Desarrollo de un algoritmo de visión por computador para el análisis biomecánico del saque de piso en arqueros de fútbol.

Laura Sofía Salamanca Barrera
Tutora: Laura Juliana Ramos Maldonado
Co-tutor: Oscar Danilo Martínez Bernal

Pontificia Universidad Javeriana
Facultad de Ciencias
Pregrado en Ciencia de Datos
Bogotá D.C
2025

Índice

Agradecimientos **10**

Resumen **11**

1. Introducción **12**
1.1. Planteamiento del problema 12
1.2. Justificación 12

2. Objetivos **14**
2.1. Objetivo general 14
2.2. Objetivos específicos 14

3. Estado del arte **15**
3.1. Descripción del movimiento. 15
3.1.1. Saque de piso 15
3.2. Ángulos biomecánicos ideales en el saque de piso de los arqueros 17
3.3. Técnicas de análisis biomecánico en el deporte 18
3.3.1. Digitalización Manual 19
3.3.2. Análisis de movimiento con marcadores 19
3.3.3. Sensores inerciales 20
3.3.4. Análisis de movimiento sin marcadores 20
3.4. Redes Neuronales Convolucionales para el procesamiento de imágenes 21
3.5. Librerías para la estimación de pose 22
3.5.1. MediaPipe 22
3.5.2. YOLOv11 23
3.5.3. Cuadro comparativo 23
3.6. Estimación de pose con MediaPipe 24
3.6.1. Fuentes de datos 24
3.6.2. Arquitectura del modelo 25
3.6.3. Pipeline de inferencia 26
3.6.4. Detector de personas 27
3.6.5. Extracción de características 28
3.6.6. Heatmaps y offsets 33
3.6.7. Regresión de puntos clave 36
3.6.8. Estimación coordenada 3D 37

4. Metodología **39**
4.1. Recolección de datos . . . 40
4.2. Preprocesamiento de datos . . . 46
4.2.1. Espacios de color . . . 46
4.2.2. Recorte de videos . . . 51
4.3. Desarrollo del algoritmo . . . 63
4.4. Comparación con una herramienta de análisis biomecánico manual . . . 68
4.5. Desarrollo de la interfaz . . . 71

5. Resultados y análisis de resultados **77**
5.1. Selección de los umbrales . . . 77
5.1.1. Vista lateral derecha . . . 78
5.1.2. Vista lateral izquierda . . . 86
5.1.3. Vista trasera . . . 92
5.2. Evaluación de consistencia y precisión del algoritmo . . . 97
5.2.1. Rodilla en la vista lateral derecha . . . 99
5.2.2. Pie en la vista lateral derecha . . . 103
5.2.3. Rodilla en la vista lateral izquierda . . . 106
5.2.4. Pie en la vista lateral izquierda . . . 110
5.2.5. Rodilla en la vista trasera . . . 114

6. Conclusiones **118**

7. Mejoras para Trabajos Futuros **119**

A. Anexos **120**

Índice de tablas

3.1. Datos cinemáticos correspondientes al instante de apoyo del pie (media ± desviación estándar). 17
3.2. Datos cinemáticos al momento del contacto con el balón (media ± desviación estándar). 17
3.3. Ángulos articulares relativos (0° corresponde a la posición anatómica). Comparación entre patadas con empeine y con curva. 18
3.4. Comparación entre la estimación de pose de YOLOv11 y de MediaPipe. 23
4.1. Clasificación del ángulo de la rodilla (momento final). 74
4.2. Clasificación del ángulo de la cadera (momento final). 74
4.3. Ángulo de la rodilla (momento inicial). 76
4.4. Ángulo de la rodilla (momento final). 76
4.5. Ángulo del tronco vertical (hombro–cadera). 76
4.6. Ángulo de inclinación del tronco (momento final). 77
5.1. Umbrales por marcador – Vista Lateral Derecha 78
5.2. Umbrales por marcador – Vista Lateral Izquierda 78
5.3. Umbrales por marcador – Vista Trasera 78
5.4. Comparación de ángulos de rodilla entre Kinovea y MediaPipe (vista lateral derecha, archivo 18). 100
5.5. Comparación del ángulo del pie entre Kinovea y MediaPipe (vista lateral derecha, archivo 18). 103
5.6. Comparación del ángulo de la rodilla entre Kinovea y MediaPipe (vista lateral izquierda, archivo 14). 107
5.7. Comparación del ángulo del pie entre Kinovea y MediaPipe (vista lateral izquierda, archivo 14). 111
5.8. Comparación del ángulo de la rodilla entre Kinovea y MediaPipe (vista trasera, archivo 55). 114

Índice de figuras

3.1. Fase de preparación en el saque de piso. 15
3.2. Fase de carga en el saque de piso. 16
3.3. Fase de golpeo en el saque de piso. 16
3.4. Fase de recuperación en el saque de piso. 16
3.5. 33 puntos clave identificados por MediaPipe (Bazarevsky et al., 2020). 24
3.6. Arquitectura del Modelo de Mediapipe (Elaboración propia). 25
3.7. Pipeline de inferencia (Bazarevsky et al., 2020). 26
3.8. Hombre de Vitruvio (Bazarevsky et al., 2020). 27
3.9. Convolución estándar (Zurister, 2019). 29
3.10. Convolución en profundidad y convolución puntual (Cheuque et al., 2022). 30
3.11. Ejemplos de desplazamientos al operar las convoluciones (Baeldung, 2023). 31
3.12. Comparación entre técnicas de reducción espacial: (a) max pooling selecciona el valor más alto en cada región; (b) average pooling calcula el promedio de los valores. (Baeldung, 2023) 32
3.13. Visualización de la salida del modelo durante el entrenamiento: (a) activación de los *heatmaps* para cada punto clave; (b) vectores de desplazamiento (*offsets*) utilizados para refinar la localización. Imágenes tomadas de Xiao et al., 2022. 33
3.14. Relación entre la resolución del *heatmap* y la imagen original. Cada píxel del *heatmap* cubre una región de múltiples píxeles en la imagen original. La coordenada de máxima activación por sí sola no es suficiente para determinar la posición exacta, por lo que se predice un vector de *offset* para refinar la ubicación dentro de la celda (Xiao et al., 2022). 34
3.15. Mallas utilizadas para el entrenamiento (Xu et al., 2020). 38
3.16. Mallas generadas durante el entrenamiento (Xu et al., 2020). 39
4.1. Diagrama de flujo de la metodología general del proyecto. 40
4.2. Diagrama de flujo de la primera etapa de la metodología: recolección de datos. . . 40
4.3. Puntos de movimiento del arquero en el saque de piso y área delimitada. 42
4.4. Parámetros fundamentales de un sistema de imágenes (Optics, s.f.). 43
4.5. Diagrama de un lente focal fijo (Optics, s.f.). 43
4.6. Configuración de las cámaras. 45
4.7. Configuración de la toma de videos utilizada durante las sesiones de grabación. . . 46
4.8. Fotogramas cargados en BGR. 47
4.9. Fotogramas cargados en RGB. 48
4.10. Fotogramas cargados en HSV. 49
4.11. Fotogramas cargados en YCrCb. 50

4.12. Fotogramas cargados en LAB. 51
4.13. Identificación de la coordenada Y de la rodilla (color azul claro) y el tobillo (color rojo), así como la diferencia entre ambas coordenadas (color azul). 52
4.14. Comparación de coordenadas x en la vista sagital derecha. 53
4.15. Comparación de coordenadas x en la vista sagital izquierda. 53
4.16. Coordenada z de los tobillos para la vista coronal. 54
4.17. Diagrama de flujo para la determinación del fotograma inicial del video según el criterio biomecánico establecido. 55
4.18. Inconsistencias en la detección del tobillo y de la rodilla desde la vista coronal. . . 56
4.19. Diagrama de flujo para la determinación del fotograma inicial del video según el criterio biomecánico auxiliar establecido. 57
4.20. Determinación del inicio del video mediante el criterio biomecánico (vista lateral derecha). 58
4.21. Determinación del inicio del video mediante el criterio biomecánico (vista lateral izquierda). 58
4.22. Determinación del inicio del video mediante el criterio biomecánico (vista trasera). 59
4.23. Inconsistencias en la detección de extremidades en fotogramas consegutivos. 60
4.24. Criterio de fin para las rodillas y los tobillos desde la vista sagital derecha. 61
4.25. Criterio de fin para las rodillas y los tobillos desde la vista sagital izquierda. 61
4.26. Criterio de fin para las rodillas y los tobillos desde la vista coronal. 62
4.27. Diagrama de flujo general del proceso de recorte de video. 63
4.28. Ejemplo de detección de cambio drástico entre fotogramas consecutivos. 64
4.29. Ángulos extraídos desde la vista lateral derecha. 67
4.30. Ángulos extraídos desde la vista lateral izquierda. 67
4.31. Ángulos extraídos desde la vista trasera. 68
4.32. Registro manual de puntos articulares en un fotograma. 68
4.33. Interfaz inicial para la carga de videos desde tres vistas distintas (derecha, izquierda y trasera). 71
4.34. Presentación final de los videos en la interfaz. 72
4.35. Vista general de la interfaz web desarrollada. 73
4.36. Clasificación resultante de la interfaz en la vista lateral derecha. 74
4.37. Clasificación resultante de la interfaz en la vista lateral izquierda para el momento inicial. 75
4.38. Clasificación resultante de la interfaz en la vista lateral izquierda para el momento final. 75
4.39. Clasificación resultante de la interfaz en la vista trasera para el momento final. . . . 77

5.2. Tabla de distancias euclidianas entre fotogramas para la vista sagital derecha. . . . 79
5.3. Boxplot de las distancias euclidianas entre fotogramas consecutivos para el marcador de la cadera derecha. 79
5.1. Secuencia de 7 fotogramas que describen el gesto técnico desde la vista sagital derecha. 80
5.4. Distribución de densidad de las distancias euclidianas para el marcador de la cadera derecha. 81
5.5. Boxplot de las distancias euclidianas entre fotogramas consecutivos para el marcador del hombro derecho. 81
5.6. Distribución de densidad de las distancias euclidianas para el marcador del hombro derecho. 82
5.7. Boxplot de las distancias euclidianas entre fotogramas consecutivos para el marcador de la rodilla derecha. 82
5.8. Distribución de densidad de las distancias euclidianas para el marcador de la rodilla derecha. 83
5.9. Boxplot de las distancias euclidianas entre fotogramas consecutivos para el marcador del talón derecho. 83
5.10. Distribución de densidad de las distancias euclidianas para el marcador del talón derecho. 84
5.11. Boxplot de las distancias euclidianas entre fotogramas consecutivos para el marcador del tobillo derecho. 84
5.12. Distribución de densidad de las distancias euclidianas para el marcador del tobillo derecho. 85
5.13. Boxplot de las distancias euclidianas entre fotogramas consecutivos para el marcador de la punta del pie derecha. 85
5.14. Distribución de densidad de las distancias euclidianas para el marcador de la punta del pie derecha. 86
5.16. Tabla de distancias euclidianas entre fotogramas para la vista sagital izquierda. . . 86
5.15. Secuencia de 8 fotogramas que describen el gesto técnico desde la vista sagital izquierda. 87
5.17. Boxplot de las distancias euclidianas entre fotogramas consecutivos para el marcador de la cadera izquierda. 88
5.18. Distribución de densidad de las distancias euclidianas para el marcador de la cadera izquierda. 88
5.19. Boxplot de las distancias euclidianas entre fotogramas consecutivos para el marcador del hombro izquierdo. 89

5.20. Distribución de densidad de las distancias euclidianas para el marcador del hombro derecho. 89
5.21. Boxplot de las distancias euclidianas entre fotogramas consecutivos para el marcador de la rodilla izquierda. 90
5.22. Distribución de densidad de las distancias euclidianas para el marcador de la rodilla izquierda. 90
5.23. Punta izquierda 91
5.24. Talon izquierdo 91
5.25. Tobillo izquierdo 91
5.26. Distribuciones de distancias euclideanas para la punta izquierda, el talón y el tobillo izquierdo, desde la vista sagital izquierda. 91
5.27. Boxplot de las distancias del pie izquierdo. 91
5.29. Tabla de distancias euclidianas entre fotogramas para la vista coronal. 92
5.30. Cadera izquierda 92
5.31. Cadera derecha 92
5.32. Distribuciones de distancias euclideanas para la cadera derecha e izquierda, desde la vista coronal. 92
5.28. Secuencia de 8 fotogramas que describen el gesto técnico desde la vista trasera. . . 93
5.33. Boxplot de las distancias las caderas desde la vista coronal. 94
5.34. Hombro derecho 94
5.35. Hombro izquierdo 94
5.36. Distribuciones de distancias euclidianas para el hombro derecho e izquierdo desde la vista coronal. 94
5.37. Boxplot de las distancias de los hombros desde la vista coronal. 95
5.38. Distribuciones de distancias euclidianas para el talón izquierdo, desde la vista coronal. 95
5.39. Boxplot de las distancias del talón desde la vista coronal. 96
5.40. Distribución de distancias euclidianas para la rodilla izquierda, desde la vista coronal. 96
5.41. Boxplot de las distancias de la rodilla izquierda desde la vista coronal. 97
5.42. Comparación visual entre Kinovea (azul) y MediaPipe (rojo) en los primeros 9 fotogramas del archivo 32, vista lateral derecha. 98
5.43. Comparación visual entre Kinovea (azul) y MediaPipe (rojo) en los primeros 9 fotogramas del archivo 42, vista lateral izquierda. 98
5.44. Comparación visual entre Kinovea (azul) y MediaPipe (rojo) en los primeros 9 fotogramas del archivo 55, vista trasera. 99
5.45. Boxplot de la diferencia de ángulo de rodilla entre Kinovea y MediaPipe. 100

5.46. Distribución de la diferencia de ángulo de rodilla. Línea roja: referencia en 0°. . . . 101
5.47. Boxplot de la diferencia de ángulo de rodilla sin valores atípicos. Línea magenta: media; línea roja: referencia en 0°. 101
5.48. Distribución de la diferencia de ángulo de rodilla sin valores atípicos. Línea magenta: media; línea roja: referencia en 0°. 102
5.49. Relación entre los ángulos de rodilla medidos por Kinovea y MediaPipe. Línea punteada roja: identidad (45°). 102
5.50. Boxplot de la diferencia de ángulo del pie entre Kinovea y MediaPipe. 103
5.51. Distribución de la diferencia de ángulo del pie. Línea roja: referencia en 0°. 104
5.52. Ejemplos de distorsiones en el pie derecho desde la vista sagital derecha. 104
5.53. Boxplot de la diferencia de ángulo del pie sin valores atípicos. Línea magenta: media; línea roja: referencia en 0°. 105
5.54. Distribución de la diferencia de ángulo del pie sin valores atípicos. Línea magenta: media; línea roja: referencia en 0°. 105
5.55. Relación entre los ángulos del pie medidos por Kinovea y MediaPipe. Línea punteada roja: identidad (45°). 106
5.56. Boxplot de la diferencia de ángulo de la rodilla entre Kinovea y MediaPipe en la vista sagital izquierda. 107
5.57. Distribución de la diferencia de ángulo de la rodilla en la vista sagital izquierda. Línea roja: referencia en 0°. 108
5.58. Problemas causados por el tipo de ropa usada por los arqueros. 108
5.59. Boxplot de la diferencia de ángulo de la rodilla izquierda sin valores atípicos. Línea magenta: media; línea roja: referencia en 0°. 109
5.60. Distribución de la diferencia de ángulo de la rodilla izquierda sin valores atípicos. Línea magenta: media; línea roja: referencia en 0°. 109
5.61. Relación entre los ángulos de la rodilla izquierda medidos por Kinovea y MediaPipe. Línea punteada roja: identidad (45°). 110
5.62. Boxplot de la diferencia de ángulo del pie entre Kinovea y MediaPipe. 111
5.63. Distribución de la diferencia de ángulo del pie. Línea roja: referencia en 0°. 112
5.64. Boxplot de la diferencia de ángulo del pie sin valores atípicos. Línea magenta: media; línea roja: referencia en 0°. 113
5.65. Distribución de la diferencia de ángulo del pie sin valores atípicos. Línea magenta: media; línea roja: referencia en 0°. 113
5.66. Boxplot de la diferencia de ángulo de la rodilla entre Kinovea y MediaPipe en la vista coronal. 115

5.67. Distribución de la diferencia de ángulo de la rodilla en la vista coronal. Línea roja: referencia en 0°. 115
5.68. Oclusiones presentadas desde la vista coronal. 116
5.69. Boxplot de la diferencia de ángulo de la rodilla izquierda sin valores atípicos desde la vista coronal. Línea magenta: media; línea roja: referencia en 0°. 116
5.70. Distribución de la diferencia de ángulo de la rodilla izquierda desde la vista coronal sin valores atípicos. Línea magenta: media; línea roja: referencia en 0°. 117

Agradecimientos

Quisiera expresar mi más profundo agradecimiento a mis padres, a mi hermana y a mis amigos de toda la vida, quienes me han acompañado y apoyado de manera incondicional en cada etapa de mi vida.

Extiendo un especial reconocimiento y agradecimiento a mis tutores, la profesora Laura Juliana Ramos Maldonado y el profesor Oscar Danilo Martínez Bernal, quienes, con su experiencia, conocimiento y paciencia, me brindaron una guía invaluable a lo largo de este proceso.

Agradezco igualmente a todas las personas que participaron en la captura de los videos y en la logística asociada, cuyo apoyo fue fundamental para el desarrollo de este proyecto.

Finalmente, a mis compañeros de universidad y a mis profesores, muchas gracias por su acompañamiento, enseñanzas y orientación constante durante esta etapa de formación académica y personal.

Resumen

Este trabajo presenta el desarrollo de un algoritmo basado en visión por computadora para el análisis biomecánico del saque de piso del arquero en fútbol, sin el uso de marcadores físicos. La propuesta busca brindar una herramienta precisa, accesible y no invasiva para contextos académicos y deportivos, permitiendo el estudio técnico del gesto sin alterar el movimiento natural del deportista.

Se diseñó un protocolo de captura con tres cámaras ubicadas en posiciones estratégicas (lateral derecha, lateral izquierda y posterior), cuyos videos fueron sincronizados para obtener un análisis multivista. Mediante técnicas de estimación de pose, el algoritmo detecta puntos clave en el cuerpo del arquero, extrayendo variables biomecánicas como los ángulos de cadera, rodilla, tobillo y tronco durante la ejecución del gesto.

El algoritmo incorpora un sistema de detección y corrección de errores, junto con una interfaz que permite visualizar el movimiento y los ángulos articulares. Estos ángulos son clasificados dentro de rangos previamente establecidos, los cuales categorizan el gesto técnico como bueno, regular o malo, facilitando así una retroalimentación clara y objetiva para el análisis biomecánico.

Para evaluar la precisión del sistema, se realizó una comparación con el software Kinovea, utilizando análisis estadísticos como la prueba t sobre las diferencias de ángulos, coeficientes de correlación de Pearson, gráficos de dispersión y la prueba no paramétrica de los signos. Los resultados evidenciaron una alta concordancia entre ambos métodos, lo que respalda la validez y fiabilidad del algoritmo propuesto.

Palabras clave: Análisis biomecánico; Visión por computadora; Estimación de pose; Análisis multivista; Saque de piso; Arqueros de fútbol.

Introducción

Planteamiento del problema

En el fútbol, el arquero desempeña un rol importante, no solo en la defensa, sino también en el inicio de las jugadas ofensivas mediante los saques. La precisión, velocidad y técnica de estos pueden influir significativamente en la dinámica del juego, proporcionando ventajas tácticas importantes. Sin embargo, muchos arqueros, especialmente a nivel universitario, no reciben el análisis detallado necesario para mejorar sus movimientos durante el saque. La falta de herramientas específicas para evaluar y corregir la técnica de los arqueros puede implicar una pérdida de oportunidades para maximizar el rendimiento tanto del jugador como del equipo.

Actualmente, no existe un software que analice la biomecánica de los saques de los arqueros en fútbol, lo que limita la capacidad de entrenadores y jugadores para identificar y corregir errores en la técnica. Por lo tanto, es necesario desarrollar un algoritmo capaz de extraer y analizar parámetros biomecánicos clave, como ángulos y posturas, de forma rápida, precisa y económica. Esto permitirá una evaluación minuciosa y una optimización efectiva de la técnica del saque del arquero, ofreciendo un análisis asequible para muchos deportistas y contribuyendo al mejoramiento del rendimiento general del equipo.

Además del impacto en el rendimiento, la técnica deficiente también puede estar relacionada con un mayor riesgo de lesiones. Un estudio indica que el 13.3 % de todas las lesiones en el fútbol se producen en el tobillo, siendo la causa principal un mal movimiento o una pisada incorrecta al correr, saltar o patear (Chorro, 2016). De igual forma, el 11.9 % de las lesiones totales son en la rodilla, resultando en esguinces o lesiones meniscales (Chorro, 2016). En estos dos casos, la causa principal puede ser el apoyo unipodal y la rotación rápida que ocurre en ciertos movimientos, de los cuales el patear un balón no es ajeno.

Por otra parte, las lesiones musculares también son prevalentes debido a la sobrecarga e hiperextensiones de los músculos durante distintos movimientos. En este caso específico, la región del aductor es la segunda zona más afectada con un 14.5 % de todas las lesiones (Chorro, 2016). El muslo posterior, o isquiotibial, frecuenta un 16.3 % de las lesiones, representando el porcentaje más alto entre todas las zonas (Chorro, 2016). El muslo anterior, o cuádriceps femoral, frecuenta un 14.2 % de las lesiones (Chorro, 2016). Estos tres grupos musculares tienen una gran influencia durante el movimiento del lanzamiento, por lo que es de gran importancia su revisión.

Justificación

El fútbol es uno de los deportes más practicados y seguidos a nivel global. Según los datos proporcionados por Nielsen, 2024, el fútbol destaca como el deporte más popular en todo el mundo y en diversas categorías demográficas, incluyendo la edad. Por ejemplo, el 40 % de las personas encuestadas entre 16 y 29 años manifestaron su afición por el fútbol, superando a otros deportes

como el baloncesto y el voleibol. A nivel global, aproximadamente el 41 % de la población sigue el fútbol, colocándolo por encima de deportes como el baloncesto y el tenis (Nielsen, 2024).

En Colombia, la tendencia es similar, siendo el fútbol el deporte más popular del país. Según él Centro Nacional de Consultoría (CNC), 2023, el 94 % de la población considera que el fútbol tiene una gran importancia a nivel nacional. En cuanto a la práctica del deporte, el 75 % de los hombres y el 35 % de las mujeres entre 18 y 24 años practican activamente el fútbol. En el ámbito del deporte universitario colombiano, según la Asociación Colombiana de Universidades (ASCUN), 2019, el fútbol predomina notablemente frente a otros deportes en ambos géneros, con aproximadamente el 12.8 % de deportistas en comparación con un total de 18 deportes.

Dada la enorme popularidad y participación en el fútbol, es importante enfocarse en mejorar el rendimiento y la seguridad de los jugadores, especialmente en el contexto universitario. Es aquí donde el análisis biomecánico juega un papel relevante.

Los arqueros de fútbol tienen un papel único y exigente en el equipo. En particular, los saques del arquero representan una oportunidad de atacar al equipo contrario de forma efectiva e incluso agresiva, dependiendo de la rapidez, del alcance y de la precisión del tiro. Esto hace que el análisis del movimiento en el saque de los porteros sea especialmente relevante para los entrenadores, pues puede representar una oportunidad de gol. Un análisis biomecánico detallado puede identificar errores en la técnica de saque, optimizando la eficiencia y potencia del lanzamiento.

Actualmente, no existe un software específico que se encargue del análisis biomecánico para los arqueros de fútbol. La mayoría de los programas disponibles se enfocan en estadísticas generales del juego, tales como la cantidad y precisión de los pases, la posesión del balón o la distancia recorrida por los jugadores.

Por ejemplo, Track160 utiliza inteligencia artificial y visión por computadora para generar estadísticas automáticas de los partidos a partir de videos, permitiendo un seguimiento táctico general de los jugadores (Track160, s.f.). Sportlogiq, por su parte, ofrece análisis avanzados mediante la extracción automatizada de datos durante los encuentros, con énfasis en el rendimiento colectivo y la toma de decisiones (Sportlogiq, s.f.). Asimismo, existen aplicaciones orientadas a mejorar el desempeño técnico de los futbolistas, como HomeCourt, que se centra en el análisis de habilidades individuales (por ejemplo, control del balón, reacción y toma de decisiones) mediante ejercicios estructurados y pruebas interactivas (HomeCourt Soccer, s.f.).

Sin embargo, estas herramientas no abordan las necesidades específicas del arquero, particularmente en lo relacionado con el análisis de los movimientos y técnicas utilizadas en los saques. Esto evidencia la necesidad de una herramienta que permita evaluar biomecánicamente la ejecución de estas acciones, facilitando su mejora técnica de manera objetiva y accesible.

Por otro lado, el análisis biomecánico con marcadores representa un reto. En primer lugar, el proceso de colocar y retirar los marcadores entre cada prueba es bastante lento, lo que puede

afectar la fluidez del entrenamiento y la recolección de datos. Además, el contacto constante de los marcadores con la piel puede ocasionar desplazamientos no deseados debido al roce, al movimiento y al sudor, generando inconsistencias en los datos. Por último, el costo asociado con la adquisición y el mantenimiento de sistemas de marcadores es elevado, lo que limita su accesibilidad para los equipos y deportistas universitarios (Mündermann et al., 2006).

Por lo anterior, el desarrollo de un algoritmo que permita analizar la biomecánica del saque del arquero sin el uso de marcadores es de suma importancia. Este enfoque permite reducir costos y mejorar la accesibilidad del análisis. En conclusión, este proyecto no solo ayudará al análisis biomecánico específico para arqueros, sino que también contribuirá al mejoramiento general de la técnica de saque, incrementando así el rendimiento de los equipos. Además, al identificar y corregir posibles defectos en la técnica de saque, se puede reducir el riesgo de lesiones en las áreas más vulnerables, como el tobillo, la rodilla, la ingle, el isquiotibial y el muslo anterior, promoviendo la seguridad de los arqueros. Esto es especialmente relevante en el contexto del deporte universitario en Colombia, donde los recursos para los equipos son limitados y la implementación de técnicas avanzadas puede marcar una diferencia significativa en el desarrollo de los deportistas.

Objetivos

Objetivo general

Desarrollar un algoritmo para extraer variables biomecánicas a partir del análisis de videos sin marcadores, con el fin de identificar y corregir aspectos técnicos del saque de piso en un arquero de fútbol.

Objetivos específicos

1. Diseñar un protocolo para la grabación de video que garantice y maximice la calidad de los datos para el análisis biomecánico sin el uso de marcadores.

2. Implementar un modelo de estimación de pose para la detección y seguimiento de las articulaciones y extremidades más relevantes del arquero durante el saque.

3. Desarrollar una técnica de corrección de desviaciones de los puntos clave durante la ejecución del movimiento.

4. Validar los resultados obtenidos por el algoritmo comparándolo con una herramienta de análisis de video convencional.

Estado del arte

Descripción del movimiento.

Los movimientos en actividades deportivas pueden dividirse en cuatro fases principales: preparación, carga, golpeo o lanzamiento y recuperación. En la fase de preparación, se establecen las condiciones iniciales necesarias para ejecutar la habilidad, como la posición del cuerpo, el equilibrio y la anticipación del movimiento. La fase de carga implica la generación de potencia mediante la coordinación de distintas partes del cuerpo, permitiendo la transferencia de energía hacia la acción deseada. La fase de golpeo o lanzamiento es el momento específico en el que se ejecuta la habilidad, como el contacto entre una herramienta (como el pie o una raqueta) y el objeto (como una pelota). Finalmente, la fase de seguimiento o recuperación implica la estabilización del cuerpo después de la ejecución, asegurando que el atleta esté listo para la siguiente acción (Field Hockey Canada, 2021).

Aplicando estas fases al movimiento específico del saque de piso de los arqueros, se observa un patrón similar. En la fase de preparación, el arquero adopta una postura que le permite estabilizar su cuerpo y controlar el balón, ajustando la posición de sus pies, tronco y brazos. En la fase de carga, acumula energía mediante la retracción de sus extremidades, la rotación del tronco y la flexión de las articulaciones. La fase de golpeo o lanzamiento corresponde a la liberación de esta energía acumulada, donde la coordinación entre las diferentes partes del cuerpo asegura que el balón siga la trayectoria deseada. Finalmente, en la fase de recuperación, el arquero disipa la energía restante y restablece su equilibrio, preparándose para la siguiente acción (Albers, 1993).

Saque de piso

En la fase de preparación, el arquero coloca el balón en el suelo, alineando su cuerpo con la dirección del saque. El tronco se inclina levemente hacia adelante, preparando el equilibrio y ajustando la postura para el inicio del movimiento (véase la Figura 3.1).



Figura 3.1. Fase de preparación en el saque de piso.

En la fase de carga, el arquero realiza una serie de pasos en línea recta o ligeramente diagonal hacia el balón, acumulando velocidad y energía. Durante la carrera, la pierna de pateo se

lleva hacia atrás en una extensión de la cadera, y flexión de rodilla y tobillo, mientras los brazos se abren para mantener el equilibrio. Esta fase también implica la rotación del tronco en sentido contrario al golpeo para maximizar la carga de energía (véase la Figura 3.2).



Figura 3.2. Fase de carga en el saque de piso.

En la fase de golpeo, la pierna dominante realiza una extensión explosiva, con una rápida alineación de cadera, rodilla y tobillo, impactando el balón con el empeine. La rotación del tronco y la extensión de la cadera son cruciales para la potencia del saque, permitiendo un golpeo limpio y dirigido (véase la Figura 3.3).



Figura 3.3. Fase de golpeo en el saque de piso.

En la fase de recuperación, el pie de golpeo sigue su recorrido hacia adelante y arriba, mientras el tronco se endereza y los brazos ayudan a estabilizar la postura del cuerpo después del golpeo, asegurando una posición equilibrada para continuar con la jugada o reposicionarse en el campo (véase la Figura 3.4).



Figura 3.4. Fase de recuperación en el saque de piso.

Ángulos biomecánicos ideales en el saque de piso de los arqueros

En el estudio realizado por Orloff et al., 2008, se analizó la mecánica del golpeo de balón con el empeine (*instep kick*) en el fútbol, centrando la atención en las diferencias en la posición de la pierna de apoyo entre jugadores masculinos y femeninos. Este análisis buscó comprender cómo los ángulos de inclinación del tronco, tanto en el plano sagital como posterior, así como el ángulo de la rodilla de apoyo en la vista sagital, impactan el rendimiento durante el golpeo.

El estudio se llevó a cabo con 23 jugadores universitarios de fútbol amateur de la División III de la NCAA. De los participantes, 11 eran hombres y 12 mujeres, con una edad promedio de 20.2 ± 1.2 años, y todos contaban con al menos 10 años de experiencia en clubes de fútbol.

Para el análisis, se utilizaron marcadores reflectantes y cámaras de video para digitalizar los datos y medir las fuerzas de reacción del suelo. Los datos fueron registrados desde dos vistas: sagital y posterior, con marcadores colocados en puntos clave del cuerpo, como el hombro, la cadera, la rodilla y los metacarpos de los pies (primer y quinto).

El objetivo principal fue obtener una comprensión más profunda de cómo los ángulos biomecánicos de la pierna de apoyo afectan el rendimiento durante el golpeo. Para ello, se recopilaron datos específicos de las inclinaciones del tronco y el ángulo de la rodilla en dos momentos clave: el instante en que el pie de apoyo contacta el suelo y el contacto con el balón. A continuación, se presentan los resultados obtenidos en ambas fases para jugadores masculinos y femeninos.

Variable	**Masculino** ($\mu \pm \sigma$)	**Femenino** ($\mu \pm \sigma$)
Inclinación del tronco, vista posterior (°)	-	-
Inclinación del tronco, vista sagital (°)	-11 ± 3	-10 ± 7
Ángulo de la rodilla (°)	157 ± 8	157 ± 9

Tabla 3.1. Datos cinemáticos correspondientes al instante de apoyo del pie (media ± desviación estándar).

Variable	**Masculino** ($\mu \pm \sigma$)	**Femenino** ($\mu \pm \sigma$)
Inclinación del tronco, vista posterior (°)	-3 ± 11	8 ± 9
Inclinación del tronco, vista sagital (°)	3 ± 8	13 ± 10
Ángulo de la rodilla (°)	138 ± 13	139 ± 19

Tabla 3.2. Datos cinemáticos al momento del contacto con el balón (media ± desviación estándar).

Por otro lado, en el estudio realizado por Alcock et al., 2012, se compararon las características cinemáticas del golpeo de balón con el empeine (*instep kick*) y con curva en jugadoras de fútbol de élite. El análisis se centró en las diferencias en los ángulos articulares durante el momento de impacto con el balón, destacando especialmente el ángulo de la rodilla, la cadera y el tobillo de la pierna de pateo en ambas técnicas.

El estudio incluyó a 15 jugadoras internacionales de fútbol femenino, que representaban al equipo clasificado en el puesto 14 a nivel mundial en el momento del análisis. Las participantes tenían una edad promedio de 20.1 ± 3.2 años, una altura de 1.69 ± 0.06 m y un peso de 61.6 ± 5.8 kg. Las pruebas consistieron en 15 patadas con el empeine y 15 patadas con efecto, realizadas tras un calentamiento estandarizado y cinco ensayos de familiarización con cada tipo de golpeo. Los resultados mostraron diferencias relevantes en los ángulos articulares, particularmente en el tobillo, al momento del impacto con el balón.

Variable	**Curve** ($\mu \pm \sigma$)	**Instep** ($\mu \pm \sigma$)	**P**
Ángulo de la cadera derecha (°)	33.2 ± 7.9	35.5 ± 7.8	0.066
Ángulo de la rodilla derecha (°)	55.3 ± 9.6	55.6 ± 6.6	0.835
Ángulo del tobillo derecho (°)	21.7 ± 8.0	31.6 ± 8.8	<0.001

Tabla 3.3. Ángulos articulares relativos (0° corresponde a la posición anatómica). Comparación entre patadas con empeine y con curva.

Es importante señalar que, aunque el análisis fue realizado en tres dimensiones, los ángulos reportados fueron extraídos desde el plano correspondiente a cada articulación, lo que implica una estimación bidimensional. Además, todos los ángulos se midieron con respecto a la posición anatómica estándar del cuerpo.

Técnicas de análisis biomecánico en el deporte

El análisis biomecánico ha evolucionado significativamente en las últimas décadas, impulsado por avances tecnológicos que han permitido una mayor precisión, velocidad y accesibilidad en la evaluación del rendimiento deportivo. Inicialmente, el análisis biomecánico se enfocaba en la observación directa de los movimientos y la medición manual, lo que limitaba la capacidad de capturar detalles complejos y afectaba la precisión de los datos obtenidos. Sin embargo, con el desarrollo de nuevas tecnologías, como la captura de movimiento, la visión artificial y los algoritmos de aprendizaje automático, el análisis biomecánico se ha transformado en una herramienta avanzada que ofrece soluciones innovadoras para mejorar el rendimiento deportivo y reducir el riesgo de lesiones.

En la biomecánica deportiva existen dos enfoques principales: el análisis de movimiento con marcadores y sin marcadores. Ambos métodos tienen ventajas y desventajas, pero ambos han sido fundamentales para entender los patrones de movimiento en distintos deportes y han allanado el camino para la creación de herramientas más sofisticadas en el análisis biomecánico. A continuación se profundizará en la progresión historica del análisis de movimiento en la biomecánica deportiva.

Digitalización Manual

La digitalización manual fue una de las técnicas más comunes para medir el movimiento en biomecánica durante varias décadas, especialmente antes de la llegada de las tecnologías digitales. En sus inicios, se utilizaban cámaras de cine debido a su alta calidad de imagen y velocidades de cuadro rápidas, pero el procesamiento de estos datos era extremadamente laborioso. Con la evolución hacia cámaras de video digitales, más rápidas y accesibles, este método se fue actualizando, aunque aún requería la localización manual de puntos clave en cada fotograma. Esta técnica sigue siendo útil en el análisis de movimiento sin la necesidad de colocar marcadores, lo que es ideal para estudios de deportes en entornos de entrenamiento o competencia sin interferir con el rendimiento del atleta (Colyer et al., 2018).

Un ejemplo es el análisis de la extensión del codo durante el lanzamiento en cricket. Entre las ventajas de la digitalización manual en cricket se encuentra su capacidad para realizar el análisis sin necesidad de colocar marcadores, permitiendo que el atleta mantenga su rendimiento normal durante la prueba. Sin embargo, entre las desventajas se destaca la dificultad para obtener resultados precisos en movimientos complejos que involucren rotaciones de las articulaciones, específicamente del hombro como ocurre con el lanzamiento en cricket, y el alto nivel de error en situaciones específicas, como el ángulo de codo abducido sin marcadores. Esto limita la precisión del análisis en contextos más complicados y sugiere la necesidad de adoptar tecnologías más avanzadas para mejorar la exactitud en la evaluación de movimientos (Elliott et al., 2007).

Análisis de movimiento con marcadores

El análisis biomecánico tradicionalmente utiliza marcadores adheridos al cuerpo del deportista y sistemas automáticos basados en cámaras infrarrojas para capturar y rastrear el movimiento. Estos sistemas emplean múltiples cámaras que emiten luz infrarroja invisible, mientras los marcadores pasivos reflejan esta luz, permitiendo deducir su posición tridimensional. Los puntos de interés se localizan en imágenes secuenciales, se convierten en coordenadas espaciales reales y se usan para inferir la pose 3D del esqueleto. Este método, ampliamente usado en laboratorios deportivos, ofrece alta precisión y eficiencia en comparación con la digitalización manual, al automatizar la localización de grandes cantidades de marcadores (Colyer et al., 2018).

Aunque los sistemas automáticos basados en marcadores ofrecen una alta precisión en la captura del movimiento, presentan varias desventajas. Uno de los principales inconvenientes es la posible incomodidad que los marcadores pueden causar al deportista, lo que puede llevar a una alteración de su técnica y, en consecuencia, afectar la calidad de los datos obtenidos. Además, los marcadores pueden moverse durante las pruebas, lo que introduce errores en los resultados y dificulta la obtención de medidas consistentes. Otro desafío significativo es el tiempo prolongado requerido para la preparación y calibración del sistema, así como la necesidad de realizar ajustes

precisos para evitar errores de colocación. Estas limitaciones pueden hacer que el uso de sistemas basados en marcadores sea menos práctico en entornos deportivos en vivo, donde la interferencia con el rendimiento del atleta y la complejidad operativa pueden ser problemáticas (Colyer et al., 2018).

Sensores inerciales

Los sensores inerciales han emergido como una herramienta clave en la biomecánica deportiva, especialmente en deportes de combate como el boxeo, karate y taekwondo. Estos sensores incluyen acelerómetros, giroscopios y magnetómetros, los cuales permiten capturar datos sobre la aceleración lineal, la velocidad angular y la orientación del cuerpo en tiempo real. Gracias a su capacidad para operar de manera inalámbrica y en diferentes entornos, los sensores inerciales ofrecen una forma detallada y precisa de medir los movimientos sin interferir en la actividad natural del atleta, lo cual es fundamental para evaluar y mejorar la técnica deportiva y reducir el riesgo de lesiones (Blanco Ortega et al., 2022).

El uso de sensores inerciales en estos deportes permite analizar variables biomecánicas críticas, como la velocidad, aceleración y potencia de los movimientos de las extremidades superiores, brindando información valiosa tanto para el entrenamiento como para la prevención de lesiones. En el boxeo, por ejemplo, se han utilizado máscaras equipadas con acelerómetros para medir la gravedad de los golpes, lo que ayuda a correlacionar la fuerza de los impactos con posibles lesiones cerebrales leves. Este tipo de análisis permite identificar patrones peligrosos y ajustar las técnicas de los atletas para optimizar su rendimiento mientras se minimizan los riesgos (Blanco Ortega et al., 2022).

