones severas (cuando el cuerpo tapa la articulación) o a la incapacidad del algoritmo para resolver la profundidad (eje Z) con una sola cámara, Los métodos clásicos (SVM) y sistemas sin optimización de profundidad suelen presentar estos errores (Maggioni et al., 2025).

A continuación, en la figura 3 se muestra la clasificación de los diferentes niveles de estimación por técnica (Alto, Medio, Bajo) basada en los rangos de error reportados en las fuentes.

Figura 3. Clasificación de los diferentes niveles de estimación

4.2.5 Tipos de movimientos más estudiados del tren superior

De los 59 documentos, la mayoría aborda el análisis de las acciones del tren superior delantero, centrándose principalmente en el hombro, el codo, el antebrazo y el tronco. Dichos movimientos constituyen un pilar para comprender el funcionamiento del motor del cuerpo y para la evaluación de procesos de rehabilitación, ergonomía o desempeño físico. Los más detallados corresponden a la extensión y flexión del codo y a la abducción y flexión del hombro, seguidos por los gestos del antebrazo relativos a la pronación y supinación, y los reajustes posturales del tronco. En total, los trabajos coinciden en que las técnicas de visión por computador permiten distinguir y medir de forma precisa y no invasiva los patrones de movimiento, lo que contribuye a un conocimiento valioso para la evaluación clínica y biomecánica. La Tabla 4 indica los principales movimientos estudiados y las fuentes que los respaldan, lo que constituye un modo fácil de identificar las tendencias más significativas en el análisis del tren superior del ser humano.

Tabla 4. Movimientos más estudiados.

N	Tipos de movimientos	Fuente
1	Codo: flexión/extensión	(Moreira et al., 2024)(Lee & Tan, 2025)(Li & Yu, 2023)(Li & Dong, 2023)(Abromavičius et al., 2025)(Ryu et al., 2025)(Ceglia et al., 2025)(Kim et al., 2025)(Boudreault-Morales et al., 2025)(Shi et al., 2024)(Abrar Ashraf et al., 2025)(Giulietti et al., 2025) (Cao et al., 2019)(Jackson et al., 2023)
2	Hombro: abducción/aducción, elevación, flexión	(Moreira et al., 2024)(Lee & Tan, 2025)(Li & Yu, 2023)(Li & Dong, 2023)(Abromavičius et al., 2025)(Ryu et al., 2025)(Ceglia et al., 2025)(Kim et al., 2025)(Boudreault-Morales et al., 2025)(Adjel et al., 2025)(Abrar Ashraf et al., 2025)(Giulietti et al., 2025) (Adolf et al., 2025)(Edriss et al., 2024)(Sun et al., 2025)(Jiang et al., 2025)(Cao et al., 2019)(Jackson et al., 2023)
3	Movimientos funcionales (alcanzar, levantar, empujar, beber, escribir)	(Kryeem et al., 2025)(Doong, 2025)(Adolf et al., 2025)(Harrison et al., 2025)(Zhou et al., 2025)(Ceglia et al., 2025)(Edriss et al., 2024)(Sun et al., 2025)(Wang et al., 2025)(Kamal et al., 2025)(A. Zhao et al., 2025)(Abrar Ashraf et al., 2025) (Giulietti et al., 2025)(Cao et al., 2019)(Warguła et al., 2025)(Jackson et al., 2023)(Suo et al., 2025)
4	Postura del tronco superior (sedestación, inclinación, alineación)	(Jin et al., 2025)(Harrison et al., 2025)(Zhou et al., 2025)(Tanthuwapathom et al., 2025)(Doong, 2025)(Adolf et al., 2025)(Warguła et al., 2025)(Jackson et al., 2023)(Cao et al., 2019)
5	Muñeca/antebrazo (pronación/supinación,	(Adolf et al., 2025)(Q. Gao et al., 2025)(Raj et al.,

flexión/extensión)

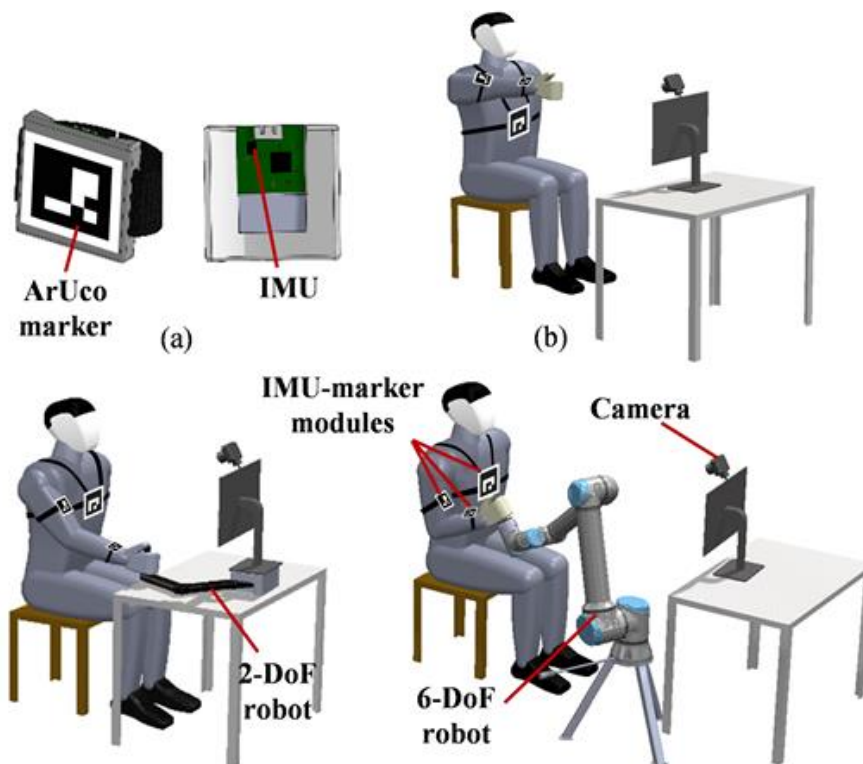
2025)(Ryu et al., 2025)(Moreira et al., 2024)(Li & Yu, 2023)(Li & Dong, 2023)(Abromavičius et al., 2025)(Ceglia et al., 2025)(Abrar Ashraf et al., 2025)(Cao et al., 2019)(Lee & Tan, 2025)

4.2.5.1 Flexión y extensión del codo

La flexión y extensión del codo es uno de los movimientos más estudiados, debido a su relevancia para evaluar la funcionalidad del tren superior y su utilidad en el ámbito clínico y ergonómico. Biomecánicamente, consiste en el desplazamiento angular del antebrazo respecto al brazo en el plano sagital, controlado principalmente por el bíceps y el tríceps braquial. Varios de los artículos aplicaron técnicas de visión por computador para estimar este movimiento con precisión: los sistemas visual-inerciales lograron correlaciones superiores a 0.98 y errores inferiores a 4° en el ángulo de medición del codo (Li & Yu, 2023); los modelos de estimación de pose basados en redes neuronales profundas lograron errores inferiores a 5° al comparar con goniómetros clínicos (Boudreault-Morales et al., 2025); las redes convolucionales con cámaras RGB-D registraron diferencias angulares menores a 5° en actividades de ciclismo (Ceglia et al., 2025).

Las técnicas de calibración visual y reconstrucción 3D presentaron errores leves de alrededor de 6° esto según el ángulo de la cámara (Kim et al., 2025). Finalmente, las comparaciones clínicas con redes ligeras demostraron una gran correlación con las mediciones manuales (Ryu et al., 2025), consagrando la extensión y flexión del codo como un movimiento de referencia para evaluar el desempeño de los sistemas de visión por computador en el análisis biomecánico del tren superior y su aplicación en el ámbito de la rehabilitación.

Figura 4. Flexión y extensión del codo(Li & Yu, 2023)



4.2.5.2 *Abducción, aducción y elevación del hombro*

Los movimientos de elevación, aducción y abducción del hombro son ampliamente analizados en la literatura revisada, debido a que se consideran importantes para la valoración funcional de la parte superior y para el análisis de los patrones cinemáticos en la rehabilitación. Estos movimientos son el resultado de un comportamiento coordinado entre la clavícula, la escápula y la articulación glenohumeral, facilitando el desplazamiento del brazo a través de diferentes planos y permite evaluar la capacidad del control neuromuscular.

De los 59 artículos, se observa que varios autores analizaron la movilidad del hombro. Por ejemplo, (Ceglia et al., 2025) utilizaron un sistema de captura RGB-D para identificar la elevación y abducción del brazo durante ejercicios de hand-cycling, demostrando la efectividad de los modelos de visión sin marcadores para representar trayectorias articulares con alta precisión. En otro estudio, (Li & Yu, 2023) desarrollaron un modelo visual-inercial que integró sensores IMU con calibración automática, logrando estimaciones estables del ángulo de abducción del hombro sin necesidad de marcadores. Finalmente, en (Kim et al., 2025) desarrollaron sistemas monoculares 3D, analizaron la posición óptima de la cámara, concluyendo que una configuración a 45° respecto

al plano frontal mejora la detección de la elevación del hombro y minimiza el error angular en la estimación del movimiento

Figura 5. Abducción, aducción y elevación del hombro(Ceglia et al., 2025)



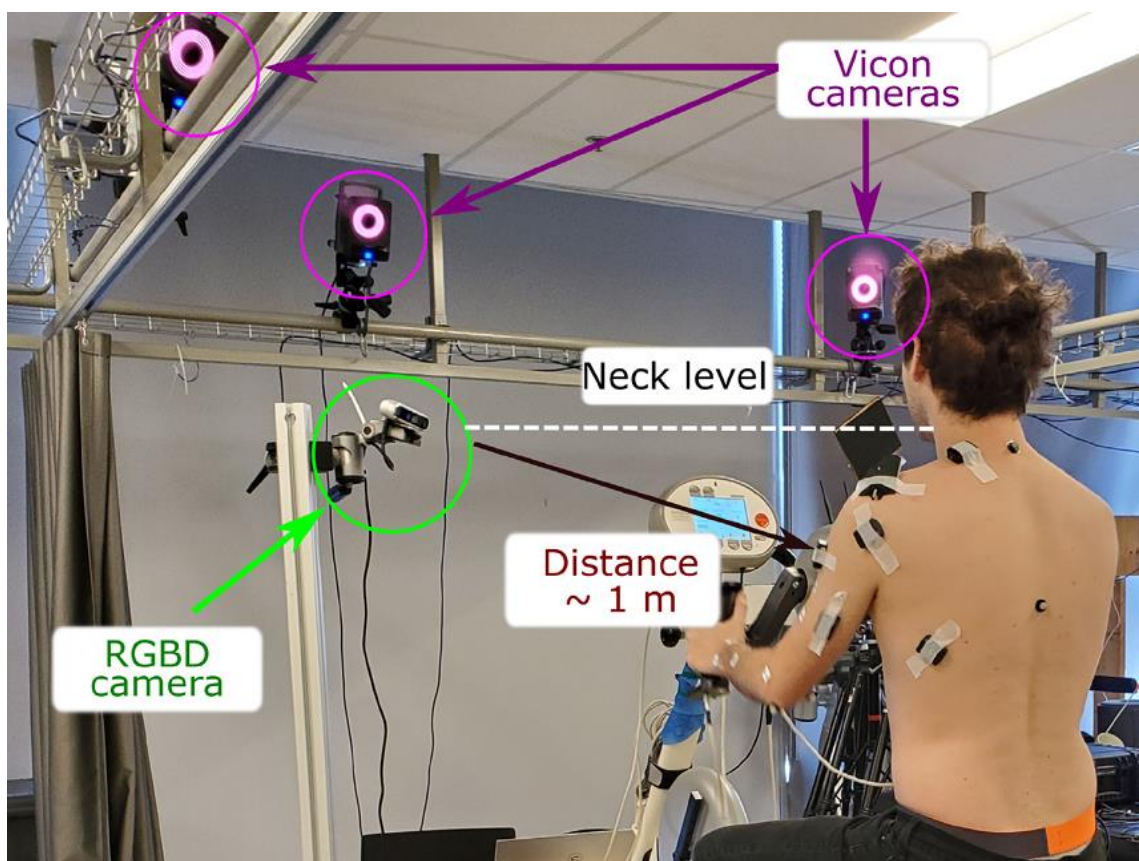
4.2.5.3 Movimientos funcionales compuestos (alcanzar, empujar, levantar, beber, escribir, gesticular)

