



UNIVERSIDAD NACIONAL DEL SUR

TESIS DE DOCTORADO EN INGENIERÍA

**VISIÓN ARTIFICIAL EN LA GESTIÓN Y APOYO A LA SEGURIDAD
DE LOS TRABAJADORES**

Ing. Manlio Miguel Massiris Fernández

BAHÍA BLANCA

ARGENTINA

2021

Prefacio

Esta Tesis es presentada como parte de los requisitos para optar al grado académico de Doctor en Ingeniería, de la Universidad Nacional del Sur, y no ha sido presentada previamente para la obtención de otro título en esta Universidad u otras. La misma contiene los resultados obtenidos en investigaciones llevadas a cabo en el Laboratorio de Ciencias de las Imágenes (UNS - CONICET) durante el período comprendido entre el día 20 de Septiembre de 2016 y el 20 de agosto de 2021, bajo la dirección del Dr. Claudio Delrieux, Profesor Titular del Departamento de Ingeniería Eléctrica e Investigador Independiente del CONICET, y bajo la codirección del Dr. Alejandro Vitale, Profesor Adjunto del Departamento de Ingeniería Eléctrica, Profesor del Departamento de Geografía y Turismo en la Carrera de Licenciatura en Oceanografía e Investigador Adjunto del CONICET.

Manlio Miguel Massiris Fernández

manlio.massiris@uns.edu.ar

DEPARTAMENTO DE INGENIERÍA ELÉCTRICA Y DE COMPUTADORAS

UNIVERSIDAD NACIONAL DEL SUR

Bahía Blanca, 20 de agosto de 2021



UNIVERSIDAD NACIONAL DEL SUR
Secretaría General de Posgrado y Educación Continua

La presente tesis ha sido aprobada el / / , mereciendo la calificación de(.....)

RESUMEN

Las evaluaciones de riesgos y seguridad en el trabajo generalmente se han realizado con observaciones efectuadas *in situ* por personal especializado. Esta evaluación de riesgos tradicional es costosa e ineficaz, especialmente en los países en desarrollo, donde hay una necesidad de recursos humanos más capacitados. Además, el estado del arte resalta que incluso los especialistas calificados carecen de precisión intra- e inter-observador, pues a menudo se equivocan en juicios de riesgos principalmente debido a sesgos subjetivos o condiciones visuales subóptimas en el lugar de trabajo, como ser iluminación reducida, oclusiones de equipos, auto-occlusiones y ángulos de video inadecuados.

En esta tesis presentamos diversas propuestas, basadas en el uso de visión artificial, para facilitar la toma de decisiones, estandarizar el proceso de evaluación y reducir el tiempo requerido para estimar el riesgo ergonómico y cuantificar el uso de equipos de protección individual. En el primer capítulo se presenta la introducción a las tecnologías de la Industria 4.0 como marco básico de requerimiento de soluciones a los problemas antes mencionados. El segundo capítulo detalla las soluciones propuestas desde la visión artificial y las redes neuronales para la estimación del riesgo ergonómico. En el tercer capítulo se presentan dos soluciones basadas en redes neuronales para la inspección y cuantificación del uso de equipos de protección personal. Finalmente, se presentan las conclusiones y el trabajo futuro. Los resultados indicaron que los métodos propuestos facilitan el proceso de evaluación de riesgos y condiciones de trabajo en aplicaciones reales en entornos desafiantes, utilizando entre otras fuentes videos obtenidos por medio de cámaras deportivas egocéntricas, teléfonos inteligentes y drones.

Abstract

Risk and safety assessment at worksites is generally carried out through on-site observations performed by specialized personnel. This traditional risk assessment procedure is costly and ineffective, especially in underdeveloped countries, where specifically trained human resources are scarce and expensive. Also, the state-of-art points out that even qualified specialists lack intra- and inter-observer precision, and often err on risk judgments, mainly due to subjective biases or workplace sub-optimal visual conditions, such as reduced illumination, equipment occlusions, self-occlusions, and inadequate video angles.