Además, los sensores inerciales se integran fácilmente con interfaces gráficas que permiten visualizar los parámetros biomecánicos en tiempo real, facilitando el feedback inmediato a los entrenadores y atletas. Estos sensores han demostrado ser efectivos no solo en la mejora del rendimiento deportivo, sino también en aplicaciones médicas, como la rehabilitación y la prevención de problemas musculoesqueléticos, al permitir un monitoreo detallado de los movimientos del atleta durante el proceso de recuperación (Blanco Ortega et al., 2022).

Análisis de movimiento sin marcadores

El análisis biomecánico sin marcadores se ha desarrollado como una solución para superar las limitaciones de los métodos tradicionales basados en marcadores. Esta técnica permite capturar y evaluar el movimiento sin la necesidad de utilizar marcadores adheridos al cuerpo del deportista, reduciendo el tiempo de preparación asociado a la colocación de los marcadores, aumentando la comodidad del deportista y disminuyendo el laborioso procesamiento requerido. Tradicionalmente, los estudios biomecánicos han dependido de estos marcadores para rastrear con precisión las

diferentes partes del cuerpo durante el movimiento. Sin embargo, el análisis sin marcadores se basa en sistemas de visión por computadora que utilizan múltiples cámaras y algoritmos avanzados para identificar y seguir las articulaciones en tiempo real. Estos sistemas pueden operar en entornos controlados (como laboratorios con luces infrarrojas) o naturales (como en el campo de juego con acciones reales), ofreciendo una mayor flexibilidad en el estudio del movimiento al no requerir intervención directa sobre el cuerpo del deportista (Colyer et al., 2018).

El funcionamiento de este método se centra en la captura de imágenes desde diferentes ángulos mediante cámaras colocadas estratégicamente alrededor del atleta. Los algoritmos de procesamiento de imágenes analizan estas capturas para identificar patrones de movimiento. Luego, estos datos se utilizan para reconstruir los movimientos en 3D, permitiendo una evaluación precisa sin necesidad de marcar el cuerpo del deportista, lo que facilita un análisis más natural y cómodo (Colyer et al., 2018).

En el análisis de la marcha, por ejemplo, el uso de sistemas sin marcadores ofrece ventajas significativas. Una de las principales es que permite estudiar la marcha en un entorno más natural, ya que los deportistas pueden caminar sin la incomodidad de los marcadores. Esto resulta en una captura de movimiento más realista y precisa, lo cual es especialmente beneficioso en el análisis clínico de la marcha, donde es crucial obtener evaluaciones exactas del movimiento para diagnosticar problemas o planificar tratamientos. Sin embargo, este método también presenta desafíos técnicos. El procesamiento de imágenes y la identificación precisa de las partes del cuerpo en movimiento pueden ser más complicados, especialmente en situaciones de oclusión o cuando las características visuales del entorno se asemejan a las del cuerpo del atleta. En tales casos, puede ser difícil para el sistema rastrear correctamente las articulaciones y, por lo tanto, los movimientos de las extremidades, como los pies o las rodillas, lo que podría afectar la precisión del análisis (Mündermann et al., 2006).

Redes Neuronales Convolucionales para el procesamiento de imágenes

Las redes neuronales convolucionales, CNNs por sus siglas en inglés, son herramientas fundamentales en el análisis biomecánico para la estimación de poses humanas, facilitando la detección y el análisis preciso de los movimientos del cuerpo sin la necesidad de marcadores físicos. Esto es especialmente valioso en la evaluación del rendimiento deportivo y en la corrección técnica de los atletas, ya que permite un análisis detallado sin interferir en la técnica natural del deportista (Mathis et al., 2018).

Un ejemplo destacado de la aplicación de CNNs es el estudio realizado por la Universidad de Zarqa, donde se utilizaron redes neuronales para reconocer poses corporales a partir de siluetas extraídas de imágenes. El proceso comienza con la detección del cuerpo humano, seguido por la extracción de la silueta y su clasificación mediante redes entrenadas con conjuntos de datos etique-

tados. Esto permite identificar poses como estar de pie o caminar, destacando la capacidad de estas redes para mapear correctamente las poses humanas a partir de la similitud entre dimensiones de las siluetas detectadas (Ahmad, 2019). Este enfoque demuestra cómo las CNNs aprenden patrones complejos a partir de datos de imágenes preprocesadas, permitiendo detectar detalles críticos en posturas y movimientos, lo que contribuye significativamente a la mejora del rendimiento y la prevención de lesiones en contextos deportivos.

Además, estudios como el de (Mathis et al., 2018) con DeepLabCut (una herramienta eficiente para la estimación de pose sin marcadores en 2D y 3D basado en el aprendizaje por transferencia con redes neuronales profundas) muestran la efectividad de las CNNs para la estimación de poses sin marcadores, utilizando modelos preentrenados que se ajustan rápidamente a tareas específicas mediante la reutilización de redes ya entrenadas en otras tareas relacionadas. Esta técnica permite una integración en tiempo real durante los entrenamientos deportivos, proporcionando retroalimentación instantánea que ayuda a ajustar movimientos y mejorar la técnica de los atletas de manera más eficiente (Mathis et al., 2018).

El enfoque en redes neuronales convolucionales, como se discute en (Jammalamadaka et al., 2017), resalta que estas redes se entrenan para reconocer configuraciones específicas del cuerpo mediante la aplicación de filtros convolucionales que extraen características complejas de las imágenes. Esta capacidad es especialmente útil en la biomecánica deportiva, donde las poses y los movimientos deben ser analizados con gran detalle para mejorar el rendimiento técnico. La robustez de las CNNs frente a condiciones desafiantes, como oclusiones y variaciones en la iluminación, las hace altamente efectivas en estudios realizados en entornos naturales, ampliando su uso más allá de los laboratorios controlados (Jammalamadaka et al., 2017).

Librerías para la estimación de pose

MediaPipe

MediaPipe es una librería de código abierto desarrollada por Google, especializada en el análisis y procesamiento de datos visuales, con un enfoque particular en la estimación de poses y movimientos humanos. Permite el seguimiento de múltiples puntos del cuerpo (hasta 33 en sus modelos más avanzados), facilitando el análisis de poses tanto en 2D como en 3D. Es utilizada en campos como el deporte, la realidad aumentada, la salud y el control de interfaces mediante gestos. MediaPipe ofrece modelos con distintos niveles de complejidad, lo que permite adaptarse a diferentes necesidades, priorizando ya sea la velocidad de ejecución o la precisión del análisis. Además, tiene buena tolerancia ante oclusiones parciales y funciona eficientemente en dispositivos con recursos limitados (Google AI Edge, 2024).

YOLOv11

YOLOv11 es una versión actualizada del modelo de detección de objetos desarrollado por Ultralytics, que también incorpora funcionalidades para la estimación de pose humana. A través de una arquitectura optimizada, permite identificar múltiples puntos del cuerpo en imágenes o secuencias de video con alta velocidad y precisión. Al igual que otras herramientas, cuenta con variantes de modelos que permiten ajustar el desempeño según las prioridades del proyecto, ya sea rapidez o exactitud en los resultados. Su enfoque flexible lo convierte en una opción adecuada para aplicaciones prácticas como el análisis deportivo o el seguimiento de movimientos en tiempo real (Ultralytics, 2024).

Cuadro comparativo

Criterio	**YOLOv11 Pose Estimation**	**MediaPipe Pose Estimation**
Cantidad de puntos clave	Detecta 17 puntos clave en articulaciones principales.	Detecta 33 puntos clave en articulaciones principales, incluyendo talones y puntas de los pies.
Estabilidad (*Jitter*)	Puede presentar variabilidad en la estimación de los puntos clave a lo largo del tiempo, especialmente en movimientos rápidos o complejos.	MediaPipe presenta menos *jitter*, proporcionando estimaciones más estables y consistentes.
Velocidad y Rendimiento	Ideal en entornos con GPU para procesamiento rápido en tiempo real.	Excelente rendimiento en CPU, óptimo para dispositivos móviles y aplicaciones ligeras sin requerir hardware avanzado.
Capacidad para varias personas	Admite detección de múltiples personas en una sola escena, aunque con limitaciones en dispositivos sin GPU.	No está optimizado para la detección de múltiples personas simultáneamente.
Tamaño de Imagen requerido	Requiere imágenes más grandes para una alta precisión	Funciona eficientemente con imágenes con resolución estándar, sin necesidad de tamaños grandes para un buen desempeño.
Configuración y Flexibilidad	Requiere ajustes técnicos para optimizar el rendimiento según el contexto de uso.	Fácil de implementar, con configuraciones sencillas.

Tabla 3.4. Comparación entre la estimación de pose de YOLOv11 y de MediaPipe.

Estimación de pose con MediaPipe

La estimación de pose con MediaPipe se basa en un sistema avanzado que utiliza redes neuronales para identificar y analizar 33 puntos clave (*landmarks*) del cuerpo humano. Estos puntos clave incluyen articulaciones principales, como hombros, caderas, rodillas y tobillos, así como características adicionales, como los ojos, las orejas y la nariz, como lo ilustra la Figura 3.5. Este enfoque permite realizar un análisis detallado del movimiento y la postura del cuerpo en tiempo real. Además de este análisis, MediaPipe se caracteriza por su diseño optimizado, lo que le permite funcionar eficientemente en dispositivos con recursos limitados, haciéndolo ideal para aplicaciones en biomecánica, deportes y otras áreas que requieren un análisis preciso del movimiento humano (Bazarevsky et al., 2020).

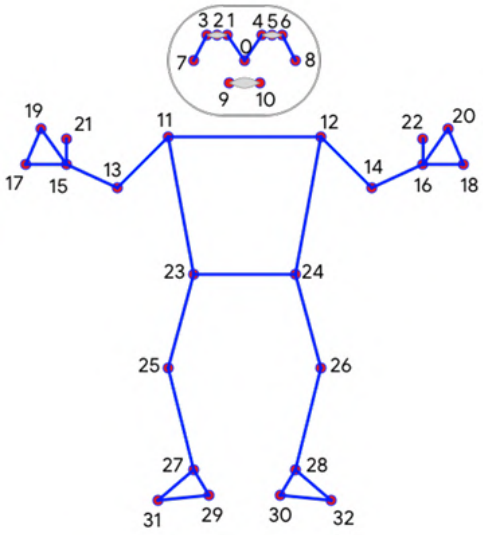


Figura 3.5. 33 puntos clave identificados por MediaPipe (Bazarevsky et al., 2020).

Fuentes de datos

Para el entrenamiento del modelo de estimación de pose de MediaPipe, se emplean tres fuentes de datos principales. La primera corresponde a un conjunto de 30.000 imágenes propias, capturadas con consentimiento informado mediante una aplicación de Realidad Aumentada en entornos naturales. La segunda fuente consiste en 85.000 imágenes con diversas poses relacionadas con el fitness provenientes de datasets públicos, como COCO y Human Pose Dataset. Todas las imágenes utilizadas en el entrenamiento han sido etiquetadas manualmente con los puntos clave *(landmarks)* correspondientes, garantizando que cada punto esté correctamente identificado para su posterior uso en el aprendizaje del modelo (Bazarevsky et al., 2021).

Arquitectura del modelo

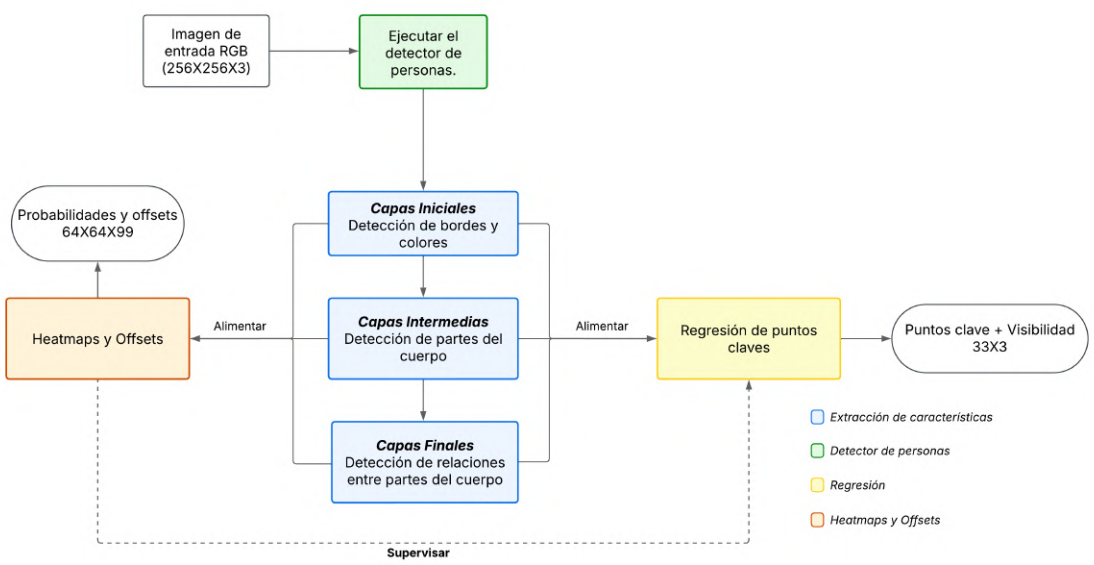


Figura 3.6. Arquitectura del Modelo de Mediapipe (Elaboración propia).

El modelo de estimación de pose de MediaPipe sigue una arquitectura optimizada para lograr un rendimiento eficiente en tiempo real y sin consumir tantos recursos. Su procesamiento se divide en varias etapas clave que permiten la identificación precisa de los puntos clave del cuerpo humano.

El proceso inicia con el preprocesamiento de la imagen, donde la entrada del modelo es una imagen en formato RGB con dimensiones $256 \times 256 \times 3$. El valor de 3 corresponde a los tres canales de color *(Red, Green, Blue)*, mientras que 256×256 indica la resolución de la imagen. Si la imagen de entrada tiene una resolución diferente, el sistema la redimensiona automáticamente, lo que contribuye a mejorar la eficiencia del procesamiento.

A continuación, se ejecuta un detector de personas que identifica la región de interés y localiza al sujeto dentro de la imagen, permitiendo centrar la atención del modelo en el cuerpo humano.

Posteriormente, la imagen es procesada por una red neuronal profunda, la cual extrae características relevantes en diferentes niveles. En las capas iniciales, se identifican bordes y colores básicos de la imagen. Luego, en las capas intermedias, el modelo detecta las partes individuales del cuerpo, y por último, en las capas finales, se establecen relaciones entre esas partes para lograr una representación estructurada de la pose humana.

A partir de las características extraídas, el modelo sigue dos caminos paralelos para la predicción de los puntos clave. Por un lado, se generan los mapas de calor *(heatmaps)*, los cuales

indican la probabilidad de que cada articulación esté presente en una determinada región de la imagen. Sin embargo, como estos mapas de calor no ofrecen una ubicación exacta, el modelo utiliza vectores de corrección *offsets* para ajustar la posición de los puntos clave con mayor precisión. Por otro lado, la regresión de puntos clave se encarga de predecir directamente las coordenadas exactas de los 33 puntos clave, utilizando como entrada las características previamente extraídas.

Durante la fase de entrenamiento del modelo, los *heatmaps y offsets* juegan un papel crucial en la supervisión de la regresión, ayudando a mejorar la precisión de la predicción. Sin embargo, en la fase de inferencia, estos elementos no se utilizan; en su lugar, la predicción de los puntos clave se basa únicamente en la regresión, lo que hace que el modelo sea más eficiente y adecuado para aplicaciones en tiempo real.

Pipeline de inferencia

A continuación se presenta el proceso que el modelo de aprendizaje de máquina sigue para procesar datos de entrada, en este caso videos, y generar resultados útiles. Este pipeline constituye uno de los diferenciales clave de MediaPipe, ya que permite reducir significativamente el costo computacional de la estimación de pose, manteniendo un alto desempeño en tiempo real.

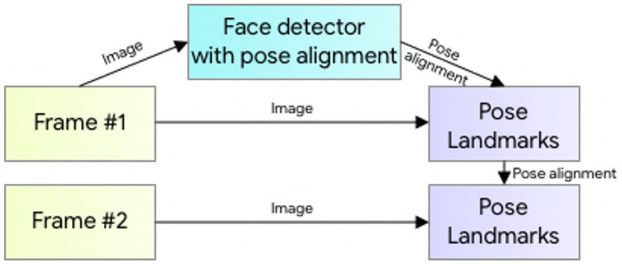


Figura 3.7. Pipeline de inferencia (Bazarevsky et al., 2020).

El pipeline de inferencia de MediaPipe emplea una configuración basada en un detector y un rastreador *(detector-tracker)*. En su primera etapa, se utiliza un detector ligero de pose corporal para procesar el primer fotograma del video. Este detector identifica al sujeto en la escena y localiza la región de interés *(Region of Interest, ROI)*, que contiene al cuerpo completo. Adicionalmente, genera los *landmarks* o puntos clave del cuerpo humano en ese fotograma inicial, asegurando que el análisis posterior se enfoque en las áreas relevantes (Bazarevsky et al., 2020).

En los fotogramas subsiguientes, el pipeline utiliza un sistema de seguimiento basado en una red neuronal. Este rastreador predice la ubicación de la ROI y los puntos clave del sujeto en el siguiente fotograma, utilizando como referencia las estimaciones del fotograma anterior. Si el rastreador logra identificar correctamente al sujeto en el nuevo fotograma, se calculan nuevamente los *landmarks* de la pose directamente sobre la región predicha, evitando así ejecutar el detector completo. Sin embargo, si el rastreador no encuentra evidencia de la presencia del sujeto en el

cuadro actual, el pipeline vuelve a ejecutar el detector de pose para localizar al sujeto y reiniciar el proceso de seguimiento (Bazarevsky et al., 2020).

El diseño de este pipeline combina eficiencia y precisión. Por un lado, el seguimiento continuo minimiza la cantidad de veces que se ejecuta el detector, reduciendo así el costo computacional. Por otro, la capacidad de ajustar dinámicamente la región de interés asegura que el modelo mantenga un alto nivel de exactitud en la detección de los puntos clave del cuerpo. Este enfoque resulta ideal para dispositivos con recursos limitados y aplicaciones que requieren tiempo real, como el análisis biomecánico de movimientos (Bazarevsky et al., 2020).

Detector de personas

El detector de personas de MediaPipe en lugar de realizar una detección completa del cuerpo, utiliza el rostro como referencia principal para estimar la posición y alineación del cuerpo completo. Esto se debe a que el rostro humano tiene características consistentes, lo que lo convierte en un punto de partida confiable para determinar la región y otras partes del cuerpo. El proceso comienza con un detector de rostros, que no solo localiza la posición del rostro, sino que también calcula parámetros clave necesarios para analizar el cuerpo completo. Entre esos parámetros se encuentran el punto medio entre las caderas, utilizado para la posición del torso, el tamaño del círculo que circunscribe a toda la persona y la inclinación del torso calculada a partir de las líneas que conectan los puntos medios de los hombros y las caderas (Bazarevsky et al., 2020).

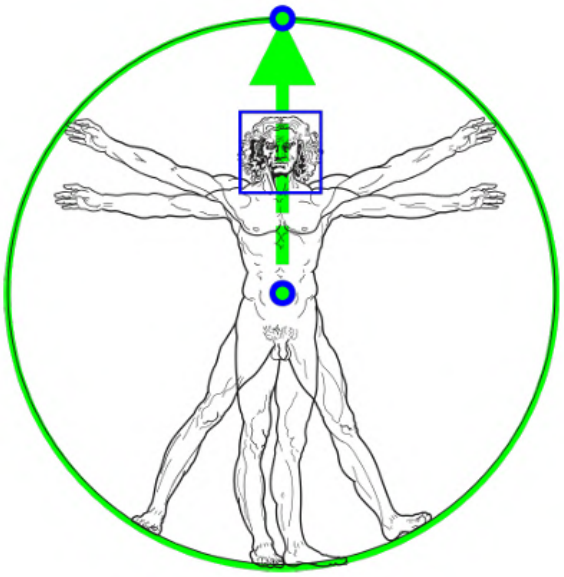


Figura 3.8. Hombre de Vitruvio (Bazarevsky et al., 2020).

Este enfoque se basa en la idea de que, en casos de uso donde solo se analiza una persona, la cabeza estará siempre visible en el video. Esta premisa simplifica el proceso de detección, ya que

permite al sistema centrar sus cálculos en una región inicial clara y confiable. Además, al utilizar el rostro como ancla, MediaPipe puede alinear el sujeto dentro de un marco predefinido, como se muestra en la figura 3.8. Esta alineación se representa mediante un círculo que abarca todo el cuerpo de la persona y una línea que define la orientación del torso. Este diseño permite que el detector genere una región de interés (ROI) precisa, donde se lleva a cabo el análisis de los puntos clave (Bazarevsky et al., 2020).

Extracción de características

En esta etapa, la red neuronal convolucional lleva a cabo la extracción de características a través de la aplicación de filtros, con el fin de transformar la imagen en diferentes niveles de representación. Tradicionalmente, este proceso se realiza mediante convoluciones estándar, donde cada filtro opera sobre la totalidad de los canales de entrada, mezclando información espacial y de profundidad en una sola operación. Sin embargo, en el caso de MediaPipe (que utiliza una arquitectura basada en MobileNetV2) se emplea un enfoque más eficiente conocido como **convoluciones separables por profundidad**, lo que permite reducir el costo computacional manteniendo una capacidad de representación comparable.

En una convolución estándar, el valor de salida en la posición (i, j) del canal c_{out} se calcula como:

$$Y(i, j, c_{\text{out}}) = \sum_{m=0}^{K-1} \sum_{n=0}^{K-1} \sum_{c_{\text{in}}=0}^{C_{\text{in}}-1} X(i+m, j+n, c_{\text{in}}) \cdot W(m, n, c_{\text{in}}, c_{\text{out}}) \tag{1}$$

donde X es la entrada con dimensiones $H \times W \times C_{\text{in}}$, W es el filtro y Y es la salida. Esta operación es computacionalmente intensiva, ya que mezcla todos los canales para producir cada salida.

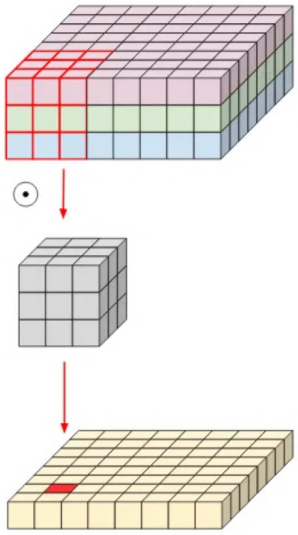


Figura 3.9. Convolución estándar (Zurister, 2019).

Por parte de la convolución separable por profundidad, la operación se divide en dos etapas:

1. **Convolución en profundidad:** se aplica un filtro $K \times K$ a cada canal de entrada por separado:

$$Y_{\text{dw}}(i, j, c) = \sum_{m=0}^{K-1} \sum_{n=0}^{K-1} X(i + m, j + n, c) \cdot W_{\text{dw}}(m, n, c) \tag{2}$$

2. **Convolución puntual (1×1):** se combinan los canales utilizando una convolución 1×1 que mezcla las salidas anteriores:

$$Y(i, j, c_{\text{out}}) = \sum_{c_{\text{in}}=0}^{C_{\text{in}}-1} Y_{\text{dw}}(i, j, c_{\text{in}}) \cdot W_{\text{pw}}(c_{\text{in}}, c_{\text{out}}) \tag{3}$$

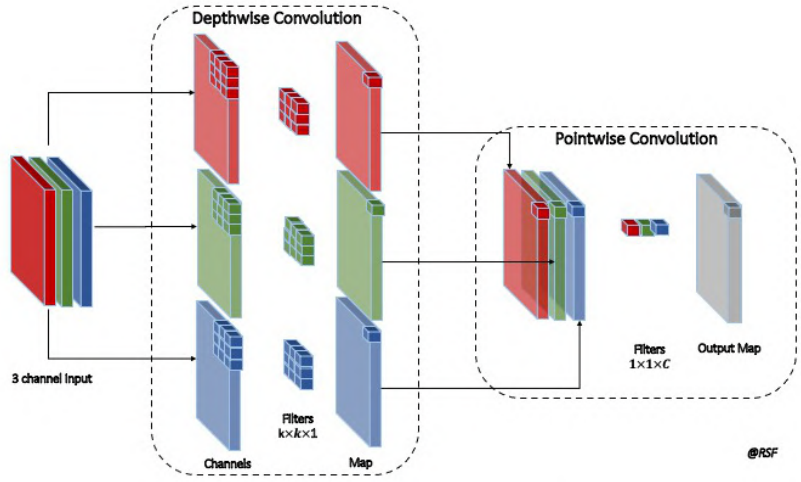


Figura 3.10. Convolución en profundidad y convolución puntual (Cheuque et al., 2022).

Esta operación permite reducir el número total de operaciones desde:

$$\text{Convolución estándar:} \quad h_i \cdot w_i \cdot d_i \cdot d_j \cdot k^2 \tag{4}$$

a

$$\text{Convolución separable por profundidad:} \quad h_i \cdot w_i \cdot d_i \cdot (k^2 + d_j) \tag{5}$$

donde:

- h_i y w_i: altura y ancho del mapa de activación de entrada.
- d_i: número de canales (profundidad) de entrada.
- d_j: número de canales de salida.
- k: tamaño del filtro convolucional (por ejemplo, $k = 3$ para un filtro de 3×3).

Como se describe en Sandler et al., 2019, esta última expresión refleja la suma del costo de una convolución en profundidad (que aplica un filtro por canal) y una convolución puntual (que combina los canales con una operación 1×1). Este cambio reduce significativamente el número de operaciones, especialmente cuando se utilizan filtros pequeños y un alto número de canales de salida, lo cual es común en redes neuronales profundas.

Luego de aplicar las operaciones de convolución en profundidad y convolución puntual, los mapas de características pueden contener información redundante. Además, su resolución puede ser demasiado grande para un procesamiento eficiente en etapas posteriores. Para solucionar esto, puede pasar por una etapa de reducción espacial, que consiste en disminuir sus dimensiones en alto

y ancho ($h \times w$) mientras se conserva la información más relevante. Esta etapa es clave para reducir el número de operaciones en capas posteriores y permitir que la red aprenda representaciones cada vez más abstractas y globales.

Existen dos mecanismos comunes para realizar esta reducción espacial:

- **Uso de *stride* en las convoluciones:** consiste en aplicar los filtros desplazándolos con un paso mayor a 1. Por ejemplo, un *stride* de 2 hace que el filtro avance dos posiciones a la vez, lo que genera una salida con aproximadamente la mitad de tamaño en cada dimensión.

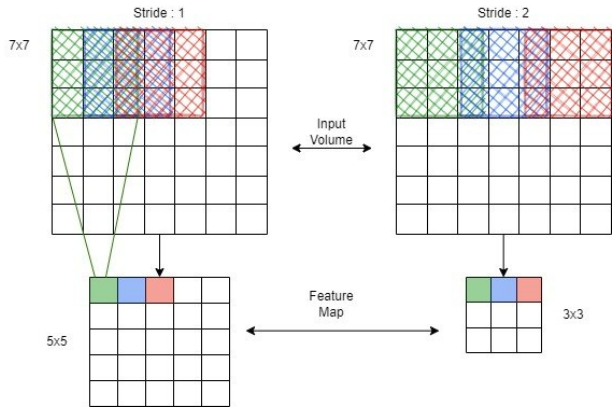


Figura 3.11. Ejemplos de desplazamientos al operar las convoluciones (Baeldung, 2023).

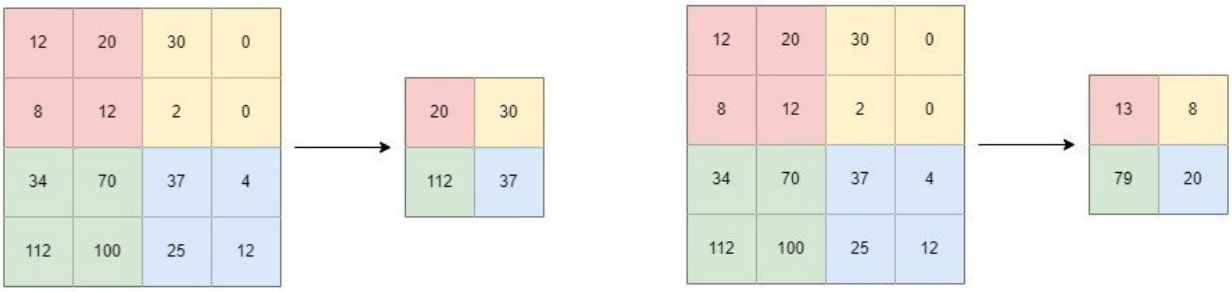
- **Técnicas de *pooling*:** estas se encargan de reducir la dimensionalidad de los mapas mientras conserva la información más relevante de ellos. Para este caso se destacan dos técnicas:

  1. **Max pooling**: consiste en seleccionar el valor máximo dentro de una matriz de tamaño $f \times f$, permitiendo que solo las características más destacadas sean retenidas (Figura 3.12a). Matemáticamente, se define como:

$$S(i,j) = \max_{a,b} I(i+a, j+b) \tag{6}$$

  2. **Average pooling**: consiste en calcular el promedio de los valores en la matriz (Figura 3.12b). Esta técnica mantiene más información global sobre la imagen.

$$S(i,j) = \frac{1}{f^2} \sum_{a,b} I(i+a, j+b) \tag{7}$$



(a) Max pooling (b) Average pooling

Figura 3.12. Comparación entre técnicas de reducción espacial: (a) max pooling selecciona el valor más alto en cada región; (b) average pooling calcula el promedio de los valores. (Baeldung, 2023)

Una vez aplicadas las capas de convolución y de reducción de dimensionalidad, la red continúa su proceso de análisis a través de una construcción jerárquica de representaciones. Este enfoque permite que la red neuronal combine información de bajo nivel con estructuras complejas, facilitando la identificación de patrones de alto nivel en la imagen. A medida que los datos avanzan a través de la red, se identifican patrones cada vez más complejos, permitiendo la detección de estructuras completas en la imagen. Esta jerarquía se puede dar de la siguiente forma:

1. **Capas iniciales**:
   - Aplican filtros convolucionales de bajo nivel que resaltan contrastes y contornos.
   - Detectan bordes y texturas básicas en la imagen.
2. **Capas intermedias**:
   - Identifican partes del cuerpo, combinando información extraída en las primeras capas.
   - Reconocen elementos más complejos, como extremidades y articulaciones.
3. **Capas finales**:
   - Estructuran la pose completa del sujeto, estableciendo relaciones espaciales entre los puntos clave del cuerpo.

Este enfoque permite que la red neuronal convolucional aprenda automáticamente qué características, bien sean de bajo o de alto nivel, reflejan una mayor relevancia. Gracias a esto, la red puede generalizar mejor en distintos escenarios y adaptarse a otras posturas.

Heatmaps y offsets

Tras la extracción de características en las capas anteriores de la red neuronal convolucional, es necesario supervisar el aprendizaje del modelo para que pueda predecir con precisión la ubicación de los puntos clave del cuerpo. Para ello, se emplean *heatmaps* (véase la Figura 3.13a) y *offsets* (véase la Figura 3.13b), los cuales no forman parte de la salida final durante la inferencia, sino que actúan como un mecanismo de supervisión durante la fase de entrenamiento.



(a) (b)

Figura 3.13. Visualización de la salida del modelo durante el entrenamiento: (a) activación de los *heatmaps* para cada punto clave; (b) vectores de desplazamiento (*offsets*) utilizados para refinar la localización. Imágenes tomadas de Xiao et al., 2022.

Los *heatmaps* se utilizan como una representación espacial de la probabilidad de la ubicación de los puntos clave del cuerpo. En lugar de predecir directamente las coordenadas (x, y) de cada articulación, la red neuronal genera una distribución de probabilidad sobre toda el área de la imagen para cada punto (véase la Figura 3.13a). Es decir, cada canal de la salida representa la probabilidad de que una articulación específica se encuentre en una determinada región de la imagen.

Durante el entrenamiento, los *heatmaps* reales (o etiquetas) se construyen a partir de funciones gaussianas bidimensionales centradas en los puntos clave anotados manualmente. Para una articulación k, cuya ubicación real es $l_k = (x_k, y_k)$, la distribución de probabilidad sobre el espacio se define como:

$$h_k(x,y) = \begin{cases} \exp\left(-\frac{(x-x_k)^2+(y-y_k)^2}{2\sigma^2}\right), & \text{si } \sqrt{(x-x_k)^2+(y-y_k)^2} \leq R \\ 0, & \text{en otro caso} \end{cases}$$

En esta fórmula, σ representa la desviación estándar de la campana gaussiana y controla la dispersión de la probabilidad alrededor del centro. El parámetro R define un radio de corte, limitando el área donde se considera la contribución de la articulación. Esto permite ignorar regiones lejanas donde la probabilidad sería prácticamente nula, optimizando así el tiempo de cómputo y reduciendo el ruido en el entrenamiento.

La red se entrena minimizando el error entre los mapas generados y los mapas de referencia. Para los *heatmaps*, se utiliza el error cuadrático medio (MSE) como función de pérdida:

$$L_h(\theta) == \frac{1}{N}\sum_{k=1}^{N}\sum_{x,y}\left(\hat{h}_k(x,y) - h_k(x,y)\right)^2$$

donde $\hat{h}_k(x,y)$ es el valor predicho por la red para el punto clave k, y $h_k(x,y)$ es el valor real de la gaussiana truncada.

Dado que los *heatmaps* tienen una resolución menor que la imagen original, la ubicación predicha por el valor máximo de cada mapa solo identifica la celda más probable. Sin embargo, esa celda abarca varios píxeles reales, por lo que no basta para determinar la posición exacta. Aquí es donde entran en juego los *offsets*: valores de corrección que indican en qué lugar específico de la celda se encuentra el punto clave, permitiendo recuperar precisión sub píxel (véase la Figura 3.14).

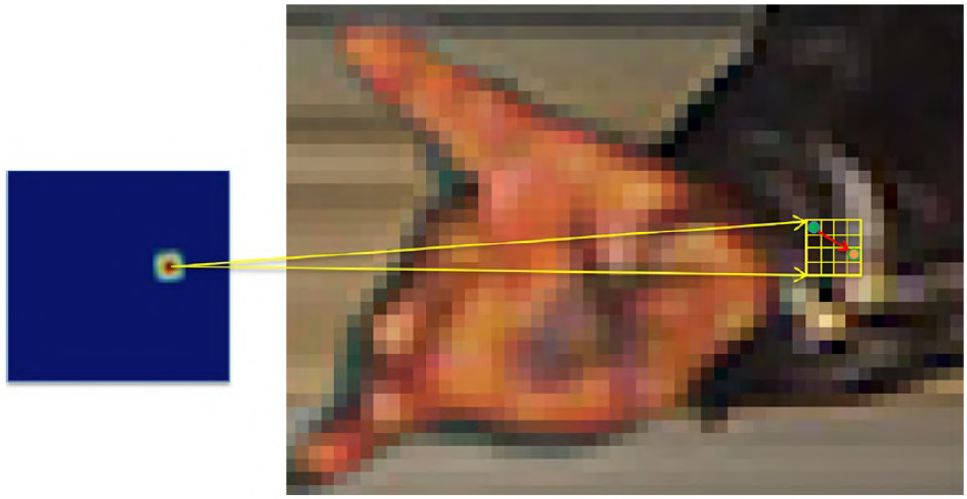


Figura 3.14. Relación entre la resolución del *heatmap* y la imagen original. Cada píxel del *heatmap* cubre una región de múltiples píxeles en la imagen original. La coordenada de máxima activación por sí sola no es suficiente para determinar la posición exacta, por lo que se predice un vector de *offset* para refinar la ubicación dentro de la celda (Xiao et al., 2022).

Cada *offset* se construye como un vector de corrección que representa la diferencia entre la ubicación real del punto clave en la imagen original y la posición central de la celda a la que

pertenece en el *heatmap*. Es decir, no se predice directamente la posición absoluta del punto clave, sino un desplazamiento relativo dentro de la celda correspondiente del mapa de calor. Esta relación puede expresarse de forma general mediante la ecuación (8), que conecta el espacio reducido del *heatmap* con el espacio original de la imagen.

$$P_k = s \cdot P_h + V_o \tag{8}$$

donde:

- P_k es la ubicación final del punto clave en coordenadas de la imagen original.
- P_h es la posición de la celda en el *heatmap* donde se encuentra la probabilidad máxima de la articulación k.
- s es el factor de escala espacial entre la imagen original y el *heatmap*.
- V_o es el vector de *offset*, que indica el desplazamiento fino dentro de la celda.

En este contexto, se define el factor de escala espacial s, que representa la relación entre las dimensiones de la imagen de entrada y las del *heatmap* de salida. Este valor puede calcularse como:

$$s = \frac{H_{\text{entrada}}}{H_{\text{heatmap}}} = \frac{W_{\text{entrada}}}{W_{\text{heatmap}}}$$

Por ejemplo, si la imagen de entrada tiene una resolución de 256×256 y el heatmap tiene una resolución de 64×64, entonces $s = 4$. Esto implica que cada celda del *heatmap* cubre un área de 4×4 píxeles en la imagen original.

Para una articulación k cuya ubicación real es $l_k = (x_k, y_k)$, la celda correspondiente en el heatmap se calcula como:

$$P_h = \left(\left\lfloor \frac{x_k}{s} \right\rfloor , \left\lfloor \frac{y_k}{s} \right\rfloor \right)$$

Y el vector de *offset* se obtiene como:

$$V_o = P_k - s \cdot P_h \tag{9}$$

$$V_o = \left(x_k - s \cdot P_h^{(x)},\ y_k - s \cdot P_h^{(y)} \right)$$

Estos valores indican con precisión dónde dentro de la celda se encuentra el punto clave.

Para entrenar correctamente la predicción de los *offsets*, se utiliza una función de pérdida conocida como *Smooth L1 Loss*. Esta pérdida se utiliza en lugar del error cuadrático medio (MSE)

tradicional porque utiliza la sensibilidad del MSE cuando los errores son pequeños y la robustez de la pérdida L1 cuando los errores son grandes. La pérdida se define de la siguiente manera:

$$L_o(\theta) = \begin{cases} 0{,}5 \cdot (o_k(x_i) - \bar{o}_k(x_i))^2, & \text{si } |o_k(x_i) - \bar{o}_k(x_i)| \leq 1 \\ |o_k(x_i) - \bar{o}_k(x_i)| - 0{,}5, & \text{si } |o_k(x_i) - \bar{o}_k(x_i)| > 1 \end{cases}$$

donde $o_k(x_i)$ es el valor del offset predicho por la red en la posición x_i para el punto clave k, y $\bar{o}_k(x_i)$ es el valor real del offset (calculado a partir de las etiquetas).

Esta función actúa de manera cuadrática cuando el error entre la predicción y el valor real es pequeño (menor o igual a 1). Esto permite que el modelo aprenda con más precisión y suavidad en las zonas cercanas al valor correcto. Sin embargo, cuando el error es mayor a 1, la función cambia a un comportamiento lineal, lo que evita que los errores grandes generen gradientes altos que podrían inestabilizar el aprendizaje.

Finalmente, el modelo de que usa MediaPipe genera una salida de dimensiones $64 \times 64 \times 99$, donde:

- 64×64 representa la resolución espacial de la malla del *heatmap*.
- 99 corresponde a los canales de salida, cada uno asignado a una predicción específica del modelo.

Dentro de estos 99 canales:

- los primeros N canales corresponden a la probabilidad de los $N = 33$ puntos clave.
- los siguientes $2N$ canales representan los offsets en x y y para cada punto clave.