Los movimientos funcionales compuestos consisten en la coordinación simultánea de múltiples articulaciones del tren superior para realizar acciones como: alcanzar un objeto a distancia, empujar una silla o levantar una herramienta. Este tipo de movimientos son descritos como secuencias de estiramiento y ajuste, dependen de una comunicación eficaz del control neuromuscular. Son de gran importancia para el análisis del desempeño físico y los procesos de rehabilitación. En los diferentes artículos revisados, estas acciones se emplean para valorar la funcionalidad global del tren superior, la fluidez del movimiento y la estabilidad postural durante tareas de la vida cotidiana.

En este orden de ideas, en (Ceglia et al., 2025) aplicaron un sistema basado en visión RGB-D para evaluar de manera cíclica la amplitud y la fuerza de empuje durante maniobras de hand-cycling, demostrando lo idóneo que es el sistema sin el uso de marcadores para identificar trayectorias articulares con alta precisión. Por otro lado, en (Adolf et al., 2025) implementaron un sistema de retroalimentación visual en tiempo real para tareas de alcance funcional y empuje, con

el objetivo de mejorar la ejecución motora y el control postural en contextos de tele-rehabilitación. En otro estudio, (Doong, 2025) utilizaron procedimientos de aprendizaje automático (machine learning) y análisis ergonómico para identificar posturas ineficientes e identificar riesgos asociados a maniobras de empuje y levantamiento, lo que impulsó el uso de la visión por computador como herramienta de evaluación preventiva.

Figura 6. Movimientos funcionales compuestos(Ceglia et al., 2025)



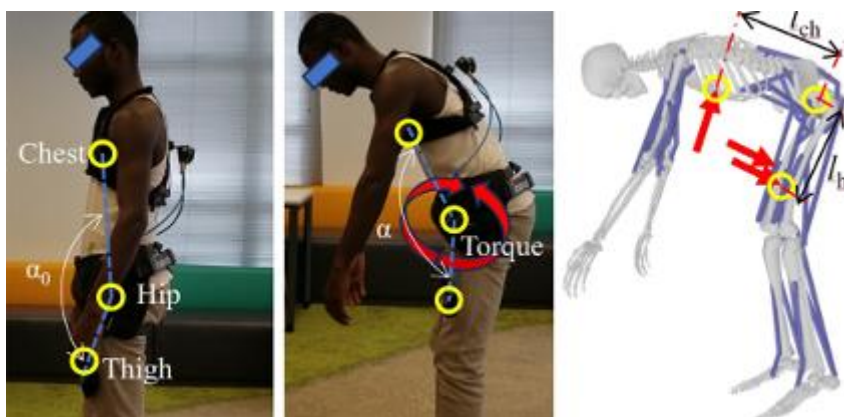
4.2.5.4 Posturas del tronco superior (sedestación, inclinación, alineación escapular/torácica)

El análisis de las posturas del tronco superior es indispensable en la investigación, por ser la base para la valoración de la estabilidad, la alineación escapular y torácica, así como para valorar la capacidad del cuerpo de mantener una postura adecuada durante la ejecución de tareas diarias o laborales. Las posturas en sedestación e inclinación son medidas pertinentes para evaluar el control postural, el equilibrio y la prevención de trastornos musculoesqueléticos.

En los análisis revisados, la visión por computador se ha mostrado como una herramienta

efectiva que ha permitido la detección automática de posturas de desviación y el seguimiento en tiempo real de la alineación del tronco. En (Adolf et al., 2025) crearon un sistema de retroalimentación visual que evalúa la postura durante ejercicios de elasticidad y estiramiento, y de fuerza y fortalecimiento, permitiendo detectar asimetrías del tronco torácico y corregir la alineación en programas de tele-rehabilitación. Para complementar, en (Doong, 2025) utilizaron técnicas de aprendizaje automático para valorar la postura en tareas ergonómicas, identificando inclinaciones y rotaciones del tronco relacionadas a riesgos causados por tareas laborales deficientes. Esto demuestra el futuro potencial de la visión por computador para mejorar la evaluación biomecánica en entornos más realistas.

Figura 7. Posturas del tronco superior(Zhou et al., 2025)



4.2.5.5 Muñeca/antebrazo (pronación/supinación, flexión/extensión)

La investigación sobre los movimientos de pronación, supinación, flexión y extensión de la muñeca y el antebrazo es relevante para entender la función de la extremidad, dado que estas acciones están presentes en la gran mayoría de las tareas diarias y son un ejemplo de la coordinación entre los segmentos distal y proximal del miembro superior. La medición precisa de dichos movimientos es clave para estimar desde la movilidad fina y la fuerza de agarre hasta el control postural en entornos clínicos y de rehabilitación.

La visión por computador se establece como una alternativa efectiva para la estimación no invasiva y precisa de los ángulos articulares del antebrazo y la muñeca. En (Lee & Tan, 2025) validaron el uso de la inteligencia artificial para medir el rango de movimiento de estas articulaciones, evidenciando una alta relación entre sus mediciones automatizadas y las pruebas

manuales tradicionales durante movimientos de pronación y flexión. En (Ceglia et al., 2025) Los investigadores aplicaron con éxito sistemas de visión sin marcadores (markerless) RGB-D para el análisis cíclico de los patrones de extensión durante ejercicios funcionales. Este registro demostró que los modelos de visión con profundidad ofrecen una estabilidad temporal continua, lo que asegura capturar con precisión la trayectoria coordinada del antebrazo y la muñeca.

Figura 8. Muñeca/antebrazo(Li & Yu, 2023)

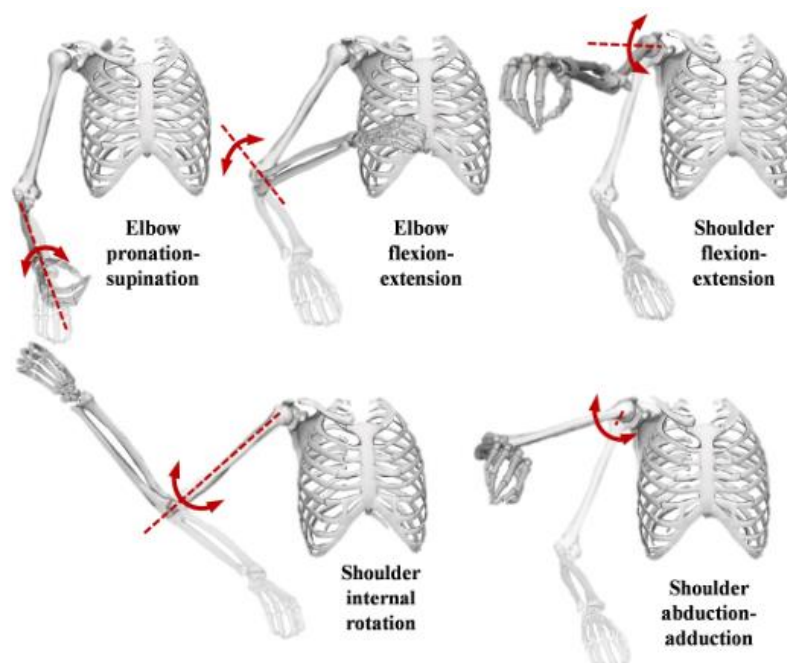


Tabla 5. Cuadro comparativo tipos de movimientos.

Tipo de movimiento	Valor de Error (RMSE/Diferencia)	Nivel de exactitud	Fuente
Flexión y extensión del codo	3.65° a 7.86°	Alto / Medio	(Yan et al., 2025)(Scott et al., 2022)(Li & Dong, 2023)
Abducción, aducción y elevación del hombro	3.21° a 7.86°	Alto / Medio	(Scott et al., 2022)(Li & Dong, 2023)
Movimientos funcionales compuestos	< 5° a 7.7° (18% - 19%/)	Alto / Medio	(Abrar Ashraf et al., 2025)
Posturas del tronco superior	< 6° a 7.7°	Alto / Medio	(Cao et al., 2019)
Muñeca/antebrazo	10.9° a 27.3° (22% - 39%)	Medio / Bajo	(Maggioni et al., 2025)

4.2.6 Evaluación de las técnicas de visión por computador para cada tipo de movimiento

1. Nivel Alto (Error < 5°):

- En esta categoría destacan los sistemas ópticos con marcadores (como Vicon y OptiTrack), que logran errores angulares habitualmente menores a 1° y una precisión espacial de menos de 1 mm. También se incluyen las Redes Neuronales Convolucionales (CNN) optimizadas que alcanzan un RMSE de entre 3.21° y 4.25° al medir rangos de movimiento en hombro y codo (Yan et al., 2025).