In this thesis we present several computer-vision- based solutions aimed to facilitate decision-making, to standardize the evaluation process, and to reduce the amount of time required for estimating ergonomic risk and quantifying the use of personal protective equipment. In the first chapter, the introduction to Industry 4.0 technologies is presented as a basic framework requiring solutions to the problems mentioned. The second chapter details the proposed solutions based on computer vision and neural networks for the estimation of ergonomic risk. In the third chapter, two solutions based on neural networks for the inspection and quantification of the use of personal protective equipment are presented. Finally, conclusions and future work are presented. The results indicated that the methods facilitated the process of assessing risks and working conditions in real applications in challenging environments, using video recorded with sports egocentric cameras, smartphones, and drones.

AGRADECIMIENTOS

- Al Pueblo Argentino por haberme abierto sus brazos y aceptado como su hijo.

ACLARACIONES

- En esta tesis se utilizará el término *Riesgo Ergonómico*, ya que la evidencia de utilización del término *Riesgo Disergonómico* solo se encuentra en revistas de bajo impacto en español. A lo largo de esta tesis se citan múltiples artículos de alto impacto, de los cuales ninguno emplea dicho término.
- En esta tesis se utilizará el término *trabajadores*, como un concepto genérico para no dificultar la lectura.

Índice general

Índice de Figuras	XII
1. Introducción	1
1.1. Problemáticas a abordar	3
1.2. Estimación de Riesgos Ergonómicos en la Industria 4.0	4
1.2.1. Soluciones propuestas	4
1.3. Monitoreo de utilización de EPP en la Industria 4.0	5
1.4. Diseminación de los resultados	8
1.4.1. Artículos en revistas indexadas	8
1.4.2. Artículos en conferencias internacionales	8
2. Estimación de riesgos Ergonómicos	9
2.1. Antecedentes de la problemática	9
2.1.1. TMEs en La Argentina	10
2.2. Evaluación de riesgos ergonómicos	12
2.3. Estimación de riesgos Ergonómicos mediante la red neuronal convolucio- nal OpenPose y el método RULA	14
2.3.1. Descripción de la metodología RULA	14
2.3.2. Descripción de Red Neuronal OpenPose	15
2.3.3. Estimación del ángulo de la articulación	16
2.3.4. Cálculo de puntuaciones de RULA	17

2.3.5. Métodos de evaluación	19
2.3.6. Pruebas en entorno controlado	22
2.3.7. Resultados en videos de secuencias de trabajo	37
2.4. Estimación de riesgos Ergonómicos mediante la red neuronal convolucio- nal VIBE y el método REBA	44
2.4.1. Descripción de Red Neuronal STAF	44
2.4.2. Descripción de Red Neuronal VIBE	44
2.4.3. Descripción de la metodología REBA.	45
2.4.4. Dataset	45
2.5. Resultados	47
3. Detección de equipos de protección personal	51
3.1. Antecedentes de la problemática	51
3.2. Detección de equipos de protección personal mediante la red neuronal convolucional YOLO	54
3.2.1. Resultados	57
3.3. Detección de equipos de protección personal mediante las redes neuro- nales convolucionales OpenPose y Resnet	65
3.3.1. Descripción de Red Neuronal OpenPose	65
3.3.2. Descripción de Red Neuronal Resnet	66
3.3.3. Implementación de la metodología	66
3.3.4. Umbrales y ROI	68
3.3.5. Conjunto de datos y entrenamiento	69
3.3.6. Resultados experimentales	72
4. Discusión, conclusiones y trabajo futuro	77
4.1. Aportes de esta tesis	77
4.2. Limitaciones de los resultados	79
4.3. Trabajo futuro	81