Es importante resaltar que los *heatmaps* y *offsets* solo se utilizan durante la fase de entrenamiento para supervisar la regresión. En esta etapa, la función de pérdida minimiza el error entre la predicción en los *heatmaps* y *offsets* con las etiquetas manuales para que el modelo aprenda a predecir con mayor precisión los puntos clave. Sin embargo, durante la inferencia, estas capas son eliminadas y la estimación de los puntos clave se realiza directamente a partir de la representación interna aprendida por la red, lo que optimiza el rendimiento computacional y evita el cálculo de *heatmaps* en tiempo de ejecución.

Regresión de puntos clave

Una vez que el modelo ha sido supervisado mediante *heatmaps* y *offsets* durante el entrenamiento, en la fase de inferencia, la predicción de los puntos clave se realiza mediante un proceso de regresión. En este enfoque, la red neuronal convolucional genera directamente las coordenadas (x, y) de cada punto clave sin necesidad de *heatmaps*.

Para realizar esta tarea, la red cuenta con una capa de salida de regresión que contiene un número de nodos igual al doble del número de puntos clave (N), ya que cada uno requiere dos coordenadas. Por ejemplo, el modelo está diseñado para predecir 33 puntos clave, entonces la capa de salida tendrá 66 valores, correspondientes a las coordenadas $x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4, ..., x_{33}, y_{33}$.

El entrenamiento de esta regresión se basa en una función de pérdida que mide la diferencia entre las coordenadas predichas y las coordenadas reales en los datos de entrenamiento. Una de las funciones más utilizadas en este contexto es el error cuadrático medio *(Mean Squared Error, MSE)*, que se define como:

$$MSE = \frac{1}{N} \sum_{i=1}^{N} \left((x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right) \tag{10}$$

donde x_i, y_i representan las coordenadas reales de los puntos clave, y $\hat{x}_i, \hat{y}_i$ son las predicciones de la red.

El modelo ajusta sus parámetros mediante algoritmos de optimización con el objetivo de minimizar esta función de pérdida. Durante el entrenamiento, este proceso se repite a lo largo de múltiples iteraciones, hasta que el modelo logra predecir las posiciones de los puntos clave con precisión.

Finalmente, en la fase de inferencia, la regresión es la única responsable de la estimación de los puntos clave. Dado que los *heatmaps* y *offsets* fueron utilizados solo para supervisar el entrenamiento, estos son eliminados del modelo en la inferencia, permitiendo que la red prediga directamente las coordenadas finales con mayor eficiencia computacional.

La salida final del modelo se representa como una matriz de dimensiones 33×3, donde cada fila corresponde a un punto clave específico del cuerpo. Cada punto clave tiene asociadas dos coordenadas espaciales x, y, que indican su ubicación en la imagen, y un valor de visibilidad v, que representa la confianza del modelo en la detección de ese punto. Este valor de visibilidad varía entre 0 y 1, donde 1 indica que el punto es completamente visible y 0 que el modelo no pudo detectar su presencia con certeza. Esta información es fundamental para descartar predicciones poco confiables y mejorar la estabilidad del modelo en aplicaciones de seguimiento de pose en tiempo real.

Estimación coordenada 3D

En la estimación de pose en 3D, la coordenada z no es obtenida directamente mediante anotaciones humanas, sino que se infiere a partir de una combinación estructural del cuerpo y datos sintéticos generados. El modelo GHUM se entrena con un conjunto amplio de configuraciones humanas en 3D (como se muestra en la Figura 3.15), obtenidas a partir de escaneos de alta resolución y técnicas de registro de pose, lo que permite construir una representación coherente de la

forma y estructura del cuerpo.

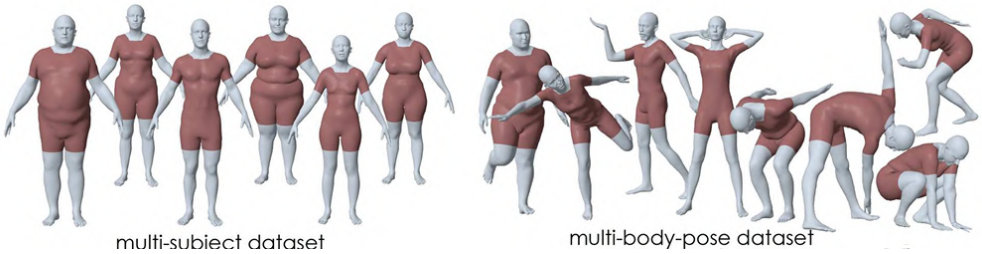


Figura 3.15. Mallas utilizadas para el entrenamiento (Xu et al., 2020).

Para predecir la coordenada z, el modelo parte de una estimación inicial basada en imágenes 2D y la ajusta utilizando restricciones biomecánicas y relaciones espaciales entre articulaciones. La profundidad de cada punto clave se calcula en relación con un sistema de referencia centrado en las caderas, donde valores negativos de z indican que el punto se encuentra más cerca de la cámara y valores positivos indican que está detrás del plano de las caderas.

Durante el entrenamiento, el modelo genera una malla tridimensional completa del cuerpo humano a partir de la imagen 2D (véase la Figura 3.16), la cual es comparada con una malla de referencia correspondiente. Para cuantificar la diferencia entre ambas, se emplea una función de pérdida basada en la distancia euclidiana promedio entre los vértices correspondientes. Esta pérdida se define como:

$$\mathcal{L}_{3D} = \frac{1}{N} \sum_{i=1}^{N} \|\hat{v}_i - v_i\|_2^2$$

donde:

- N es el número total de vértices en la malla,
- $\hat{v}_i$ es la posición del vértice i predicho por el modelo,
- v_i es la posición del vértice correspondiente en la malla de referencia.

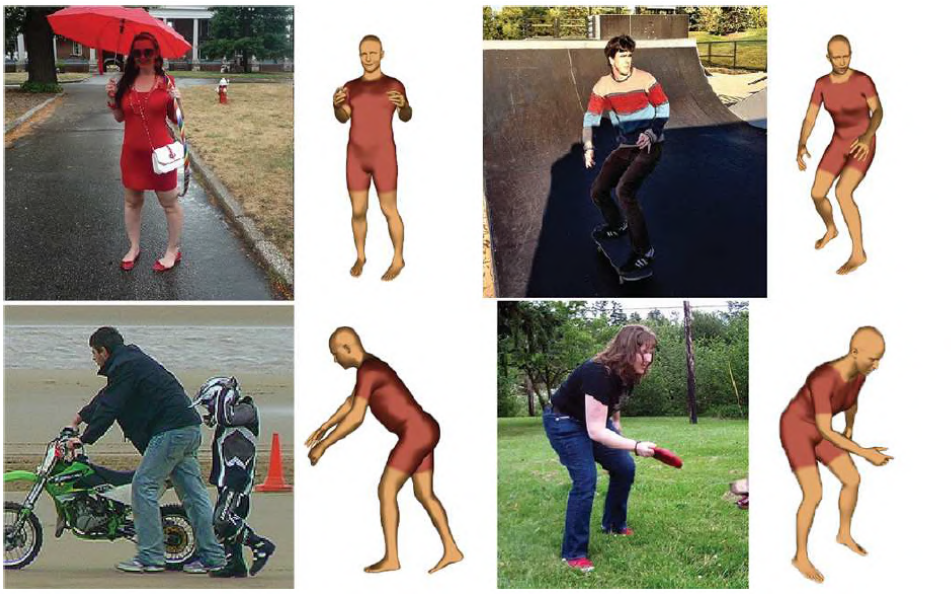


Figura 3.16. Mallas generadas durante el entrenamiento (Xu et al., 2020).

Este enfoque permite que el modelo aprenda a generar representaciones tridimensionales articuladas sin necesidad de contar con información explícita de profundidad en los datos de entrenamiento, aprovechando en su lugar inferencia estructurada sobre el cuerpo humano y supervisión basada en mallas 3D sintéticas.

Metodología

El proceso para el desarrollo de un sistema de análisis biomecánico de arqueros en fútbol se estructuró en varias etapas. En primer lugar, se realizó la recolección de datos mediante videos que capturaron los movimientos de los arqueros al ejecutar el gesto técnico del saque de piso. Posteriormente, dichos videos pasaron por una fase de preprocesamiento, en la cual se ajustaron las escalas de color para garantizar una calidad adecuada, y se sincronizaron las grabaciones provenientes de las vistas de las cámaras.

Una vez preprocesados los datos, se desarrolló un algoritmo de extracción de coordenadas para los puntos articulares de interés. A partir de estas coordenadas se realizaron los cálculos de los ángulos articulares necesarios para el análisis biomecánico.

Con el algoritmo implementado se analizó la información extraída con el objetivo de evaluar la precisión y consistencia de los resultados. Para ello, los valores obtenidos se compararon con los generados por otra herramienta de análisis biomecánico, lo que permitió validar la efectividad del método propuesto. Finalmente, se desarrolló una interfaz de usuario que permitió acceder de forma sencilla al algoritmo y facilitó su uso tanto para los análisis posteriores como para su implementación en entornos prácticos.

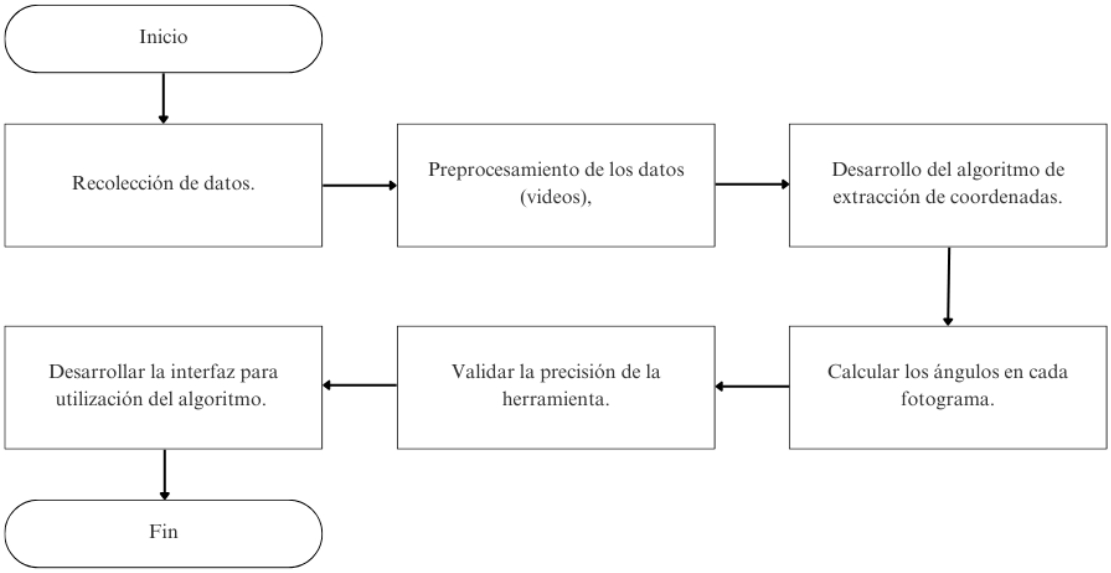


Figura 4.1. Diagrama de flujo de la metodología general del proyecto.

Recolección de datos

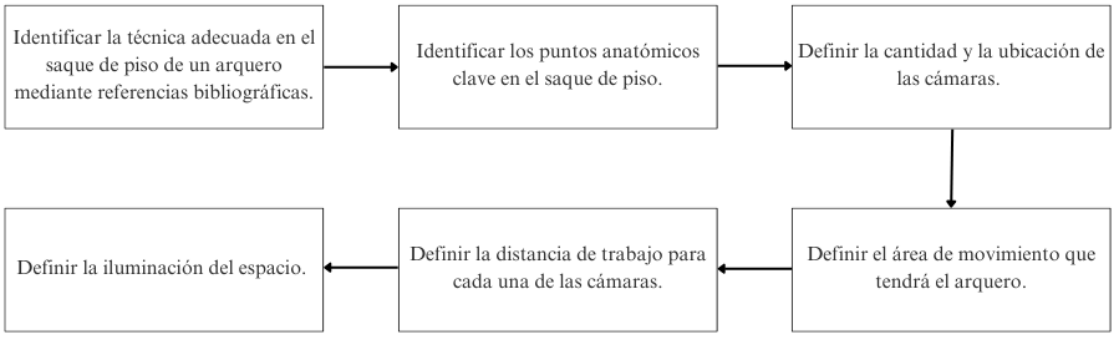


Figura 4.2. Diagrama de flujo de la primera etapa de la metodología: recolección de datos.

Para el saque de piso, que incluye un movimiento fundamental del tren inferior, se analizaron ángulos clave como el tobillo, la rodilla y la cadera para evaluar la estabilidad y eficiencia del tren inferior. Además, se tomaron ángulos que representaron la inclinación del cuerpo tanto desde la perspectiva trasera como lateral.

Por lo anterior, se consideró necesario utilizar 3 cámaras que permitieron capturar adecuadamente el movimiento del arquero desde diferentes ángulos. Una cámara se ubicó en la parte trasera para observar los ángulos y movimientos en el plano coronal, proporcionando una visión clara de la alineación del cuerpo durante la ejecución de los saques. Esta cámara se colocó en la parte trasera para evitar que los arqueros golpearan accidentalmente la cámara. Las otras dos cámaras se colocaron en los lados, ya que el movimiento del arquero no es simétrico; es decir, los

movimientos del lado derecho no son iguales a los del izquierdo. Con estas dos perspectivas se aseguró un análisis biomecánico preciso de ambos lados. Es importante mencionar que, aunque una mayor cantidad de cámaras podrían haber capturado más detalles, tres cámaras fueron suficientes para obtener los ángulos necesarios para un análisis biomecánico preciso. Este número de cámaras se ajustó además al objetivo de mantener el sistema práctico y accesible.

Para esta investigación se utilizaron tres dispositivos iPhone modelo 14 o superior, asegurando que todos contaran con las mismas especificaciones técnicas esenciales para el análisis, incluso si no pertenecían al mismo modelo exacto. Cada dispositivo fue configurado para grabar en resolución 1080p a 60 cuadros por segundo (*fps*), lo cual garantizó un nivel de detalle suficiente para el análisis biomecánico sin comprometer el rendimiento en la grabación. Además, en los tres celulares se utilizó de forma consistente la cámara gran angular (con un ángulo de visión de 120°), que permite capturar un campo visual amplio y mantener la coherencia entre las diferentes tomas.

Contar con cámaras con las mismas especificaciones fue importante, ya que aseguró que las imágenes obtenidas tuvieran la misma resolución y nivel de detalle, lo cual fue indispensable para un análisis biomecánico consistente y preciso. Cada cámara se colocó en un trípode para garantizar la estabilidad durante la grabación, evitando cualquier movimiento o vibración que pudiera comprometer la calidad de la captura. La estabilidad proporcionada por los trípodes aseguró que las grabaciones fueran nítidas y sin distorsiones, lo que fue fundamental para que el algoritmo de reconocimiento de articulaciones funcionara de manera correcta.

Por otro lado, para asegurar una recolección de datos precisa, fue fundamental delimitar adecuadamente el área de movimiento del arquero. En este estudio, se realizó una fase preliminar de toma de datos para determinar las dimensiones máximas del espacio necesario. Se tomaron un total de 20 intentos de saque de piso, distribuidos entre cuatro arqueros: tres hombres y una mujer, cada uno con cinco intentos. Durante esta fase, se registraron las posiciones iniciales y finales de cada saque, lo que permitió definir el espacio efectivo en el que se desplazan los arqueros. Los resultados obtenidos fueron los siguientes:

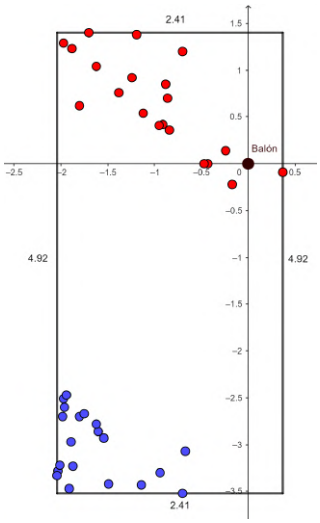


Figura 4.3. Puntos de movimiento del arquero en el saque de piso y área delimitada.

De igual forma, con el fin de lograr una recolección de datos precisa, fue importante aplicar conceptos fundamentales de fotografía que influyen directamente en la calidad y precisión de las imágenes, afectando así la exactitud del algoritmo. En este contexto, conceptos como el ángulo de visión, el campo de visión (FOV por sus siglas en inglés) y la distancia de trabajo (WD por sus siglas en inglés) jugaron un papel clave en la metodología. El ángulo de visión determina el rango que la cámara puede capturar, esencial para posicionar las cámaras correctamente y cubrir todo el cuerpo del deportista en su área de movimiento. El campo de visión se refiere a la extensión del área visible que la cámara puede observar en su entorno, asegurando que la zona de actividad del arquero esté completamente encuadrada. La distancia de trabajo es fundamental para garantizar que el campo de visión cubra adecuadamente la totalidad del área de interés, permitiendo capturar imágenes claras y detalladas que faciliten un análisis biomecánico preciso.

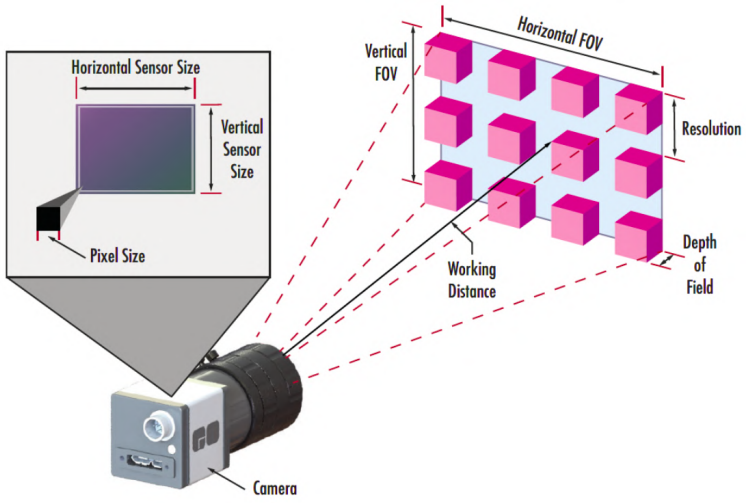


Figura 4.4. Parámetros fundamentales de un sistema de imágenes (Optics, s.f.).

Para determinar la posición exacta de las cámaras, se calculó la distancia de trabajo (WD por sus siglas en inglés), que es la distancia entre el lente de la cámara y el sujeto (en este caso, el arquero). Esta distancia aseguró que se capturara todo el campo de visión deseado. La fórmula utilizada para encontrar la distancia de trabajo fue la siguiente:

$$\tan\left(\frac{\theta}{2}\right) = \frac{\text{FOV}}{2 \times \text{WD}}$$

Despejando para encontrar la distancia de trabajo:

$$\text{WD} = \frac{\text{FOV}}{2 \times \tan\left(\frac{\theta}{2}\right)}$$

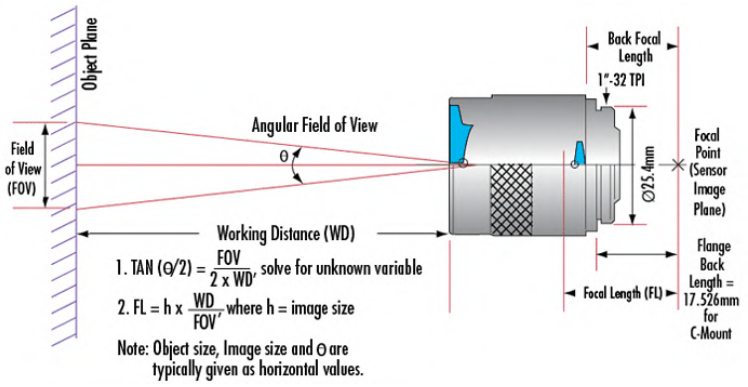


Figura 4.5. Diagrama de un lente focal fijo (Optics, s.f.).

Aplicando esta fórmula se obtuvieron las distancias específicas para cada configuración de las cámaras:

1. La cámara trasera se ubicó detrás del arquero para capturar un área de movimiento de 1.92 metros de altura (que corresponde a la altura máxima de los arqueros de la muestra) y 2.41 metros de ancho (campo de movimiento horizontal). Esta configuración aseguró que se capturara el cuerpo completo del arquero durante toda la ejecución del saque. El ángulo de visión de la cámara utilizada fue de 120 grados.

$$FOV = 2{,}41m$$

$$\theta = 120\circ$$

$$\tan\left(\frac{120^\circ}{2}\right) = \frac{2{,}41}{2 \times \mathrm{WD}}$$

$$\mathrm{WD} = \frac{2{,}41}{2 \times \tan(60^\circ)}$$

$$\mathrm{WD} = \frac{2}{2 \times \sqrt{3}}$$

$$\mathrm{WD} = 0{,}696\mathrm{m} \approx 0{,}7\mathrm{m}$$

2. Las cámaras laterales se colocaron a ambos lados del arquero para cubrir un campo de movimiento horizontal de 4.92 metros. Esto se debe a que, durante la ejecución del saque, el arquero se desplaza frontalmente, y es necesario capturar todo ese movimiento en detalle. Al igual que la cámara trasera, las cámaras laterales también tuvieron un ángulo de visión de 120 grados.

$$FOV = 4{,}92m$$

$$\theta = 120\circ$$

$$\tan\left(\frac{120^\circ}{2}\right) = \frac{4{,}92}{2 \times \mathrm{WD}}$$

$$\mathrm{WD} = \frac{4{,}92}{2 \times \tan(60^\circ)}$$

$$\text{WD} = \frac{4{,}92}{2 \times \sqrt{3}}$$

$$\text{WD} = 1{,}42\text{m} \approx 1{,}5\text{m}$$

Con todo lo anterior, la configuración de las cámaras quedó de la siguiente manera: tres cámaras iPhone modelo 14 o superior, configuradas a 1080p y a 60 cuadros por segundo, dispuestas estratégicamente alrededor del área de movimiento del arquero para asegurar la captura completa de su cuerpo y el análisis biomecánico preciso en cada fase del saque. La cámara trasera se ubicó a una distancia de aproximadamente $0{,}7$ metros detrás del arquero, mientras que las cámaras laterales estuvieron a una distancia de aproximadamente $1{,}5$ metros a cada lado del área de movimiento, garantizando una cobertura óptima del campo visual.

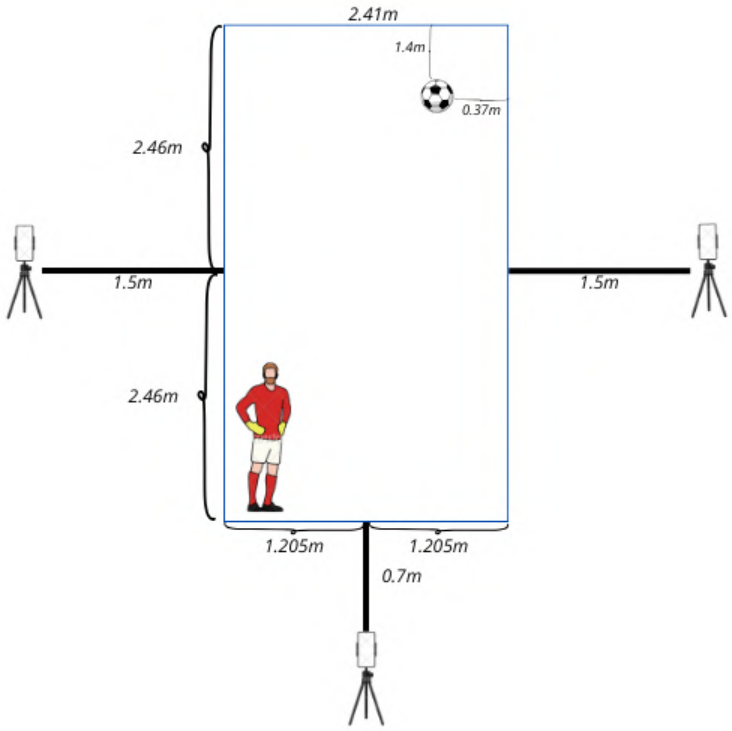


Figura 4.6. Configuración de las cámaras.

Los videos fueron grabados en la cancha sintética de la Pontificia Universidad Javeriana. Las grabaciones se realizaron en un horario comprendido entre las 9:00 a.m. y las 11:00 a.m., con el objetivo de garantizar una iluminación natural óptima. Este intervalo permitió evitar tanto los excesos de luz y sombras intensas que suelen presentarse al mediodía, como la baja luminosidad característica de la tarde-noche, lo cual fue fundamental para asegurar la calidad visual de los

videos.

Por otro lado, se ejerció un control básico sobre la vestimenta de los arqueros, procurando que el tronco presentara un color distinto al del resto del cuerpo, con el fin de facilitar la identificación de los puntos clave durante el procesamiento y análisis de los videos.

A continuación, se presentan una fotografía de la configuración real utilizada durante las sesiones de grabación.



Figura 4.7. Configuración de la toma de videos utilizada durante las sesiones de grabación.

En total, se recolectaron 99 videos correspondientes a las ejecuciones del gesto técnico del saque de piso, realizados por arqueros de la Pontificia Universidad Javeriana. La muestra incluyó tres mujeres y cinco hombres, con un rango de 12 a 13 videos grabados por cada participante. Todas las grabaciones se realizaron después de que los arqueros completaran una fase de calentamiento y firmaran el consentimiento informado, garantizando así las condiciones físicas adecuadas y el cumplimiento de los requisitos éticos del estudio. Esta cantidad de registros proporcionó una base de datos suficientemente robusta para el desarrollo y validación del sistema de análisis biomecánico.

Preprocesamiento de datos

Espacios de color

En esta etapa, se realizó el preprocesamiento de los datos obtenidos a partir de los videos capturados, en los cuales se observan los movimientos específicos de los arqueros durante el saque de piso. Para definir el tratamiento adecuado de estos datos, se llevaron a cabo pruebas preliminares con el objetivo de determinar las condiciones óptimas para la detección de la pose y la segmentación del movimiento.

Uno de los aspectos evaluados en estas pruebas fue la representación del cuerpo en distintos espacios de color, con el fin de identificar cuál resaltaba mejor la silueta del arquero en comparación con el fondo. Para ello, se tomó un único fotograma de los videos de prueba y se revisaron diferentes espacios de color.

En primer lugar, la imagen fue cargada en ***OpenCV***, el cual por defecto lee las imágenes en el espacio de color ***BGR*** *Blue, Green, Red* (véase la Figura 4.8).



Figura 4.8. Fotogramas cargados en BGR.

Posteriormente, la imagen fue convertida al espacio de color ***RGB*** *(Red, Green, Blue)* por medio de la función *cv2.cvtColor(imagen, cv2.COLOR_BGR2RGB)* (véase la Figura 4.9). Esta conversión simplemente reorganiza los canales de color. En este caso, logramos ver que el cuerpo se resalta un poco más sobre el fondo.



Figura 4.9. Fotogramas cargados en RGB.

A continuación, la imagen fue transformada al espacio de color ***HSV*** *(Hue, Saturation, Value)* (véase la Figura 4.10). En este caso, la conversión a HSV evidencia una limitación importante para la segmentación del cuerpo del arquero. Se observa que los tonos de piel y los contornos del cuerpo se confunden con el fondo. Además, la cabeza del arquero se pierde entre las sombras proyectadas por los árboles. Esta falta de contraste entre el arquero y el fondo indica que el espacio de color ***HSV*** no es el más adecuado.

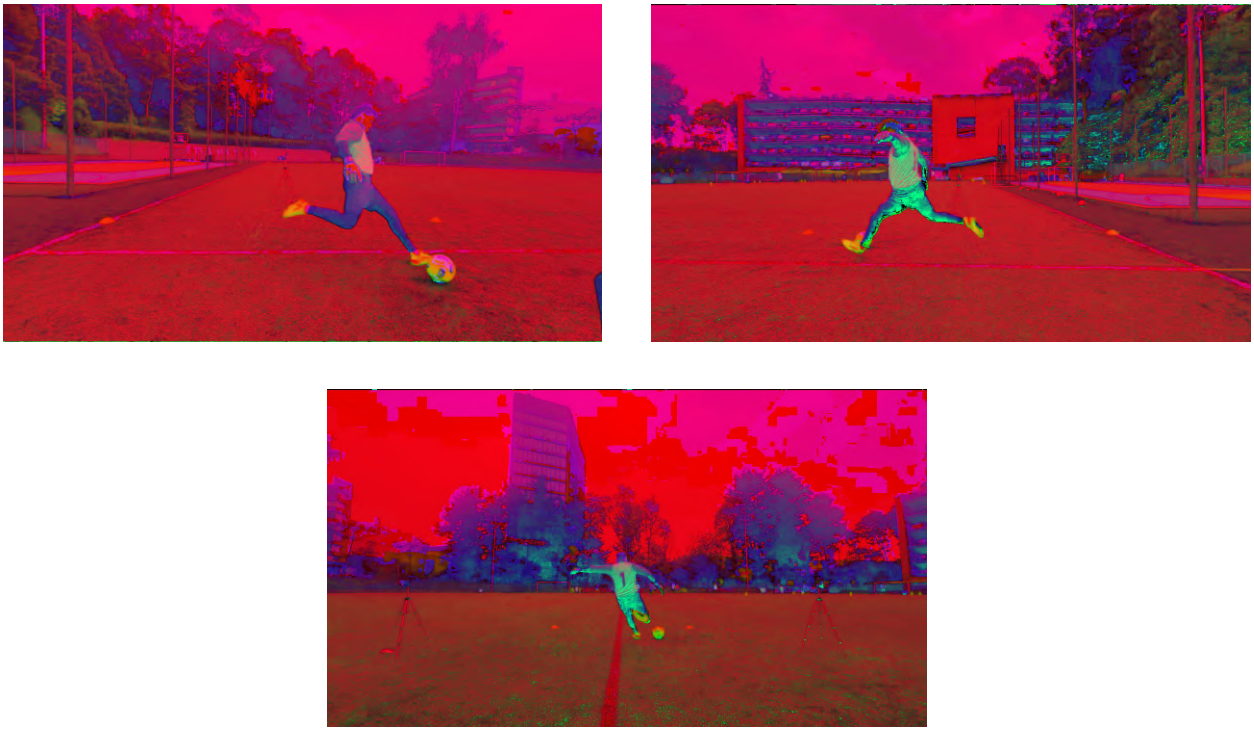


Figura 4.10. Fotogramas cargados en HSV.

Posteriormente, la imagen fue transformada al espacio de color **YCrCb** (véase la Figura 4.11). En este caso, se logran distinguir tres regiones principales en la imagen: el cielo, el suelo y el conjunto del arquero con los árboles. Si bien logra resaltar diferencias en el entorno, no logra separar la totalidad del cuerpo del arquero del fondo, lo que indica que el espacio de color ***YCrCb*** no es el más adecuado.



Figura 4.11. Fotogramas cargados en YCrCb.

Finalmente, la imagen fue transformada al espacio de color **LAB**. Al igual que en el espacio de color **YCrCb**, se logran distinguir tres regiones principales, sin embargo, no se logra separar el cuerpo del arquero del fondo, lo que indica que el espacio ***LAB*** no es el más adecuado.



Figura 4.12. Fotogramas cargados en LAB.

Finalmente, tras evaluar las diferentes conversiones de color, se concluyó que el espacio de color RGB es el más adecuado para la detección de la pose del arquero. A diferencia de los modelos HSV, YCrCb y LAB, en los cuales el arquero tiende a perderse entre los elementos del fondo debido a la similitud cromática, RGB permite una mejor diferenciación del cuerpo respecto al entorno. Esto se debe a que el color de la vestimenta contrasta de manera más clara con el verde del campo, facilitando la identificación de los puntos clave en la estimación de pose.

Por lo tanto, en el preprocesamiento de los datos, se utilizó el modelo RGB para optimizar la segmentación del arquero y garantizar un análisis más preciso del movimiento.

Recorte de videos

Para garantizar que los videos capturados desde diferentes ángulos (vista coronal y vistas sagitales) estuvieran sincronizados y permitieran un análisis preciso de la ejecución del movimiento, fue necesario definir un punto de inicio y un punto final común para todos ellos. El inicio se determinó a partir de un criterio biomecánico basado en la posición del tobillo de la pierna de golpeo en relación con la rodilla de la misma pierna.

Con este objetivo, se desarrolló un algoritmo que recorrió los videos fotograma por fotograma y analizó la información espacial de los puntos clave extraídos mediante el modelo de estimación de pose de MediaPipe. Este modelo utilizó un sistema de coordenadas donde el punto $(0, 0)$ se encontraba en la esquina superior izquierda, por lo que valores mayores en el eje Y

representaban posiciones más cercanas al piso (véase la Figura 4.13).

En cada fotograma, el algoritmo calculó la distancia vertical entre el tobillo y la rodilla de la pierna de pateo para identificar el momento en que esta extremidad comenzaba a despegar del suelo. Esta distancia fue usada como parámetro principal para determinar el inicio del gesto técnico (véase la Figura 4.13).



Figura 4.13. Identificación de la coordenada Y de la rodilla (color azul claro) y el tobillo (color rojo), así como la diferencia entre ambas coordenadas (color azul).

Además, se incorporaron dos condiciones adicionales:

- Que la pierna de pateo se encontrara por detrás de la pierna de apoyo: para verificar esta condición, se compararon las posiciones espaciales de ambas extremidades inferiores, dependiendo de la orientación de la cámara.
    - En la vista sagital derecha se emplearon las coordenadas en x de la punta del pie de pateo y el talón de la pierna de apoyo. En este caso, la punta del pie de pateo debía estar por detrás del talón de la pierna de apoyo, lo cual implicaba que la coordenada en x de la punta del pie de pateo fuera **menor** que la del tobillo de la pierna de apoyo, como se muestra en la Figura 4.14.

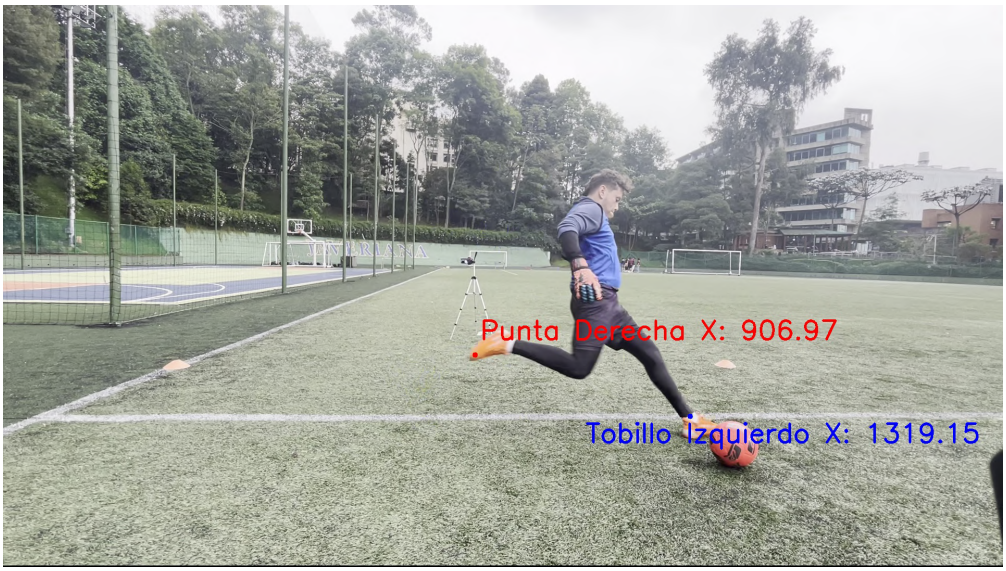


Figura 4.14. Comparación de coordenadas x en la vista sagital derecha.

- En la vista sagital izquierda se emplearon las coordenadas en x de la punta del pie de pateo y el talón de la pierna de apoyo. En este caso, la punta del pie de pateo debía estar por detrás del talón de la pierna de apoyo, lo cual implicaba que la coordenada en x de la punta del pie de pateo fuera **mayor** que la del tobillo de la pierna de apoyo, como se muestra en la Figura 4.15.



Figura 4.15. Comparación de coordenadas x en la vista sagital izquierda.

- En la vista coronal, se emplearon las coordenadas en el eje z de la punta del pie y del tobillo de la pierna de pateo, así como del tobillo de la pierna de apoyo. Adicionalmen-

te, se incorporó un criterio basado en las coordenadas en el eje y para verificar que la rodilla de la pierna de pateo se encontrara más cerca del suelo que la cadera de la pierna de apoyo. Se consideró que la pierna de pateo estaba detrás cuando sus coordenadas en z eran más negativas que las de la pierna de apoyo, teniendo en cuenta que el sistema de referencia estaba centrado en las caderas (véase la Figura 4.16).



Figura 4.16. Coordenada z de los tobillos para la vista coronal.

- Que el arquero se viera a cuerpo completo: para validar que el arquero se encontrara completamente visible, se evaluó la visibilidad de tres puntos clave: la nariz y ambos talones. Solo se consideraron válidos los fotogramas en los que los tres puntos presentaban una visibilidad superior a $0{,}7$, garantizando así que el cuerpo estuviera completamente dentro del encuadre.

Inicialmente, se estableció un valor de referencia llamado $mejor_distancia$, el cual se actualizaba únicamente si el fotograma actual superaba dicha distancia y cumplía con las dos condiciones descritas. Si se detectaban tres cuadros consecutivos que satisfacían simultáneamente estas condiciones, se seleccionaba el último de ellos como el fotograma de inicio común para los distintos ángulos de grabación. Esto permitió que el recorte del video fuera automático y se basara en un criterio biomecánico replicable.

A continuación, en la Figura 4.17 se presenta un diagrama de flujo que describe claramente este proceso:

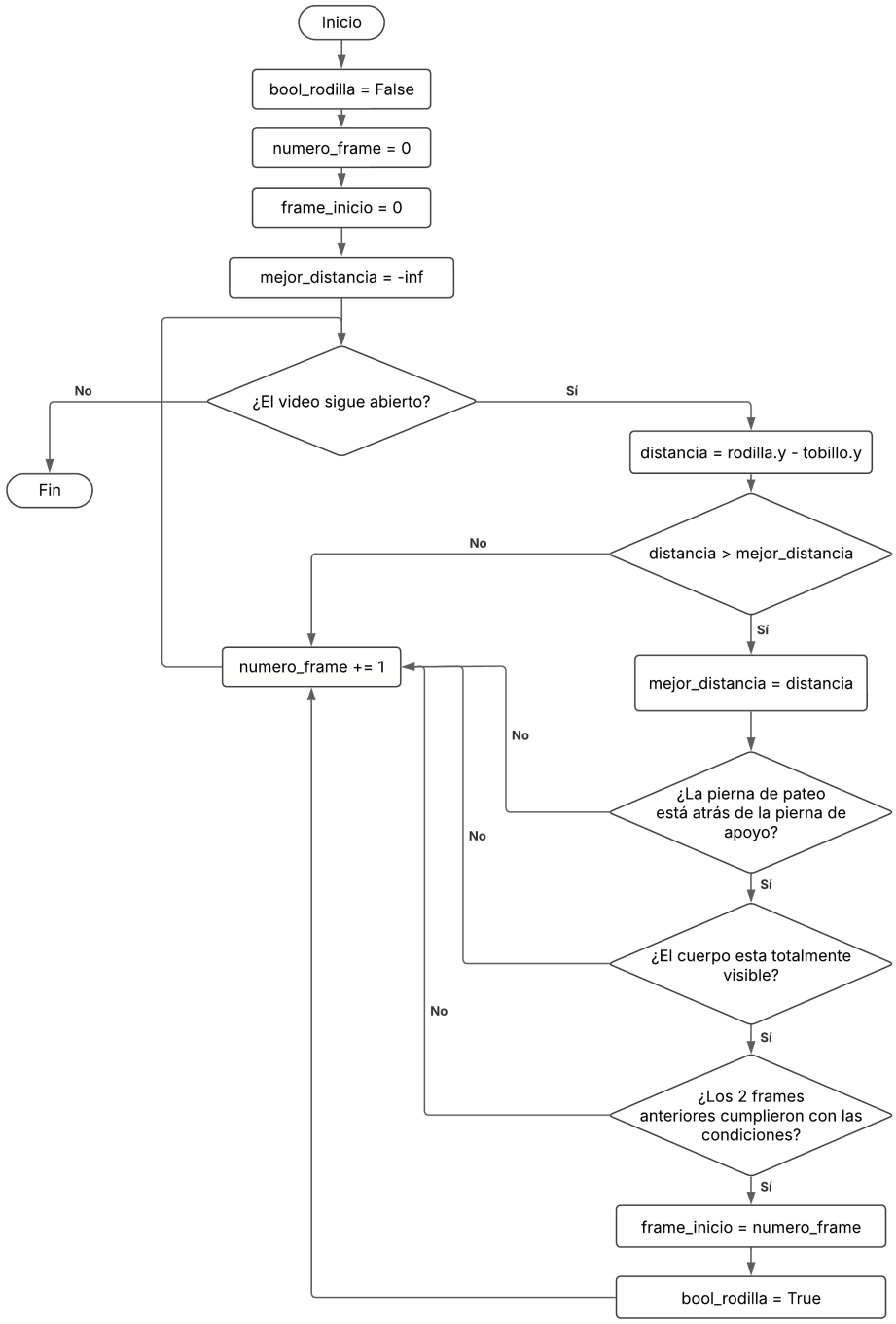


Figura 4.17. Diagrama de flujo para la determinación del fotograma inicial del video según el criterio biomecánico establecido.