2. Nivel Medio (Error entre 5° y 15°):

- Se considera un rango aceptable para la tele-rehabilitación, el monitoreo de ejercicios en casa y evaluaciones ergonómicas donde la tendencia del movimiento es más importante que la medida absoluta. Las técnicas de estimación de pose sin marcadores (Markerless) son las más representativas de este nivel; por ejemplo, OpenPose reporta un RMSE de 7.7°, mientras que MediaPipe presenta una desviación promedio de 10.9°. También se incluyen los sistemas de fusión visual-inercial (Cámara + IMU + ArUco), que logran un RMSE entre 5.13° y 7.86°. Arquitecturas avanzadas como Transformers y GCN operan en un rango general de 2.0° a 9.0° para estimación 3D, y modelos temporales como RNN/LSTM

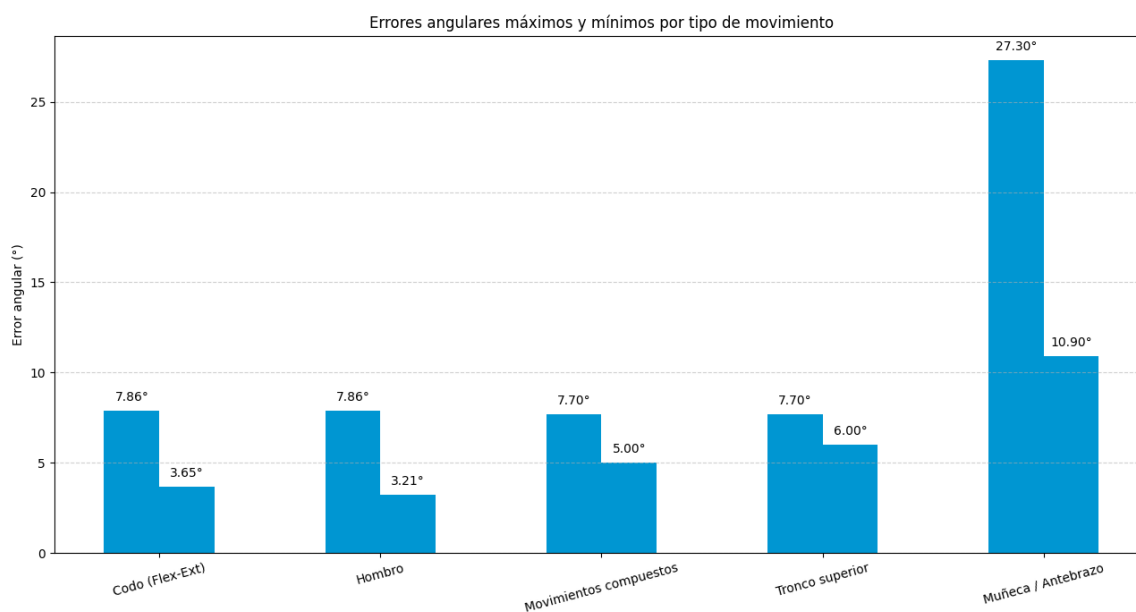
registran un error estimado en fusión de datos de entre 5.0° y 12.0° (Abrar Ashraf et al., 2025).

3. Nivel Bajo (Error > 15° o 20°):

- Este nivel ocurre cuando la tecnología enfrenta limitaciones por oclusiones severas, falta de datos de profundidad o movimientos de baja amplitud, resultando inadecuado para diagnósticos cinemáticos precisos. Las cámaras RGB monoculares estándar sin asistencia de profundidad pueden presentar errores de hasta 25° debido a la ambigüedad en el eje Z. Los métodos clásicos (SVM) y los filtros de Kalman suelen superar los 15° de error cuando no se optimizan para grados precisos. Regiones de alta densidad articular como la muñeca son las más afectadas, registrando un Error Absoluto Medio (MAE) de 27.3° debido a la complejidad de sus grados de libertad y constantes oclusiones manuales. Finalmente, la exactitud se considera baja cuando el desplazamiento es menor a 10 cm, capturando solo el 36% de la varianza real del movimiento (Ceglia et al., 2025).

A continuación, en la figura 9 se muestra la clasificación de los diferentes niveles de exactitud (Alto, Medio, Bajo) basada en los rangos de error reportados en las fuentes.

Figura 9. Clasificación de niveles de exactitud



4.2.7 Tecnologías que se utilizaron para registrar el movimiento del tren superior del cuerpo.

En los 59 artículos se han logrado reconocer cuatro tipos de tecnologías de captura utilizadas para registrar y estudiar el movimiento del tren superior, como se indica en la Tabla 6. Estas tecnologías corresponden a como cámaras RGB monoculares, cámaras RGB-D/Kinect, sistemas ópticos con marcadores y sistemas visual-inerciales (cámara + IMU/ArUco), y ofrecen distintos métodos para la recolección de datos biomecánicos. Cada tecnología presenta sus propias ventajas y desventajas en relación con la precisión, el costo, la complejidad de uso y la adaptabilidad a diversas circunstancias.

Tabla 6. Tipos de tecnologías para registrar el movimiento del tren superior.

N	Tipos de tecnologías	Fuente
1	Cámara RGB monocular	(Ryu et al., 2025)(Moreira et al., 2024)(Yan et al., 2025)(Adolf et al., 2025)(Kim et al., 2025)(Giulietti et al., 2025)(Z. Gao et al., 2025)(Edriss et al., 2024)(Behera et al., 2020)(A. Zhao et al., 2025)(Scott et al., 2022) (Y. Gao et al., 2025)(Shahid et al., 2025)(van den Hoorn et al., 2025)(Boudreault-Morales et al., 2025) (Wu et al., 2025)
2	Cámaras RGB-D / Kinect	(Ceglia et al., 2025)(Jin et al., 2025)(Jia et al., 2025)(Kwon, 2025)(Boudreault-Morales et al., 2025)(W. Zhao et al., 2024)(Abrar Ashraf et al., 2025)(Kryeem et al., 2025)(Jackson et al., 2023)(Yan et al., 2025)(Lee & Tan, 2025) (Cao et al., 2019)(Doong, 2025)(Dellepiane et al., 2025)(Zaher et al., 2025)
3	Sistemas ópticos con marcadores (Vicon/Visualeyez/OptiTrack)	(Ceglia et al., 2025)(Jia et al., 2025)(Maggioni et al., 2025)(Herdiman et al., 2025)(Kim et al., 2025)(Koul & Novembre, 2025)(Suo et al., 2025)
4	Sistemas visual-inerciales (cámara + IMU / ArUco)	(Li & Yu, 2023)(Li & Dong, 2023)(Adjel et al., 2025)(Wang et al., 2025)(Zhou et al., 2025)(Shi

et al., 2024)(Abromavičius et al., 2025)(Dalimarta et al., 2025)(Chang et al., 2021)(Lovanshi et al., 2024)(Xu & Xu, 2024) (Kuang et al., 2025)(Murillo et al., 2025)(Q. Gao et al., 2025)(Lin et al., 2025)(Jiang et al., 2025)(Ryu et al., 2025)(Harrison et al., 2025)(Tanthuwapathom et al., 2025)(Esmail et al., 2025)(Sun et al., 2025)(Paiva et al., 2025) (Xu & Xu, 2025)(Raj et al., 2025)(Kamal et al., 2025)(Warguła et al., 2025)

4.2.7.1 Cámara RGB monocular

Las cámaras monoculares RGB son dispositivos que no pueden brindar información espacial de forma directa a diferencia de los sistemas que ofrecen percepción en tres dimensiones o sensores de profundidad; en cambio, utilizan algoritmos de visión artificial para deducir la ubicación y el movimiento de las articulaciones del cuerpo a partir de imágenes planas. Su mayor ventaja es su bajo costo, alta disponibilidad y la facilidad con la que se pueden implementar en espacios comunes, como hospitales, aulas o lugares de trabajo. Son dispositivos que capturan imágenes en dos dimensiones utilizando los colores RGB (Red, Green, Blue) o rojo, verde y azul respectivamente. En los artículos analizados, los autores de (Kim et al., 2025) evidenciaron que los sistemas monoculares pueden lograr una precisión similar a la de equipos ópticos profesionales, siempre y cuando la cámara esté ubicada adecuadamente en relación con el área de movimiento, lo que permite un seguimiento fiable del rango de movimiento en el tren superior del cuerpo. Por su parte, en (Adolf et al., 2025) utilizaron una cámara RGB estándar para crear un sistema de retroalimentación visual en tiempo real que facilitó el monitoreo de ejercicios de estiramiento y la corrección de la postura sin la necesidad de ocupar sensores adicionales.

4.2.7.2 Cámaras RGB-D / Kinect

Las cámaras RGB-D son dispositivos más avanzados que integran una cámara de color estándar (RGB) y un sensor de profundidad, lo que les permite tomar al mismo tiempo tanto la información visual como la distancia de los objetos respecto al sensor. Esta tecnología utiliza luz

estructurada o tiempos de vuelo (ToF) para lograr crear un mapa 3D del entorno, lo que permite la reconstrucción precisa de la posición y el movimiento del cuerpo humano sin necesitar marcadores físicos. Entre estos sistemas se encuentran el Kinect de Microsoft y el Intel RealSense, que son muy utilizados en investigaciones de rehabilitación, biomecánica y ergonomía, gracias a su balance entre precisión, accesibilidad y facilidad de uso.

Entre los artículos revisados, los autores de (Ceglia et al., 2025) usaron una cámara Intel RealSense D455 para examinar la musculatura del tren superior durante ejercicios de ciclismo manual, mostrando que la información de profundidad mejora la precisión en la estimación de los ángulos de las articulaciones en comparación con modelos que solo utilizan RGB. En otro estudio, (Jia et al., 2025) se verificó la exactitud del sistema Kinect v2 usando métricas de similitud coseno entre las trayectorias articulares estimadas y los datos de referencia, concluyendo que los sensores Kinect presentan un margen de error angular menor a cinco grados en movimientos controlados del brazo. Por último, en (Jin et al., 2025) utilizó la cámara Azure Kinect DK para detectar posturas en tiempo real, demostrando que la información de profundidad ayuda a identificar desviaciones posturales y permite corregir automáticamente la alineación de la parte superior del torso mientras se está sentado.

4.2.7.3 *Sistemas ópticos con marcadores (Vicon/Visualeyz/OptiTrack)*

Los sistemas ópticos con marcadores son modalidades de captura de movimiento que utilizan varias cámaras infrarrojas y marcadores, ya sean reflejantes o activos, colocados en zonas específicas del cuerpo humano. Mediante el principio de triangulación óptica, estos dispositivos dan a lugar la reconstrucción con gran precisión de las coordenadas tridimensionales de cada marcador, entregando una representación detallada del movimiento. Entre estos, se destacan los sistemas Vicon, OptiTrack y Visualeyz, que son considerados el estándar de excelencia en la investigación biomecánica debido a su alta frecuencia de muestreo, precisión milimétrica y capacidad para captar movimientos complejos del tren superior.

En los estudios revisados, los autores de (Ceglia et al., 2025) emplearon un sistema Vicon de 14 cámaras como referencia para comprobar los resultados obtenidos con una cámara RGB-D, mostrando que los sistemas ópticos son superiores en términos de precisión espacial. Finalmente, en (Jia et al., 2025) llevaron a cabo la validación del sensor Kinect utilizando el sistema Visualeyz, subrayando su eficacia como estándar de referencia para medir la exactitud de los métodos sin

marcadores en la estimación de movimientos articulares.

4.2.7.4 *Sistemas visual-inerciales (cámara + IMU / ArUco)*

Los sistemas visual-inerciales son sistemas que se basan en la combinación de una o más cámaras RGB o RGB-D con sensores inerciales (IMU), uniendo tanto la información visual como los datos sobre aceleración y orientación. Este tipo de tecnología combina sensores que ayudan a superar las limitaciones de cada uno: las cámaras entregan precisión espacial, mientras que las IMUs brindan estabilidad temporal y continuidad en el seguimiento del movimiento, incluso en situaciones con obstrucción. En algunas ocasiones, los módulos inerciales incluyen marcadores Aumented Reality University of Córdoba (ArUco) para facilitar la calibración automática entre las partes del cuerpo y la cámara.

En el artículo, (Li & Yu, 2023) los autores crearon un sistema visual-inercial para la estimación automática de la posición del tren superior, empleando una cámara RGB y tres sensores IMU que están sincronizados mediante marcadores ArUco. Los indicadores ArUco son figuras cuadradas de elevado contraste que tienen un código particular, lo cual simplifica que las cámaras los identifiquen. Esto hace posible que las cámaras calculen su posición y orientación al proyectar las esquinas de tales indicadores. Por consiguiente, el sistema puede unificar cada IMU con el modelo visual, creando una referencia común entre los sensores y la cámara. El uso de estos elementos agiliza la calibración, minimiza las equivocaciones en el ajuste y perfecciona la coherencia del movimiento registrado. Su estudio mostró una mejora notable en la medición de los ángulos del hombro y el codo al unir la precisión visual con la estabilidad inercial.