Índice de figuras

1.1. Trabajadores que experimentaron cambios tecnológicos en su puesto de trabajo en Argentina.	2
1.2. Percepción sobre el cambio tecnológico en el lugar de trabajo.	2
1.3. Temas de investigación en sistemas de la industria 4.0 centrados en el ser humano (Rauch, Linder, y Dallasega, 2019).	3
2.1. Zonas del cuerpo donde los trabajadores argentinos expresaron algún tipo de malestar relacionado con su trabajo.	11
2.2. Indicadores de bienestar o situaciones nocivas según trabajadores expuestos a posturas forzadas (Gadea y cols., 2019).	11
2.3. Porcentaje de valoraciones con errores y categoría de error. (Diego-Mas, Alcaide-Marzal, y Poveda-Bautista, 2017).	13
2.4. articulaciones del esqueleto de OpenPose 25.	16
2.5. Regiones de oclusión (Gupta, Mittal, y Davis, 2006).	20
2.6. Semi-esfera 3D de Hammersley.	21
2.7. Las imágenes de la persona simulada en el dataset.	21
2.8. Confianza \bar{p} en la detección del esqueleto frente a la cámara vs (a) latitud y (b) azimut.	22
2.9. La confianza (escala en el eje radial) de la detección de punto antropométrico en función del azimuth de la cámara para (a,b) nariz, (c,d) cuello y (e,f) hombro izquierdo.	24
2.10. La confianza de la detección de punto antropométrico en función del azimuth de la cámara para (a,b) codo izquierdo, (c,d) mano izquierda y (e,f) hombro derecho.	25

2.11. La confianza de la detección de punto antropométrico en función del azimuth de la cámara para (a,b) codo derecho, (c,d) mano derecha y (e,f) centro de cadera.	26
2.12. La confianza de la detección de punto antropométrico en función del azimuth de la cámara para (a,b) cadera izquierda, (c,d) rodilla izquierda y (e,f) tobillo izquierdo.	27
2.13. La confianza de la detección de punto antropométrico en función del azimuth de la cámara para (a,b) cadera derecha, (c,d) rodilla derecha y (e,f) pie derecho.	28
2.14. La confianza de la detección de punto antropométrico en función del azimuth de la cámara para (a,b) ojo izquierdo, (c,d) ojo derecho y (e,f) oído izquierdo.	29
2.15. La confianza de la detección de punto antropométrico en función del azimuth de la cámara para (a,b) oído derecho, (c,d) dedo pulgar del pie derecho y (e,f) dedo meñique del pie derecho.	30
2.16. La confianza de la detección de punto antropométrico en función del azimuth de la cámara para (a,b) talón del pie derecho, (c,d) dedo pulgar del pie izquierdo y (e,f) dedo meñique del pie izquierdo.	31
2.17. La confianza de la detección de punto antropométrico en función del azimuth de la cámara para (a,b) talón del pie izquierdo	32
2.18. Prueba de laboratorio de torsión del tronco: (a) esqueleto OpenPose 25 superpuesto en una persona, (b,c) muestras de la simulación del conjunto de datos renderizados en 3D y esqueletos computados y (d) gráfico de ángulos del tronco inferidos TFA (naranja) y TTA (azul) vs fotograma, la acción de torsión ejecutada una vez para cada lado del cuerpo.	34
2.19. Prueba de laboratorio de flexión del tronco: (a) esqueleto OpenPose 25 superpuesto en una persona y (b,c) muestras de la simulación del conjunto de datos renderizados en 3D y esqueletos computados y (d) gráfico de ángulos del tronco inferidos TFA (naranja) y TTA (azul) vs fotograma, la acción de torsión ejecutada una vez para cada lado del cuerpo.	35
2.20. Muestras de la secuencia de trabajos con signos de Marshall.	36
2.21. Muestras de la secuencia de trabajos de revoque de paredes.	40
2.22. Muestras de secuencia de trabajo de martilleo.	41
2.23. Muestras de secuencia de trabajo de tala de árboles.	42
2.24. Muestras de la secuencia del trabajo de perforación.	43