Este criterio biomecánico constituye el método principal para determinar el inicio del video. Sin embargo, durante las pruebas se identificaron casos en los que no todas las tomas lograban cumplir con este criterio de manera consistente. Específicamente, en algunas grabaciones (particu-

larmente en la vista posterior) se presentaron dificultades en la identificación precisa del tobillo y la rodilla de la pierna de golpeo. Esto ocurría cuando la conexión entre estos dos puntos clave se solapaba visualmente con la conexión del fémur (véase la Figura 4.18), dificultando su detección por parte del modelo.



Figura 4.18. Inconsistencias en la detección del tobillo y de la rodilla desde la vista coronal.

Dado que lo ideal es que las tres vistas se rijan por el mismo criterio, y para evitar inconsistencias causadas por estas limitaciones en la vista trasera, se optó por incorporar un criterio auxiliar que garantizara una detección confiable sin depender de ese solapamiento.

Este criterio auxiliar consistió en utilizar la posición del tobillo y la rodilla de la pierna de golpeo para determinar el inicio del gesto. El proceso se basó en un análisis fotograma por fotograma, en el que se identificó el primer instante en que el tobillo de la pierna de golpeo superaba, en la coordenada Y, a la rodilla de la misma pierna. Una vez se cumplía esta condición, se seleccionaba ese mismo fotograma como el punto de inicio del video recortado. Cabe resaltar que, además de este criterio, también se evaluaban las condiciones de que la pierna de golpeo se encontrara por detrás y que el cuerpo completo del arquero fuera visible en el cuadro.

A continuación, en la Figura 4.19 se presenta un diagrama de flujo que describe claramente este proceso:

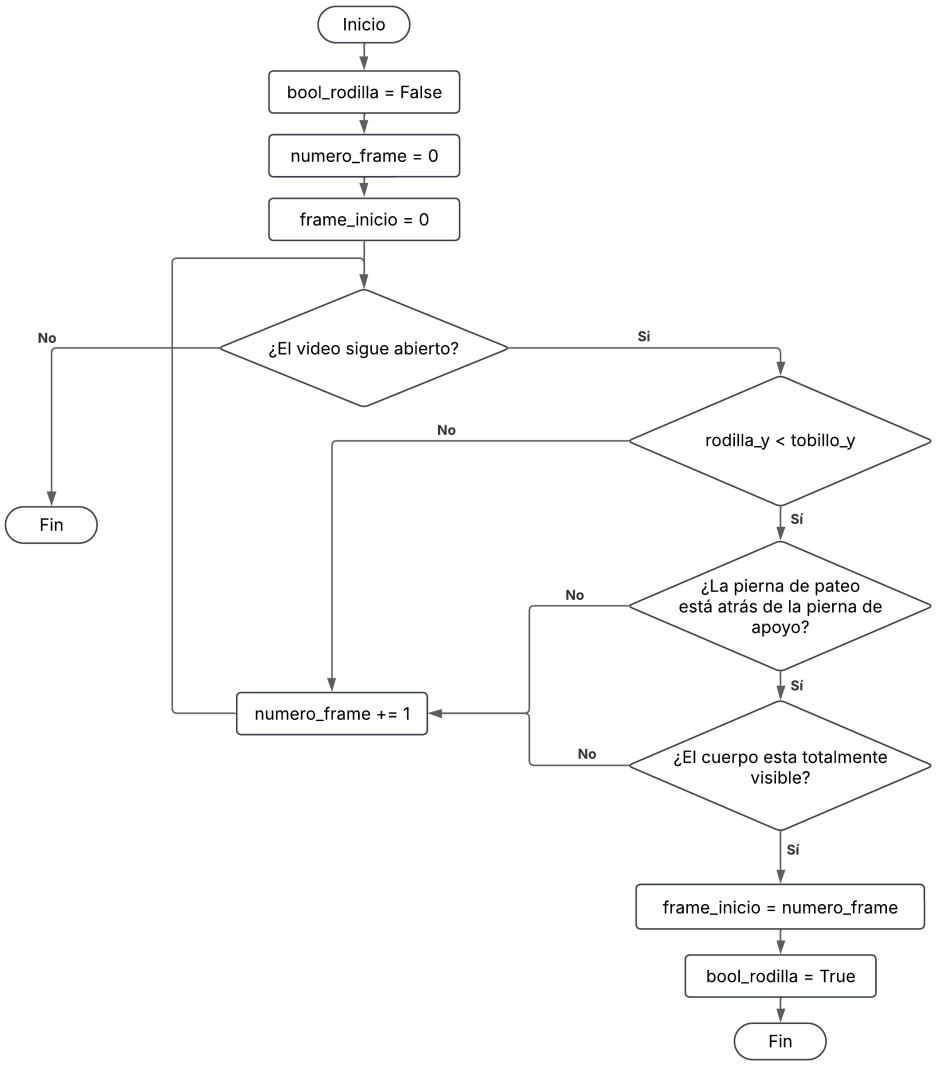


Figura 4.19. Diagrama de flujo para la determinación del fotograma inicial del video según el criterio biomecánico auxiliar establecido.

Seguidamente, las figuras 4.20, 4.21 y 4.22 ilustran visualmente cómo se aplica este criterio en las tomas laterales y traseras, mostrando un ejemplo concreto del instante previo, actual y posterior al cumplimiento de la condición biomecánica auxiliar establecida.



Figura 4.20. Determinación del inicio del video mediante el criterio biomecánico (vista lateral derecha).



Figura 4.21. Determinación del inicio del video mediante el criterio biomecánico (vista lateral izquierda).



Figura 4.22. Determinación del inicio del video mediante el criterio biomecánico (vista trasera).

Por otro lado, durante las pruebas se identificaron inconsistencias en la detección de las extremidades entre fotogramas consecutivos, lo que podría afectar el recorte de video. La figura 4.23 ilustra este problema.

En algunos fotogramas, el sistema de estimación de pose reconocía correctamente la pierna de apoyo y la pierna de pateo, mientras que en otros se observaba una inversión en la identificación de las extremidades. Este error generaba variaciones abruptas en la detección de la pose, afectando la coherencia de los datos extraídos y la sincronización de los videos.



Figura 4.23. Inconsistencias en la detección de extremidades en fotogramas consegutivos. .

Para corregir esta situación, se implementó un mecanismo de verificación que detectara diferencias drásticas en la asignación de las extremidades entre fotogramas consecutivos. En caso de que el modelo asignara incorrectamente las piernas, el programa ignoraba ese fotograma y no lo tomaba en cuenta para el análisis del recorte de video.

Por otro lado, el final del video se determinó con base en un criterio biomecánico relacionado con la posición relativa de la pierna de pateo respecto a la pierna de apoyo. Para ello, se recorrieron los fotogramas **desde el fotograma de inicio del recorte identificado** hasta encontrar el primer instante en el que la pierna de golpeo hubiera completado el movimiento de impacto y superado en coordenadas específicas a la pierna de apoyo. De igual forma, acá se verificaba la condición de visibilidad descrita anteriormente.

Dado que las grabaciones incluían tres perspectivas distintas (lateral derecha, lateral izquierda y trasera), se establecieron criterios diferenciados para cada una:

- Para la toma **lateral derecha**, el criterio de finalización se definió en el fotograma siguiente en el que el tobillo, la rodilla, el talón y la punta del pie de la pierna de golpeo superaran en coordenada X al tobillo, la rodilla, el talón y la punta del pie de la pierna de apoyo. En la Figura 4.24 se puede observar este criterio aplicado a los tobillos y a las rodillas.

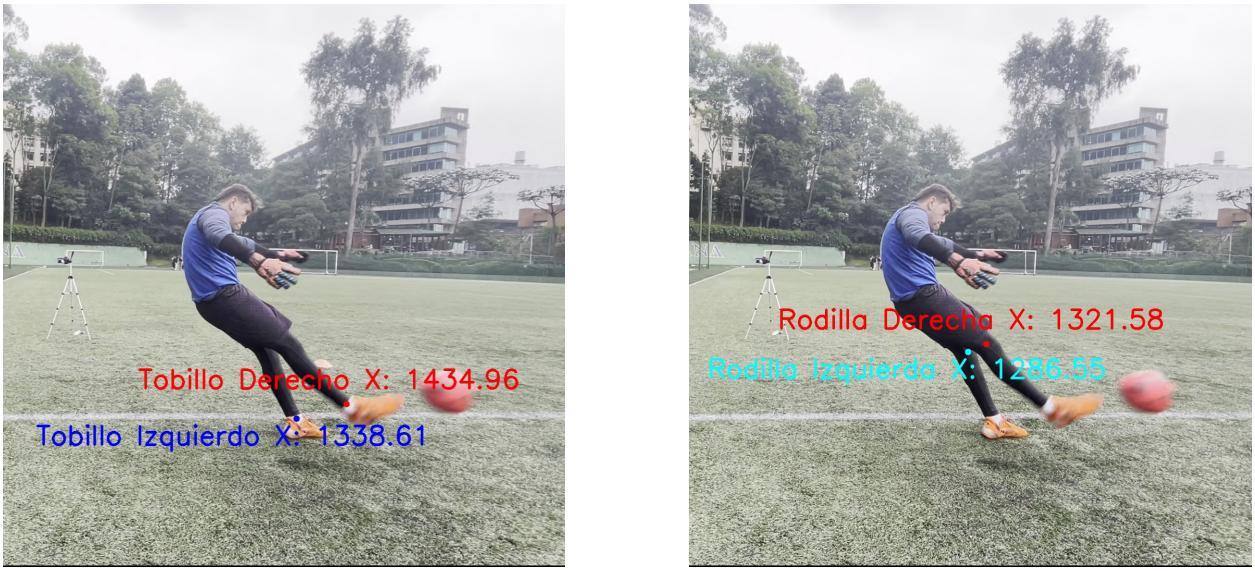


Figura 4.24. Criterio de fin para las rodillas y los tobillos desde la vista sagital derecha.

- Para la toma **lateral izquierda**, el criterio de finalización se definió en el fotograma siguiente en el que el tobillo, la rodilla, el talón y la punta del pie de la pierna de golpeo estuvieran en una posición inferior en la coordenada X respecto al tobillo, la rodilla, el talón y la punta del pie de la pierna de apoyo. En la Figura 4.25 se puede observar este criterio aplicado a los tobillos y a las rodillas.

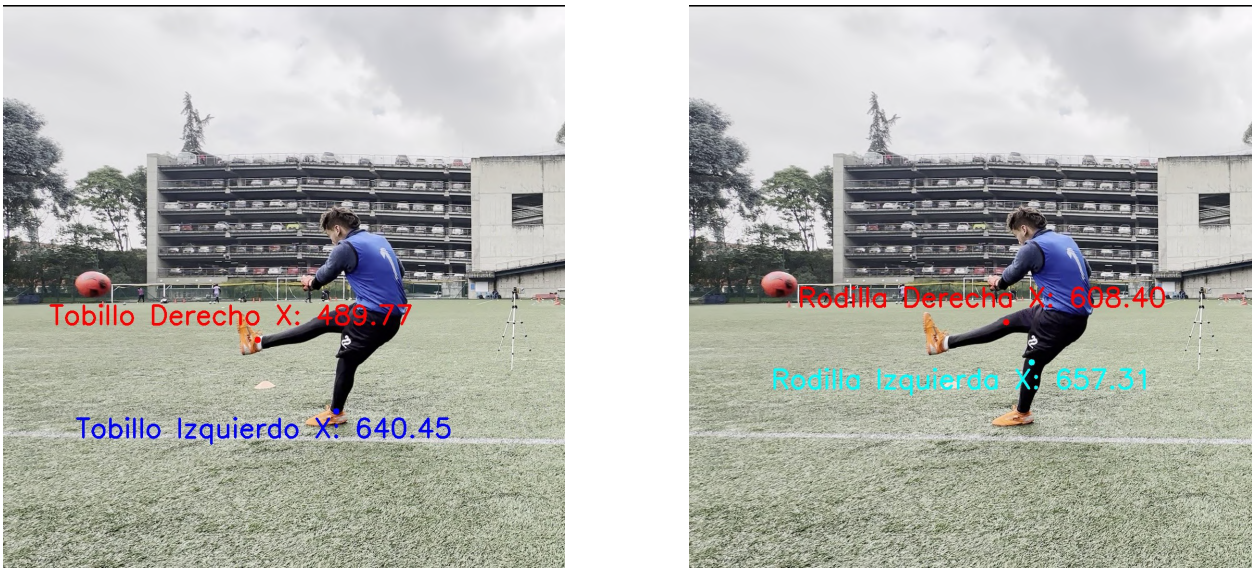


Figura 4.25. Criterio de fin para las rodillas y los tobillos desde la vista sagital izquierda.

- Para la toma **trasera**, el criterio de finalización se definió en el fotograma siguiente en el que el tobillo, la rodilla, el talón y la punta del pie de la pierna de golpeo superaran en coordenada

Z al tobillo, la rodilla, el talón y la punta del pie de la pierna de apoyo. En la Figura 4.26 se puede observar este criterio aplicado a los tobillos y a las rodillas.

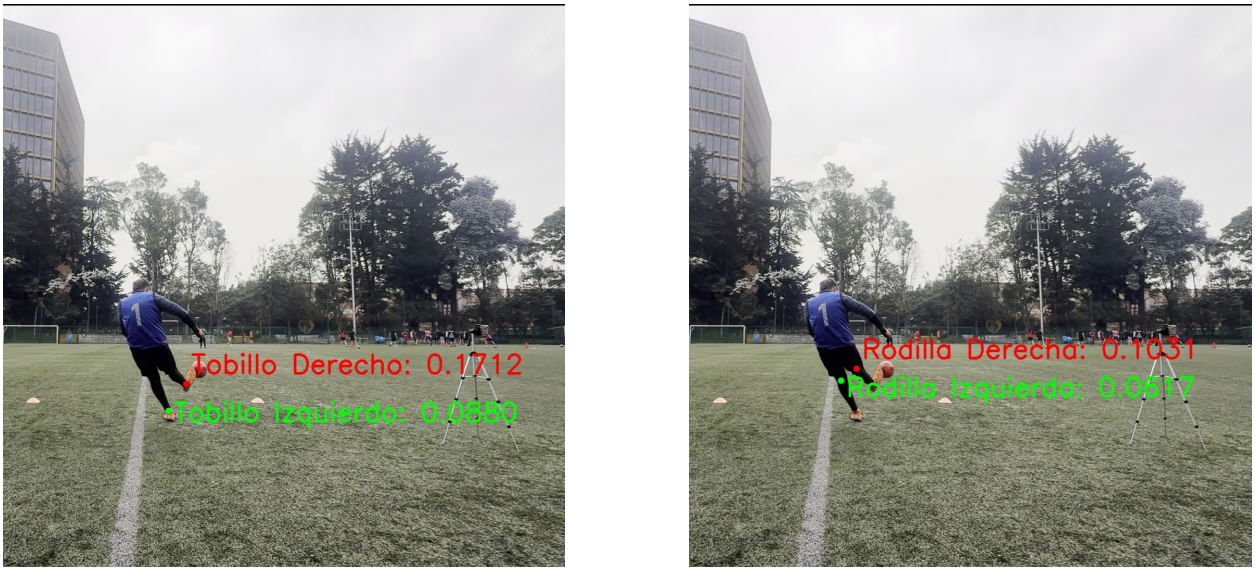


Figura 4.26. Criterio de fin para las rodillas y los tobillos desde la vista coronal.

Este proceso garantizó que el video se recortara en el momento preciso en que la patada había sido completada. Una vez se detectaba el primer fotograma en el que se cumplía el criterio de finalización, se establecía el fotograma siguiente como el fotograma final del video.

Finalmente, una vez determinados los fotogramas de inicio y fin, se recortó cada video en función de esos valores. En la Figura 4.27 se logra observar el proceso general de recorte de video.

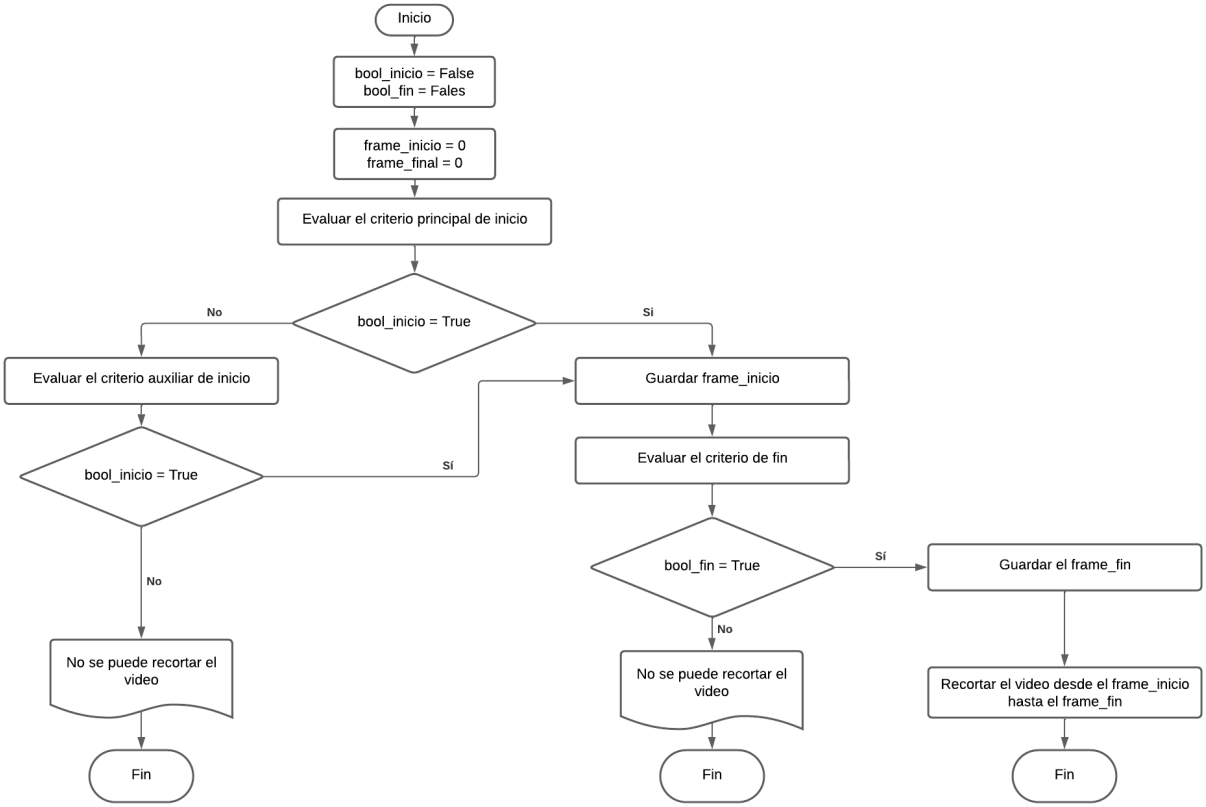


Figura 4.27. Diagrama de flujo general del proceso de recorte de video.

Desarrollo del algoritmo

Luego de obtener los videos recortados, se desarrolló el algoritmo encargado de extraer los puntos clave de las extremidades de interés en cada uno de los videos. A diferencia del procedimiento utilizado para el recorte, cuyo objetivo principal era identificar el inicio y final del gesto técnico, esta etapa se centró en obtener las coordenadas precisas de los puntos anatómicos relevantes a lo largo de toda la secuencia de movimiento. Además de la extracción, el algoritmo incluyó mecanismos de verificación y corrección que permitieron detectar inconsistencias en la estimación de pose y subsanarlas de forma automática, garantizando así la coherencia y calidad de los datos obtenidos.

En primer lugar, se definieron los puntos clave que debían extraerse según la vista del video, ya que cada perspectiva permitió analizar distintos aspectos del gesto técnico del arquero:

- **Vista trasera:** esta vista se empleó principalmente para evaluar la inclinación del cuerpo durante todo el movimiento. Para ello, se consideraron como puntos de interés los hombros y las caderas, utilizando el punto medio entre ambos para representar la inclinación del tronco superior e inferior, respectivamente. Adicionalmente, se incluyó el talón de la pierna

de apoyo, ya que sirvió como punto de referencia para el análisis de estabilidad.

- **Vista lateral izquierda:** desde esta perspectiva se analizó tanto la inclinación del cuerpo como el comportamiento de la pierna de apoyo, prestando especial atención al ángulo de la rodilla y del tobillo. Por tanto, los puntos extraídos fueron el hombro, la cadera, la rodilla, el tobillo, el talón y la punta del pie izquierdos.
- **Vista lateral derecha:** esta vista permitió observar en detalle el gesto técnico de la pierna de golpeo, por lo que se extrajeron los puntos correspondientes al hombro, cadera, rodilla, tobillo, talón y punta del pie derechos. Estos puntos fueron fundamentales para calcular los ángulos articulares implicados en el movimiento de golpeo.

Una vez definidos los puntos clave para cada una de las vistas, se procedió con la extracción y corrección de los mismos a partir de los videos previamente recortados. Para ello, se recorrió cada video cuadro por cuadro, con el objetivo de extraer los puntos anatómicos de interés y de detectar posibles inconsistencias en la estimación de ellos.

En cada iteración se extrajeron los puntos clave del fotograma actual y se compararon con los del fotograma anterior, haciendo énfasis en las coordenadas del tobillo de la pierna de golpeo y de la pierna de apoyo. Si se detectaba una diferencia drástica en la posición horizontal (eje X) de alguno de estos puntos, se consideraba que la estimación del fotograma actual podía ser incorrecta. De forma más específica, este tipo de cambio solía indicar que el modelo había intercambiado las piernas, como se muestra en la Figura 4.28. En tal caso, se conservaba temporalmente la información del fotograma anterior para reemplazar momentáneamente al fotograma erróneo y así mantener la continuidad del movimiento. Sin embargo, este reemplazo no se consideraba una corrección definitiva, ya que los fotogramas afectados serían revisados y corregidos posteriormente en una etapa específica del proceso. En caso contrario, se guardaba la información del cuadro como válida.



Figura 4.28. Ejemplo de detección de cambio drástico entre fotogramas consecutivos.

En los videos se identificaron tres casos de error en la estimación de puntos clave:

- Cuando el modelo no fue capaz de reconocer bien el primer fotograma.
- Cuando había un cambio entre piernas.
- Cuando había cambios drásticos entre la coordenada de un punto entre fotogramas.

Para abordar estos casos, se realizó un análisis posterior a la extracción de los puntos clave, basado en las distancias euclidianas entre fotogramas consecutivos. Para cada punto clave, se calculaba la distancia euclidiana 11 entre su posición en un fotograma y en el siguiente. Este cálculo permitió identificar comportamientos anómalos en el movimiento.

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{11}$$

A partir de este análisis, se distinguieron principalmente dos tipos de inconsistencias:

- Distancias excesivamente grandes entre fotogramas consecutivos: cuando la posición de un punto variaba más de lo esperado, podía indicar un error de estimación en uno de los fotogramas, especialmente si dicha variación no se repetía en los fotogramas siguientes.
- Distancia igual a cero entre dos fotogramas: este caso sugería que los valores del segundo fotograma eran exactamente iguales a los del anterior, lo cual no es común en una secuencia de movimiento y puede deberse a una copia errónea de datos. Este comportamiento solía ser consecuencia del reemplazo aplicado durante la primera etapa del proceso.

Estas inconsistencias fueron tratadas mediante mecanismos específicos de corrección, según el tipo de error identificado:

- Corrección del primer fotograma: en los casos donde se detectó que el primer fotograma presentaba una distancia excesiva respecto al segundo, pero las distancias entre los siguientes fotogramas eran pequeñas y coherentes, se asumía que el error estaba en el primer fotograma. Por lo tanto, se reemplazaron sus coordenadas por las del segundo fotograma, asumiendo que este contenía una estimación más precisa y representativa del inicio del gesto técnico.
- Corrección por valores duplicados: cuando se identificó que un fotograma presentaba una distancia euclidiana de cero respecto al anterior, se interpretaba que los valores fueron duplicados como parte del de reemplazo en la primera parte del proceso. En este caso, se consideraba que dicho fotograma requería una corrección definitiva. Para ello, sus coordenadas fueron recalculadas como el promedio entre las coordenadas del fotograma anterior y el fotograma siguiente, con el objetivo de suavizar la transición y preservar la continuidad del movimiento.

- Corrección por variaciones extremas: si se identificaba una variación anormal entre fotogramas consecutivos, y la distancia euclidiana entre un punto en ambos fotogramas superaba un umbral, se asumía que la estimación del fotograma actual era incorrecta. En ese caso, el fotograma afectado se corrigió reemplazando sus coordenadas por el promedio entre el fotograma anterior y el siguiente. Este umbral no se estableció de forma arbitraria en la metodología, sino que se determinó posteriormente durante el análisis de los datos, tal como se describe en la sección de resultados y análisis de los mismos. De esta manera, se garantizó que el criterio para identificar variaciones anormales se fundamentó en evidencia empírica. Cabe resaltar que, como consecuencia de esta corrección, podía ocurrir que dos fotogramas consecutivos quedaran con coordenadas idénticas o muy similares. En dichos casos, la anomalía fue tratada posteriormente mediante la lógica de corrección por valores duplicados, asegurando una revisión completa en etapas sucesivas del algoritmo.

Una vez completado el proceso de extracción y corrección de los puntos clave, se procedió al cálculo de los ángulos articulares a partir de las coordenadas obtenidas. Estos ángulos se determinaron en función de la posición entre tres puntos, como se puede observar en el siguiente código:

```
def calcular_angulo(p1, p2, p3):
    a = np.array(p1) - np.array(p2)
    b = np.array(p3) - np.array(p2)
    angulo_rad = math.acos(np.dot(a, b) / (np.linalg.norm(a) * np.linalg.norm(
        b)))

    return round(np.degrees(angulo_rad), 3)
```

Código 1. Cálculo de ángulo entre tres puntos

Los ángulos extraídos fueron los siguientes:

- Vista lateral derecha:

  1. Ángulo de cadera: se definió como el ángulo formado entre el hombro derecho, la cadera derecha y la rodilla derecha.
  2. Ángulo de rodilla: se definió como el ángulo formado entre la cadera derecha, la rodilla derecha y el tobillo derecho.
  3. Ángulo de pie: se definió como el ángulo formado entre el tobillo derecho, el talón derecho y la punta del pie derecha.



Figura 4.29. Ángulos extraídos desde la vista lateral derecha.

- Vista lateral izquierda:

    1. Ángulo de inclinación: se definió como el ángulo entre el vector que une la cadera izquierda con el hombro izquierdo y el eje vertical Y. Este ángulo se calculó en el plano sagital (lateral) utilizando las coordenadas en x y y. Se consideró negativo si el hombro se encontraba por delante de la cadera (inclinación del tronco hacia adelante) y positivo si el hombro estaba por detrás de la cadera (inclinación hacia atrás).
    2. Ángulo de rodilla: se definió como el ángulo formado entre la cadera izquierda, la rodilla izquierda y el tobillo izquierdo.
    3. Ángulo de pie: se definió como el ángulo formado entre el tobillo izquierdo, el talón izquierdo y la punta del pie izquierda.



Figura 4.30. Ángulos extraídos desde la vista lateral izquierda.

- Vista trasera:

    1. Ángulo de inclinación: se calculó a partir del punto medio entre ambos hombros, el punto medio entre ambas caderas y el talón de la pierna de apoyo. Un valor positivo

indicó que el tronco estaba inclinado hacia la derecha del arquero, mientras que un valor negativo indicó una inclinación hacia la izquierda.



Figura 4.31. Ángulos extraídos desde la vista trasera.

Comparación con una herramienta de análisis biomecánico manual

Con el fin de validar la precisión y confiabilidad del algoritmo desarrollado, se llevó a cabo un análisis comparativo utilizando la herramienta Kinovea, un software de anotación de video diseñado para el análisis deportivo (Charmant, 2023). Esta herramienta permite realizar mediciones biomecánicas en dos dimensiones (2D), mediante la identificación manual de puntos articulares directamente sobre los fotogramas del video (véase la Figura 4.32).

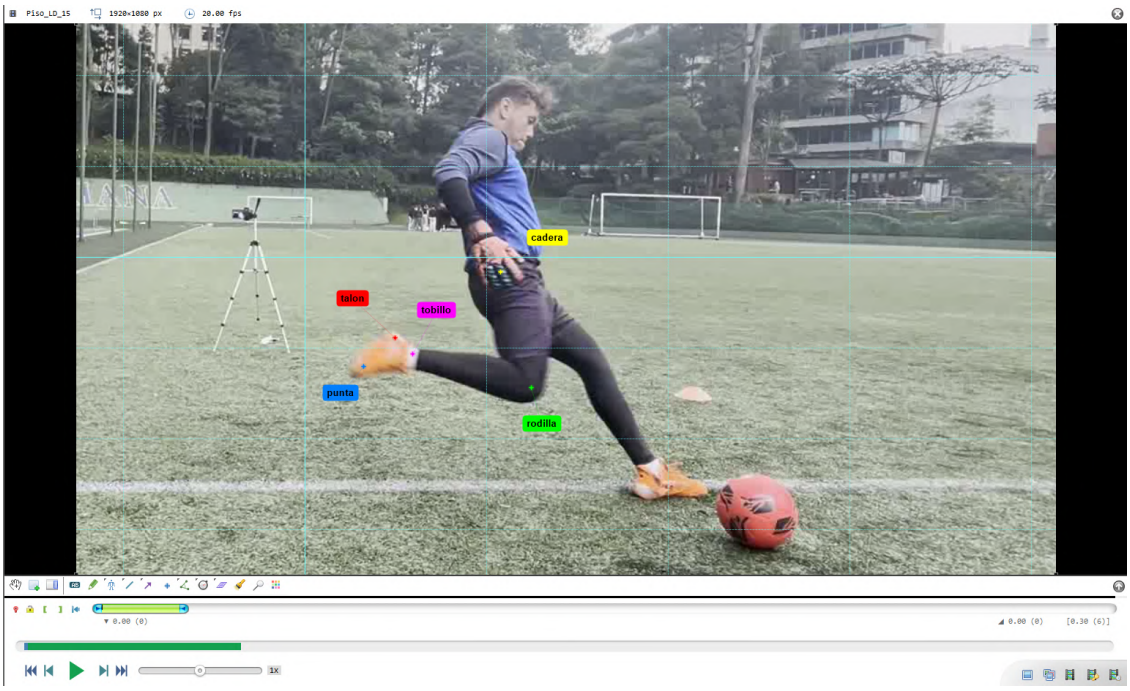


Figura 4.32. Registro manual de puntos articulares en un fotograma.

A partir de los videos previamente recortados, se realizó un análisis cuadro a cuadro en

Kinovea, identificando manualmente los puntos articulares relevantes según la vista del video. En la vista derecha, se marcaron la cadera, rodilla, tobillo, punta del pie y talón del lado derecho; en la vista izquierda, se marcaron los mismos puntos pero del lado izquierdo. Para la vista trasera, se registraron las posiciones de la cadera, rodilla y tobillo del lado de pateo. Este procedimiento se aplicó a todos los fotogramas de cada video, con el fin de obtener los puntos articulares necesarios para calcular posteriormente los ángulos.

Luego, se aplicó el algoritmo propuesto sobre cada uno de los videos recortados, obteniendo las coordenadas (x, y) de los puntos articulares previamente identificados manualmente con Kinovea en cada fotograma.

Después de extraer las coordenadas (x, y) mediante ambas herramientas, se procedió al cálculo de los ángulos articulares de interés. Cabe resaltar que, con el fin de poder comparar ambas herramientas, los ángulos a calcular fueron en dos dimensiones. En el caso de las vistas laterales (derecha e izquierda), se utilizaron los puntos de la cadera, rodilla y tobillo para calcular el ángulo de la rodilla, y los puntos del tobillo, talón y punta del pie para calcular el ángulo del pie; por parte de la vista trasera, se utilizaron los puntos de la cadera izquierda, rodilla izquierda y tobillo izquierdo del lado de apoyo para calcular el ángulo de la rodilla. Estos ángulos fueron calculados para cada fotograma tanto de Kinovea como de MediaPipe, utilizando la función mostrada en el Código 2, permitiendo una comparación detallada entre los valores obtenidos manualmente y los generados automáticamente por el algoritmo.

```
def calcular_angulo(p1, p2, p3):
    a = np.array([p1[0] - p2[0], p1[1] - p2[1]])
    b = np.array([p3[0] - p2[0], p3[1] - p2[1]])
    angulo_rad = math.acos(np.dot(a,b)/(np.linalg.norm(a)*np.linalg.norm(b)))

    return round(np.degrees(angulo_rad),3)
```

Código 2. Cálculo de ángulo entre tres puntos 2D

Seguidamente, se obtuvo la diferencia entre los ángulos de Kinovea y los ángulos calculados con MediaPipe. Esta diferencia se halló con el fin de cuantificar la precisión del algoritmo en relación con otra herramienta.

Estas diferencias fueron analizadas en cuatro etapas:

1. Análisis descriptivo: se calcularon la media, la mediana, el rango intercuartílico (IQR), y los valores mínimo y máximo de las diferencias, con el fin de describir el comportamiento de las mismas.

2. Identificación de valores atípicos: se utilizó el método del rango intercuartílico para detectar diferencias anómalas. Las observaciones fuera del intervalo $[Q1 - 1{,}5 \times IQR,\ Q3 + 1{,}5 \times$

$IQR]$ fueron etiquetadas como atípico. Se verificó manualmente si estos correspondían a errores de marcación o de predicción. Posteriormente, se repitió el análisis descriptivo sin los valores atípicos.

3. Análisis estadístico de significancia:

   - Primero se evaluó la normalidad de la distribución de diferencias mediante la prueba de Lilliefors. Las hipótesis a probar fueron:

   $$H_0 : D \sim \mathcal{N}(\mu, \sigma^2)$$

   $$H_1 : D \not\sim \mathcal{N}(\mu, \sigma^2)$$

     donde D representa la variable de diferencia entre los ángulos de Kinovea y MediaPipe.

   - Si las diferencias seguían una distribución normal, se aplicó la prueba t de una muestra para evaluar si la media era igual a cero:

   $$H_0 : \mu_D = 0$$

   $$H_1 : \mu_D \neq 0$$

     donde μ_D representa la media de las diferencias.

   - Si no seguían una distribución normal, se utilizó la prueba no paramétrica de los signos para evaluar si la mediana era igual a cero:

   $$H_0 : \tilde{D} = 0$$

   $$H_1 : \tilde{D} \neq 0$$

     donde $\tilde{D}$ representa la mediana poblacional de las diferencias.

4. Correlación entre herramientas: se calculó el coeficiente de correlación de Pearson r entre los valores angulares estimados por Kinovea y MediaPipe, para evaluar la relación lineal entre ambos métodos. Además, se graficó la dispersión de los datos, incluyendo la línea de identidad de 45° como referencia de concordancia perfecta.

Este procedimiento se replicó a cada uno de los ángulos definidos y cada vista registrada, permitiendo una comparación robusta entre las estimaciones manuales y automáticas.

Desarrollo de la interfaz

Como parte del presente trabajo, se desarrolló una interfaz web utilizando el framework Django, con el propósito de facilitar el acceso al análisis biomecánico del saque de piso por parte de entrenadores y jugadores. Esta herramienta permitió aplicar de manera práctica el algoritmo propuesto, brindando una experiencia intuitiva y accesible para el usuario final.

La Figura 4.33 muestra la vista inicial de la interfaz, donde el usuario puede cargar los videos correspondientes a las tres vistas necesarias para el análisis: lateral derecha, lateral izquierda y vista trasera.

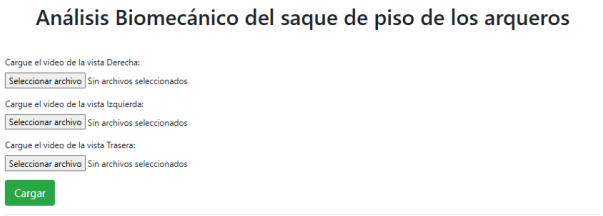


Figura 4.33. Interfaz inicial para la carga de videos desde tres vistas distintas (derecha, izquierda y trasera).

Una vez cargados los videos, el backend ejecutó automáticamente el algoritmo de análisis biomecánico planteado en este trabajo. Este proceso incluyó el recorte y la sincronización de los videos, la detección de puntos clave corporales, la corrección de posibles errores de estimación y el cálculo de los ángulos relevantes para el análisis del movimiento.

Tras el procesamiento, la interfaz presentó los tres videos con anotaciones visuales que mostraron tanto los puntos detectados como los ángulos calculados en cada fotograma (véase la Figura 4.34). Asimismo, se brindó la posibilidad de descargar un reporte en formato Excel que contenía la información numérica de los ángulos extraídos, permitiendo un análisis detallado por parte de los usuarios.



Figura 4.34. Presentación final de los videos en la interfaz.

A continuación, se presenta una vista general de la interfaz desarrollada.