Tabla 7. Cuadro comparativo tipos de tecnologías.

Tecnologías de captura (hardware)	Valor de Error (RMSE/Diferencia)	Nivel de exactitud	Fuentes
Cámara RGB monocular	11° a 25°	Medio/Bajo	(Kim et al., 2025)
Cámaras RGB-D / Kinect	3.2° a 10.9°	Alto/Medio	(Ceglia et al., 2025)(Maggioni et al., 2025)
Sistemas ópticos con marcadores (Vicon/Visualeyez/OptiTrack)	< 1°	Alto	(Adjel et al., 2025)(Li & Dong, 2023)
Sistemas visual-inerciales (cámara + IMU / ArUco)	5.13° a 9.6°	Medio	(Ceglia et al., 2025)(Kim et al., 2025)

4.2.7.5 Evaluación del Nivel de Exactitud de herramientas para el la análisis de movimiento del tren superior del cuerpo humano.

1. Nivel Alto (Error < 5°): Sistemas Ópticos con Marcadores y Modelos ROM Optimizados

Este nivel representa la máxima precisión tecnológica disponible, siendo el referente para todas las demás mediciones.

- **Sistemas ópticos con marcadores (Vicon, Visualeyez, OptiTrack):** Son considerados el estándar de oro (gold standard) en biomecánica. Poseen una precisión espacial para la posición de los marcadores de menos de 1 mm en condiciones estáticas y aproximadamente 3.3 mm en dinámicas. Sus errores angulares son habitualmente menores a 1°, superando la capacidad de cualquier otro método (Harrison et al., 2025).
- **Modelos de IA de una sola cámara optimizados para ROM:** Ciertas arquitecturas de redes neuronales ligeras (como el modelo Hourglass reducido y cuantizado) han logrado un RMSE entre 3.21° y 4.25° al medir el rango de movimiento (ROM) de hombro, codo y cadera (Yan et al., 2025).
- **Sistemas de profundidad específicos:** En tareas de rehabilitación de miembros superiores,

mediciones de elevación del hombro con sensores como Kinect v1 han registrado diferencias de apenas 3.32° , lo cual es "clínicamente aceptable" (Scott et al., 2022).

2. Nivel Medio (Error entre 5° y 15°):

Estas tecnologías ofrecen un equilibrio entre facilidad de uso y una precisión suficiente para la mayoría de las aplicaciones clínicas y de monitoreo.

- Cámaras RGB-D / Kinect: Los sensores de profundidad (Kinect v2, Azure) muestran un acuerdo con los sistemas de referencia dentro de un 7% de error (Harrison et al., 2025). En mediciones de ángulos articulares como los del codo o el hombro, presentan un RMSE de aproximadamente 7.7° (Abrar Ashraf et al., 2025).
- Sistemas Visual-Inerciales (Cámara + IMU / ArUco): La fusión de una cámara monocular con sensores inerciales y marcadores ArUco permite calibraciones que logran un RMSE entre 5.13° y 7.86° en los ángulos del hombro y el codo (Li & Dong, 2023).
- Cámara RGB Monocular (Markerless avanzado): El uso de herramientas como MediaPipe para el análisis técnico (ej. natación artística o rehabilitación) reporta desviaciones angulares de 8.13° en comparación con herramientas de medición profesional (Edriss et al., 2024).
- Articulaciones específicas: La medición del cuello mediante MediaPipe se sitúa en este rango con un MAE de 13.76° , debido a la falta de puntos de referencia esqueléticos cervicales específicos (Kim et al., 2025).

3. Nivel Bajo (Error $> 15^\circ$ o 20°):

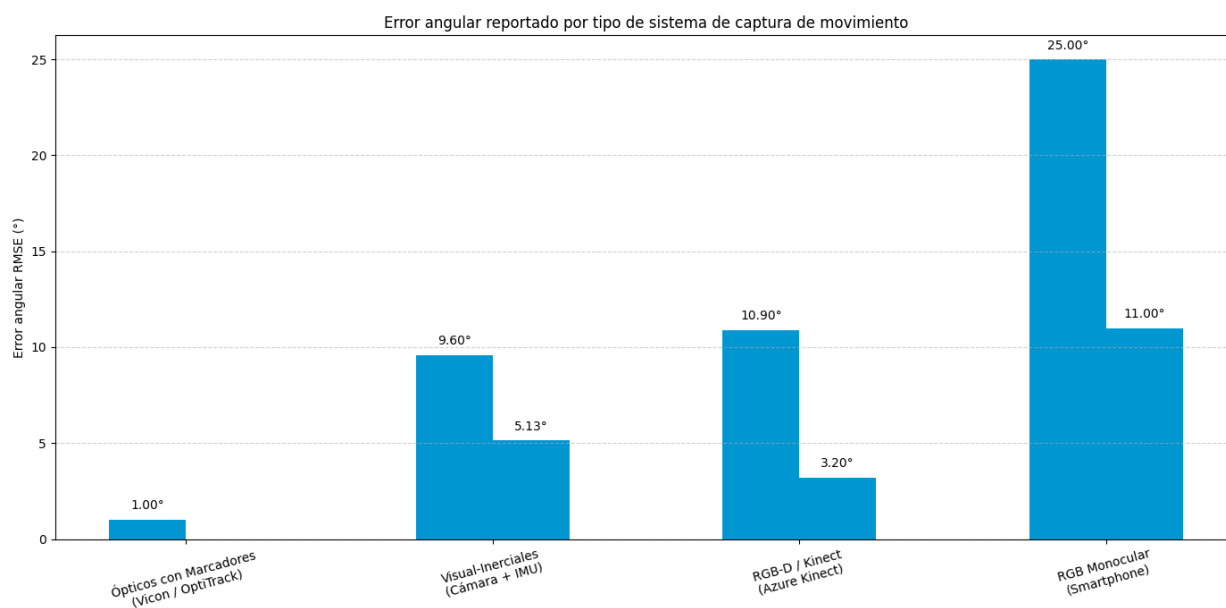
Este nivel ocurre cuando la tecnología se enfrenta a limitaciones geométricas, oclusiones o falta de datos de profundidad.

- Cámara RGB Monocular (Sin asistencia de profundidad): La tecnología de una sola cámara estándar puede presentar errores de hasta 25° en movimientos del tren superior, debido a la dificultad de localizar el centro de la articulación y a la ambigüedad de la profundidad en imágenes 2D (Ceglia et al., 2025).
- Zonas con oclusión y alta densidad articular: La muñeca es la articulación con menor precisión, registrando un MAE de 27.3° debido a la complejidad de sus grados de libertad y a que la mano suele oscurecer los puntos clave (Maggioni et al., 2025).

- Impacto de la distancia y perspectiva: Los errores superan los 20° cuando el sujeto se encuentra a distancias no óptimas (más de 3.5 metros) o en ángulos de visión oblicuos donde el cuerpo genera auto-occlusiones (Adjel et al., 2025).
- Movimientos de baja amplitud: La exactitud se considera "baja" cuando el movimiento es menor a 10 cm, capturando solo el 36% de la varianza real del movimiento comparado con sistemas Vicon (Koul & Novembre, 2025).

A continuación, en la figura 10 se muestra la clasificación de los diferentes niveles de exactitud (Alto, Medio, Bajo) basada en los rangos de error reportados por tipo de sistema de captura de movimiento en las fuentes.

Figura 10. Error angular de cada tecnología



4.2.8 Oportunidades de investigación

Tras analizar los 59 artículos se identificaron múltiples oportunidades mencionadas por los propios autores, tanto a partir de las limitaciones reconocidas en sus estudios como del análisis crítico realizado en esta revisión sistemática. A pesar de los avances logrados en exactitud, accesibilidad y automatización, aún existen vacíos relacionados con la validación clínica, la

combinación de tecnologías, la robustez en entornos no controlados y la representación de las muestras poblacionales. Estas oportunidades reflejan la necesidad de investigación, la integración entre la ingeniería, la biomecánica y la rehabilitación, impulsando a desarrollos más exactos, inclusivos y aplicables en contextos clínicos y cotidianos.

Tabla 8. Oportunidades de investigación.

N	Oportunidades de investigación	Fuente
1	Validación clínica estandarizada y generalización	(Moreira et al., 2024)(Lee & Tan, 2025)(Jia et al., 2025)(Ceglia et al., 2025)(Scott et al., 2022)(Jackson et al., 2023)(Ryu et al., 2025)(Tanthuwapathom et al., 2025)(Boudreault-Morales et al., 2025)(Kim et al., 2025)
2	Fusión multimodal y calibración robusta (visión + inercial + profundidad)	(Li & Yu, 2023)(Li & Dong, 2023)(Adjel et al., 2025)(Wang et al., 2025)(Ceglia et al., 2025)
3	Robustez en condiciones reales: vistas, oclusiones y tiempo real	(Jin et al., 2025)(Sun et al., 2025)(Yan et al., 2025)(Chang et al., 2021)(Y. Gao et al., 2025)(Lin et al., 2025)(Harrison et al., 2025)(Giulietti et al., 2025)
4	Datos y cobertura funcional: cohortes diversas y tareas con significado clínico	(Boudreault-Morales et al., 2025)(Zaher et al., 2025)(Kuang et al., 2025)(Kryeem et al., 2025)(Harrison et al., 2025)(Jin et al., 2025)(Edriss et al., 2024)(Ceglia et al., 2025)(Ryu et al., 2025)(Tanthuwapathom et al., 2025)

4.2.8.1 Validación clínica estandarizada y generalización

Una futura oportunidad de investigación tiene que ver con la validación clínica estandarizada y de qué manera se pueden implementar los modelos de visión por computador al análisis del movimiento del tren superior. Aunque actualmente técnicas tales como las Redes Neuronales

Convolucionales (CNN), la estimación de postura sin marcadores en 2D/3D o las Redes Neuronales de Grafos (GNN), han demostrado una precisión en condiciones controladas, existen limitaciones al usarlas en situaciones clínicas reales o con personas que tienen diferentes características anatómicas y funcionales. Los autores están de acuerdo en que es necesario establecer protocolos de validación uniformes, de modo que se puedan comparar resultados entre distintos sistemas, métodos y grupos clínicos.

Por esto, en (Boudreault-Morales et al., 2025) señalan que la discapacidad del miembro superior afecta la precisión de los modelos RGB monoculares, sugiriendo que la validación debería incluir a grupos con diversas condiciones motoras. Además, en (Ceglia et al., 2025) sugieren que se debería incluir la validación biomecánica en tiempo real a través de análisis musculoesqueléticos, señalando la importancia de comparar estos resultados con mediciones clínicas de referencia. Finalmente, en (Lee & Tan, 2025) destacan la importancia de unificar los protocolos de medición del rango de movimiento que usan inteligencia artificial para asegurar que los resultados sean consistentes entre diferentes sujetos y se puedan aplicar en instituciones de salud.