2.25. Metodología propuesta.	44
2.26. Fotogramas de un ejemplo de simulación tomados con una posición de cámara aleatoria.	46
2.27. Error del ángulo del tronco vs altura y azimuth de la cámara.	47
2.28. Error del ángulo del cuello vs altura y azimuth de la cámara.	47
2.29. Error del ángulo del codo izquierdo vs altura y azimuth de la cámara.	48
2.30. Error del ángulo del codo derecho vs altura y azimuth de la cámara.	48
2.31. Error del ángulo del hombro izquierdo vs altura y azimuth de la cámara.	48
2.32. Error del ángulo del hombro derecho vs altura y azimuth de la cámara.	49
2.33. Error del ángulo de la rodilla izquierda vs. altura y azimuth de la cámara.	49
2.34. Error del ángulo de la rodilla derecha vs. altura y azimuth de la cámara.	49
3.1. Utilización de EPP de los trabajadores que expresan necesitarlos (Gadea y cols., 2019).	52
3.2. Utilización de EPP de los trabajadores que expresan necesitarlos según tamaño de empresa (Gadea y cols., 2019).	52
3.3. Utilización de EPP de los trabajadores que expresan necesitarlos según sector productivo.	53
3.4. Proceso Propuesto, de izquierda a derecha: (1) dividir en bloques de 13x13, (2) hacer predicciones con YOLO re-entrenado y (3) establecer umbrales para obtener solo las detecciones más confiables.	54
3.5. Curvas precisión-sensibilidad para cada una de las clases.	57
3.6. Ejemplo de verdadero Positivo en el fotograma 17956.	58
3.7. Ejemplo de falso negativo en el fotograma 17871.	58
3.8. Ejemplo de falso positivo en el fotograma 02376.	59
3.9. Ejemplo de etiquetado múltiple en el fotograma 17786.	59
3.10. Ejemplo de etiquetado múltiple en el fotograma 00015.	60
3.11. Ejemplo de etiquetado múltiple en el fotograma 20448.	60
3.12. Ejemplo de impacto del rango visual 00187.	61

3.13. Ejemplo de impacto del rango visual 00095	61
3.14. Ejemplo de impacto de las oclusiones 15950.	62
3.15. Ejemplo de impacto de las oclusiones 19532.	62
3.16. Ejemplo de impacto de las oclusiones 19560.	63
3.17. Ejemplo de impacto de la postura individual y de la distorsión generada por el tipo de lente 19764.	63
3.18. impacto de la postura individual y de la distorsión generada por el tipo de lente 18141.	64
3.19. impacto de la postura individual y de la distorsión generada por el tipo de lente 18380.	64
3.20. Puntos antropométricos en OP (Cao, Simon, Wei, y Sheikh, 2017).	65
3.21. Determinación de las regiones de interés.	66
3.22. Puntos antropométricos de <i>E.Rca</i>	68
3.23. <i>Dataset E.R^{ca}</i> con etiquetas de “Casco” (a)-(e) No y (f)-(j) Sí.	69
3.24. <i>Dataset E.R^{tr}</i> con etiquetas de “Chaleco” (a)-(f) No y (g)-(l) Sí.	70
3.25. Resultados del video de evaluación 1 <i>Dron</i>	73
3.26. Resultados del video de evaluación 2 <i>Atardecer</i>	74
3.27. Resultados del video de evaluación 3 (<i>smartphone</i>).	75

Capítulo 1

Introducción

La irrupción de las industrias 4.0 ha llegado al ámbito académico. Sin embargo, hay poca evidencia de su continua aplicación en las industrias Latinoamericanas, por lo cual son un nicho de investigación creciente y que evoluciona a pasos agigantados. La reducción en los costos del hardware y la democratización de las tecnologías de software ha permitido vislumbrar un futuro guiado por los sistemas colaborativos, donde los trabajadores como agentes se integran a sistemas inteligentes ciberfísicos (Rauch y cols., 2019).