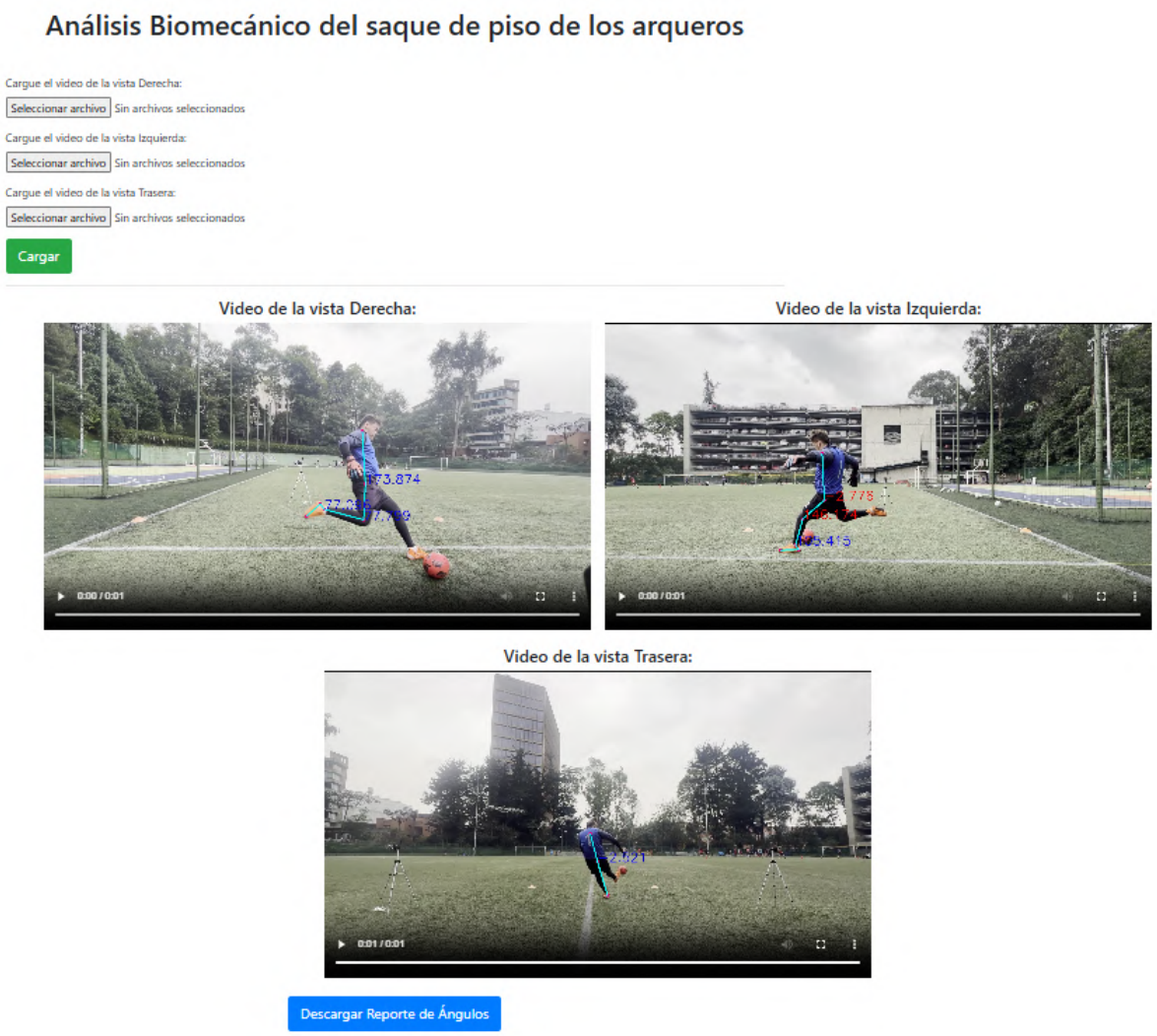


Figura 4.35. Vista general de la interfaz web desarrollada.

Por otro lado, para apoyar la interpretación del análisis, la interfaz implementada incluyó un sistema de visualización por colores, que permitió a entrenadores y jugadores identificar rápidamente el grado de adecuación de los ángulos biomecánicos observados durante la ejecución del saque. Esta funcionalidad se aplicó en los fotogramas clave del movimiento y utilizó la siguiente codificación de colores:

- Verde: el valor del ángulo se encuentra dentro del rango considerado óptimo o deseable, según la literatura consultada.
- Amarillo: el ángulo se encuentra en una zona intermedia o de advertencia, ligeramente alejado del rango óptimo.
- Rojo: el valor está fuera del rango aceptable, indicando una posible desviación técnica.

- Azul oscuro: se utilizó en los fotogramas en los que no aplica la clasificación (por ejemplo, aquellos que no corresponden al inicio o final del gesto técnico), pero aun así se visualizó el valor del ángulo.

Vista lateral derecha

A continuación, se presentan los rangos utilizados para la clasificación, apoyados en los resultados dados por la interfaz:

Rango (°)	**Clasificación**	**Color**
[140,6, 148,4]	Bueno	Verde
[136,7, 140,6)	Regular	Amarillo
(148,4, 152,3]	Regular	Amarillo
Fuera de esos rangos	Malo	Rojo
Otros fotogramas	Sin clasificar	Azul oscuro

Tabla 4.1. Clasificación del ángulo de la rodilla (momento final).

Rango (°)	**Clasificación**	**Color**
[121,1, 127,7]	Bueno	Verde
[117,8, 121,1)	Regular	Amarillo
(127,7, 130,9]	Regular	Amarillo
Fuera de esos rangos	Malo	Rojo
Otros fotogramas	Sin clasificar	Azul oscuro

Tabla 4.2. Clasificación del ángulo de la cadera (momento final).



Figura 4.36. Clasificación resultante de la interfaz en la vista lateral derecha.

Vista lateral izquierda



Figura 4.37. Clasificación resultante de la interfaz en la vista lateral izquierda para el momento inicial.



Figura 4.38. Clasificación resultante de la interfaz en la vista lateral izquierda para el momento final.

Tabla 4.3. Ángulo de la rodilla (momento inicial).

Rango (°)	**Clasificación**	**Color**
$[152{,}7, 161{,}2]$	Bueno	Verde
$[148{,}5, 152{,}7)$	Regular	Amarillo
$(161{,}2, 165{,}5]$	Regular	Amarillo
Fuera de esos rangos	Malo	Rojo
Otros fotogramas	Sin clasificar	Azul oscuro

Tabla 4.4. Ángulo de la rodilla (momento final).

Rango (°)	**Clasificación**	**Color**
$[130{,}35, 146{,}65]$	Bueno	Verde
$[122{,}2, 130{,}35)$	Regular	Amarillo
$(146{,}65, 154{,}8]$	Regular	Amarillo
Fuera de esos rangos	Malo	Rojo
Otros fotogramas	Sin clasificar	Azul oscuro

Tabla 4.5. Ángulo del tronco vertical (hombro–cadera).

Momento	**Rango (°)**	**Color**
Inicial	$[-13{,}2, -7{,}8]$	Verde
	$[-15{,}9, -13{,}2)$	Amarillo
	$[-7{,}8, -5{,}1]$	Amarillo
	Fuera de esos rangos	Rojo
Final	$[3{,}5, 13{,}5]$	Verde
	$[-1, 3{,}5)$	Amarillo
	$(13{,}5, 18]$	Amarillo
	Fuera de esos rangos	Rojo
	Otros fotogramas	Azul oscuro

Vista trasera



Figura 4.39. Clasificación resultante de la interfaz en la vista trasera para el momento final.

Tabla 4.6. Ángulo de inclinación del tronco (momento final).

Rango (°)	**Clasificación**	**Color**
$[1{,}5, 11{,}5]$	Bueno	Verde
$[-3{,}5, 1{,}5)$	Regular	Amarillo
$(11{,}5, 16{,}5]$	Regular	Amarillo
Fuera de esos rangos	Malo	Rojo
Otros fotogramas	Sin clasificar	Azul oscuro

Resultados y análisis de resultados

Selección de los umbrales

Con el fin de definir el umbral para la corrección automática de errores en la estimación de puntos clave, se llevó a cabo un análisis exploratorio de las distancias euclidianas entre puntos anatómicos en fotogramas consecutivos. Este análisis permitió establecer umbrales empíricos que distinguen entre el movimiento natural del cuerpo y las variaciones anómalas que requieren corrección. Este proceso se realizó con cada una de las vistas (lateral derecha, lateral izquierda y trasera), pues el movimiento y los puntos analizados son diferentes para cada una de ellas.

A continuación se muestra un resumen de los umbrales definidos, los cuales serán justificados con detalle en esta sección.

Marcador	Umbral (pixeles)
Cadera derecha	22
Rodilla derecha	60
Tobillo derecho	130
Talón derecho	150
Punta derecha	150
Hombro derecho	18

Tabla 5.1. Umbrales por marcador – Vista Lateral Derecha

Marcador	Umbral (pixeles)
Cadera izquierda	30
Rodilla izquierda	25
Tobillo izquierdo	15
Talón izquierdo	15
Punta izquierda	15
Hombro izquierdo	30

Tabla 5.2. Umbrales por marcador – Vista Lateral Izquierda

Marcador	Umbral (pixeles)
Hombro derecho	13
Cadera izquierda	10
Talón izquierdo	10
Hombro izquierdo	13
Cadera derecha	10
Rodilla izquierda	18

Tabla 5.3. Umbrales por marcador – Vista Trasera

Vista lateral derecha

Se identificó que la vista lateral derecha, correspondiente al perfil de la pierna de golpeo, presenta las mayores variaciones espaciales entre fotogramas consecutivos. Esta vista permite observar con mayor detalle la trayectoria de la pierna de pateo durante la fase de golpeo, la cual se caracteriza por un desplazamiento rápido y amplio, especialmente en los segmentos distales como el tobillo y el pie.

Como se muestra en la Figura 5.1, la amplitud del movimiento de la pierna derecha genera desplazamientos significativos de los marcadores correspondientes a la cadera, la rodilla, el tobillo y el pie. En consecuencia, los valores de distancia euclidiana obtenidos en esta vista tienden a ser más elevados en comparación con otras perspectivas. La Figura 5.2 presenta las distancias euclidianas entre fotogramas consecutivos correspondientes a la secuencia de movimiento mostrada. Los

resultados confirman que las articulaciones con mayor desplazamiento son las del pie, seguidas por la rodilla, la cadera y, en menor medida, el hombro.

Frames	cadera_derecha	hombro_derecho	rodilla_derecha	tobillo_derecho	talon_derecho	punta_derecha
0-1	11.392	9.765	45.948	56.100	44.265	96.041
1-2	10.916	13.890	50.363	89.390	86.130	132.473
2-3	16.566	4.510	20.341	107.172	122.625	150.562
3-4	12.028	8.052	15.313	112.523	138.102	139.593
4-5	1.502	5.512	12.393	77.009	88.315	91.787
5-6	0.728	5.712	3.990	49.766	55.923	64.624

Figura 5.2. Tabla de distancias euclidianas entre fotogramas para la vista sagital derecha.

Por lo anterior, se realizó un análisis individual por articulación, con el fin de establecer umbrales específicos de variación para cada una de ellas. Este enfoque permite adaptar los criterios de corrección a las características particulares del movimiento en cada punto anatómico. De forma general, se consideraron como candidatos a corrección aquellos valores que se identificaron como atípicos extremos dentro de la distribución de distancias euclidianas. No obstante, la decisión final no se basó únicamente en criterios estadísticos: cada uno de estos valores fue contrastado directamente con la secuencia de video original para determinar si correspondía a un error de estimación o si, por el contrario, formaba parte de un movimiento natural esperado. La presencia de estos valores atípicos sugiere estimaciones que no corresponden al movimiento real y, por tanto, pueden ser indicativas de errores en la detección de pose.

Cadera derecha Se observó una dispersión moderada en la distribución de las distancias y varios valores clasificados como atípicos por el boxplot (Figuras 5.4 y 5.3). Sin embargo, tras su revisión visual, se concluyó que muchas de estas variaciones corresponden a movimientos reales del gesto técnico. Se estableció un umbral de 22 píxeles como límite de corrección.

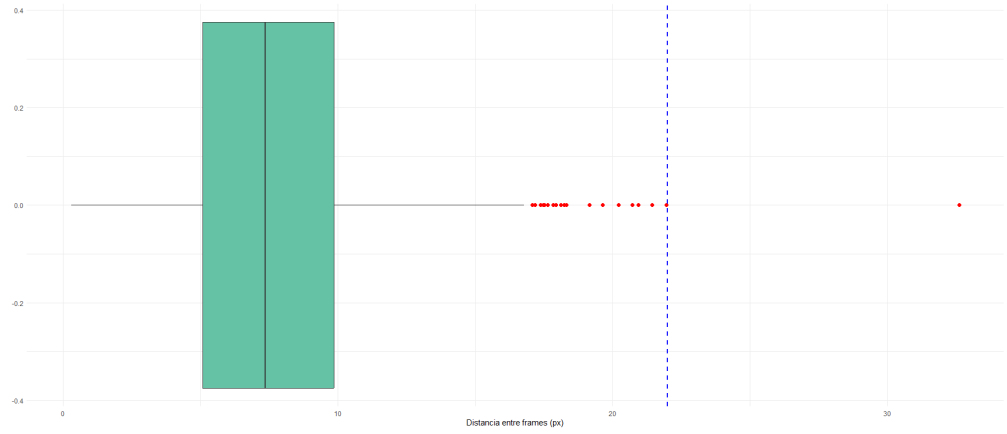


Figura 5.3. Boxplot de las distancias euclidianas entre fotogramas consecutivos para el marcador de la cadera derecha.

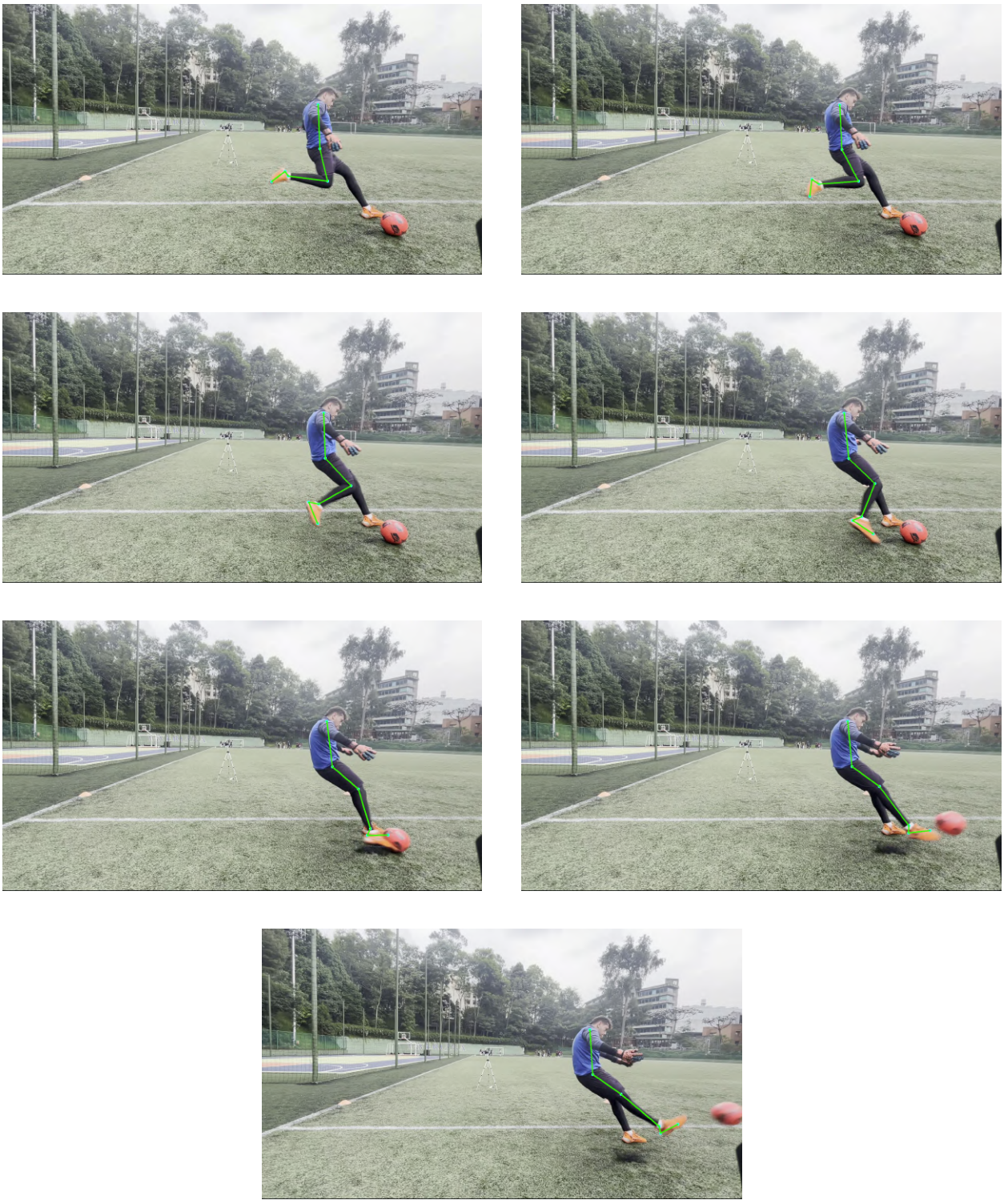


Figura 5.1. Secuencia de 7 fotogramas que describen el gesto técnico desde la vista sagital derecha.

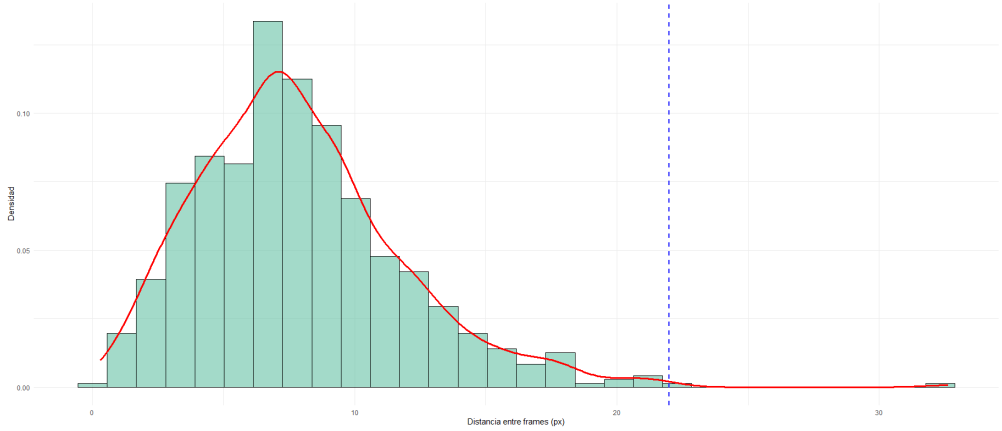


Figura 5.4. Distribución de densidad de las distancias euclidianas para el marcador de la cadera derecha.

Hombro derecho El análisis (Figuras 5.5 y 5.6) mostró dos valores atípicos. La validación visual reveló que uno de ellos era un error de estimación, mientras que el otro se asociaba a una variación legítima del movimiento. Se definió un umbral de 18 píxeles.

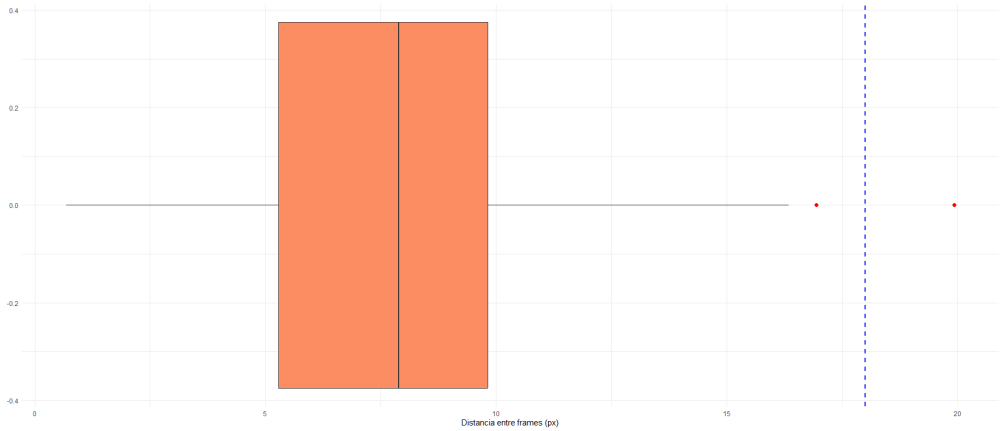


Figura 5.5. Boxplot de las distancias euclidianas entre fotogramas consecutivos para el marcador del hombro derecho.

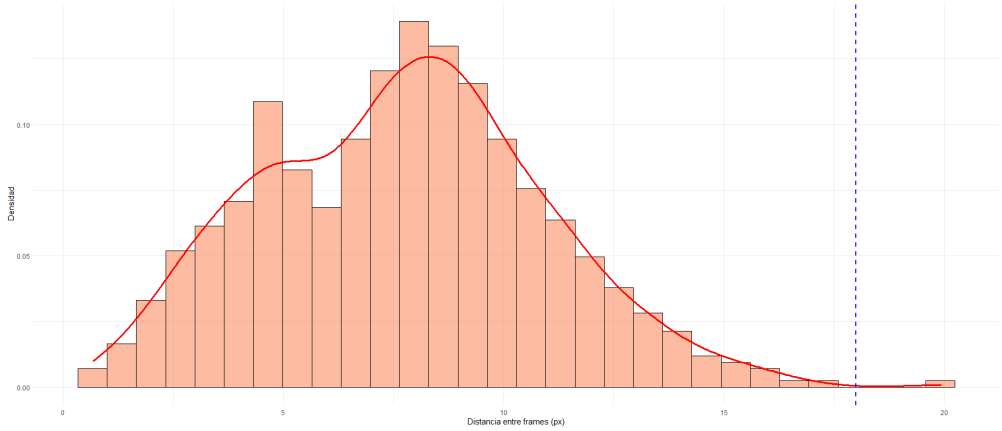


Figura 5.6. Distribución de densidad de las distancias euclidianas para el marcador del hombro derecho.

Rodilla derecha En este caso, los valores atípicos observados sí reflejan errores en la estimación del modelo, confirmados mediante revisión de los fotogramas. Se estableció un umbral de 60 píxeles (Figuras 5.7 y 5.8).

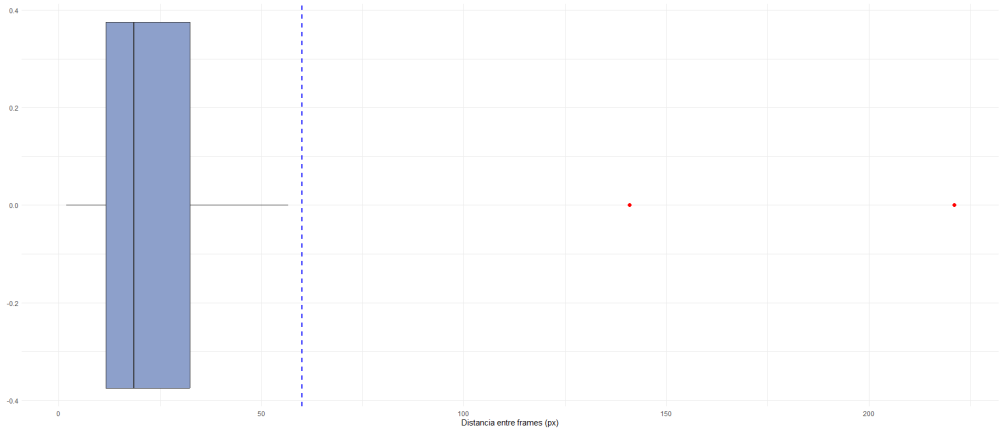


Figura 5.7. Boxplot de las distancias euclidianas entre fotogramas consecutivos para el marcador de la rodilla derecha.

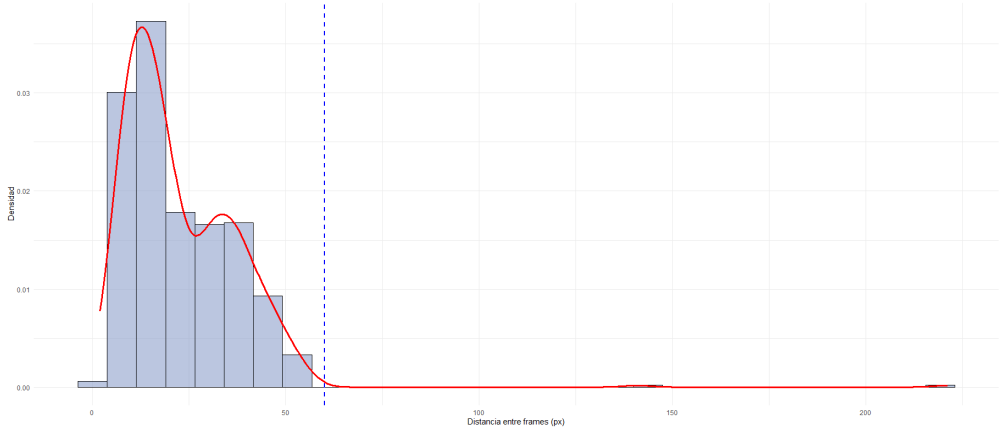


Figura 5.8. Distribución de densidad de las distancias euclidianas para el marcador de la rodilla derecha.

Talón derecho Las Figuras 5.9 y 5.10 muestran una amplia dispersión en las distancias. Sin embargo, gran parte de estos desplazamientos corresponden a movimientos reales y no a errores, por lo que se fijó un umbral alto de 150 píxeles.

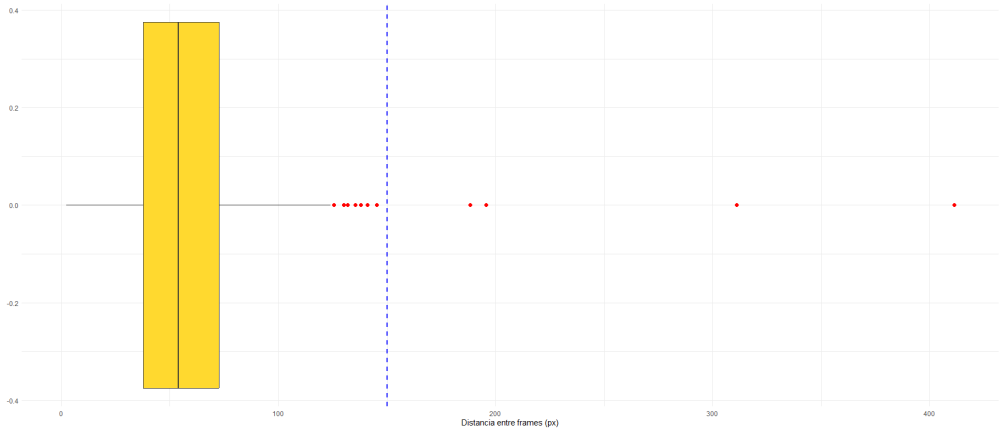


Figura 5.9. Boxplot de las distancias euclidianas entre fotogramas consecutivos para el marcador del talón derecho.

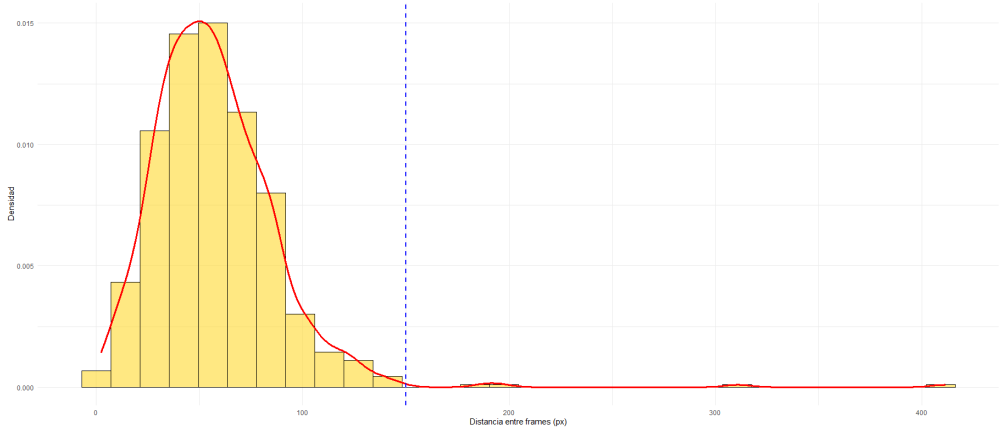


Figura 5.10. Distribución de densidad de las distancias euclidianas para el marcador del talón derecho.

Tobillo derecho El análisis reveló que, a pesar de la presencia de múltiples valores atípicos (Figuras 5.11 y 5.12), muchos de ellos se relacionan con el movimiento dinámico natural del tobillo. Se adoptó un umbral de 130 píxeles.

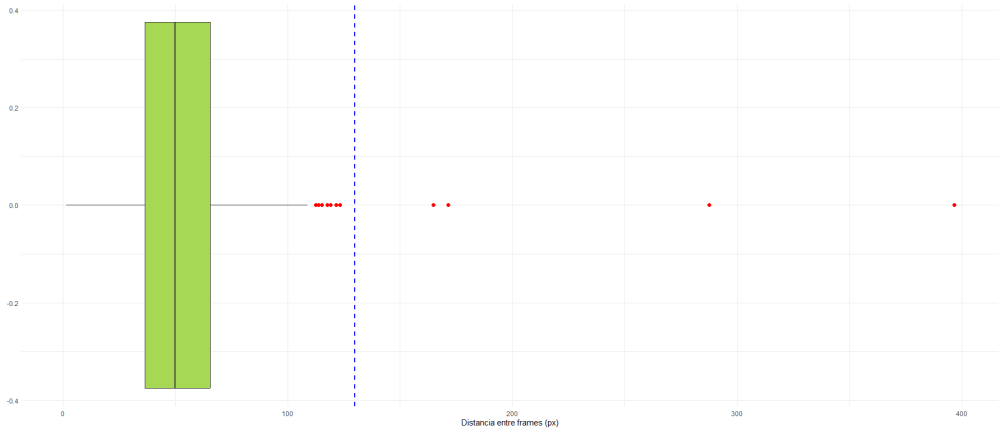


Figura 5.11. Boxplot de las distancias euclidianas entre fotogramas consecutivos para el marcador del tobillo derecho.

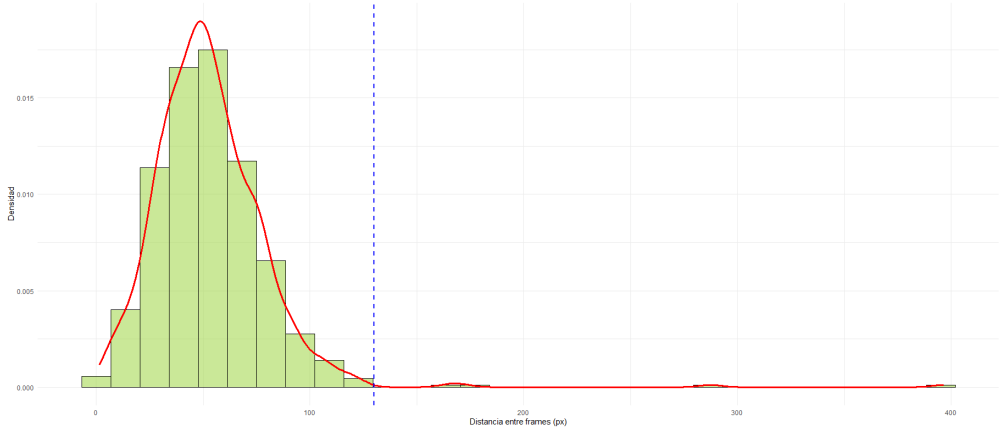


Figura 5.12. Distribución de densidad de las distancias euclidianas para el marcador del tobillo derecho.

Punta del pie derecha Se observó una alta variabilidad en las distancias entre fotogramas (Figuras 5.13 y 5.14). Tras la verificación manual, se concluyó que la mayoría de los valores atípicos no representaban errores, por lo que se estableció un umbral de 150 píxeles.

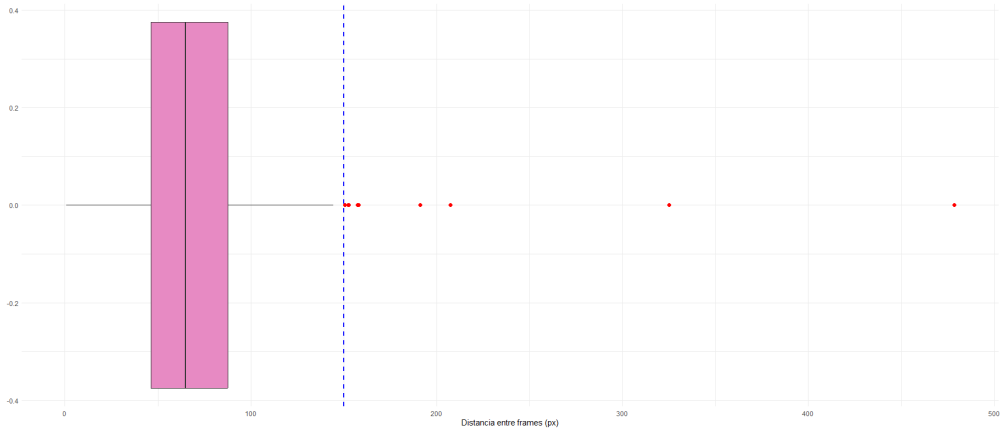


Figura 5.13. Boxplot de las distancias euclidianas entre fotogramas consecutivos para el marcador de la punta del pie derecha.

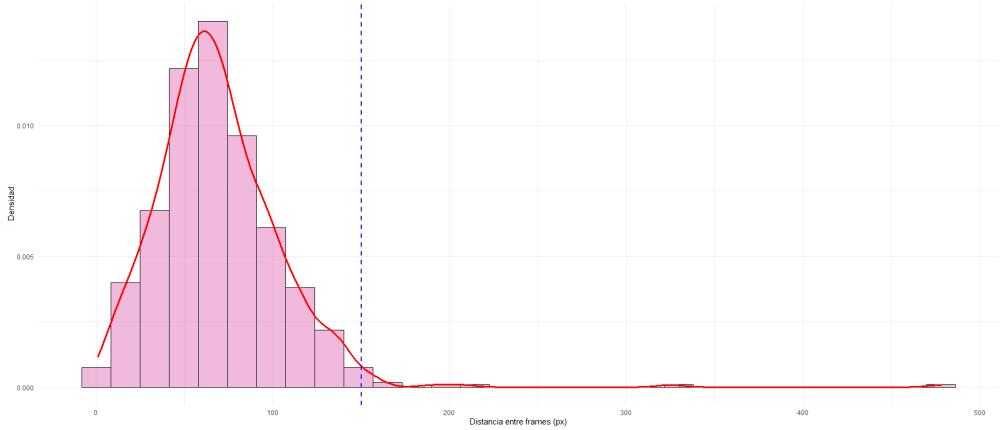


Figura 5.14. Distribución de densidad de las distancias euclidianas para el marcador de la punta del pie derecha.

Vista lateral izquierda

Se identificó que la vista lateral izquierda, correspondiente al perfil de la pierna de apoyo, presentó variaciones más moderadas entre fotogramas consecutivos. Esta vista permite observar con mayor claridad la estabilidad del gesto técnico, ya que la pierna de apoyo mantiene una posición relativamente constante durante la fase de golpeo.

Como se muestra en la Figura 5.15, los desplazamientos de los marcadores son menores en comparación con los observados en la pierna de golpeo. En consecuencia, los valores de distancia euclidiana obtenidos en esta vista tienden a ser más bajos que los registrados desde otras perspectivas. La Figura 5.16 presenta las distancias euclidianas entre fotogramas consecutivos correspondientes a la secuencia de movimiento mostrada. Los resultados confirman que los desplazamientos más relevantes se concentran en la articulación de la cadera, seguidos por la rodilla y el hombro, mientras que el pie mantiene una trayectoria prácticamente estable.

Frames	cadera_izquierda	hombro_izquierdo	rodilla_izquierda	tobillo_izquierdo	talon_izquierdo	punta_izquierda
0-1	12.875	9.689	5.502	0.863	0.540	6.022
1-2	18.411	22.867	15.571	0.267	2.217	5.183
2-3	22.657	13.993	14.782	0.451	1.312	0.094
3-4	14.371	12.277	7.925	0.130	0.545	0.520
4-5	16.882	18.214	6.067	0.504	1.632	0.790
5-6	13.146	23.732	15.193	2.041	1.341	3.219
5-7	17.042	18.131	6.873	1.862	4.254	1.666
5-8	9.944	14.998	6.469	2.230	2.709	1.365

Figura 5.16. Tabla de distancias euclidianas entre fotogramas para la vista sagital izquierda.

Cadera izquierda Se observó una dispersión moderada en la distribución de las distancias y varios valores clasificados como atípicos por el boxplot (Figuras 5.18 y 5.17). Sin embargo, tras su revisión visual, se concluyó que muchas de estas variaciones corresponden a movimientos reales. Se estableció un umbral de 30 píxeles como límite de corrección.

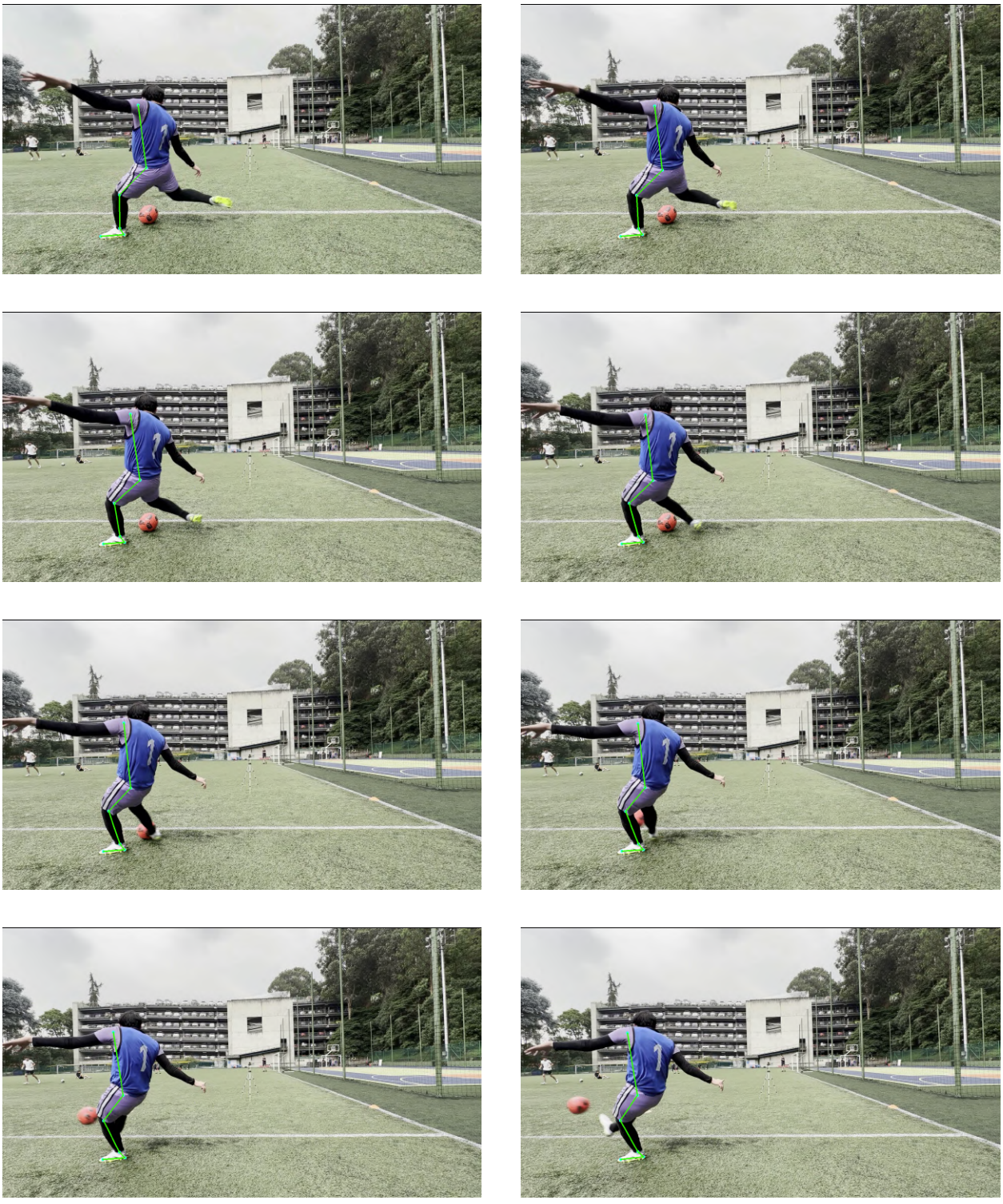


Figura 5.15. Secuencia de 8 fotogramas que describen el gesto técnico desde la vista sagital izquierda.