4.2.8.2 Fusión multimodal y calibración robusta (visión + inercial + profundidad)

Una de las oportunidades más claras que se encontró en el análisis es la que tiene que ver con juntar diferentes tipos de sensores y lograr una calibración correcta entre las cámaras y los dispositivos inerciales. Los estudios revisados demuestran que unir distintos tipos de información, como imágenes en color, datos de profundidad y señales inerciales, ayuda a tener una imagen más completa y precisa del movimiento humano. Sin embargo, todavía hay problemas con la sincronización del tiempo, la alineación geométrica y la consistencia entre los diferentes tipos de datos, lo cual puede influir en la exactitud del análisis biomecánico.

En (Li & Dong, 2023) mencionan que para calibrar en tiempo real la cámara y los sensores inerciales, es necesario seguir pasos estandarizados para disminuir los errores en la estimación de los ángulos. En (Adjel et al., 2025) muestran que una mala alineación entre los sensores visuales e inerciales afecta directamente la precisión de los ángulos en las articulaciones, sugiriendo al uso de algoritmos de ajuste automático que sean más confiables. En (Wang et al., 2025), utilizando redes neuronales que manejan varios modos, demuestran que combinar visión e inercia aumenta

la sensibilidad del sistema a los cambios en el movimiento, aunque señalan que hay que mejorar la sincronización y la velocidad de cálculo para que un sistema que emplea esta combinación funcione en tiempo real.

4.2.8.3 Robustez en condiciones reales: vistas, oclusiones y tiempo real

Una de las principales oportunidades de estudio señaladas en la revisión está relacionada con crear sistemas de visión artificial que sean fuertes y eficientes en situaciones reales. Estos sistemas deberían operar de manera instantánea y con poco retraso en dispositivos locales. Aunque los modelos actuales consiguen una alta precisión en entornos controlados, su eficacia baja cuando hay cambios en la iluminación, bloqueos, diferentes ángulos de la cámara o limitaciones en la capacidad de procesamiento. Esto resalta la necesidad de mejorar los algoritmos para asegurar un rendimiento confiable y constante fuera del entorno del laboratorio.

En múltiples artículos se encontró que varios autores coinciden en este desafío, por ejemplo en (Jin et al., 2025) sugieren un modelo que facilita la detección de posturas en tiempo real, destacando la importancia de mantener la precisión en condiciones dinámicas y cambiantes. De manera similar, en (Sun et al., 2025) desarrollan una red espaciotemporal para la estimación de pose 3D en el entrenamiento físico, enfocándose en un funcionamiento de baja latencia. Finalmente, en (Yan et al., 2025) indican que los sistemas ajustados para dispositivos locales pueden realizar análisis de movimiento con alta eficiencia energética y exactitud, abriendo la puerta a aplicaciones clínicas portátiles.

4.2.8.4 Datos y cobertura funcional: cohortes diversas y tareas con significado clínico

Por último, una de las oportunidades más importantes que se encontraron en la revisión está relacionada con la necesidad de aumentar y diversificar los tipos de datos que se utilizan para entrenar y validar los modelos de visión por computador en el análisis del movimiento del tren superior. Los estudios actuales muestran que la mayoría de las bases de datos son creadas a partir de muestras pequeñas, similares entre sí y en situaciones controladas, lo que reduce la capacidad de aplicar los resultados a otras situaciones. Por eso, varios autores sugieren la creación de bases de datos más grandes, variadas y que sean representativas de situaciones reales, incluyendo tareas reales y personas con diferentes características físicas, médicas y laborales. En (Boudreault-Morales et al., 2025) mencionan que los modelos que estiman la postura deben ser evaluados en

grupos diversos, que incluyan a personas con limitaciones motoras, para garantizar que se puedan aplicar en la clínica y mejorar la precisión en situaciones reales. De la misma manera, en (Kuang et al., 2025) resaltan la importancia de incluir tareas de rehabilitación y movimientos complejos, lo que ayudaría a los modelos a aprender patrones biomecánicos más representativos.

5. Discusión de resultados

El análisis sistemático de los 59 artículos permitió comprender de manera integral la evolución de las técnicas de visión por computador aplicadas al estudio del movimiento del tren superior, así como también sus limitaciones, alcances y tendencias de desarrollo. En términos generales, los resultados obtenidos evidencian una transición progresiva desde métodos tradicionales (como los basados en técnicas de procesamiento de imagen y modelos estadísticos simples) hacia enfoques basados en aprendizaje profundo, y que a día de hoy dominan en la literatura. Esta transición ha permitido mejorar la precisión de los modelos, automatizar la identificación de articulaciones y reducir la dependencia de sistemas de captura con marcadores, lo que ha mejorado significativamente la aplicabilidad práctica de estas herramientas en entornos clínicos y deportivos.

Dentro de la cantidad de técnicas identificadas, las redes neuronales convolucionales (CNN) y la estimación de pose sin marcadores (Pose Estimation Markerless) fueron las más utilizadas, consolidándose como las bases de los sistemas modernos de análisis del movimiento. Las CNN demostraron ser altamente efectivas a la hora de la extracción de características espaciales complejas y en la identificación de patrones de movimiento, lo que les muestra una notable robustez frente a variaciones de iluminación o fondo. Sin embargo, su principal limitación radica en la alta demanda computacional y la necesidad de grandes cantidades de datos etiquetados para entrenamiento, lo que restringe su adaptación en entornos de bajo costo o en aplicaciones en tiempo real. Por su parte, las técnicas markerless se destacan por su carácter no invasivo y su accesibilidad, al prescindir de sensores de contacto o marcadores físicos. Sin embargo, su desempeño puede verse comprometido por las oclusiones parciales del cuerpo, la posición de la cámara o la calibración insuficiente del sistema, factores que inciden en la fiabilidad de las mediciones articulares.

Al analizar el comportamiento de los distintos tipos de movimientos, se observa que la flexión y extensión del codo y las posturas del tronco superior son los patrones más frecuentes en los estudios revisados. Esto se debe a su simplicidad biomecánica, su baja variabilidad intersujeto y la facilidad con la que los algoritmos pueden detectar los cambios angulares asociados. Estos movimientos son particularmente útiles para validar modelos de captura y segmentación, además de tener un valor clínico importante en la rehabilitación de lesiones del miembro superior. En cambio, los movimientos del hombro, como la abducción o la elevación, presentan mayor complejidad estructural debido al número de grados de libertad implicados y a la superposición de

segmentos, lo que incrementa la probabilidad de error en la reconstrucción tridimensional. Los movimientos funcionales compuestos, que integran varios segmentos corporales, aportan una evaluación más realista de la funcionalidad motora, aunque resultan más exigentes para la segmentación automática y la generalización del modelo.

Asimismo, los resultados revelan que existe una correlación entre el tipo de movimiento analizado y la técnica de visión por computador empleada. Por ejemplo, los estudios basados en CNN y Graph Convolutional Networks (GCN) tienden a enfocarse en movimientos de mayor complejidad biomecánica, como la abducción de hombro o los patrones funcionales, debido a su capacidad para modelar relaciones espaciales jerárquicas entre articulaciones. En contraste, los métodos markerless simples o basados en filtros estadísticos suelen utilizarse en tareas más delimitadas, como la flexión de codo o la evaluación postural, donde la precisión angular es más controlable. Esta correspondencia sugiere que la elección del enfoque técnico no solo responde al objetivo del estudio, sino también a la viabilidad computacional y al nivel de control experimental disponible.

Otro hallazgo relevante es el creciente interés por la integración de sistemas visuales e inerciales, donde las cámaras RGB o RGB-D se complementan con unidades de medición inercial (IMU) para mejorar la precisión y estabilidad de las mediciones. Aunque esta combinación aumenta la robustez del sistema y permite registrar movimientos en condiciones más realistas, también introduce mayores requerimientos de calibración y sincronización, lo que explica por qué su uso aún no es generalizado. Aun así, las investigaciones que aplican estos enfoques híbridos demuestran un avance importante hacia la consolidación de sistemas portátiles, económicos y clínicamente aplicables, lo que amplía el potencial de la visión por computador más allá de los laboratorios especializados.

En cuanto a la aplicabilidad, los resultados confirman que el entorno de uso condiciona tanto la selección de la técnica como el tipo de movimiento analizado. En contextos clínicos, las investigaciones tienden a priorizar la precisión angular y la reproducibilidad de las mediciones, mientras que en el ámbito deportivo se privilegia la capacidad de análisis en tiempo real y la detección de patrones de rendimiento. En ambos casos, se destaca la ventaja de las soluciones

markerless por su facilidad de implementación, aunque persisten retos vinculados con la variabilidad de las condiciones ambientales y la necesidad de calibración dinámica del sistema.

Finalmente, las oportunidades de investigación identificadas apuntan hacia tres direcciones estratégicas. Primero, el desarrollo de modelos más robustos ante oclusiones y variaciones de entorno, mediante arquitecturas más profundas y mecanismos de atención espacial. Segundo, la optimización del procesamiento en tiempo real, con modelos más ligeros y eficientes que puedan ejecutarse en dispositivos de bajo costo. Y tercero, la integración de aprendizaje multimodal y adaptativo, que permita combinar información visual, cinemática e inercial para capturar la complejidad del movimiento humano de forma holística. Estas perspectivas refuerzan la idea de que el análisis del tren superior mediante visión por computador se encuentra en un punto de madurez tecnológica que permite su transición hacia aplicaciones prácticas en salud, deporte y ergonomía, aunque aún requiere superar limitaciones metodológicas y de validación empírica.

En relación con el análisis desarrollado, a continuación, se presenta una síntesis integradora de los principales hallazgos en relación con las preguntas de investigación planteadas.

El análisis sistemático del movimiento del tren superior indica una clara transición desde métodos estadísticos tradicionales hacia arquitecturas de aprendizaje profundo, donde las Redes Neuronales Convolucionales (CNN) y la estimación de pose sin marcadores (Markerless) se consolidan como las bases tecnológicas más robustas para reducir la subjetividad en evaluaciones clínicas, deportivas y ergonómicas. En respuesta a las preguntas de investigación, las técnicas empleadas incluyen adicionalmente Transformers, RNN/LSTM, GCN/GNN y reconstrucción 3D, utilizando tecnologías que van desde cámaras RGB monoculares, RGB-D y sensores de profundidad (Kinect) hasta sistemas ópticos profesionales y visual-inerciales. Los movimientos más estudiados fueron flexión de codo, abducción de hombro, posturas del tronco y tareas funcionales compuestas. Se investigan principalmente para mejorar la rehabilitación musculoesquelética, el desempeño deportivo y el monitoreo ergonómico. Si bien existen oportunidades latentes en la fusión multimodal de sensores, la validación clínica estandarizada y el procesamiento en tiempo real, la investigación enfrenta limitaciones críticas como la escasez de estudios centrados exclusivamente en el tren superior, la alta sensibilidad al error por oclusiones

corporales, la falta de datasets diversos que representen a poblaciones con discapacidades y la baja exactitud reportada en movimientos finos de la muñeca y el antebrazo.