Esta cuarta revolución industrial podría señalar una mejoría de calidad de los procesos productivos al vincular trabajadores, máquinas y materiales, formando un nuevo sistema de producción que permite un intercambio de información más rápido y específico a lo largo de la cadena de suministro (Rauch y cols., 2019). Como actores en este proceso de cambio de paradigma, los científicos deberíamos tener una visión holística de las interacciones socio-tecnológicas, para plantear soluciones y aplicar nuevos enfoques centrados en el ser humano, para facilitar la adopción de las nuevas tecnologías digitales y diseminar los sistemas de trabajo habilitados para la industria 4.0 (Kadir, Broberg, y Souza da Conceição, 2019).

Para alentar el proceso de integración de trabajadores, equipo y materiales han surgido los sistemas ciberfísicos. Dichos sistemas tienen como meta la integración de procesos donde los computadores y redes de sensores monitorean y controlan los procesos físicos. Generalmente la integración se logra a lo largo de varios ciclos de retroalimentación donde los procesos físicos afectan la toma de decisiones y viceversa (Igelmo, Syberfeldt, Högberg, García Rivera, y Pérez Luque, 2020).

Particularmente en la Argentina se tiene poca evidencia de la integración de las tecnologías 4.0 al entramado productivo. Como se muestra en la Figura 1.1, en la última Encuesta Nacional de Trabajadores sobre Condiciones de Empleo, Trabajo, Salud y Seguridad de la Superintendencia de Riesgos de Trabajo (Gadea y cols., 2019), el 80.3% de los trabajadores argentinos expresaron no haber presenciado la adopción de nuevas tecnologías en su lugar de trabajo. Además, de los trabajadores que sí manifestaron te-

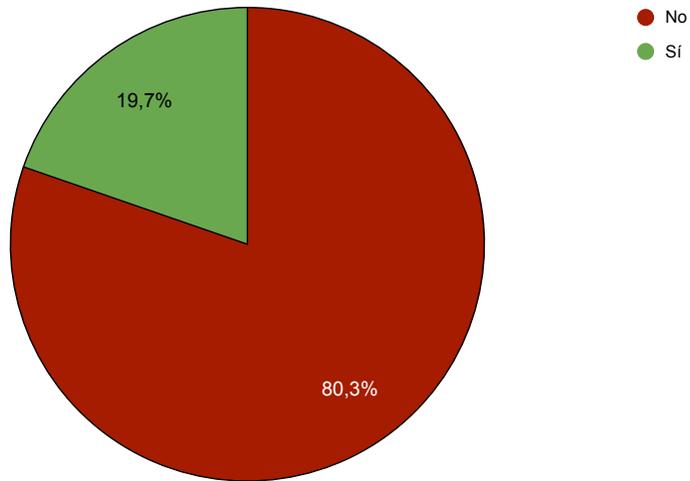


Figura 1.1: Trabajadores que experimentaron cambios tecnológicos en su puesto de trabajo en Argentina.

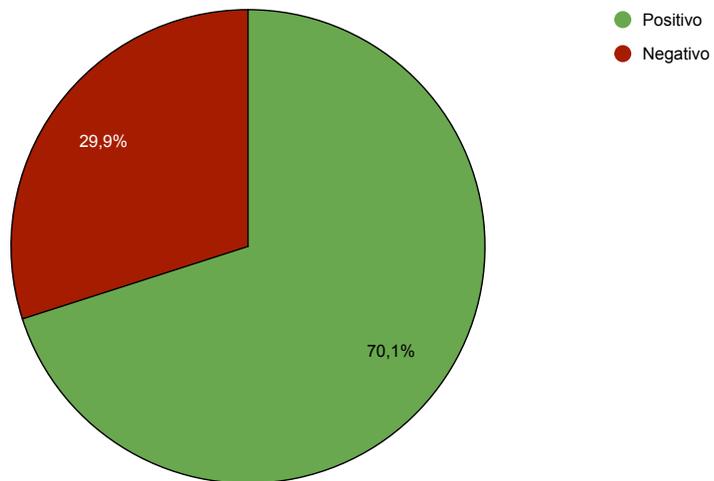


Figura 1.2: Percepción sobre el cambio tecnológico en el lugar de trabajo.