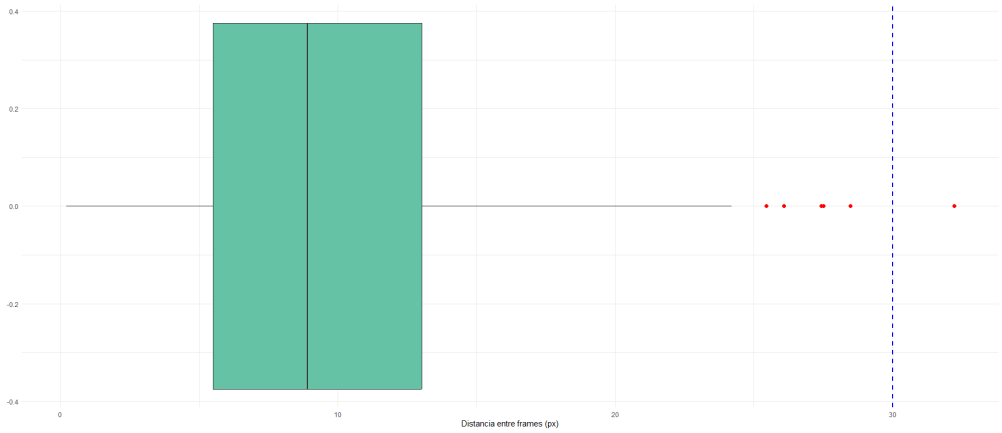


Figura 5.17. Boxplot de las distancias euclidianas entre fotogramas consecutivos para el marcador de la cadera izquierda.

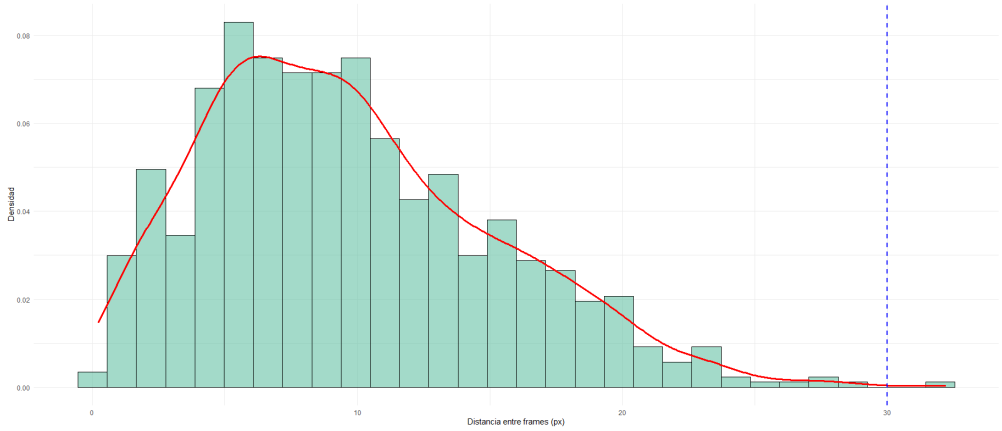


Figura 5.18. Distribución de densidad de las distancias euclidianas para el marcador de la cadera izquierda.

Hombro izquierdo Al igual que en la cadera izquierda, la dispersión es moderada pero con varios valores atípicos (Figuras 5.19 y 5.20). Sin embargo, tras su revisión visual, se concluyó que muchas de estas variaciones corresponden a movimientos reales. Se estableció un umbral de 30 píxeles como límite de corrección.

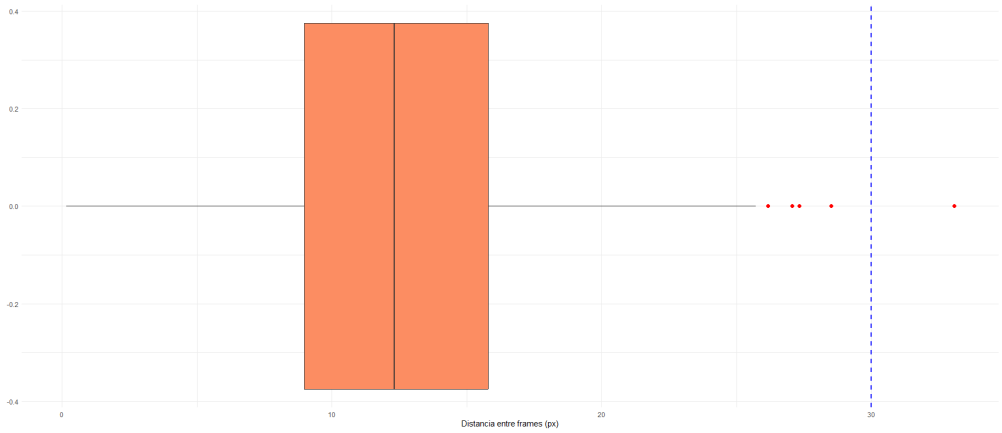


Figura 5.19. Boxplot de las distancias euclidianas entre fotogramas consecutivos para el marcador del hombro izquierdo.

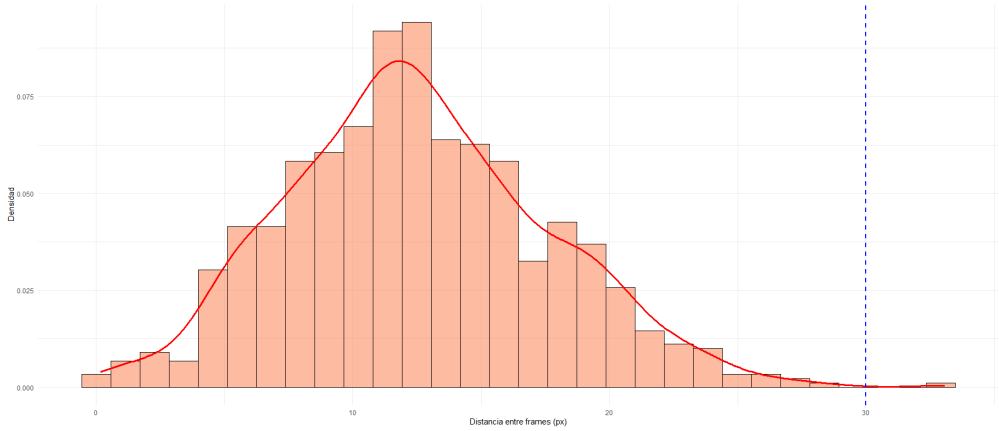


Figura 5.20. Distribución de densidad de las distancias euclidianas para el marcador del hombro derecho.

Rodilla izquierda La rodilla izquierda presentó un movimiento más moderado que el de la cadera izquierda y el hombro izquierdo. Sin embargo, siguen existiendo valores atípicos como los muestra las Figuras 5.21 y 5.22. Al momento de hacer una revisión exhaustiva, se concluyó que las distancias con más de 25 píxeles no representan el movimiento natural de la rodilla, por eso se estableció un umbral de 25 píxeles.

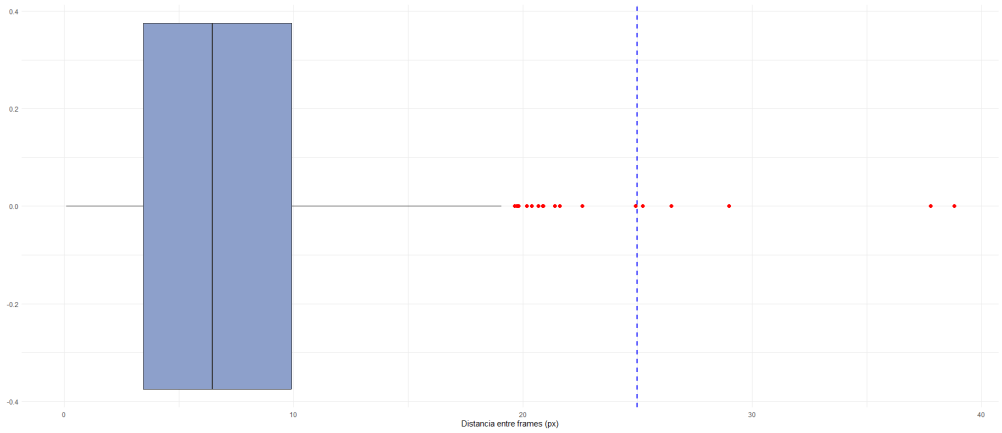


Figura 5.21. Boxplot de las distancias euclidianas entre fotogramas consecutivos para el marcador de la rodilla izquierda.

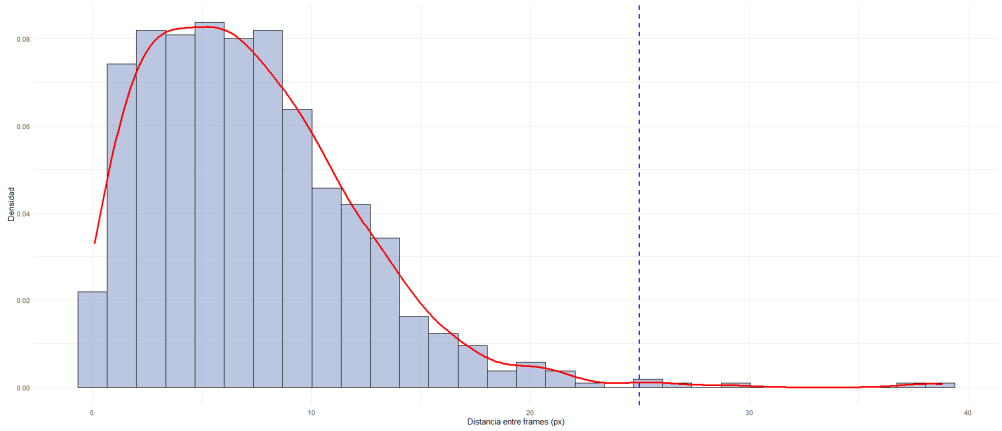


Figura 5.22. Distribución de densidad de las distancias euclidianas para el marcador de la rodilla izquierda.

Pie izquierdo Dado que la pierna izquierda actúa como punto de apoyo durante el gesto técnico, se observa que las articulaciones del pie (tobillo, talón y punta del pie) presentan un comportamiento muy similar. Las distribuciones de distancias euclidianas para estos tres marcadores son estrechas y con valores bajos (véase la Figura 5.26), lo que indica un movimiento mínimo entre fotogramas consecutivos. Además, la revisión visual de la secuencia confirma que estos puntos permanecen prácticamente fijos a lo largo del gesto.

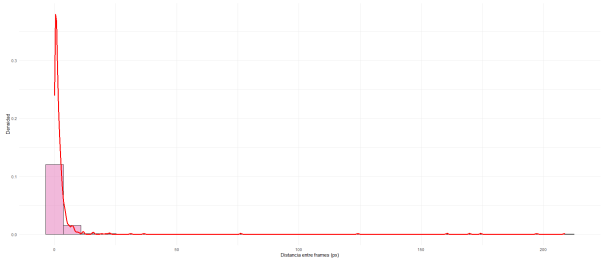


Figura 5.23. Punta izquierda

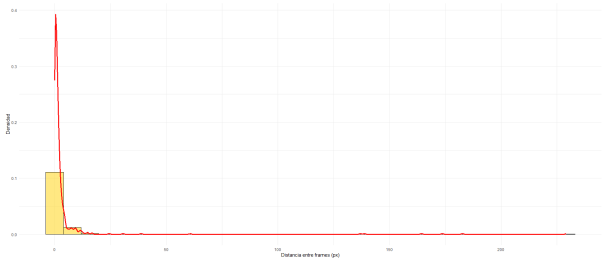


Figura 5.24. Talon izquierdo

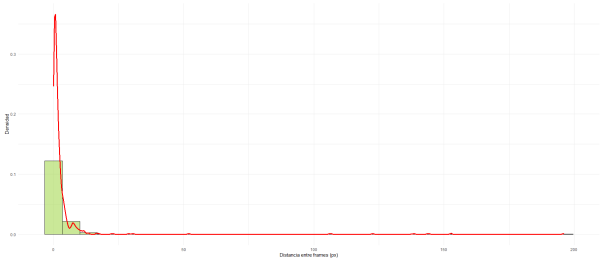


Figura 5.25. Tobillo izquierdo

Figura 5.26. Distribuciones de distancias euclideanas para la punta izquierda, el talón y el tobillo izquierdo, desde la vista sagital izquierda.

Por esta razón, se optó por agrupar estos tres marcadores en un único segmento de análisis, ya que sus patrones de variación no difieren significativamente.

El la Figura 5.27 logramos ver que hay bastantes valores atípicos. Sin embargo, al momento de analizarlos detalladamente, vemos que las distancias menores a 15 representan un movimiento válido dentro del gesto de apoyo. Por este motivo, se establece un umbral de 15 para el tobillo, el talón y la punta izquierda.

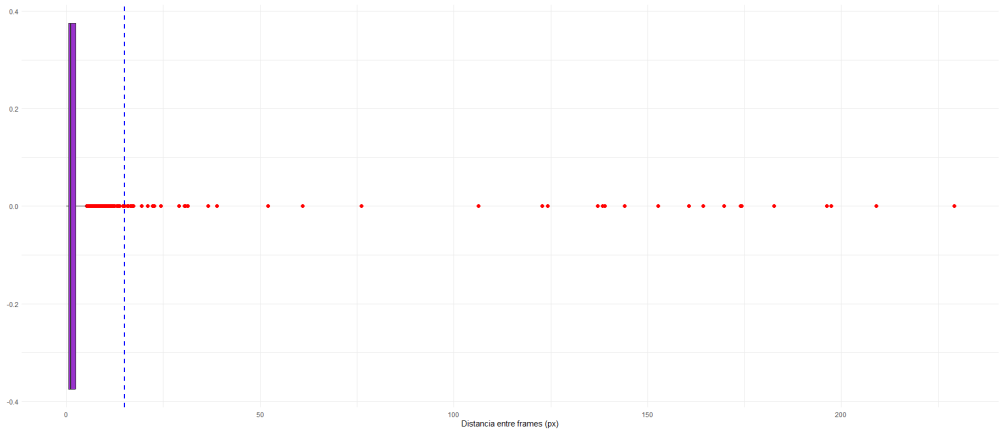


Figura 5.27. Boxplot de las distancias del pie izquierdo.

Vista trasera

Se identificó que la vista trasera, en la cual se analiza la inclinación con respecto a la pierna de apoyo, presenta variaciones más moderadas entre fotogramas consecutivos. Esta vista permite observar con mayor claridad la estabilidad del gesto técnico, ya que la pierna de apoyo mantiene una posición relativamente constante durante la fase de golpeo.

Como se muestra en la Figura 5.28, los desplazamientos de los marcadores son mínimos al tener como referencia la pierna de apoyo. La Figura 5.29 presenta las distancias euclidianas entre fotogramas consecutivos correspondientes a la secuencia de movimiento mostrada. Los resultados confirman que los desplazamientos son estables durante todo el movimiento.

Frames	cadera_derecha	cadera_izquierda	hombro_derecho	hombro_izquierdo	rodilla_izquierda	talon_izquierdo
0-1	1.698	2.325	1.639	2.169	2.039	0.958
1-2	0.538	1.091	2.177	2.464	1.898	1.861
2-3	1.093	0.418	2.691	1.524	0.266	4.483
3-4	4.371	3.063	1.885	2.039	3.742	1.145
4-5	2.495	2.850	1.493	3.647	0.558	0.509
5-6	2.229	2.580	3.594	3.534	2.955	0.709
6-7	5.108	9.078	2.800	4.962	1.693	1.379
7-8	5.641	5.387	1.175	3.895	1.167	0.137

Figura 5.29. Tabla de distancias euclidianas entre fotogramas para la vista coronal.

Caderas Desde la vista trasera se logra observar que ambas caderas presentan un movimiento muy similar. Las distribuciones de distancias euclidianas para estos dos marcadores son estrechas y con valores bajos (véase la Figura 5.32), lo que indica un movimiento mínimo entre fotogramas consecutivos. Además, la revisión visual de la secuencia confirma que estos puntos no presentan un movimiento pronunciado.

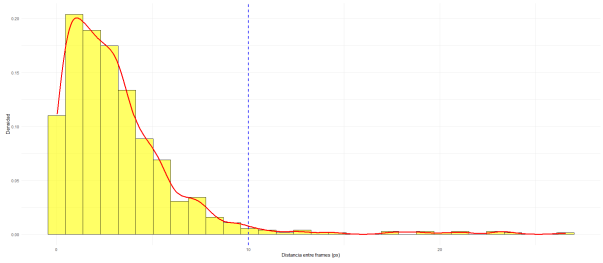


Figura 5.30. Cadera izquierda

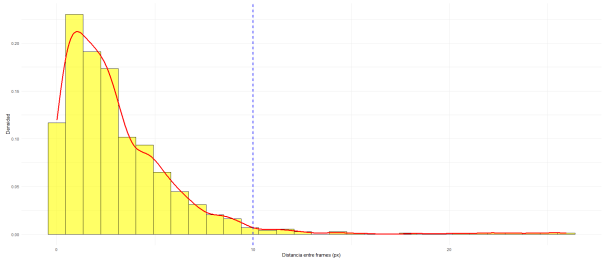


Figura 5.31. Cadera derecha

Figura 5.32. Distribuciones de distancias euclideanas para la cadera derecha e izquierda, desde la vista coronal.

Por esta razón, se optó por agrupar estos dos marcadores en un único segmento de análisis, ya que sus patrones de variación no difieren significativamente.



Figura 5.28. Secuencia de 8 fotogramas que describen el gesto técnico desde la vista trasera.

El la Figura 5.33 logramos ver que hay bastantes valores atípicos. Sin embargo, al momento de analizarlos detalladamente, vemos que las distancias menores a 10 representan un movimiento válido dentro de la inclinación del cuerpo. Por este motivo, se establece un umbral de 10 para la cadera derecha e izquierda en la vista coronal.

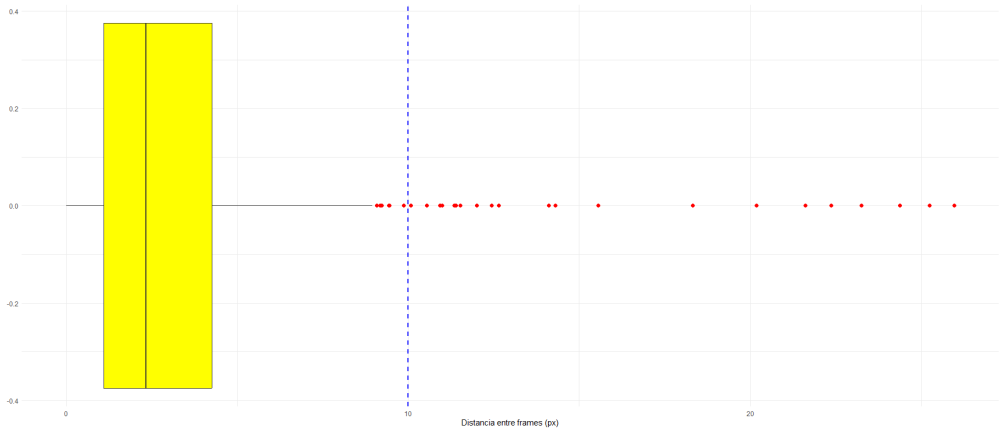


Figura 5.33. Boxplot de las distancias las caderas desde la vista coronal.

Hombros Desde la vista trasera se logra observar que ambos hombros presentan un movimiento muy similar. Las distribuciones de distancias euclidianas para estos dos marcadores son estrechas y con valores bajos (véase la Figura 5.36), lo que indica un movimiento mínimo entre fotogramas consecutivos. Además, la revisión visual de la secuencia confirma que estos puntos no presentan un movimiento pronunciado.

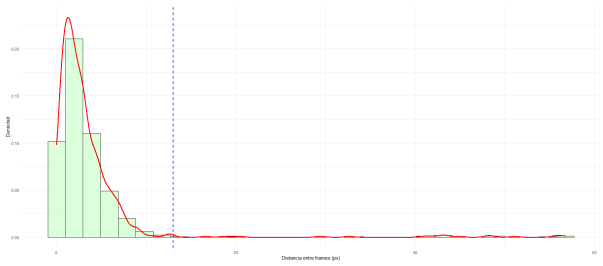


Figura 5.34. Hombro derecho

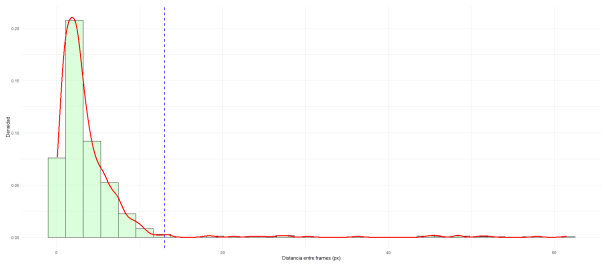


Figura 5.35. Hombro izquierdo

Figura 5.36. Distribuciones de distancias euclidianas para el hombro derecho e izquierdo desde la vista coronal.

Por esta razón, se optó por agrupar estos dos marcadores en un único segmento de análisis, ya que sus patrones de variación no difieren significativamente.

El la Figura 5.37 logramos ver que hay bastantes valores atípicos. Sin embargo, al momento de analizarlos detalladamente, vemos que las distancias menores a 13 representan un movimiento

válido dentro de la inclinación del cuerpo. Por este motivo, se establece un umbral de 13 para el hombro derecho e izquierdo.

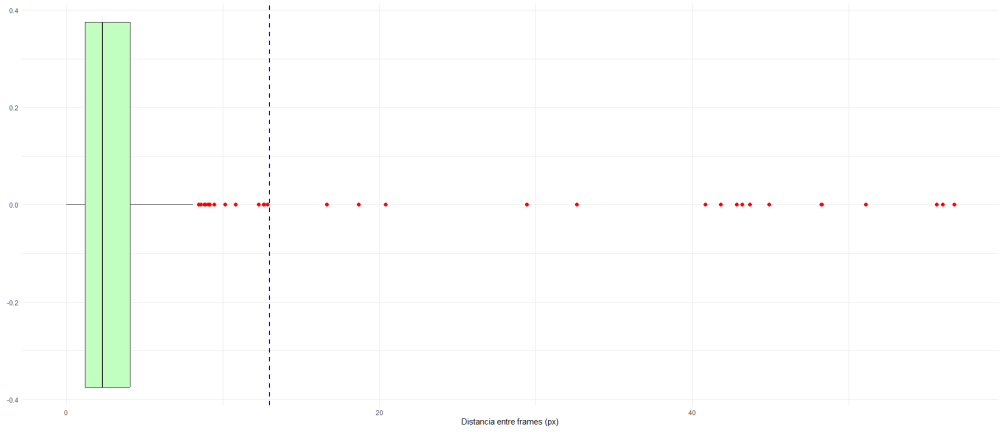


Figura 5.37. Boxplot de las distancias de los hombros desde la vista coronal.

Talón Desde la vista trasera se analiza el talón del pie de apoyo. En este caso, esta articulación presenta un movimiento casi nulo, en los únicos momentos en que cambia su distancia es al final, cuando el gesto finaliza los arqueros suelen elevar ligeramente su talón y apoyarse en la punta. Es por esto que la distribución de distancias para este marcador es bastante estrecha (véase la Figura 5.38), lo que indica un movimiento mínimo entre fotogramas consecutivos. Además, la revisión visual de la secuencia confirma que estos puntos no presentan un movimiento pronunciado.

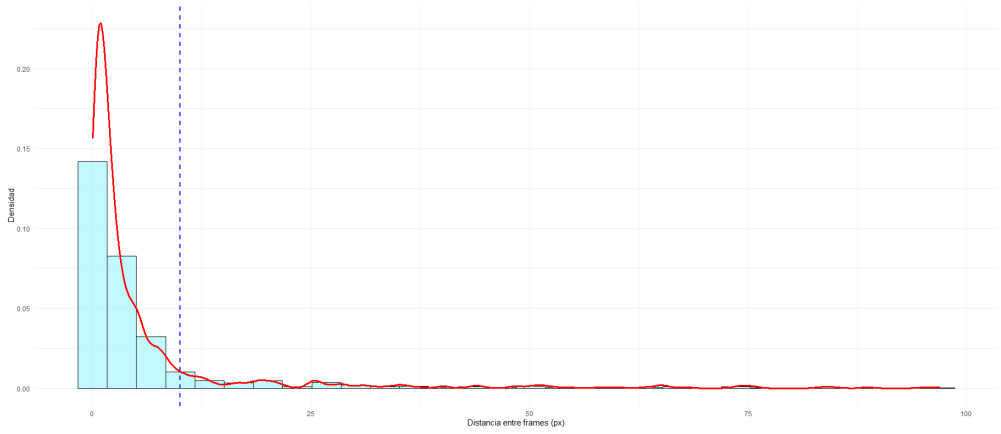


Figura 5.38. Distribuciones de distancias euclidianas para el talón izquierdo, desde la vista coronal.

El la Figura 5.39 logramos ver que hay bastantes valores atípicos. Sin embargo, al momento de analizarlos detalladamente, vemos que las distancias menores a 10 representan un movimiento válido. Por este motivo, se establece un umbral de 10 para el talón derecho en la vista coronal.

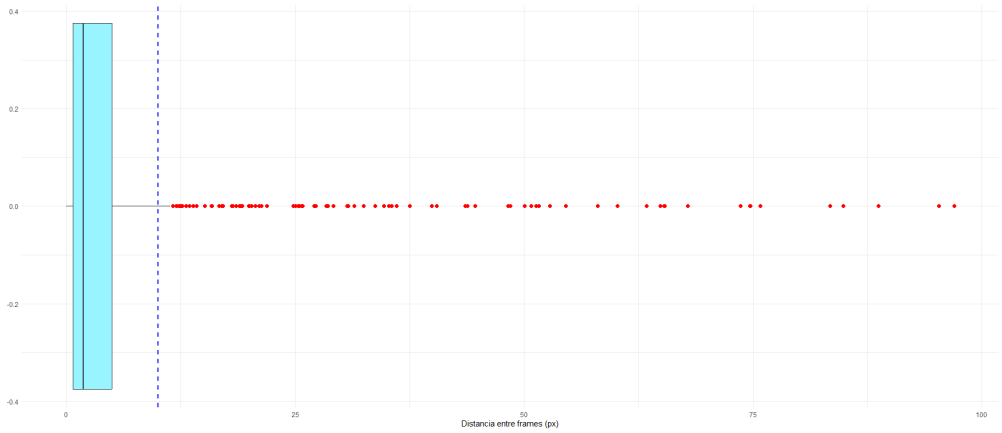


Figura 5.39. Boxplot de las distancias del talón desde la vista coronal.

Rodilla Desde la vista trasera se analiza la rodilla del pie de apoyo. En este caso, esta articulación presenta un movimiento moderado; sin embargo, hay arqueros que realizan un movimiento de rodilla muy horizontal y como consecuencia las distancias no se concentran del todo en un rango inferior. La distribución de distancias euclidianas, aunque presenta valores bajos, no es igual de estrecha a las distribuciones anteriormente mencionadas (véase la Figura 5.40), lo que indica un movimiento si existe un movimiento moderado.

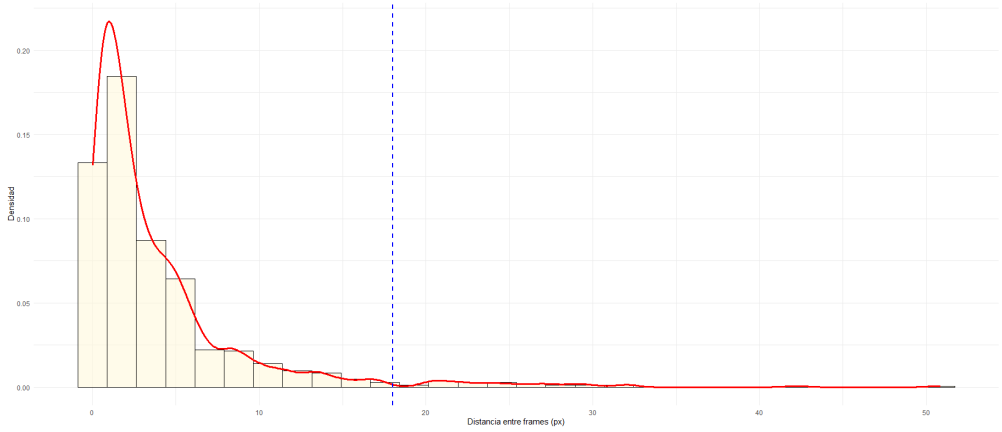


Figura 5.40. Distribución de distancias euclidianas para la rodilla izquierda, desde la vista coronal.

El la Figura 5.41 logramos ver que hay bastantes valores atípicos. Sin embargo, al momento de analizarlos detalladamente, vemos que las distancias menores a 18 representan un movimiento válido. Por este motivo, se establece un umbral de 18 para la rodilla de apoyo en la vista coronal.

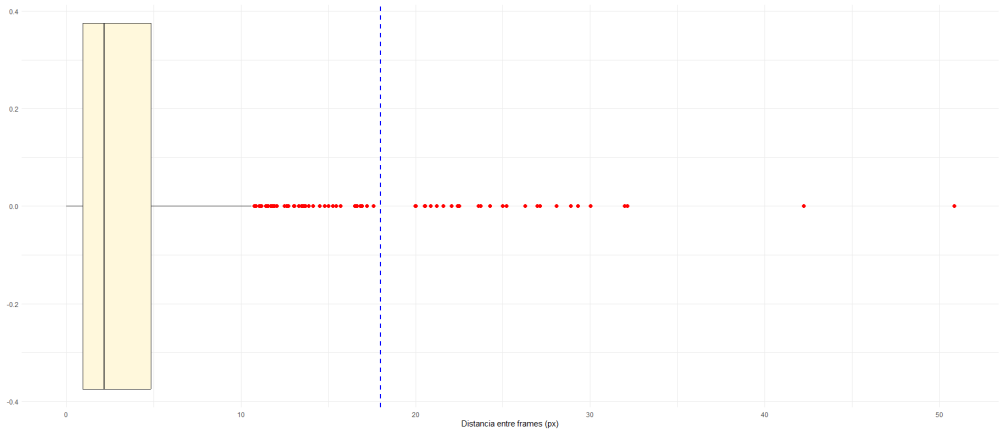


Figura 5.41. Boxplot de las distancias de la rodilla izquierda desde la vista coronal.

Evaluación de consistencia y precisión del algoritmo

La evaluación de la consistencia y precisión del algoritmo se realizó a través de un análisis visual y cuantitativo para las tres vistas del movimiento. En primer lugar, se compararon las estimaciones de los puntos articulares generados por Kinovea y MediaPipe en las tres vistas del cuerpo (lateral derecha, lateral izquierda y trasera) de manera visual.

A continuación se presenta la comparación visual entre los puntos articulares anotados manualmente en Kinovea y los estimados automáticamente por MediaPipe para las tres vistas del cuerpo: lateral derecha, lateral izquierda y trasera (véanse las Figuras 5.42,5.43 y5.44, respectivamente). En estas imágenes, los puntos de Kinovea se representan en azul, mientras que los de MediaPipe se muestran en rojo.

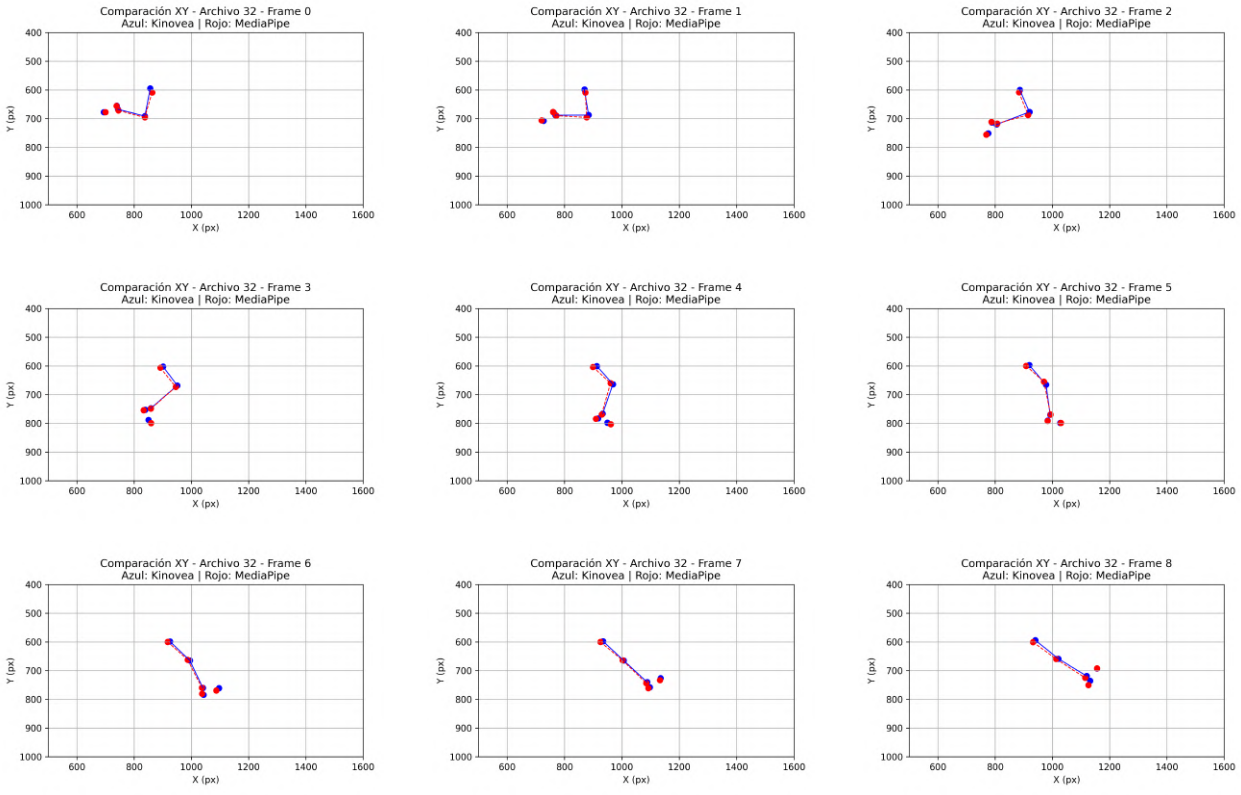


Figura 5.42. Comparación visual entre Kinovea (azul) y MediaPipe (rojo) en los primeros 9 fotogramas del archivo 32, vista lateral derecha.

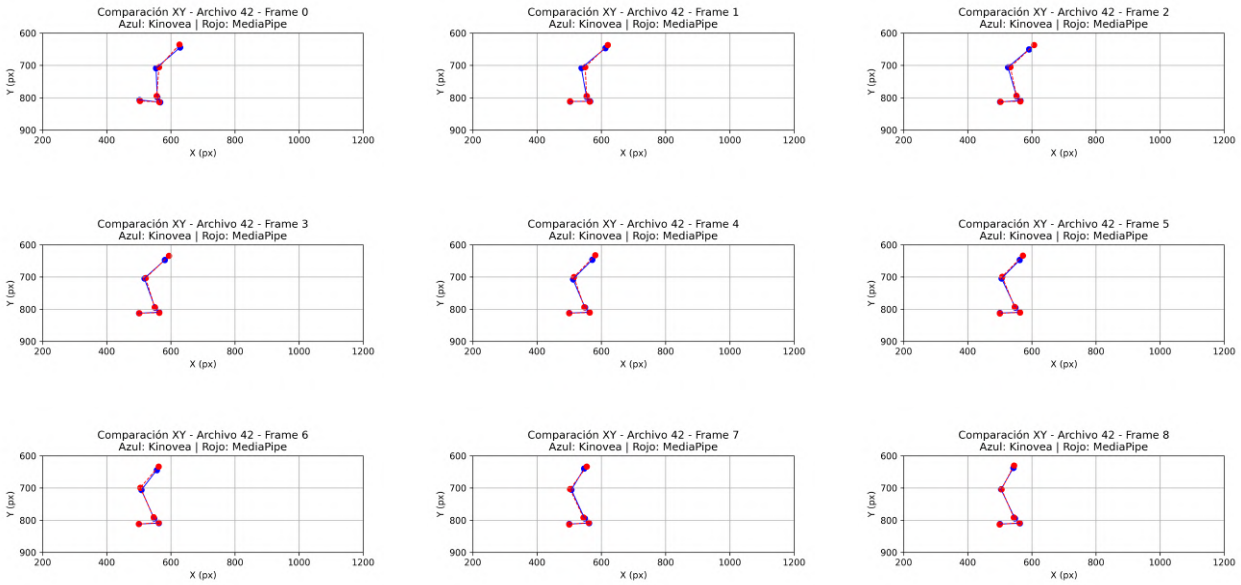


Figura 5.43. Comparación visual entre Kinovea (azul) y MediaPipe (rojo) en los primeros 9 fotogramas del archivo 42, vista lateral izquierda.

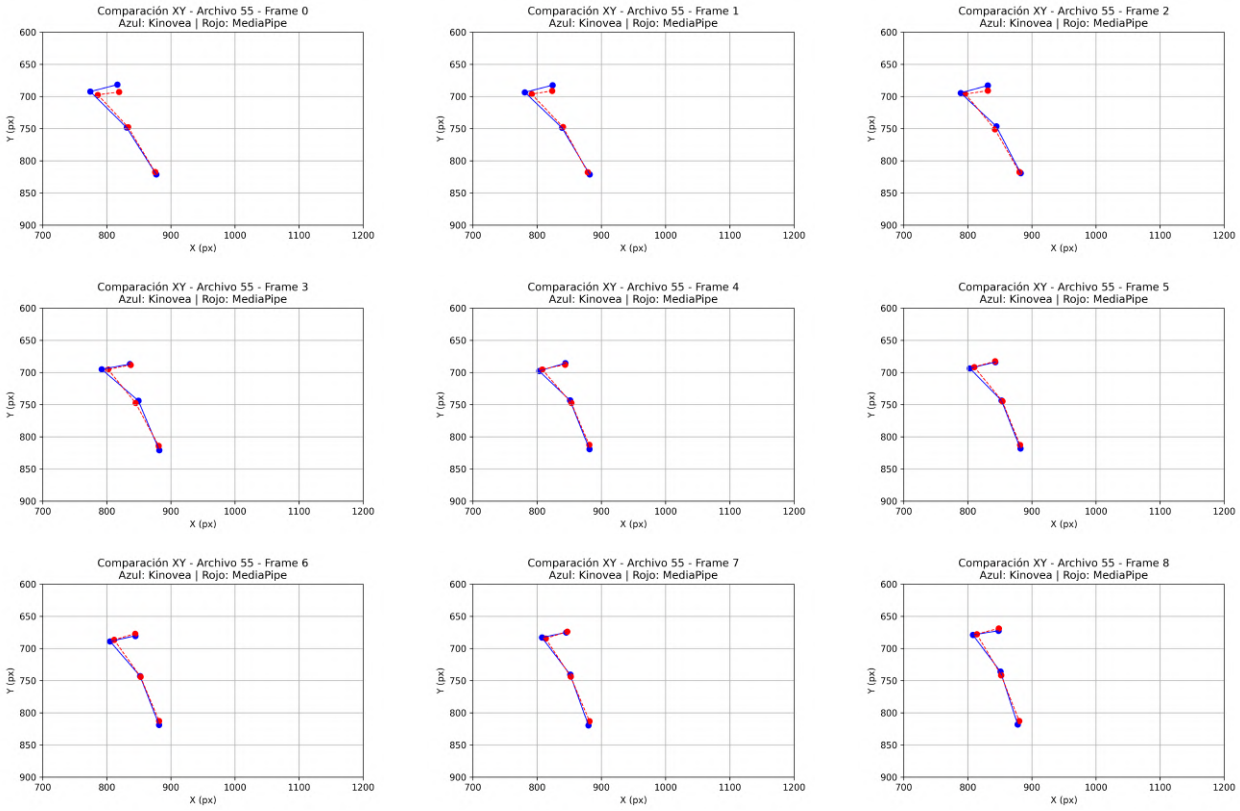


Figura 5.44. Comparación visual entre Kinovea (azul) y MediaPipe (rojo) en los primeros 9 fotogramas del archivo 55, vista trasera.