Conclusiones

A través de una revisión sistemática de literatura exhaustiva se logró identificar, organizar y analizar en detalle las técnicas de visión por computador aplicadas al estudio del movimiento del tren superior del cuerpo humano, así como los tipos de movimientos más investigados, las tecnologías utilizadas para su captura y las oportunidades de investigación emergentes en este campo. A partir del análisis de los 59 artículos incluidos, se evidencia que las metodologías basadas en deep learning (especialmente la estimación de pose markerless), las Redes Neuronales Convolucionales (CNN), las arquitecturas Transformers y los modelos basados en grafos constituyen la base tecnológica más sólida para la caracterización del movimiento humano, debido a su capacidad para extraer patrones complejos y proporcionar resultados biomecánicamente precisos. Asimismo, los resultados muestran que los movimientos del hombro, codo, antebrazo y tronco son los más estudiados, principalmente por su relevancia clínica y funcional en procesos de rehabilitación, ergonomía y desempeño físico. Sin embargo, también se observa una falta de coherencia en la selección de las tareas funcionales, así como un menor enfoque a movimientos finos como lo son; los movimientos de muñeca y antebrazo, lo que las señala como áreas aún poco exploradas.

En cuanto a las tecnologías de captura, se identificó una clara inclinación por el uso de cámaras RGB, sistemas RGB-D y plataformas combinadas visual-inerciales, cada una con ventajas propias en costo, precisión, complejidad y adaptabilidad. Este panorama muestra una tendencia clara hacia sistemas no invasivos, accesibles y de bajo costo, sin que se vea la necesidad de sacrificar la calidad de la medición, aunque aún hay limitaciones por oclusiones, cambios en el entorno y restricciones geométricas en la captura.

Por último, la revisión identifica oportunidades de investigación en cuatro áreas principales: la necesidad de validación clínica estandarizada, la integración multimodal más robusta, la mejora de la robustez y funcionamiento en tiempo real en el borde (edge computing), y la creación de datasets representativos que incluyan grupos con diferentes capacidades. Estas oportunidades confirman que, a pesar de que se han logrado avances significativos, el campo aún requiere esfuerzos orientados a la generalización, el uso en entornos reales y la aplicabilidad clínica, abriendo un gran camino para futuros desarrollos en ingeniería electrónica y ciencias aplicadas al movimiento humano.

Recomendaciones

A partir del trabajo desarrollado en esta revisión sistemática, se han identificado varios caminos que pueden orientar futuras investigaciones sobre el análisis del movimiento del tren superior mediante visión por computador. En primer parte, se recomienda realizar estudios con poblaciones más amplias y diversas, incluyendo personas de distintas situaciones médicas, edades, niveles de aptitud y entornos de trabajo. Esto permitirá aumentar la capacidad de generalización de los modelos y fortalecer su validez externa. Del mismo modo, resulta imprescindible progresar en la formulación de procedimientos de validación clínica normalizados que integren indicadores equiparables, métodos de referencia constantes y criterios de calidad biomecánica lógicos. Esto favorecerá la comparación entre tecnologías y facilitará su adopción en entornos terapéuticos y médicos.

Otro aspecto a tener en cuenta es promover la integración combinada entre cámaras RGB, sensores inerciales y sistemas RGB-D, empleando algoritmos de calibración más sólidos y precisos. La combinación de diferentes fuentes de datos puede mejorar significativamente la estabilidad temporal, la precisión angular y la resistencia a oclusiones en situaciones reales. También se recomienda avanzar hacia sistemas optimizados para operar en tiempo real y computación en el borde, reduciendo la dependencia de infraestructuras de alto costo y posibilitando la futura creación de aplicaciones portátiles, asequibles y adaptables en ámbitos como la rehabilitación, la ergonomía y el deporte. Para terminar, se sugiere la creación de datasets funcionales y contextualizados, que incluyan movimientos trascendentes desde el punto de vista clínico, actividades diarias y registros en condiciones no reguladas. Esto facilitará el entrenamiento de modelos que sean más representativos, favoreciendo un desempeño más consistente y provechoso en situaciones prácticas.

Referencias

- Abrar Ashraf, M., Najam, S., Sadiq, T., Algamdi, S., Aljuaid, H., Rahman, H., & Jalal, A. (2025). A Novel Telerehabilitation System for Physical Exercise Monitoring in Elderly Healthcare. *IEEE Access*, *13*(January), 9120–9133. <https://doi.org/10.1109/ACCESS.2025.3526710>
- Abromavičius, V., Gisleris, E., Daunoravičienė, K., Žižienė, J., Serackis, A., & Maskeliūnas, R. (2025). Enhanced Human Skeleton Tracking for Improved Joint Position and Depth Accuracy in Rehabilitation Exercises. *Applied Sciences (Switzerland)*, *15*(2). <https://doi.org/10.3390/app15020906>
- Adjel, M., Dumas, R., Mohammed, S., & Bonnet, V. (2025). Influence of Visual-Inertial Sensor-to-Segment Calibration on Upper Limb Joint Angles Estimation From Multiple Inverse Kinematics Methods. *IEEE Transactions on Automation Science and Engineering*, *22*, 11519–11528. <https://doi.org/10.1109/TASE.2025.3535857>
- Adolf, J., Kán, P., Feuchtner, T., Adolfová, B., Doležal, J., & Lhotská, L. (2025). Offstretch: camera-based real-time feedback for daily stretching exercises. *Visual Computer*, *41*(3), 1555–1571. <https://doi.org/10.1007/s00371-024-03450-y>
- Alexander, F., & Urbina, Z. (2024). *Cesmag 2024*.
- Ayma Quirita, V. H., Achanccaray Diaz, P. M., Arauco Canchumuni, S. W., & Soto Vega, P. J. (2022). Desafíos del Aprendizaje Profundo en la Visión por Computador. *Entornos Híbridos En La Pospandemia: Posibilidades Para Las Nuevas Tecnologías. Congreso Internacional de Ingeniería de Sistemas 2022*, 49–53. <https://doi.org/10.26439/ciis2022.6070>
- Behera, A., Matthew, P., Keidel, A., Vangorp, P., Fang, H., & Canning, S. (2020). Associating Facial Expressions and Upper-Body Gestures with Learning Tasks for Enhancing Intelligent Tutoring Systems. *International Journal of Artificial Intelligence in Education*, *30*(2), 236–270. <https://doi.org/10.1007/s40593-020-00195-2>
- Boudreault-Morales, G. E., Marquez-Chin, C., Liu, X., & Zariffa, J. (2025). The effect of depth data and upper limb impairment on lightweight monocular RGB human pose estimation models. *BioMedical Engineering Online*, *24*(1). <https://doi.org/10.1186/s12938-025-01347-y>
- Cao, W., Zhong, J., Cao, G., & He, Z. (2019). Physiological Function Assessment Based on Kinect V2. *IEEE Access*, *7*, 105638–105651. <https://doi.org/10.1109/ACCESS.2019.2932101>
- Carrizo, D., & Moller, C. (2018). Methodological structures of systematic literature review in

- software engineering: A systematic mapping study. *Ingeniare*, 26, 45–54. <https://doi.org/10.4067/S0718-33052018000500045>
- Ceglia, A., Facon, K., Begon, M., & Seoud, L. (2025). Real-time, accurate, and open source upper-limb musculoskeletal analysis using a single RGBD camera — An exploratory hand-cycling study. *Computers in Biology and Medicine*, 184(November 2024), 109434. <https://doi.org/10.1016/j.combiomed.2024.109434>
- Chai, T. (2014). *Root mean square error (RMSE) or mean absolute error (MAE)?* 1525–1534. <https://doi.org/10.5194/gmdd-7-1525-2014>
- Chang, M. C., Ke, L., Qi, H., Wen, L., & Lyu, S. (2021). Fast Online Video Pose Estimation by Dynamic Bayesian Modeling of Mode Transitions. *IEEE Transactions on Cybernetics*, 51(1), 2–15. <https://doi.org/10.1109/TCYB.2018.2884216>
- Colyer, S. L., Evans, M., Cosker, D. P., & Salo, A. I. T. (2018). Colyer2018_Article_AReviewOfTheEvolutionOfVision-.pdf. *Sports Medicine-Open*, 4(1), 24.
- Dalimarta, F. F., Andono, P. N., Soeleman, M. A., & Hasibuan, Z. A. (2025). Enhancing Motoric Impulsivity Detection in Children through Deep Learning and Body Keypoint Recognition. *International Journal on Informatics Visualization*, 9(1), 275–285. <https://doi.org/10.62527/joiv.9.1.2779>
- David, A., Alonso, C., & Mart, A. (n.d.). *Vision por computadora : identificacion, clasificacion y seguimiento de objetos*.
- Dellepiane, S. G., Ferraro, F., Baffigo, C., & Simonini, M. (2025). Signal Processing and Feature Extraction in Markerless Telerehabilitation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 33, 911–924. <https://doi.org/10.1109/TNSRE.2025.3541153>
- Doong, S. H. (2025). Predicting postural risk level with computer vision and machine learning on multiple sources of images. *Engineering Applications of Artificial Intelligence*, 143(August 2024), 109981. <https://doi.org/10.1016/j.engappai.2024.109981>
- Edriss, S., Caprioli, L., Campoli, F., Manzi, V., Padua, E., Bonaiuto, V., Romagnoli, C., & Annino, G. (2024). Advancing Artistic Swimming Officiating and Performance Assessment: A Computer Vision Study Using MediaPipe. *International Journal of Computer Science in Sport*, 23(2), 35–47. <https://doi.org/10.2478/ijcss-2024-0010>
- Electrónica, P. de I. (2015). Proyecto Educativo del programa. Pasto: Universidad CESMAG.

Línea y Sublínea de Investigación U. CESMAG, 67.

- Esmail, M. A., Wang, J., Wang, Y., Sun, L., Zhu, G., & Zhang, G. (2025). Resource-aware strategies for real-time multi-person pose estimation. *Image and Vision Computing*, 155(August 2024), 105441. <https://doi.org/10.1016/j.imavis.2025.105441>
- Gao, Q., Zhang, M., & Ju, Z. (2025). LGF-SLR: Hand Local–Global Fusion Network for Skeleton-Based Sign Language Recognition. *IEEE Sensors Journal*, 25(5), 8586–8597. <https://doi.org/10.1109/JSEN.2025.3527198>
- Gao, Y., Wang, S., & Zha, Z. (2025). Human pose estimation via inter-view image similarity with adaptive weights. *Displays*, 87(January). <https://doi.org/10.1016/j.displa.2025.102972>
- Gao, Z., Chen, J., Liu, Y., Jin, Y., & Tian, D. (2025). A systematic survey on human pose estimation: upstream and downstream tasks, approaches, lightweight models, and prospects. *Artificial Intelligence Review*, 58(3). <https://doi.org/10.1007/s10462-024-11060-2>
- Giulietti, N., Todesca, D., Carnevale, M., & Giberti, H. (2025). A Real-Time Human Pose Measurement System for Human-In-The-Loop Dynamic Simulators. In *IEEE Access* (Vol. 13, pp. 24954–24969). <https://doi.org/10.1109/ACCESS.2025.3538332>
- Harrison, S. M., Cohen, R. C. Z., O’Hanlon, J., Cece, T., Fisher, N., & Thrall, P. H. (2025). Proof-of-concept system evaluation of Ergomechanic for non-invasive estimation of upper-body posture and body load exposure in the workplace. *International Journal of Occupational Safety and Ergonomics*, 3548. <https://doi.org/10.1080/10803548.2025.2458442>
- Herdiman, L., Susmartini, S., Rochman, T., & Artanty, F. (2025). Videography Techniques For Gait Assessment In Children With Diplegia Cerebral Palsy In Indonesian Rehabilitation Center Clinics. *Journal of Applied Science and Engineering*, 28(5), 1073–1093. [https://doi.org/10.6180/jase.202505_28\(5\).0014](https://doi.org/10.6180/jase.202505_28(5).0014)
- Host, K., & Ivašić-Kos, M. (2022). An overview of Human Action Recognition in sports based on Computer Vision. *Heliyon*, 8(6). <https://doi.org/10.1016/j.heliyon.2022.e09633>
- Jackson, K. L., Durić, Z., Engdahl, S. M., Santago, A. C., DeStefano, S., & Gerber, L. H. (2023). Computer-assisted approaches for measuring, segmenting, and analyzing functional upper extremity movement: a narrative review of the current state, limitations, and future directions. *Frontiers in Rehabilitation Sciences*, 4(April), 1–18. <https://doi.org/10.3389/fresc.2023.1130847>
- Jia, W., Wang, H., Chen, Q., Bao, T., & Sun, Y. (2025). Analysis of Kinect-Based Human Motion