ner algún cambio tecnológico en su trabajo, dos de cada tres expresaron percibir dicho cambio tecnológico como positivo (ver Figura 1.2), lo cual abre claras posibilidades para nuevos desarrollos tecnológicos en la Argentina.

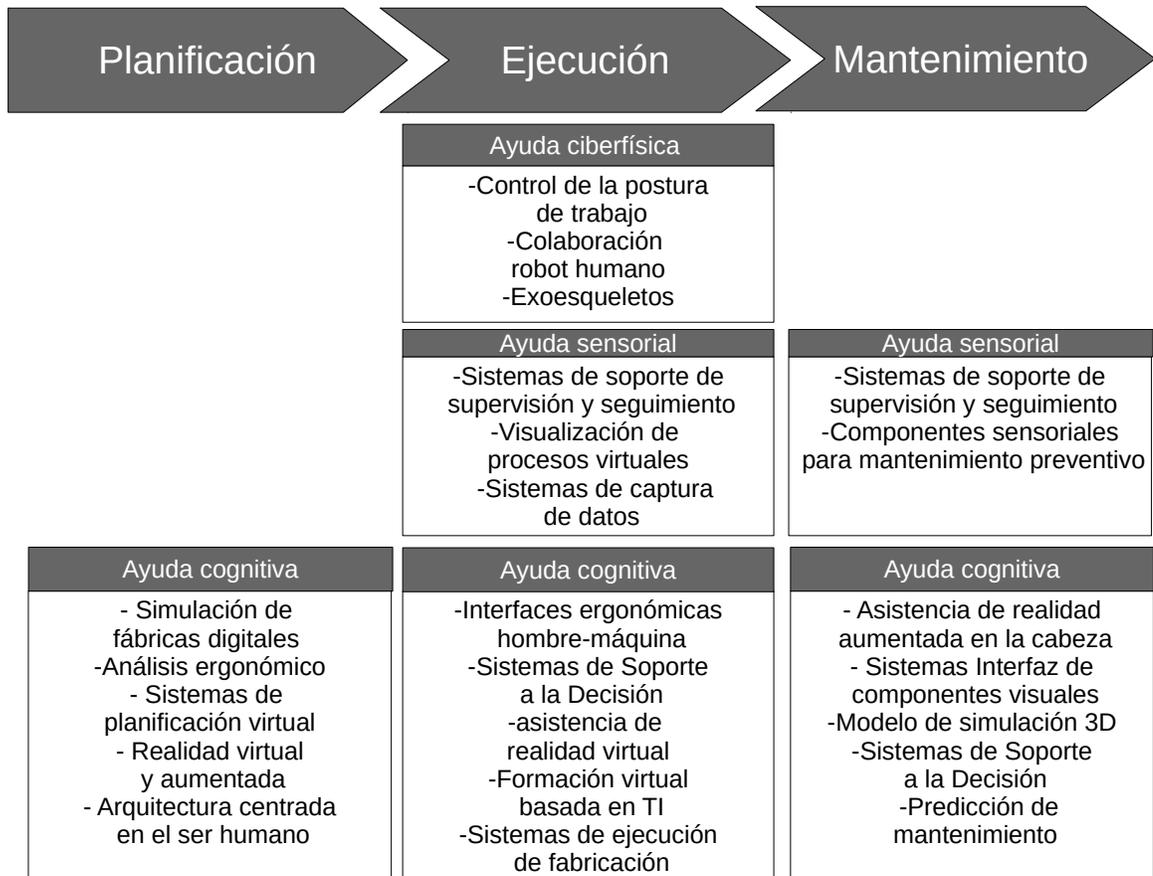


Figura 1.3: Temas de investigación en sistemas de la industria 4.0 centrados en el ser humano (Rauch y cols., 2019).