En la mayoría de los cuadros analizados se observa un solapamiento entre ambas herramientas, lo cual sugiere una buena correspondencia punto a punto y brinda una primera indicación favorable sobre la precisión angular. No obstante, también se evidencian algunas diferencias mínimas, especialmente en los puntos correspondientes a la cadera y, en menor medida, a la rodilla. Estas diferencias visuales pueden repercutir en la estimación de los ángulos, particularmente en el análisis del ángulo de la rodilla.

Seguidamente, se realiza la evaluación cuantitativa de estas estimaciones para cada una de las vistas consideradas.

Rodilla en la vista lateral derecha

A continuación se presenta un ejemplo representativo de la tabla construida para comparar los ángulos de rodilla entre Kinovea y MediaPipe, correspondiente al archivo 18.

archivo	frame	ángulo rodilla Kinovea	ángulo rodilla MediaPipe	diferencia
18	0	73.107	77.105	-3.998
18	1	69.843	75.184	-5.341
18	2	76.239	81.499	-5.260
18	3	102.478	104.650	-2.172
18	4	136.637	142.359	-5.722
18	5	162.377	157.137	5.240
18	6	171.811	173.477	-1.666

Tabla 5.4. Comparación de ángulos de rodilla entre Kinovea y MediaPipe (vista lateral derecha, archivo 18).

En el conjunto total de observaciones, la diferencia promedio fue de $-1{,}41°$ y la mediana de $-1{,}03°$, con un IQR entre $-5{,}08°$ y $2{,}43°$. El mínimo fue de $-31{,}52°$ y el máximo de $28{,}43°$, indicando cierta presencia de valores extremos. El boxplot (Figura 5.45) y la distribución de densidad (Figura 5.46) muestran que la mayoría de las diferencias se concentran cerca de cero, con ligera asimetría negativa.

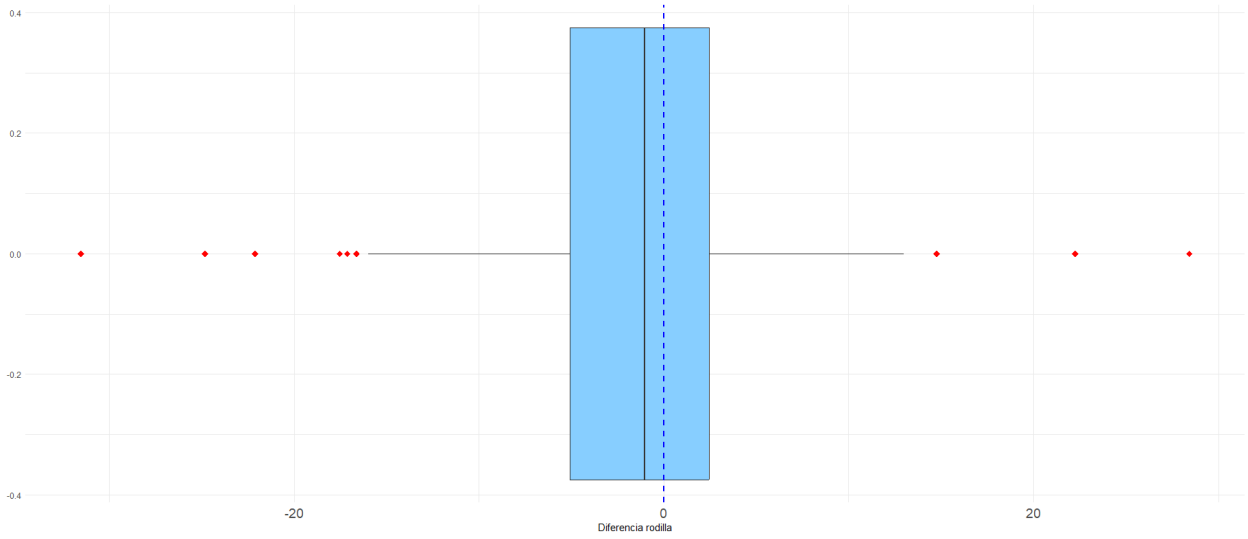


Figura 5.45. Boxplot de la diferencia de ángulo de rodilla entre Kinovea y MediaPipe.

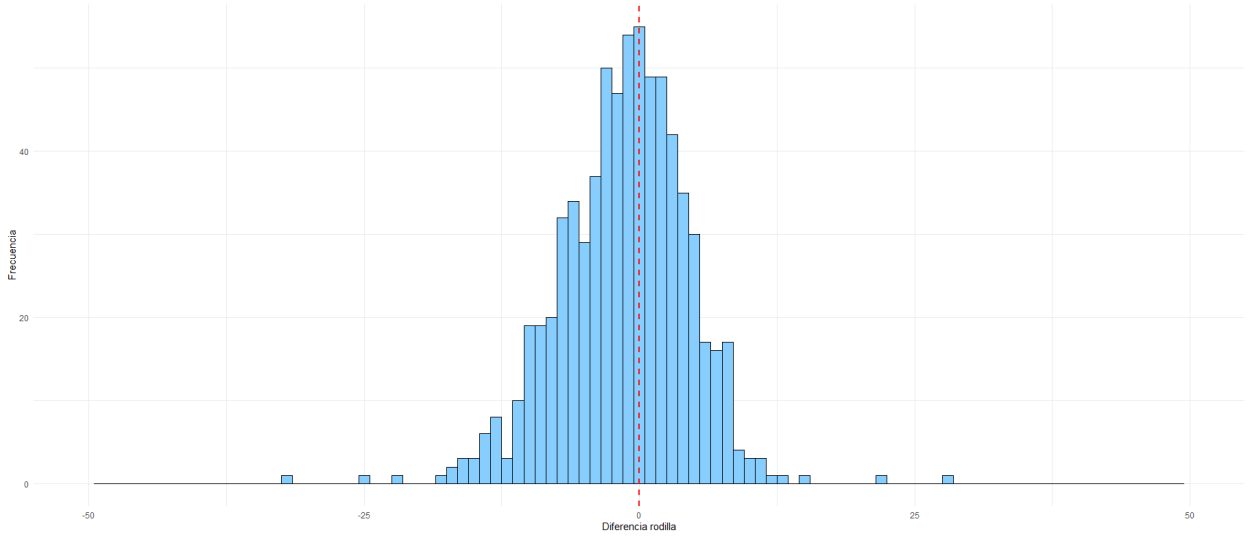


Figura 5.46. Distribución de la diferencia de ángulo de rodilla. Línea roja: referencia en 0°.

El análisis de valores atípicos, basado en el IQR, identificó como umbrales $-16{,}35°$ y $13{,}71°$. Se detectaron 9 valores atípicos (1.13 % del total), de los cuales 3 fueron atribuibles a errores en Kinovea y 6 a fallos de MediaPipe en momentos donde ambas piernas se superponían. Dado que estos errores eran identificables y escasos, se procedió a su eliminación.

Sin los valores atípicos, la media de las diferencias fue de $-1{,}34°$, la mediana de $-0{,}97°$, y el nuevo IQR quedó entre $-4{,}94°$ y $2{,}43°$. El rango total se redujo a $[-16{,}00°,\ 12{,}98°]$. Los gráficos actualizados (Figuras 5.47, 5.48) muestran una dispersión más acotada y simetría ligeramente mejorada.

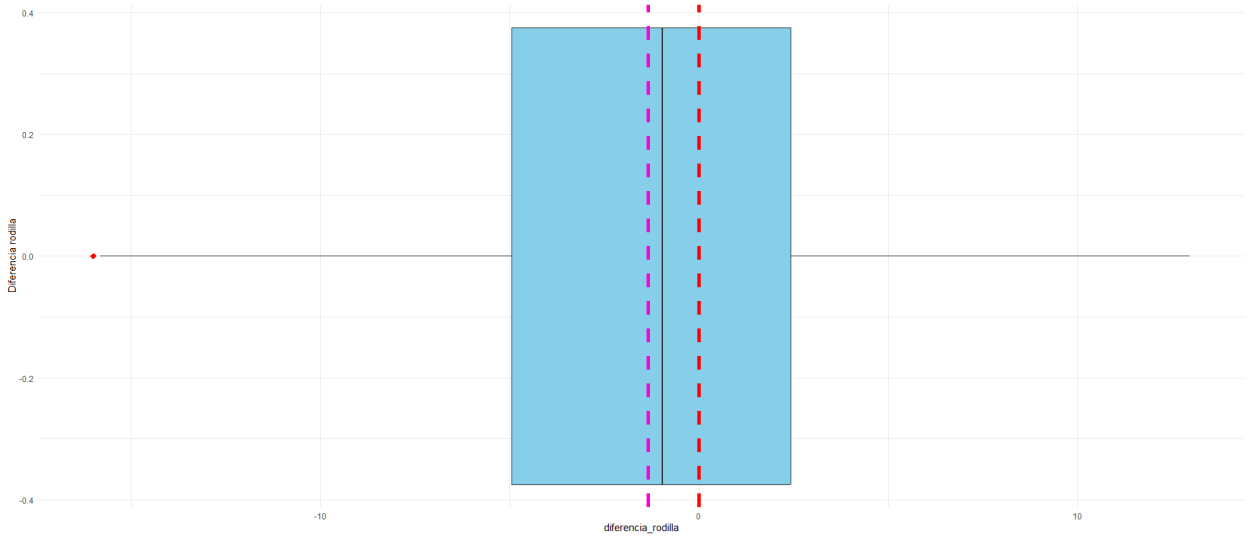


Figura 5.47. Boxplot de la diferencia de ángulo de rodilla sin valores atípicos. Línea magenta: media; línea roja: referencia en 0°.

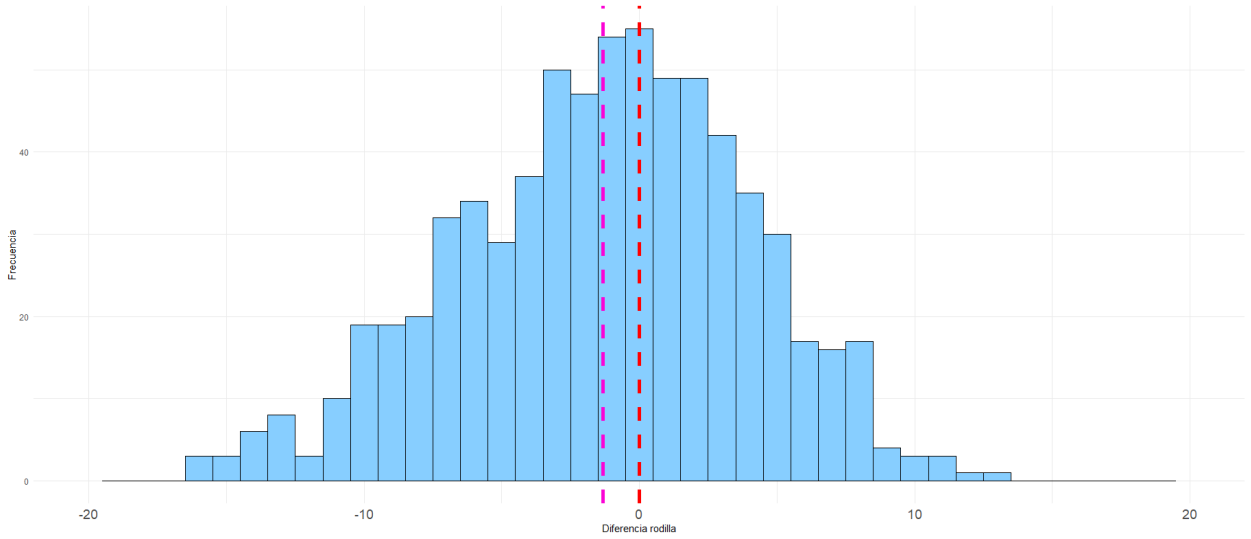


Figura 5.48. Distribución de la diferencia de ángulo de rodilla sin valores atípicos. Línea magenta: media; línea roja: referencia en 0°.

Dado que la distribución no fue normal (según la prueba de Lilliefors), se utilizó la prueba no paramétrica de los signos. El resultado fue significativo ($p = 0{,}0001709$), lo que indica que la mediana de las diferencias es distinta de cero. El intervalo de confianza del 95 % para la mediana estuvo entre $-1{,}39°$ y $-0{,}50°$, confirmando que Kinovea tiende a reportar valores de rodilla levemente menores que MediaPipe.

La correlación de Pearson entre ambos métodos fue de $r = 0{,}992$, evidenciando una relación lineal extremadamente fuerte. La Figura 5.49 muestra que la mayoría de los puntos se agrupan cerca de la línea de identidad de 45°, confirmando una alta concordancia.

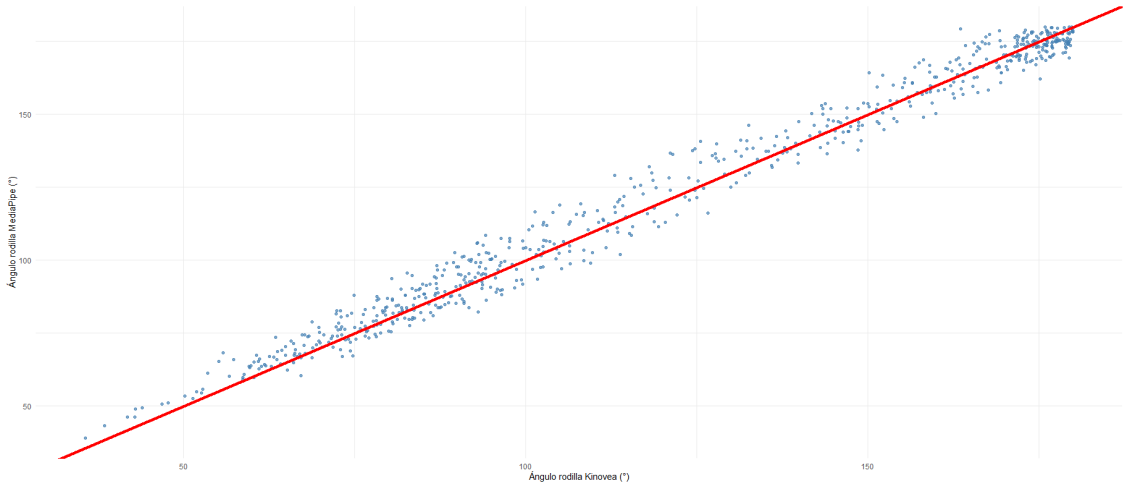


Figura 5.49. Relación entre los ángulos de rodilla medidos por Kinovea y MediaPipe. Línea punteada roja: identidad (45°).

En conjunto, los resultados indican que si bien existen diferencias significativas, estas son moderadas y pueden atribuirse principalmente a errores de identificación manual en Kinovea, especialmente en el punto de la cadera, cuya localización puede verse afectada por la vestimenta. El algoritmo automático mostró mayor consistencia y una fuerte correlación con los registros manuales.

Pie en la vista lateral derecha

A continuación se presenta un ejemplo representativo de la tabla construida para comparar los ángulos del pie entre Kinovea y MediaPipe, correspondiente al archivo 18.

Archivo	Frame	Ángulo pie Kinovea	Ángulo pie MediaPipe	Diferencia
18	0	88.247	81.007	7.240
18	1	81.428	63.493	17.935
18	2	68.795	59.442	9.353
18	3	70.646	60.069	10.577
18	4	78.405	46.030	32.375
18	5	103.040	74.787	28.253
18	6	97.061	69.881	27.180

Tabla 5.5. Comparación del ángulo del pie entre Kinovea y MediaPipe (vista lateral derecha, archivo 18).

En el conjunto total de observaciones, la diferencia promedio fue de $6{,}320°$ y la mediana de $7{,}008°$, con un IQR entre $-1{,}539°$ y $15{,}682°$. El mínimo fue de $-118{,}642°$ y el máximo de $104{,}816°$, indicando la presencia de datos atípicos tanto positivos como negativos. El boxplot (Figura 5.50) y la distribución de densidad (Figura 5.51) muestran que la mayoría de las diferencias son positivas, indicando que los ángulos calculados en Kinovea más altos que los de MediaPipe.

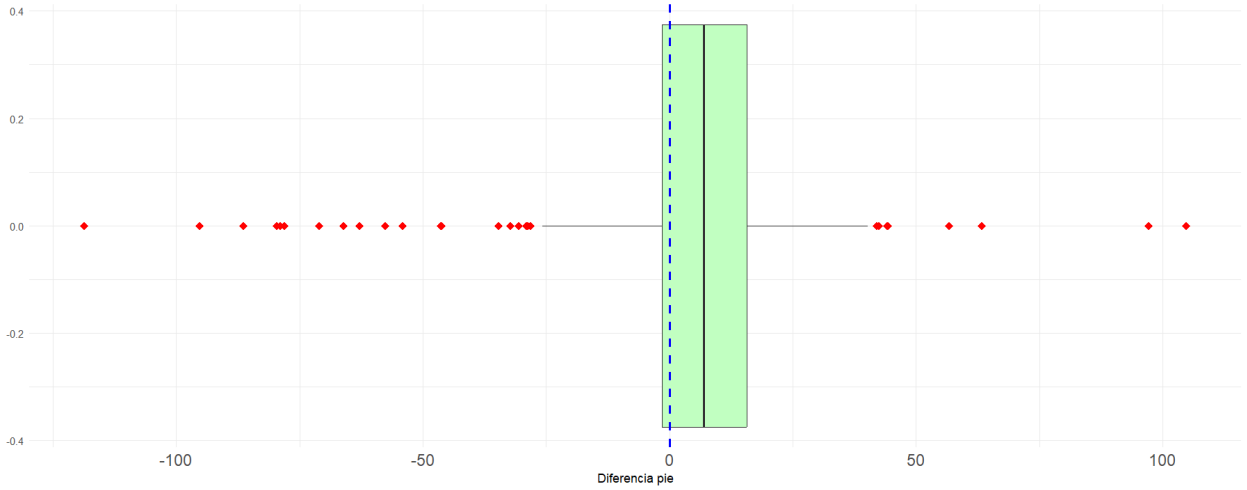


Figura 5.50. Boxplot de la diferencia de ángulo del pie entre Kinovea y MediaPipe.

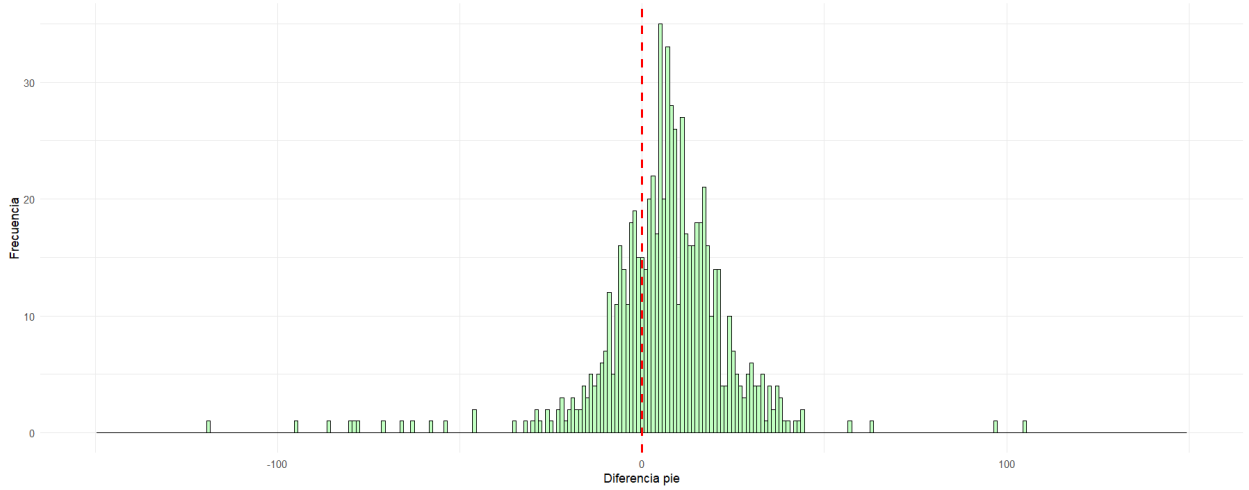


Figura 5.51. Distribución de la diferencia de ángulo del pie. Línea roja: referencia en 0°.

El análisis de valores atípicos, basado en el IQR, identificó como umbrales $-27{,}3705°$ y $41{,}5135°$. Se detectaron 27 valores atípicos (3.82 % del total).

Después de una revisión exhaustiva, se logró encontrar un patrón para estos valores atípicos. En aquellos fotogramas donde el arquero rota el pie (lo cual influye en una coordenada z), es más complejo identificar el punto tanto para MediaPipe como para la anotación en Kinovea. Especialmente la anotación manual se ve distorsionada por la calidad del video, lo que puede generar errores al ubicar el punto exacto de la punta del pie. Estas discrepancias explican las diferencias elevadas registradas en algunos fotogramas, particularmente en los instantes clave del golpeo donde el movimiento es más rápido y tridimensional (véase la Figura 5.52).

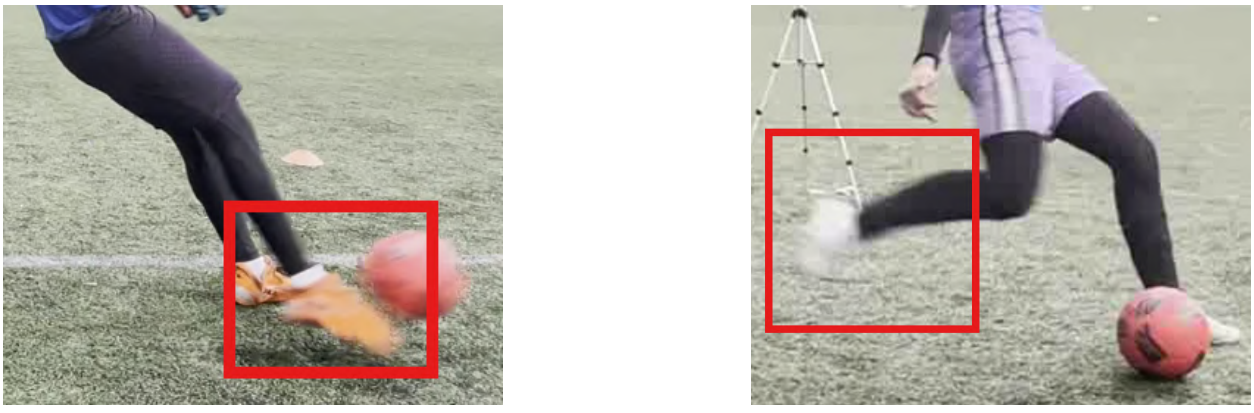


Figura 5.52. Ejemplos de distorsiones en el pie derecho desde la vista sagital derecha.

Dado que estos errores eran identificables y representan únicamente el 3.82 % de los datos, se procedió a su eliminación.

Sin los valores atípicos, la media de las diferencias fue de $7{,}4988°$, la mediana de $7{,}11°$, y

el nuevo IQR quedó entre $-0{,}6422°$ y $15{,}5467°$. El rango total se redujo a $[-25{,}8050°,\ 40{,}1940°]$. Los gráficos actualizados (Figuras 5.53, 5.54) muestran menos dispersión en los datos, sin embargo, se nota más la asimetría a la derecha de la distribución de los datos.

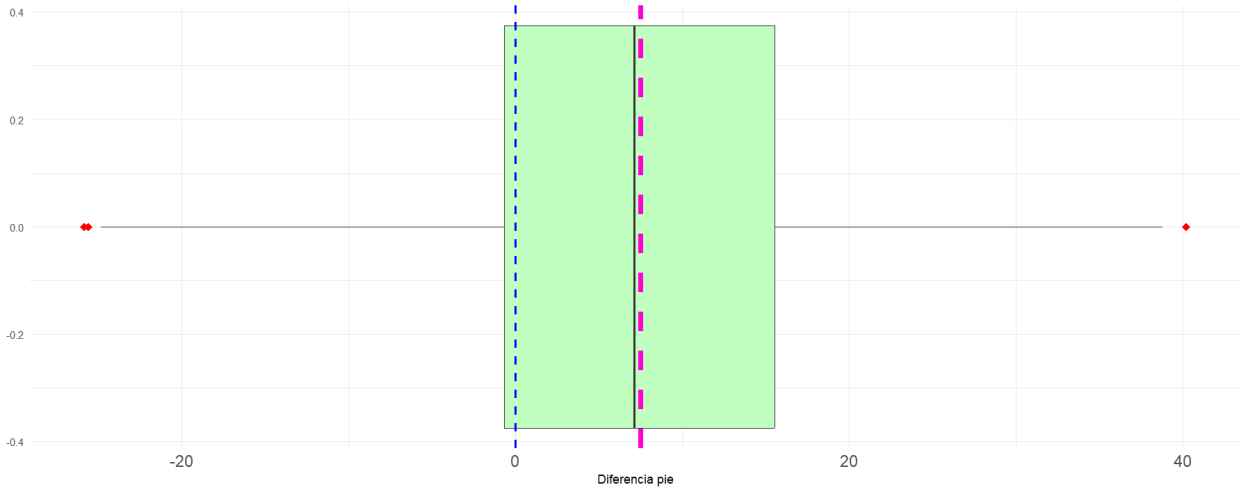


Figura 5.53. Boxplot de la diferencia de ángulo del pie sin valores atípicos. Línea magenta: media; línea roja: referencia en 0°.

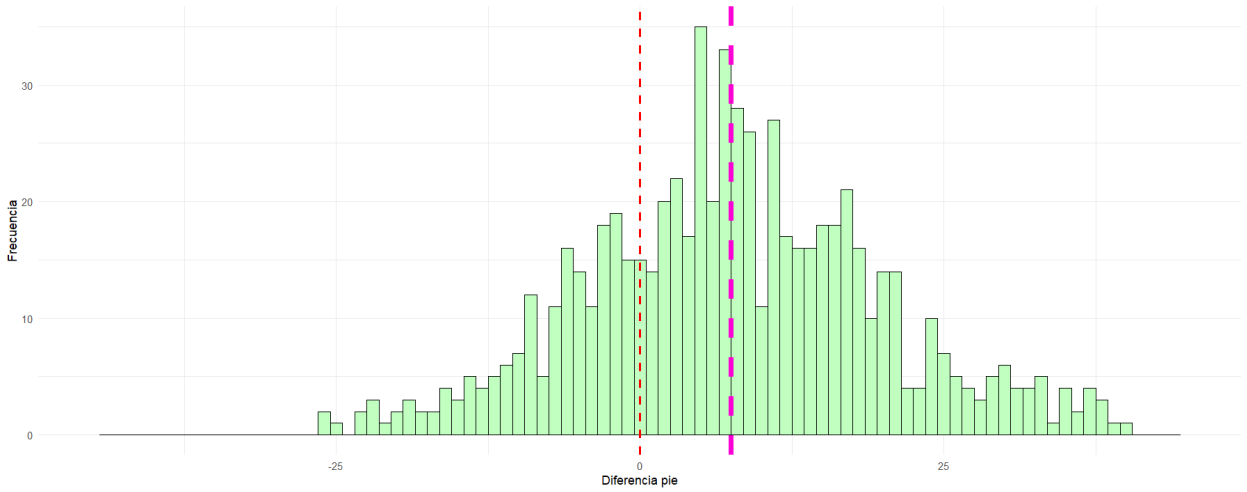


Figura 5.54. Distribución de la diferencia de ángulo del pie sin valores atípicos. Línea magenta: media; línea roja: referencia en 0°.

Según la prueba de normalidad de Lilliefors, las diferencias entre los ángulos estimados por Kinovea y MediaPipe para el pie derecho siguen una distribución normal. El valor $p = 0{,}1002$ indica que, al utilizar un nivel de significancia del 5 %, no se rechaza la hipótesis nula de normalidad. Por esta razón, se considera adecuado aplicar una prueba t para la diferencia de medias con el fin de evaluar esta es igual a 0.

Los resultados de la prueba t arrojan un valor $p < 2{,}2 \times 10^{-16}$, lo cual permite rechazar la hipótesis nula y concluir que existe una diferencia significativa entre ambas herramientas. El intervalo de confianza del 95 % para la media de las diferencias se encuentra entre $6{,}576°$ y $8{,}422°$, lo que refuerza esta conclusión.

La correlación de Pearson entre ambos métodos fue de $r = 0{,}636$, evidenciando una relación lineal relativamente baja. La Figura 5.55 muestra que los datos son muy dispersos. Existen más datos abajo de la identidad, lo que soporta lo dicho anteriormente sobre una estimación mayor por parte de Kinovea.



Figura 5.55. Relación entre los ángulos del pie medidos por Kinovea y MediaPipe. Línea punteada roja: identidad (45°).

En conjunto, los resultados indican que sí existen diferencias significativas entre ambas herramientas. Tanto la prueba t como el intervalo de confianza reflejan que las diferencias en los ángulos medidos son considerables. Adicionalmente, la correlación de Pearson sugiere que la relación entre ambas variables no es lo suficientemente fuerte como para concluir que Kinovea y MediaPipe arrojan resultados equivalentes. Estas discrepancias pueden atribuirse al movimiento complejo de la punta del pie y a la dificultad inherente de identificar con precisión este punto en ambas herramientas, especialmente durante las fases rápidas del gesto técnico.

Rodilla en la vista lateral izquierda

A continuación se presenta un ejemplo representativo de la tabla construida para comparar los ángulos del pie entre Kinovea y MediaPipe, correspondiente al archivo 14.

Archivo	Frame	Ángulo rodilla Kinovea	Ángulo rodilla MediaPipe	Diferencia
14	0	134.990	136.339	-1.349
14	1	131.412	136.626	-5.214
14	2	132.886	133.028	-0.142
14	3	136.989	133.393	3.596
14	4	142.430	135.236	7.194
14	5	142.330	145.440	-3.110
14	6	147.022	144.238	2.784
14	7	150.471	144.777	5.694

Tabla 5.6. Comparación del ángulo de la rodilla entre Kinovea y MediaPipe (vista lateral izquierda, archivo 14).

En el conjunto total de observaciones, la diferencia promedio fue de $-2{,}985°$ y la mediana de $-3{,}162°$, con un IQR entre $-7{,}786°$ y $1{,}606°$. El mínimo fue de $-29{,}484°$ y el máximo de $17{,}375°$, indicando cierta presencia de valores extremos (sobre todo negativos). El boxplot (Figura 5.56) y la distribución de densidad (Figura 5.57) muestran que la mayoría de las diferencias son negativas, indicando que los ángulos calculados en Mediapipe son más altos que los de Kinovea.

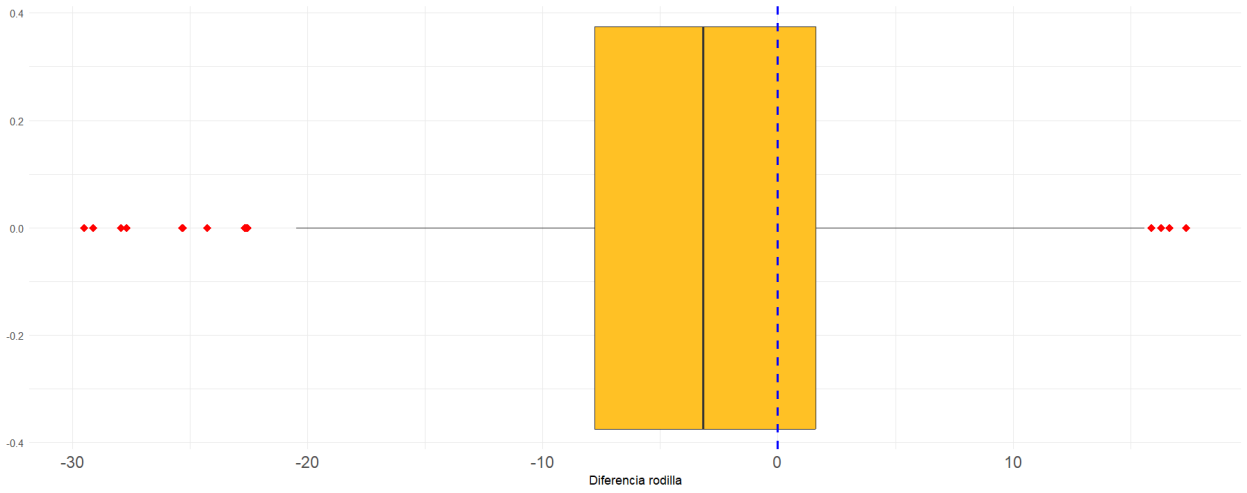


Figura 5.56. Boxplot de la diferencia de ángulo de la rodilla entre Kinovea y MediaPipe en la vista sagital izquierda.

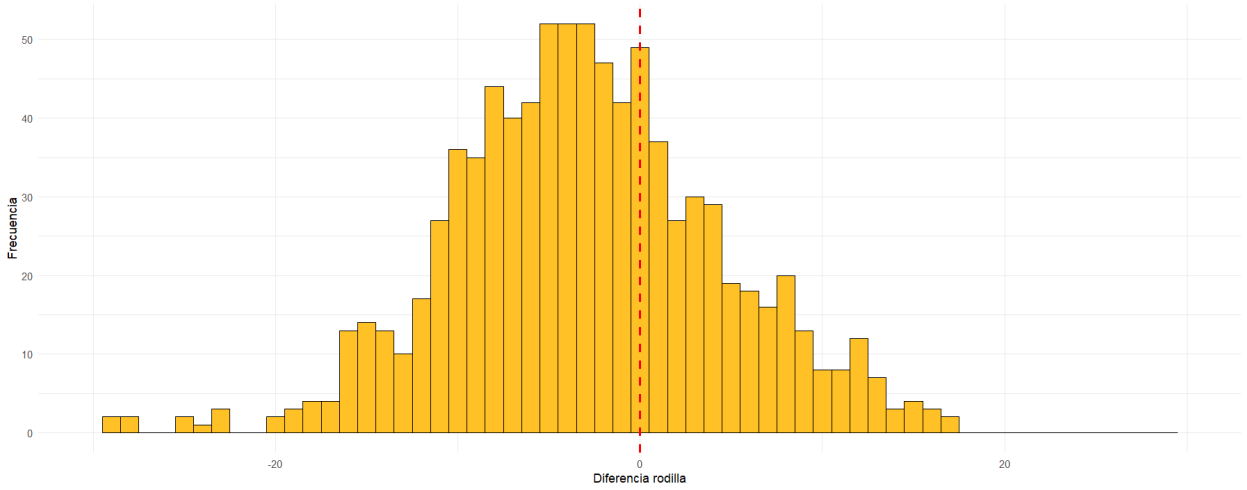


Figura 5.57. Distribución de la diferencia de ángulo de la rodilla en la vista sagital izquierda. Línea roja: referencia en 0°.

El análisis de valores atípicos, basado en el IQR, identificó como umbrales $-21{,}87°$ y $15{,}695°$. Se detectaron 14 valores atípicos (1.6 % del total).

Después de una revisión exhaustiva, se identificó que las diferencias más extremas se presentan principalmente al inicio o al final del video. En siete casos, el valor atípico ocurre en el primer fotograma, lo cual se atribuye al comportamiento del algoritmo de MediaPipe, que reemplaza el primer fotograma por el segundo cuando no logra reconocer correctamente la pose inicial. En otros casos, las discrepancias aparecen al final del video, aunque son menos extremas. En estos se observa una ligera diferencia entre las anotaciones manuales y el movimiento real. Estas discrepancias pueden deberse a la dificultad de identificar con claridad la articulación de la rodilla en algunas prendas de vestir utilizadas por los arqueros (véase la Figura 5.58).



Figura 5.58. Problemas causados por el tipo de ropa usada por los arqueros.

Dado que estos errores eran identificables y representan únicamente el $1.6\,\%$ de los datos, se procedió a su eliminación.

Sin los valores atípicos, la media de las diferencias fue de $-2{,}810°$, la mediana de $-3{,}123°$, y el nuevo IQR quedó entre $-7{,}633°$ y $1{,}605°$. El rango total se redujo a $[-20{,}478°,\ 15{,}597°]$. Los gráficos actualizados (Figuras 5.59, 5.60) muestran menos dispersión en los datos, sin embargo, se nota más la asimetría a la derecha de la distribución de los datos.

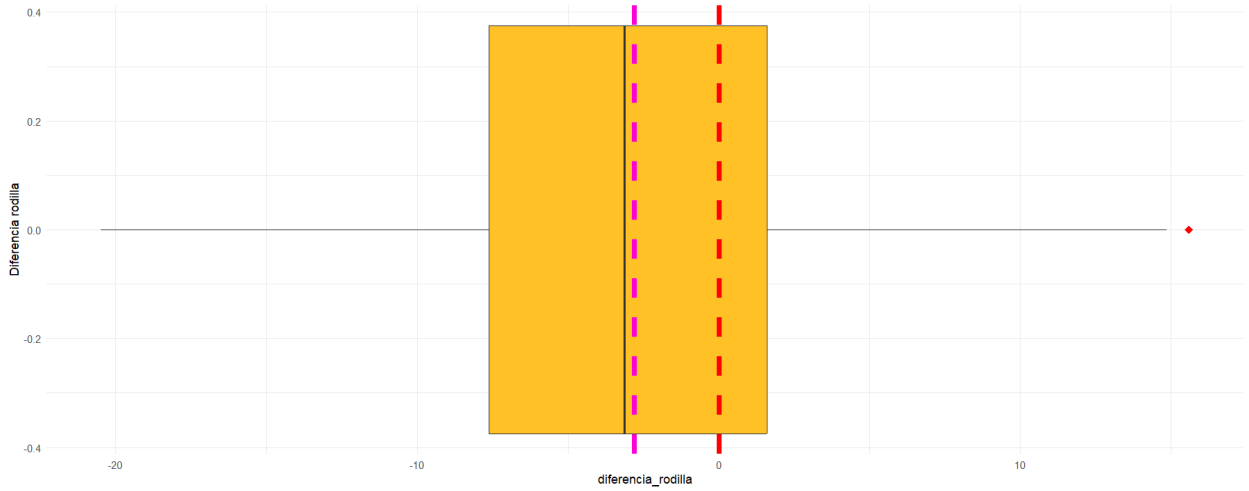


Figura 5.59. Boxplot de la diferencia de ángulo de la rodilla izquierda sin valores atípicos. Línea magenta: media; línea roja: referencia en 0°.

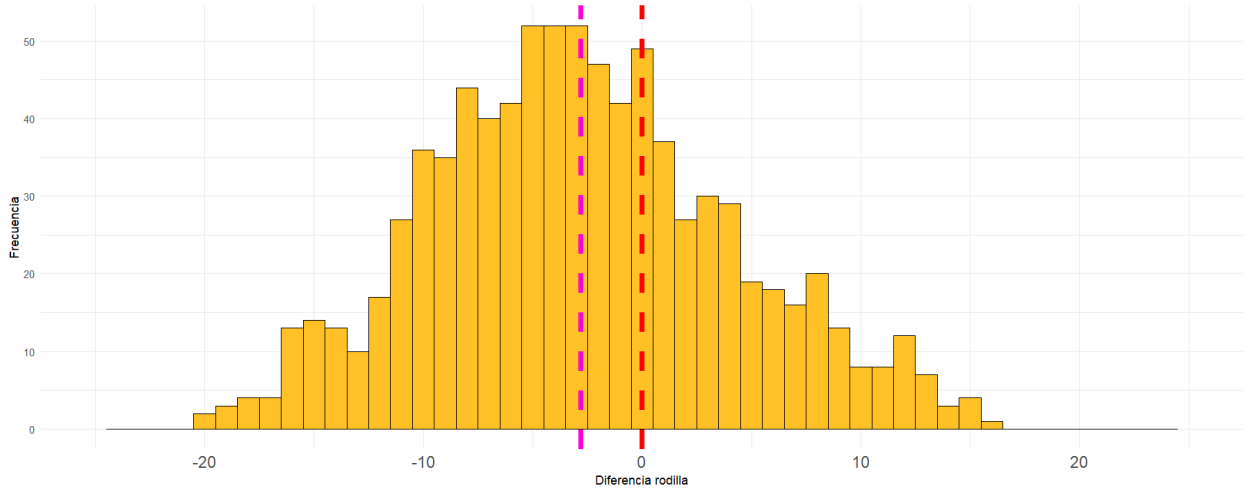


Figura 5.60. Distribución de la diferencia de ángulo de la rodilla izquierda sin valores atípicos. Línea magenta: media; línea roja: referencia en 0°.