- Capture Accuracy Using Skeletal Cosine Similarity Metrics. *Sensors*, 25(4). <https://doi.org/10.3390/s25041047>
- Jiang, J., Strelcić, P., Luo, X., Gebhardt, C., & Holz, C. (2025). MANIKIN: Biomechanically Accurate Neural Inverse Kinematics for Human Motion Estimation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 15060 LNCS, 128–146. https://doi.org/10.1007/978-3-031-72627-9_8
- Jin, H., He, X., Wang, L., Zhu, Y., Jiang, W., & Zhou, X. (2025). SitPose: Real-Time Detection of Sitting Posture and Sedentary Behavior Using Ensemble Learning With Depth Sensor. *IEEE Sensors Journal*, 25(7), 12444–12454. <https://doi.org/10.1109/JSEN.2025.3541821>
- José, D. F., & Peñalvo, G. (2017). *Índice*.
- Kamal, S., Alhassan, H. F., Alnusayri, M., Alatiyyah, M., Aljuaid, H., Jalal, A., & Liu, H. (2025). Vision Sensor for Automatic Recognition of Human Activities via Hybrid Features and Multi-Class Support Vector Machine. *Sensors*, 25(1). <https://doi.org/10.3390/s25010200>
- Karim, M., Khalid, S., Aleryani, A., Ullah, I., & Ali, Z. (2024). *Human Action Recognition Systems : A Review of the Trends and State-of-the-Art*. 12(March).
- Kim, T., Jung, M. C., & Mo, S. M. (2025). Suggestion for Camera Location in Monocular Markerless 3D Motion Capture System: Focused on Accuracy Comparison With Marker-Based System for Upper Limb Joints. *IEEE Access*, 13(March), 47605–47616. <https://doi.org/10.1109/ACCESS.2025.3547715>
- Kitchenham, B., Charters, S., Budgen, D., Brereton, P., Turner, M., Linkman, S., Jørgensen, M., Mendes, E., & Visaggio, G. (2007). Guidelines for performing systematic literature reviews in software engineering. EBSE Technical Report EBSE-2007-01, Keele U, D. U. (2007). *Guidelines for performing Systematic Literature Reviews in Software Engineering: Vol. EBSE Techn.*
- Koul, A., & Novembre, G. (2025). How accurately can we estimate spontaneous body kinematics from video recordings? Effect of movement amplitude on OpenPose accuracy. *Behavior Research Methods*, 57(1), 1–13. <https://doi.org/10.3758/s13428-024-02546-6>
- Kryeem, A., Boutboul, N., Bear, I., Raz, S., Eluz, D., Itah, D., Hel-Or, H., & Shimshoni, I. (2025). Action assessment in rehabilitation: Leveraging machine learning and vision-based analysis. *Computer Vision and Image Understanding*, 251(November 2024), 1–11.

- <https://doi.org/10.1016/j.cviu.2024.104228>
- Kuang, Z., Wang, J., Sun, D., Zhao, J., Shi, L., & Zhu, Y. (2025). Hierarchical Contrastive Representation for Accurate Evaluation of Rehabilitation Exercises via Multi-View Skeletal Representations. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *33*, 201–211. <https://doi.org/10.1109/TNSRE.2024.3523906>
- Kwon, J. (2025). Synchrony Vision: Capturing Body Motion Synchrony Through Phase Difference Using the Kinect. *IEEE Access*, *13*(March), 41658–41669. <https://doi.org/10.1109/ACCESS.2025.3548142>
- Lee, G., & Tan, E. W. (2025). Exploring the Potential of AI-Assisted Technology in Joint Range-of-Motion Measurements: A Reliability Study. *Medicina (Lithuania)*, *61*(1), 1–21. <https://doi.org/10.3390/medicina61010119>
- Li, T., & Dong, T. (2023). Monocular camera-based online sensor-to-segment calibration for upper body pose estimation. *Sensors and Actuators A: Physical*, *364*(July), 114829. <https://doi.org/10.1016/j.sna.2023.114829>
- Li, T., & Yu, H. (2023). Upper Body Pose Estimation Using a Visual-Inertial Sensor System With Automatic Sensor-to-Segment Calibration. *IEEE Sensors Journal*, *23*(6), 6292–6302. <https://doi.org/10.1109/JSEN.2023.3241084>
- Lin, R., Zong, H., Xu, P., Gao, S., & Ge, S. (2025). Low-cost 3D human body reconstruction under limited views and girth measurement for the apparel customisation. *Ergonomics*, *0*(0), 1–18. <https://doi.org/10.1080/00140139.2024.2449113>
- Lovanshi, M., Tiwari, V., Ingle, R., & Jain, S. (2024). Fusion of Temporal Transformer and Spatial Graph Convolutional Network for 3-D Skeleton-Parts-Based Human Motion Prediction. *IEEE Transactions on Human-Machine Systems*, *54*(6), 788–797. <https://doi.org/10.1109/THMS.2024.3452133>
- Maggioni, V., Azevedo-Coste, C., Durand, S., & Bailly, F. (2025). Optimisation and Comparison of Markerless and Marker-Based Motion Capture Methods for Hand and Finger Movement Analysis. *Sensors*, *25*(4). <https://doi.org/10.3390/s25041079>
- Moreira, R., Teixeira, S., Fialho, R., Miranda, A., Lima, L. D. B., Carvalho, M. B., Alves, A. B., Bastos, V. H. V., & Teles, A. S. (2024). Validity Analysis of Monocular Human Pose Estimation Models Interfaced with a Mobile Application for Assessing Upper Limb Range of Motion. *Sensors*, *24*(24), 1–18. <https://doi.org/10.3390/s24247983>

- Morillo-Baro, J. P., Reigal, R. E., & Hernández-Mendo, A. (2015). Análisis del ataque posicional de balonmano playa masculino y femenino mediante coordenadas polares. *RICYDE: Revista Internacional de Ciencias Del Deporte*, *11*(41), 226–244. <https://doi.org/10.5232/ricyde>
- Murillo, P. U., Jiménez-Moreno, R., & Baquero, J. E. M. (2025). Imitation of the human upper limb by convolutional neural networks. *IAES International Journal of Artificial Intelligence*, *14*(1), 193–203. <https://doi.org/10.11591/ijai.v14.i1.pp193-203>
- Paiva, P. V. V., Ramos, J. J. G., Gavrilova, M., & Carvalho, M. A. G. de. (2025). SkelETT - Skeleton-to-Emotion Transfer Transformer. *IEEE Access*, *13*(January), 23344–23358. <https://doi.org/10.1109/ACCESS.2025.3534145>
- Perales, F. (2001). Human motion analysis and synthesis using computer vision and graphics techniques. State of art and applications. *Proc. World Multiconf. on Systemics, Cybernetics and ...*
http://pdf.aminer.org/000/351/715/analysis_and_synthesis_of_human_motion_from_external_measurements.pdf
- Raj, V., Mittal, K., Shenbagam, M., Perunthattil, J., Venkataraman, S., & Mukherjee, B. (2025). Toward Markerless, Noncontact Optical Myography for Prosthetic Control of Multiple Degrees of Freedom. *IEEE Sensors Journal*, *25*(4), 7154–7162. <https://doi.org/10.1109/JSEN.2024.3512454>
- Ramzan, M., & Dawn, S. (2019). Temporal measures for analysis of emotional states from human electroencephalography signals. *2019 12th International Conference on Contemporary Computing, IC3 2019*, 1–6. <https://doi.org/10.1109/IC3.2019.8844893>
- Ryu, S. M., Shin, K., Doh, C. H., Ben, H., Park, J. Y., Koh, K. H., Shin, H., & Jeon, I. H. (2025). Orthopedic surgeon level joint angle assessment with artificial intelligence based on photography: a pilot study. *Biomedical Engineering Letters*, *15*(1), 131–142. <https://doi.org/10.1007/s13534-024-00432-w>
- Scott, B., Seyres, M., Philp, F., Chadwick, E. K., & Blana, D. (2022). Healthcare applications of single camera markerless motion capture: a scoping review. *PeerJ*, *10*, 1–27. <https://doi.org/10.7717/peerj.13517>
- Shahid, M. A., Raza, M., Sharif, M., Alshenaifi, R., & Kadry, S. (2025). Pedestrian POSE estimation using multi-branched deep learning pose net. In *PLoS ONE* (Vol. 20, Issue 1 January). <https://doi.org/10.1371/journal.pone.0312177>

- Shi, L., Wang, R., Zhao, J., Zhang, J., & Kuang, Z. (2024). Detection of Rehabilitation Training Effect of Upper Limb Movement Disorder Based on MPL-CNN. *Sensors*, 24(4), 1–19. <https://doi.org/10.3390/s24041105>
- Sun, Y., Deng, R., & Wei, D. (2025). ST-LineNet: A spatiotemporal network for real-time 3D Pose estimation in martial arts training. *Alexandria Engineering Journal*, 117(January), 136–147. <https://doi.org/10.1016/j.aej.2024.12.097>
- Suo, X., Tang, W., Mao, L., & Li, Z. (2025). Digital human and embodied intelligence for sports science: advancements, opportunities and prospects. *Visual Computer*, 41(4), 2477–2493. <https://doi.org/10.1007/s00371-024-03547-4>
- Tanineh, W., & Halaweh, H. (2023). Cardiorespiratory Fitness, Motor Coordination, and Academic Achievement in School Students (11-13 years). *Global Pediatric Health*, 10. <https://doi.org/10.1177/2333794X231207311>
- Tanthuwapathom, R., Manupibul, U., Jarumethitanont, W., Limroongreungrat, W., Ongwattanakul, S., & Charoensuk, W. (2025). Reliability of sitting posture between physical therapist video-based evaluation and SMART IMU system using rapid upper limb assessment (RULA). *Scientific Reports*, 15(1), 1–11. <https://doi.org/10.1038/s41598-025-85159-z>
- van den Hoorn, W., Fabre, A., Nardese, G., Su, E. Y. S., Cutbush, K., Gupta, A., & Kerr, G. (2025). The Future of Clinical Active Shoulder Range of Motion Assessment, Best Practice, and Its Challenges: Narrative Review. *Sensors*, 25(3), 1–24. <https://doi.org/10.3390/s25030667>
- Wang, Y., Niu, X., Lv, X., & Yu, C. (2025). FFCI: A Camera and IMU Sensors Based Multi-modal Neural Network for Activity Recognition in Smart Factory. *IEEE Transactions on Consumer Electronics*, PP, 1. <https://doi.org/10.1109/TCE.2025.3534236>
- Warguła, Ł., Wieczorek, B., Giedrowicz, M., Kukla, M., & Nati, C. (2025). Wood Chippers: Influence of Feed Channel Geometry on Possibility of Musculoskeletal System Overload. *Croatian Journal of Forest Engineering*, 46(1), 59–76. <https://doi.org/10.5552/crojfe.2025.2501>
- Wu, J., Chen, J., Luo, Q., Liu, S., Xue, Y., & Ma, H. (2025). Upper-Body Hierarchical Graph for Skeleton Based Emotion Recognition in Assistive Driving. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 15084 LNCS, 1–17. https://doi.org/10.1007/978-3-031-73347-5_1
- Xu, Z., & Xu, J. (2024). GR-Former: Graph-reinforcement transformer for skeleton-based driver