Según la prueba de normalidad de Lilliefors, las diferencias entre los ángulos estimados por Kinovea y MediaPipe para el pie derecho siguen una distribución normal. El valor $p = 0{,}08512$

indica que, al utilizar un nivel de significancia del 5 %, no se rechaza la hipótesis nula de normalidad. Por esta razón, se considera adecuado aplicar una prueba t para la diferencia de medias con el fin de evaluar esta es igual a 0.

Los resultados de la prueba t arrojan un valor $p < 2{,}2 \times 10^{-16}$, lo cual permite rechazar la hipótesis nula y concluir que existe una diferencia significativa entre ambas herramientas. El intervalo de confianza del 95 % para la media de las diferencias se encuentra entre $-3{,}278659°$ y $-2{,}340691°$, lo que refuerza esta conclusión.

La correlación de Pearson entre ambos métodos fue de $r = 0{,}9170888$, evidenciando una relación lineal alta. La Figura 5.61 muestra que los datos son ligeramente dispersos, sin embargo, se logra ver la relación lineal existente.

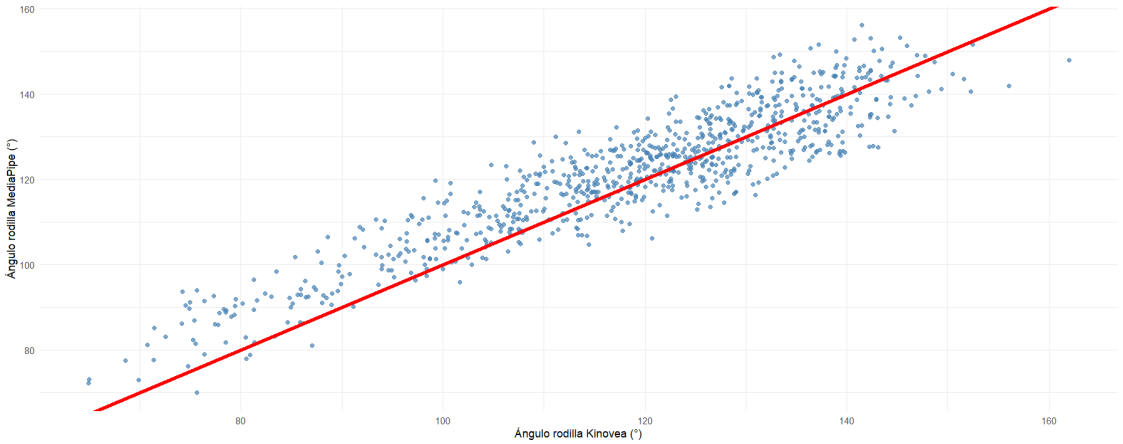


Figura 5.61. Relación entre los ángulos de la rodilla izquierda medidos por Kinovea y MediaPipe. Línea punteada roja: identidad (45°).

En conjunto, los resultados indican que, aunque existen diferencias significativas entre ambas herramientas, estas se encuentran relacionadas de forma individual fotograma a fotograma. Además, el intervalo de confianza sugiere que la magnitud de dicha diferencia es mínima. Sin embargo, se identificaron posibles fuentes de error que explican estas discrepancias, como la vestimenta utilizada por los arqueros, que dificulta una marcación precisa de los puntos articulares, así como la corrección automática aplicada por MediaPipe en el primer fotograma, cuando no se logra una detección confiable.

Pie en la vista lateral izquierda

A continuación se presenta un ejemplo representativo de la tabla construida para comparar los ángulos del pie entre Kinovea y MediaPipe, correspondiente al archivo 14.

Archivo	Frame	Ángulo pie Kinovea	Ángulo pie MediaPipe	Diferencia
14	0	46.048	51.013	-4.965
14	1	42.916	52.256	-9.340
14	2	38.843	51.170	-12.327
14	3	38.843	49.852	-11.009
14	4	41.621	44.968	-3.347
14	5	43.114	45.446	-2.332
14	6	43.114	45.777	-2.663
14	7	43.114	47.946	-4.832

Tabla 5.7. Comparación del ángulo del pie entre Kinovea y MediaPipe (vista lateral izquierda, archivo 14).

En el conjunto total de observaciones, la diferencia promedio fue de $2{,}728°$ y la mediana de $2{,}031°$, con un IQR entre $-1{,}928°$ y $6{,}822°$. El mínimo fue de $-16{,}115°$ y el máximo de $38{,}109°$, indicando la presencia de datos positivos atípicos. El boxplot (Figura 5.62) y la distribución de densidad (Figura 5.63) muestran que la mayoría de las diferencias son positivas, indicando que los ángulos calculados en Kinovea más altos que los de MediaPipe.

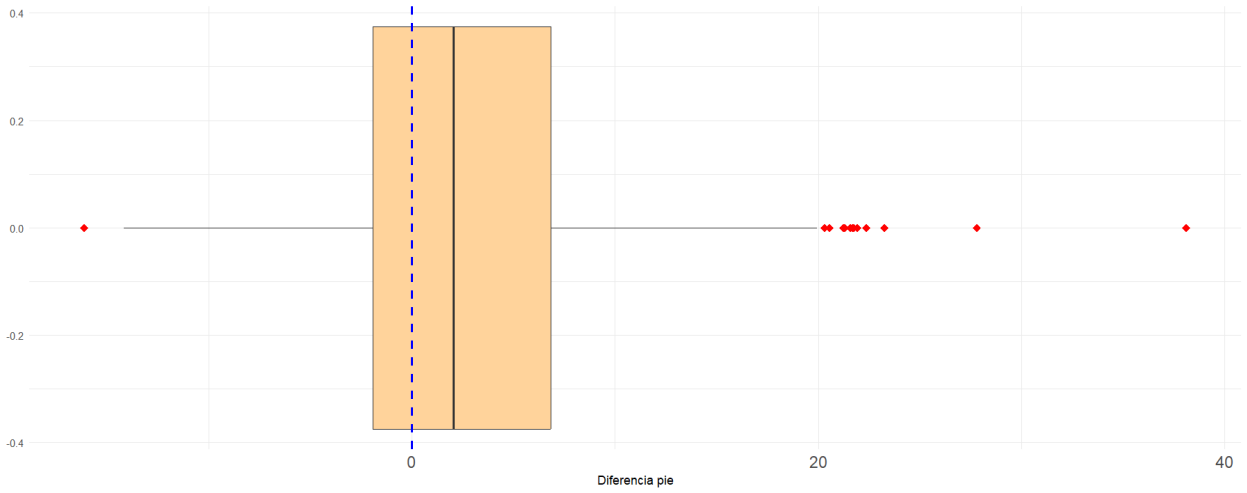


Figura 5.62. Boxplot de la diferencia de ángulo del pie entre Kinovea y MediaPipe.

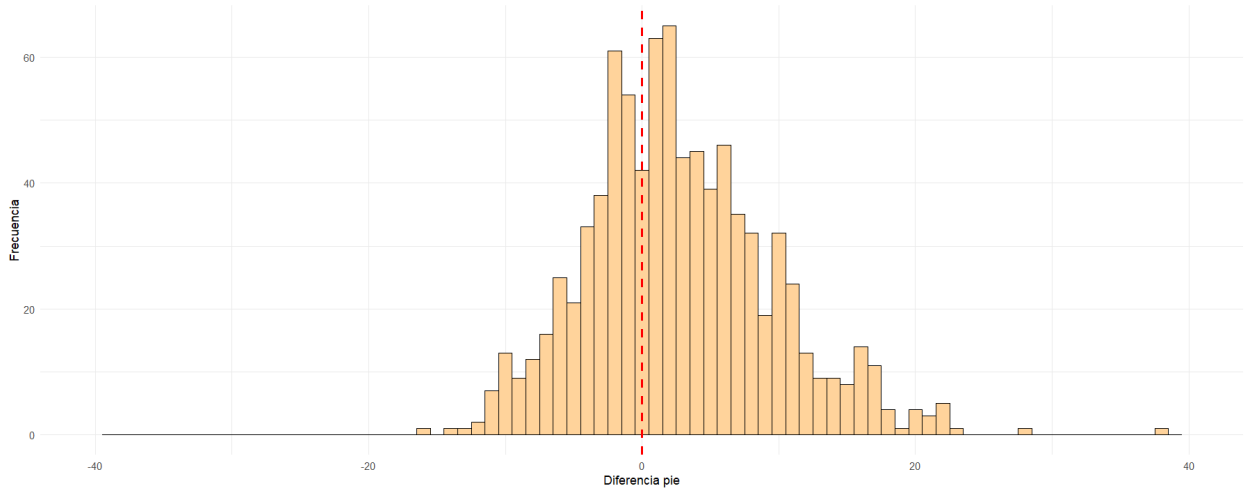


Figura 5.63. Distribución de la diferencia de ángulo del pie. Línea roja: referencia en 0°.

El análisis de valores atípicos, basado en el IQR, identificó como umbrales $-15{,}05288°$ y $19{,}94613°$. Se detectaron 13 valores atípicos (1.5 % del total).

Después de hacer la respectiva revisión, se identificó que 9 de los 13 valores atípicos se presentan en el momento del golpeo del balón, instancia en la que, por efecto de la inercia, el talón izquierdo se eleva ligeramente del suelo. En estos casos, ambas herramientas muestran discrepancias en la estimación del ángulo, que, aunque pequeñas en apariencia, pueden generar diferencias superiores a $20°$. Por otro lado, los 4 fotogramas restantes corresponden a los primeros instantes del video y, al igual que lo observado en el análisis de la rodilla en la vista lateral izquierda, dichas discrepancias pueden explicarse por la corrección automática que MediaPipe aplica al primer fotograma cuando no logra detectar adecuadamente la pose inicial.

Dado que estos errores eran identificables y representan únicamente el 1.5 % de los datos, se procedió a su eliminación.

Sin los valores atípicos, la media de las diferencias fue de $2{,}457°$, la mediana de $1{,}945°$, y el nuevo IQR quedó entre $-1{,}942°$ y $6{,}516°$. El rango total se redujo a $[-14{,}164°,\ 19{,}945°]$. Los gráficos actualizados (Figuras 5.64, 5.65) muestran menos dispersión en los datos, sin embargo, se nota una ligera asimetría a la derecha de la distribución de los datos.

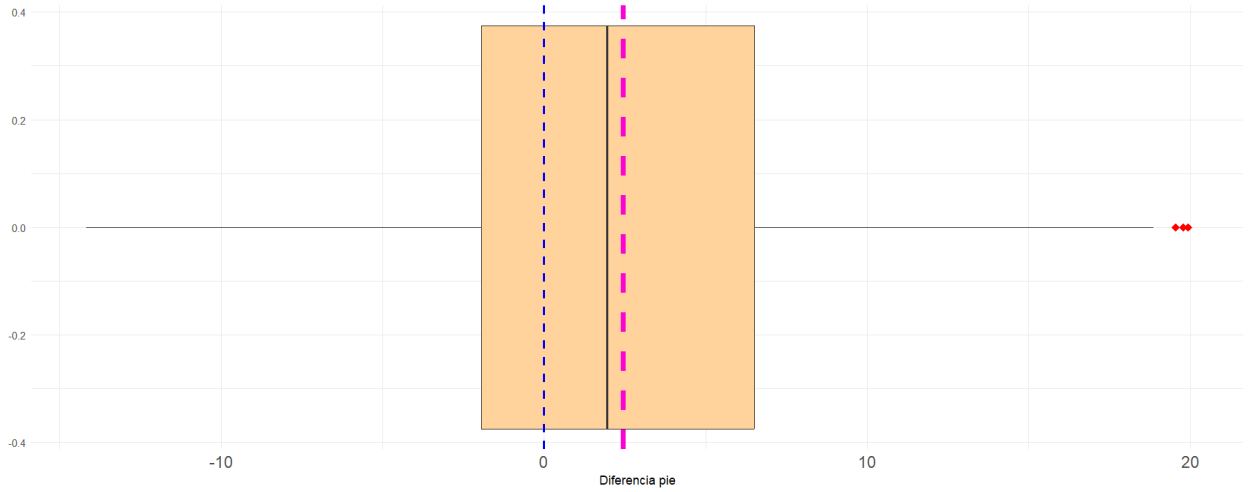


Figura 5.64. Boxplot de la diferencia de ángulo del pie sin valores atípicos. Línea magenta: media; línea roja: referencia en 0°.

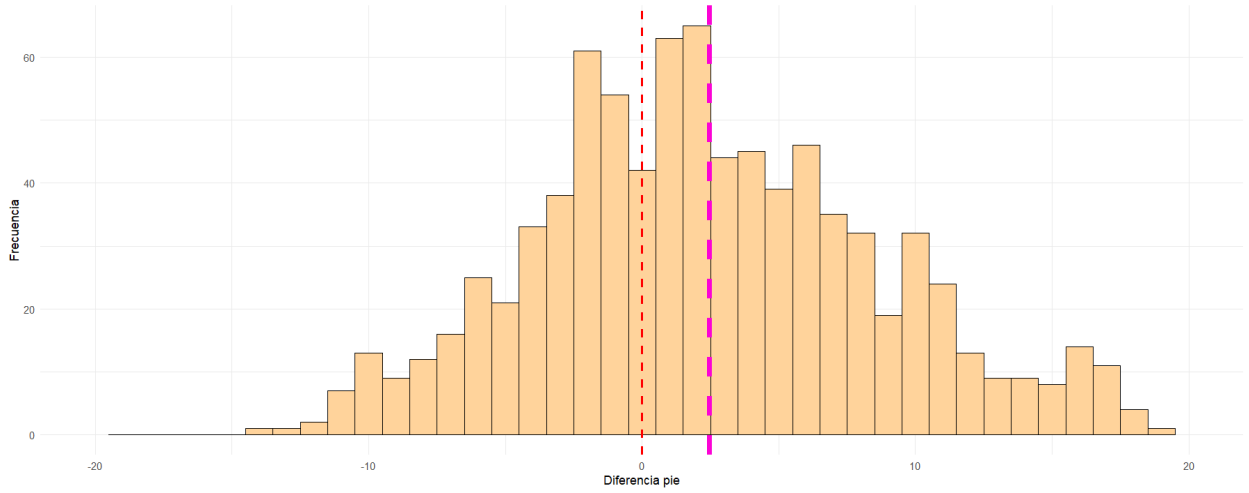


Figura 5.65. Distribución de la diferencia de ángulo del pie sin valores atípicos. Línea magenta: media; línea roja: referencia en 0°.

Dado que la distribución no fue normal (según la prueba de Lilliefors), se utilizó la prueba no paramétrica de los signos. El resultado fue significativo ($p = 3{,}553 \times 10^{-15}$), lo que indica que la mediana de las diferencias es distinta de cero. El intervalo de confianza del 95 % para la mediana estuvo entre $1{,}465920°$ y $2{,}370733°$, confirmando que Kinovea tiende a reportar valores de rodilla levemente mayores que MediaPipe.

En este caso particular, no se analiza la correlación de Pearson, ya que los valores de los ángulos se mantienen relativamente constantes durante la secuencia observada. Esta escasa variabilidad provoca que los datos en la gráfica de correlación se concentren en una región estrecha, lo

cual limita la utilidad de evaluar la linealidad de forma visual o estadística. Por tanto, se considera que dicho análisis no aporta información relevante en este contexto específico.

Finalmente, se concluye que, para el pie en la vista lateral izquierda, las diferencias entre ambas herramientas, aunque estadísticamente significativas, son mínimas.

Rodilla en la vista trasera

A continuación se presenta un ejemplo representativo de la tabla construida para comparar los ángulos del pie entre Kinovea y MediaPipe, correspondiente al archivo 53.

Archivo	Frame	Ángulo rodilla Kinovea	Ángulo rodilla MediaPipe	Diferencia
55	0	167.035	167.871	-0.836
55	1	164.408	165.140	-0.732
55	2	160.631	169.884	-9.253
55	3	153.589	168.835	-15.246
55	4	155.280	162.104	-6.824
55	5	156.943	162.604	-5.661
55	6	160.313	167.232	-6.919
55	7	162.583	169.624	-7.041
55	8	161.151	170.948	-9.797
55	9	160.865	169.975	-9.110
55	10	160.754	168.946	-8.192
55	11	163.822	174.848	-11.026

Tabla 5.8. Comparación del ángulo de la rodilla entre Kinovea y MediaPipe (vista trasera, archivo 55).

En el conjunto total de observaciones, la diferencia promedio fue de $1{,}481°$ y la mediana de $1{,}135°$, con un IQR entre $-3{,}392°$ y $4{,}724°$. El mínimo fue de $-35{,}621°$ y el máximo de $141{,}530°$, indicando cierta presencia de valores extremos (sobre todo positivos). El boxplot (Figura 5.66) y la distribución de densidad (Figura 5.67) muestran que hay una cantidad similar de datos positivos y negativos.

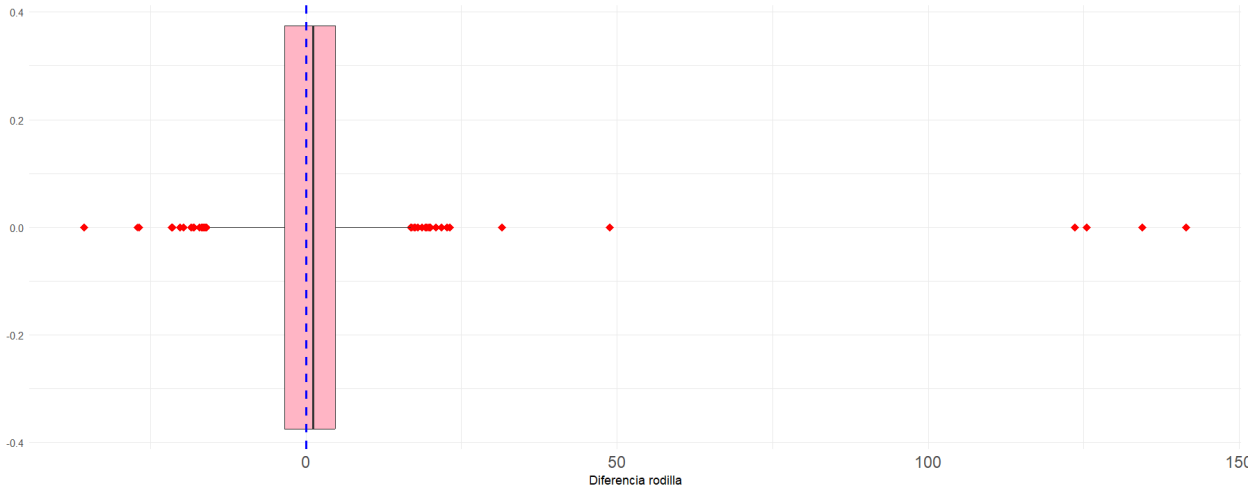


Figura 5.66. Boxplot de la diferencia de ángulo de la rodilla entre Kinovea y MediaPipe en la vista coronal.

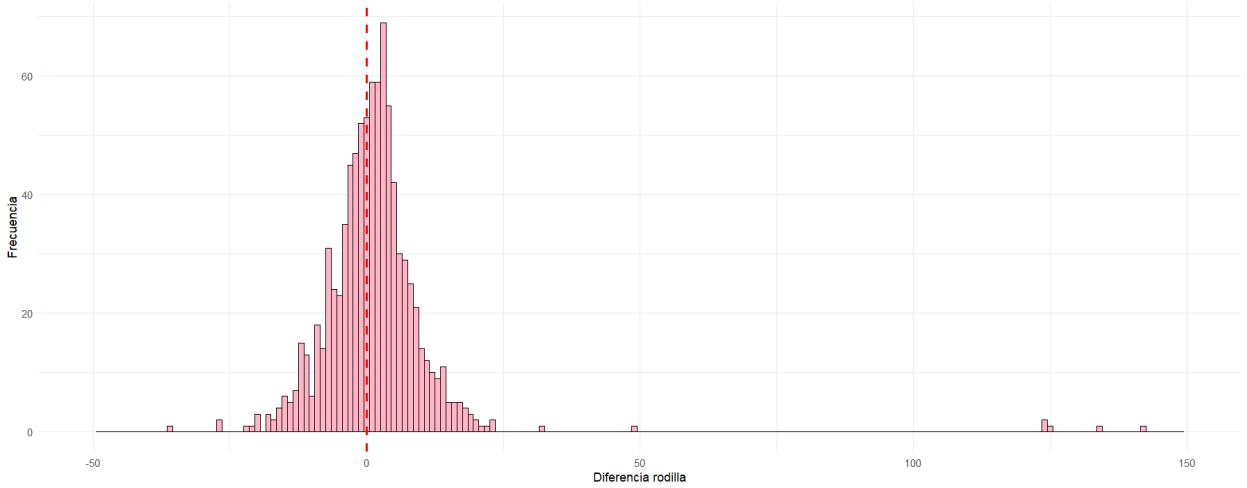


Figura 5.67. Distribución de la diferencia de ángulo de la rodilla en la vista coronal. Línea roja: referencia en 0°.

El análisis de valores atípicos, basado en el IQR, identificó como umbrales $-15{,}56675°$ y $16{,}89925°$. Se detectaron 41 valores atípicos (4.6 % del total).

Después de una revisión exhaustiva, se identificó que las diferencias más extremas se deben a las oclusiones presentadas en la vista trasera, como se logra ver en la Figura 5.68. Estas oclusiones limitan tanto el etiquetado manual de Kinovea, como la identificación automática de MediaPipe.



Figura 5.68. Oclusiones presentadas desde la vista coronal.

Dado que estos errores eran identificables y representan únicamente el 4.6 % de los datos, se procedió a su eliminación.

Sin los valores atípicos, la media de las diferencias fue de $0{,}7175°$, la mediana de $1{,}0875°$, y el nuevo IQR quedó entre $-3{,}2117°$ y $4{,}5168°$. El rango total se redujo a $[-15{,}4350°,\ 16{,}8550°]$. Los gráficos actualizados (Figuras 5.69, 5.70) muestran menos dispersión en los datos, pero con una ligera asimetría a izquierda.

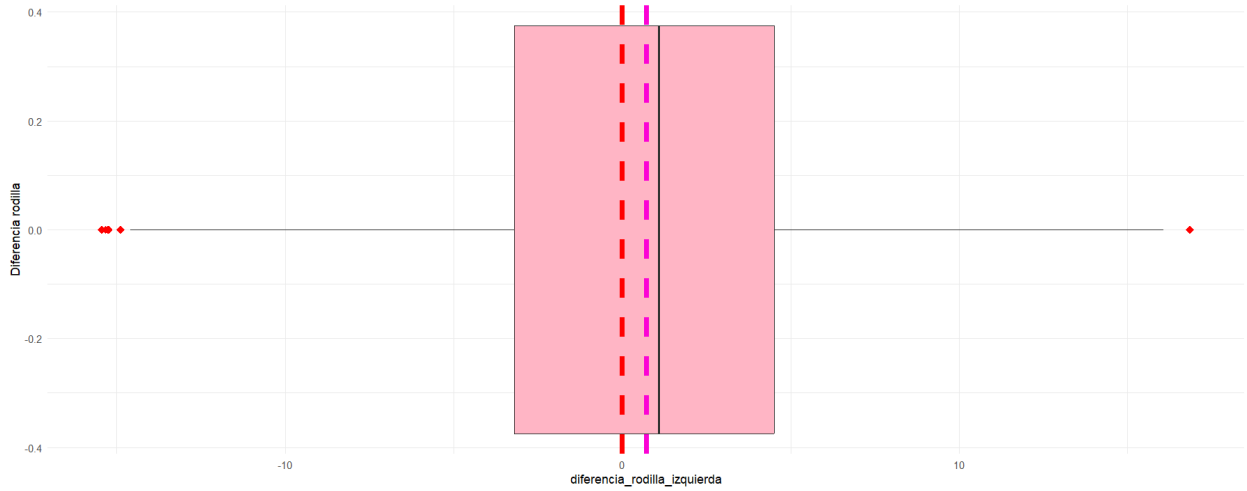


Figura 5.69. Boxplot de la diferencia de ángulo de la rodilla izquierda sin valores atípicos desde la vista coronal. Línea magenta: media; línea roja: referencia en 0°.

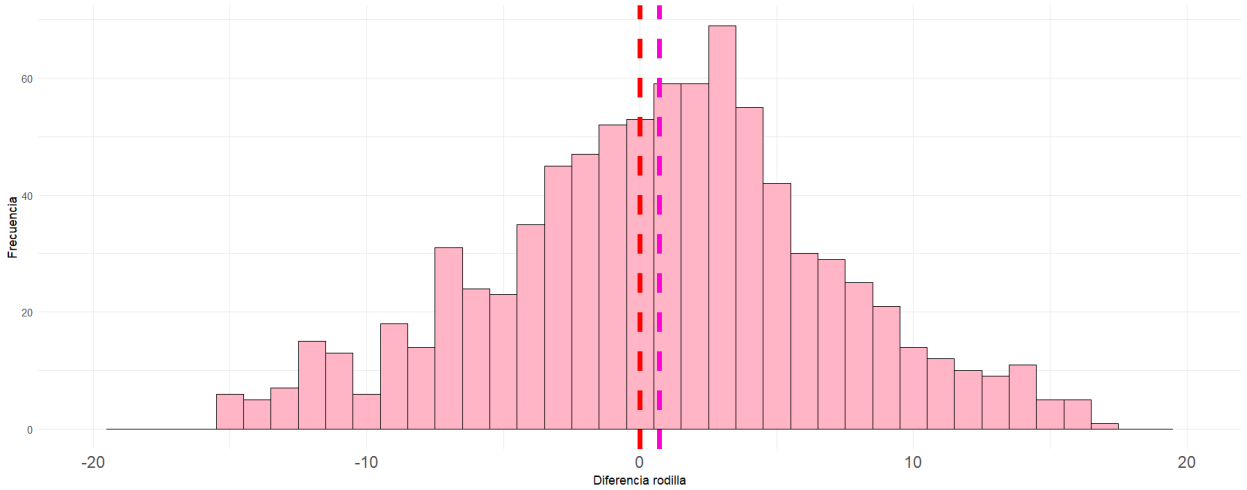


Figura 5.70. Distribución de la diferencia de ángulo de la rodilla izquierda desde la vista coronal sin valores atípicos. Línea magenta: media; línea roja: referencia en 0°.

Según la prueba de normalidad de Lilliefors, las diferencias entre los ángulos estimados por Kinovea y MediaPipe para el pie derecho siguen una distribución normal. El valor $p = 0{,}08007$ indica que, al utilizar un nivel de significancia del 5 %, no se rechaza la hipótesis nula de normalidad. Por esta razón, se considera adecuado aplicar una prueba t para la diferencia de medias con el fin de evaluar esta es igual a 0.

Los resultados de la prueba t arrojan un valor $p = 0{,}0008643$, lo cual permite rechazar la hipótesis nula y concluir que existe una diferencia significativa entre ambas herramientas. El intervalo de confianza del 95 % para la media de las diferencias se encuentra entre $0{,}2962518°$ y $1{,}1386753°$, lo que refuerza esta conclusión.

Al igual que en el pie en la vista lateral izquierda, no se analiza la correlación de Pearson, ya que los valores de los ángulos se mantienen relativamente constantes durante la secuencia observada. Esta escasa variabilidad provoca que los datos en la gráfica de correlación se concentren en una región estrecha, lo cual limita la utilidad de evaluar la linealidad de forma visual o estadística. Por tanto, se considera que dicho análisis no aporta información relevante en este contexto específico.

Finalmente, se concluye que, para la rodilla en la vista coronal las diferencias entre ambas herramientas, aunque estadísticamente significativas, son mínimas.

Después de analizar todas las perspectivas, se observó que el uso de vistas bidimensionales independientes, si bien permitió obtener resultados precisos y accesibles, presentó ciertas limitaciones al momento de representar movimientos con componentes en profundidad. Por ejemplo, gestos como las rotaciones del pie no pudieron analizarse con total fidelidad al depender exclusivamente del plano observado. Esta limitación no impidió la validación de la herramienta, pero resalta

la importancia de integrar una representación más completa del movimiento en futuros desarrollos.

Conclusiones

En este trabajo se desarrolló un sistema de análisis biomecánico para el gesto del saque de piso de los arqueros, sin la utilización de marcadores, lo que evitó interferencias con el movimiento natural del deportista.

El protocolo de grabación implementado permitió identificar los parámetros relevantes para el análisis biomecánico del gesto técnico utilizando recursos accesibles, como cámaras no especializadas y dispositivos móviles. Este enfoque representa una ventaja significativa para la replicabilidad del proyecto en distintos contextos.

Además, se cumplió el objetivo de desarrollar una técnica de corrección de desviaciones en los puntos clave durante la ejecución del movimiento. Esta permitió ajustar inconsistencias entre fotogramas consecutivos, manteniendo la coherencia espacial y temporal de los datos. Aunque se identificaron dificultades en casos de solapamiento entre articulaciones, la técnica resultó efectiva en la mayoría de las situaciones, mejorando la estabilidad del análisis biomecánico a partir de video.

Los resultados obtenidos en las distintas vistas confirman que el algoritmo desarrollado presenta una alta concordancia con las mediciones realizadas en Kinovea. La correlación de Pearson alcanzó valores de hasta $r = 0{,}992$, evidenciando una relación lineal extremadamente fuerte entre ambas herramientas. Asimismo, las diferencias promedio entre las mediciones realizadas por ambas herramientas fueron, en su mayoría, inferiores a $3°$, salvo en el ángulo del pie en la vista lateral derecha, donde la complejidad del movimiento elevó la diferencia promedio a aproximadamente $7{,}5°$.

Si bien se identificaron diferencias estadísticamente significativas, en la mayoría de los casos estas fueron de baja magnitud y atribuibles a factores específicos como oclusiones, errores en la anotación manual o la calidad del video. En conjunto, estos resultados respaldan la fiabilidad y aplicabilidad del sistema desarrollado para el análisis biomecánico del saque de piso en arqueros.

Finalmente, se identificó como limitación del presente trabajo el uso de vistas bidimensionales independientes para el análisis del gesto técnico. Aunque esta metodología permitió validar la herramienta y obtener resultados confiables utilizando recursos accesibles, restringe la capacidad de representar movimientos con componentes en profundidad, lo cual puede limitar la interpretación de ciertos patrones biomecánicos complejos que ocurren fuera del plano observado (como las rotaciones del pie).

Mejoras para Trabajos Futuros

En primer lugar, aunque el uso de cámaras convencionales responde al objetivo de diseñar una herramienta de bajo costo, se evidenciaron limitaciones en la calidad de imagen, como la pérdida de información entre fotogramas y distorsiones en algunas regiones del cuerpo, particularmente en el pie. Por ello, se recomienda que, para aplicaciones que requieran un análisis más preciso, se utilicen cámaras con mayor resolución y con una tasa de cuadros por segundo más alta.

En segundo lugar, se identificó que la vestimenta del deportista influye en la precisión del algoritmo. La ropa suelta dificultó la detección de algunas articulaciones. Se sugiere, por tanto, que los deportistas utilicen prendas ajustadas al cuerpo, como licras o pantalonetas, con el fin de mejorar la visibilidad de los puntos clave durante el análisis.

Una tercera mejora, en respuesta a la limitación previamente mencionada, consiste en implementar un sistema de emparejamiento de puntos en tres dimensiones. Esta herramienta permitiría fusionar de manera coherente la información proveniente de múltiples vistas, generando una reconstrucción tridimensional del movimiento. Al trabajar en un espacio 3D (X,Y,Z), sería posible analizar con mayor fidelidad movimientos que incluyen desplazamientos diagonales y rotaciones, los cuales no pueden observarse completamente en 2 dimensiones.

Además, la reconstrucción 3D permitiría calcular ángulos articulares, trayectorias y distancias entre articulaciones con mayor precisión, ya que considera la orientación real de los segmentos en el espacio tridimensional (a diferencia del análisis 2D, donde los ángulos pueden verse afectados por la perspectiva de la cámara). Esto permite una medición más fiel de las rotaciones articulares, especialmente aquellos movimientos complejos que combinan planos. Este avance también abriría la posibilidad de integrar variables biomecánicas con mayor grado de confianza, además de generar visualizaciones más completas e interactivas del gesto técnico.

Siguiendo la mejora planteada anteriormente, se debería incorporar una herramienta de validación tridimensional que permita integrar la coordenada z en el análisis. Esto permitiría realizar comparaciones más completas con sistemas de captura de movimiento avanzados.

Asimismo, se observó que el tiempo de procesamiento del algoritmo podría representar una barrera en contextos donde se requiere retroalimentación inmediata. Como mejora, se sugiere optimizar el rendimiento del sistema para reducir los tiempos de ejecución.

Otra propuesta de mejora consiste en implementar un sistema de gestión de historial para cada arquero, que almacene los resultados obtenidos en diferentes sesiones de análisis. De esta manera, sería posible generar reportes automáticos que indiquen la evolución del deportista a lo largo del tiempo, identificando tendencias de mejora o retrocesos en parámetros biomecánicos específicos. Esta funcionalidad permitiría realizar un seguimiento continuo y estructurado del rendimiento individual, apoyando la toma de decisiones en los procesos de entrenamiento.

Finalmente, se recomienda ampliar la aplicación del sistema a otros gestos técnicos del

arquero, como el saque de volea y el saque de mano. Esto permitiría validar la versatilidad del algoritmo en diferentes contextos del juego y aumentar su impacto como herramienta de análisis técnico.

Anexos

A continuación se presenta el repositorio del proyecto, el cual contiene todo el código fuente, el consentimiento informado, documentación adicional y recursos utilizados durante el desarrollo del trabajo de grado.

- **Repositorio GitHub:** https://github.com/lauraSSB/AnalisisBiomecanicoArqueros.git

Referencias

Ahmad, J. (2019). Human pose recognition using deep learning. *Journal of Artificial Intelligence Research*.

Albers, K. (1993). *The Biomechanical Analysis of Two Types of Place Kicks: The Toe and Instep Kicks* [Tesis de maestría, Western Michigan University]. https://scholarworks.wmich.edu/masters_theses/3310/

Alcock, A. M., Gilleard, W., Hunter, A. B., Baker, J., & and, N. B. (2012). Curve and instep kick kinematics in elite female footballers [PMID: 22214481]. *Journal of Sports Sciences*, *30*(4), 387-394. https://doi.org/10.1080/02640414.2011.643238

Asociación Colombiana de Universidades (ASCUN). (2019). XXVIII Juegos Universitarios Nacionales Barranquilla 2019.

Baeldung. (2023). Understanding channels in CNNs [Accessed: 2025-04-16]. https://www.baeldung.com/cs/cnn-channels

Bazarevsky, V., Grishchenko, I., & Bazavan, E. G. (2021). Model card: BlazePose GHUM 3D. https://storage.googleapis.com/mediapipe-assets/Model%20Card%20BlazePose%20GHUM%203D.pdf

Bazarevsky, V., Zhu, T., Grishchenko, I., Zhang, F., Raveendran, K., & Grundmann, M. (2020). BlazePose: On-device real-time body pose tracking. https://arxiv.org/abs/2006.10204

Blanco Ortega, A., Isidro Godoy, J., Szwedowicz Wasik, D. S., Martínez Rayón, E., Cortés García, C., Ramón Azcaray Rivera, H., & Gómez Becerra, F. A. (2022). Biomechanics of the Upper Limbs: A Review in the Sports Combat Ambit Highlighting Wearable Sensors. *Sensors (Basel, Switzerland)*, *22*(13), 4905. https://doi.org/10.3390/s22134905

Centro Nacional de Consultoría (CNC). (2023). El poder del fútbol. https://encuestacnc.com/attachments/article/108/El_poder_del_futbol.pdf

Charmant, J. (2023). Kinovea: Video analysis software [Computer software] [Versión 0.9.5, software de código abierto para análisis de movimiento en 2D.]. https://www.kinovea.org/

Cheuque, C., Querales, M., León, R., Salas, R., & Torres, R. (2022). An efficient multi-level convolutional neural network approach for white blood cells classification. *Diagnostics*, *12*(2). https://doi.org/10.3390/diagnostics12020248

Chorro, D. (2016). Lesiones en el fútbol: Diagnóstico, tratamiento y prevención. https://www.davidchorro.com/lesiones-en-el-futbol

Colyer, S. L., Evans, M., Cosker, D. P., & Salo, A. I. T. (2018). A review of the evolution of vision-based motion analysis and the integration of advanced computer vision methods towards developing a markerless system. *Sports Medicine - Open*, *4*, 24. https://doi.org/10.1186/s40798-018-0139-y

Elliott, B. C., Alderson, J. A., & Denver, E. R. (2007). System and modelling errors in motion analysis: Implications for the measurement of the elbow angle in cricket bowling. *Journal of Biomechanics*, (12). https://doi.org/10.1016/j.jbiomech.2006.12.012

Field Hockey Canada. (2021). Four phases of movement and making intervention: FHC coach resource [Version 1.0]. *Field Hockey Canada Resources*, *1*.

Google AI Edge. (2024). Pose landmark detection guide. https://ai.google.dev/edge/mediapipe/solutions/vision/pose_landmarker

HomeCourt Soccer. (s.f.). HomeCourt — Football Performance App. https://soccerstars.com/homecourt/

Jammalamadaka, N., Zisserman, A., & Jawahar, C. (2017). Human pose search using deep networks. *Image and Vision Computing*, *59*.

Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., & Bethge, M. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, *21*(9), 1281-1289.

Mündermann, L., Corazza, S., & Andriacchi, T. P. (2006). The evolution of methods for the capture of human movement leading to markerless motion capture for biomechanical applications. *Journal of NeuroEngineering and Rehabilitation*. https://doi.org/10.1186/1743-0003-3-6

Nielsen. (2024). Paris 2024 Olympic Games Hub. https://www.nielsen.com/es/data-center/paris-2024-olympic-games-hub/

Optics, E. (s.f.). 6 fundamental parameters of an imaging system [Accessed: 2024-10-06]. https://www.edmundoptics.com/knowledge-center/application-notes/imaging/6-fundamental-parameters-of-an-imaging-system/

Orloff, H., Sumida, B., Chow, J., Habibi, L., Fujino, A., & Kramer, B. (2008). Ground reaction forces and kinematics of plant leg position during instep kicking in male and female collegiate soccer players. *Sports Biomechanics*, *7*(2). https://doi.org/10.1080/14763140701841704

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2019). MobileNetV2: Inverted Residuals and Linear Bottlenecks. https://arxiv.org/abs/1801.04381

Sportlogiq. (s.f.). Next-Generation Sports Analytics. https://www.sportlogiq.com

Track160. (s.f.). AI-Powered Football Analytics. https://www.track160.com

Ultralytics. (2024). Ultralytics YOLOv11. https://docs.ultralytics.com/models/yolo11/

Xiao, J., Li, H., Qu, G., Fujita, H., Cao, Y., Zhu, J., & Huang, C. (2022). HOPE: Heatmap and offset for pose estimation. *Journal of Ambient Intelligence and Humanized Computing*, *13*(6). https://doi.org/10.1007/s12652-021-03124-w

Xu, H., Bazavan, E. G., Zanfir, A., Freeman, W. T., Sukthankar, R., & Sminchisescu, C. (2020). GHUM & GHUML: Generative 3D human shape and articulated pose models. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*,

6184-6193. https://openaccess.thecvf.com/content_CVPR_2020/html/Xu_GHUM_GHUML_Generative_3D_Human_Shape_and_Articulated_Pose_CVPR_2020_paper.html

Zurister. (2019). Depth-wise convolution and depth-wise separable convolution [Medium article]. https://medium.com/@zurister/depth-wise-convolution-and-depth-wise-separable-convolution-37346565d4ec