- action recognition. *IET Computer Vision*, 18(7), 982–991. <https://doi.org/10.1049/cvi2.12298>
- Xu, Z., & Xu, J. (2025). Spatiotemporal decoupling attention transformer for 3D skeleton-based driver action recognition. *Complex and Intelligent Systems*, 11(4), 1–12. <https://doi.org/10.1007/s40747-025-01811-1>
- Yan, X., Zhang, L., Liu, B., & Qu, G. (2025). Efficient Limb Range of Motion Analysis from a Monocular Camera for Edge Devices †. *Sensors*, 25(3), 1–20. <https://doi.org/10.3390/s25030627>
- Zaher, M., Ghoneim, A. S., Abdelhamid, L., & Atia, A. (2025). Unlocking the potential of RNN and CNN models for accurate rehabilitation exercise classification on multi-datasets. *Multimedia Tools and Applications*, 84(3), 1261–1301. <https://doi.org/10.1007/s11042-024-19092-0>
- Zhao, A., Tang, C., Wang, L., Li, Y., Dave, M., Tao, L., Twigg, C. D., & Wang, R. Y. (2025). EgoBody3M: Egocentric Body Tracking on a VR Headset using a Diverse Dataset. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 15137 LNCS, 375–392. https://doi.org/10.1007/978-3-031-72986-7_22
- Zhao, W., Wang, L., Li, Y., Liu, X., Zhang, Y., Yan, B., & Li, H. (2024). A Multi-Scale and Multi-Stage Human Pose Recognition Method Based on Convolutional Neural Networks for Non-Wearable Ergonomic Evaluation. *Processes*, 12(11). <https://doi.org/10.3390/pr12112419>
- Zhou, Y., Seo, J. O., Gong, Y., Heung, K. H. L., Khan, M., & Lei, T. (2025). Biomechanical assessment of a passive back exoskeleton using vision-based motion capture and virtual modeling. *Automation in Construction*, 172(September 2024), 106035. <https://doi.org/10.1016/j.autcon.2025.106035>

 <p>UNIVERSIDAD CESMAG NIT: 800.109.387-7 VIGILADA MINEDUCACIÓN</p>	CARTA DE ENTREGA TRABAJO DE GRADO O TRABAJO DE APLICACIÓN – ASESOR(A)	CÓDIGO: AAC-BL-FR-032
		VERSIÓN: 1
		FECHA: 09/JUN/2022

San Juan de Pasto, 23 de abril de 2026

Biblioteca
REMIGIO FIORE FORTEZZA OFM. CAP.
Universidad CESMAG
Pasto

Saludo de paz y bien.

Por medio de la presente se hace entrega del Trabajo de Grado / Trabajo de Aplicación denominado “Revisión sistemática de literatura del análisis de movimiento del tren superior del cuerpo humano mediante visión por computador”, presentado por los autores Daniel Esteban Quintero Unigarro y Juan Diego Rivera Caicedo del Programa Académico Ingeniería Electrónica al correo electrónico biblioteca.trabajosdegrado@unicesmag.edu.co. Manifiesto como asesor(a) Mg. Mario Fernando Henao Rosero, que su contenido, resumen, anexos y formato PDF cumple con las especificaciones de calidad, guía de presentación de Trabajos de Grado o de Aplicación, establecidos por la Universidad CESMAG, por lo tanto, se solicita el paz y salvo respectivo.

Atentamente,




Mario Fernando Henao Rosero
c.c. 4612719
Ingeniería Electrónica
3163455039
mfhenao@unicesmag.edu.co

 UNIVERSIDAD CESMAG <small>NIT: 800.109.387-7 VIGILADA MINEDUCACIÓN</small>	AUTORIZACIÓN PARA PUBLICACIÓN DE TRABAJOS DE GRADO O TRABAJOS DE APLICACIÓN EN REPOSITORIO INSTITUCIONAL	CÓDIGO: AAC-BL-FR-031
		VERSIÓN: 1
		FECHA: 09/JUN/2022

INFORMACIÓN DEL (LOS) AUTOR(ES)	
Nombres y apellidos del autor: Juan Diego Rivera Caicedo	Documento de identidad: 1081052263
Correo electrónico: Juan20rivera@outlook.com	Número de contacto: 3143389174
Nombres y apellidos del autor: Daniel Esteban Quintero Unigarro	Documento de identidad: 1004234835
Correo electrónico: unigarrodaniel3@gmail.com	Número de contacto: 3057213509
Nombres y apellidos del autor:	Documento de identidad:
Correo electrónico:	Número de contacto:
Nombres y apellidos del autor:	Documento de identidad:
Correo electrónico:	Número de contacto:
Nombres y apellidos del asesor: Mario Fernando Henao Rosero	Documento de identidad: 4612719
Correo electrónico: mfhenao@unicesmag.edu.co	Número de contacto: 3163455039
Título del trabajo de grado: Revisión sistemática de literatura del análisis de movimiento del tren superior del cuerpo humano mediante visión por computador	
Facultad y Programa Académico: Sistemas - Ingeniería Electrónica	

En mi (nuestra) calidad de autor(es) y/o titular (es) del derecho de autor del Trabajo de Grado o de Aplicación señalado en el encabezado, confiero (conferimos) a la Universidad CESMAG una licencia no exclusiva, limitada y gratuita, para la inclusión del trabajo de grado en el repositorio institucional. Por consiguiente, el alcance de la licencia que se otorga a través del presente documento, abarca las siguientes características:

- a) La autorización se otorga desde la fecha de suscripción del presente documento y durante todo el término en el que el (los) firmante(s) del presente documento conserve (mos) la titularidad de los derechos patrimoniales de autor. En el evento en el que deje (mos) de tener la titularidad de los derechos patrimoniales sobre el Trabajo de Grado o de Aplicación, me (nos) comprometo (comprometemos) a informar de manera inmediata sobre dicha situación a la Universidad CESMAG. Por consiguiente, hasta que no exista comunicación escrita de mi(nuestra) parte informando sobre dicha situación, la Universidad CESMAG se encontrará debidamente habilitada para continuar con la publicación del Trabajo de Grado o de Aplicación dentro del repositorio institucional. Conozco(conocemos) que esta autorización podrá revocarse en cualquier momento, siempre y cuando se eleve la solicitud por escrito para dicho fin ante la Universidad CESMAG. En estos eventos, la Universidad CESMAG cuenta con el plazo de un mes después de recibida la

 UNIVERSIDAD CESMAG <small>NIT: 800.109.387-7 VIGILADA MINEDUCACIÓN</small>	AUTORIZACIÓN PARA PUBLICACIÓN DE TRABAJOS DE GRADO O TRABAJOS DE APLICACIÓN EN REPOSITORIO INSTITUCIONAL	CÓDIGO: AAC-BL-FR-031
		VERSIÓN: 1
		FECHA: 09/JUN/2022

petición, para desmarcar la visualización del Trabajo de Grado o de Aplicación del repositorio institucional.




- b) Se autoriza a la Universidad CESMAG para publicar el Trabajo de Grado o de Aplicación en formato digital y teniendo en cuenta que uno de los medios de publicación del repositorio institucional es el internet, acepto(amos) que el Trabajo de Grado o de Aplicación circulará con un alcance mundial.
- c) Acepto (aceptamos) que la autorización que se otorga a través del presente documento se realiza a título gratuito, por lo tanto, renuncio(amos) a recibir emolumento alguno por la publicación, distribución, comunicación pública y/o cualquier otro uso que se haga en los términos de la presente autorización y de la licencia o programa a través del cual sea publicado el Trabajo de grado o de Aplicación.
- d) Manifiesto (manifestamos) que el Trabajo de Grado o de Aplicación es original realizado sin violar o usurpar derechos de autor de terceros y que ostento(amos) los derechos patrimoniales de autor sobre la misma. Por consiguiente, asumo(asumimos) toda la responsabilidad sobre su contenido ante la Universidad CESMAG y frente a terceros, manteniéndose indemne de cualquier reclamación que surja en virtud de la misma. En todo caso, la Universidad CESMAG se compromete a indicar siempre la autoría del escrito incluyendo nombre de(los) autor(es) y la fecha de publicación.
- e) Autorizo(autorizamos) a la Universidad CESMAG para incluir el Trabajo de Grado o de Aplicación en los índices y buscadores que se estimen necesarios para promover su difusión. Así mismo autorizo (autorizamos) a la Universidad CESMAG para que pueda convertir el documento a cualquier medio o formato para propósitos de preservación digital.


NOTA: En los eventos en los que el trabajo de grado o de aplicación haya sido trabajado con el apoyo o patrocinio de una agencia, organización o cualquier otra entidad diferente a la Universidad CESMAG. Como autor(es) garantizo(amos) que he(hemos) cumplido con los derechos y obligaciones asumidos con dicha entidad y como consecuencia de ello dejo(dejamos) constancia que la autorización que se concede a través del presente escrito no interfiere ni transgrede derechos de terceros.

Como consecuencia de lo anterior, autorizo(autorizamos) la publicación, difusión, consulta y uso del Trabajo de Grado o de Aplicación por parte de la Universidad CESMAG y sus usuarios así:

- Permiso(permitimos) que mi(nuestro) Trabajo de Grado o de Aplicación haga parte del catálogo de colección del repositorio digital de la Universidad CESMAG, por lo tanto, su contenido será de acceso abierto donde podrá ser consultado, descargado y compartido con otras personas, siempre que se reconozca su autoría o reconocimiento con fines no comerciales.

En señal de conformidad, se suscribe este documento en San Juan de Pasto a los 23 días del mes de abril del año 2026

	
Nombre del autor: Juan Diego Rivera Caicedo	Nombre del autor: Daniel Esteban Quintero Unigarro
Firma del autor	Firma del autor
Nombre del autor:	Nombre del autor:
	

 <p>UNIVERSIDAD CESMAG NIT: 800.109.387-7 VIGILADA MINEDUCACIÓN</p>	AUTORIZACIÓN PARA PUBLICACIÓN DE TRABAJOS DE GRADO O TRABAJOS DE APLICACIÓN EN REPOSITORIO INSTITUCIONAL	CÓDIGO: AAC-BL-FR-031
		VERSIÓN: 1
		FECHA: 09/JUN/2022

Nombre del asesor: Mario Fernando Henao Rosero