1.1. Problemáticas a abordar

Entre las investigaciones enmarcadas dentro de la industria 4.0, el análisis de los riesgos laborales ha crecido recientemente como una sub-área de interés, donde nuevas tecnologías son diseñadas para la ayudar en la reducción de accidentes laborales. Dichas tecnologías hacen énfasis en la reducción de riesgos específicos a lo largo de las diferentes etapas de un ciclo productivo. Un ciclo productivo integrado a la industria 4.0 puede ser dividido en las etapas de planeamiento, ejecución y mantenimiento. Las herramientas de la industria 4.0 realizan medición continua durante las etapas de ejecución y mantenimiento y facilitan con evidencia cuantitativa el proceso de toma de decisiones en la etapa de planeamiento (ver Figura 1.3).

En esta tesis nos enfocaremos en proponer soluciones a dos problemáticas de investigación:

- Sistemas de apoyo al proceso de estimación de riesgos ergonómicos (ERE), cuya meta es ofrecer soporte a todas las etapas del sistema productivo.
- Sistemas para el monitoreo de utilización de equipo de protección personal (EPP) mediante visión artificial, relacionados con ejecución y el mantenimiento del sistema productivo, por medio de ayuda sensorial.

1.2. Estimación de Riesgos Ergonómicos en la Industria 4.0

El impacto de la recopilación y el análisis automatizados de datos en la industria 4.0 está dando forma a un nuevo grupo de aplicaciones, donde los avances tecnológicos en sensores de hardware y software abren caminos novedosos para la ergonomía (Hegde y Rokseth, 2020; Kadir y cols., 2019; Rauch, Linder, y Dallasega, 2020).

Los métodos utilizados en la Industria 4.0 están soportados en sensores inerciales, textiles inteligentes y visión por computadora, o una mezcla de ellos (Hegde y Rokseth, 2020; Kadir y cols., 2019; Rauch y cols., 2020). Podemos mencionar, entre otras tecnologías, el uso de redes de sensores inerciales y electrogoniómetros corporales (Mura y cols., 2012), sensores RGB-D (Diego-Mas, Poveda-Bautista, y Garzon-Leal, 2015; Krüger y Nguyen, 2015; Plantard, Shum, Le Pierres, y Multon, 2017; Xu y McGorry, 2015), inclinómetros y acelerómetros (Lee, Seto, Lin, y Migliaccio, 2017; Nath, Akhavian, y Behzadan, 2017; Peppoloni, Filippeschi, Ruffaldi, y Avizzano, 2014), cámaras estéreo (Liu, 2019) e incluso cámaras RGB (Figlali y cols., 2015; Yan y cols., 2017; Zhang, Yan, y Li, 2018). Además, debido al rápido desarrollo de los métodos y técnicas de visión por computadora (CV) en la última década, la identificación y el análisis de la postura humana se pueden automatizar para ayudar en los métodos de estimación de riesgos ergonómicos (ERE) basados en la observación de elección (Battini, Persona, y Sgarbossa, 2014; Vignais, Bernard, Touvenot, y Sagot, 2017; Yan y cols., 2017; Zhang y cols., 2018).

1.2.1. Soluciones propuestas

En el capítulo 2 seguimos las ideas propuestas recientemente por Yan y cols. (2017) y Zhang y cols. (2018), que utilizan cámaras RGB combinadas con modelos de aprendizaje de máquina para generar estimaciones de ERE siguiendo en video el sistema de evaluación de la postura de Ovako (2009). A diferencia de estas ideas pioneras nosotros proponemos la siguientes soluciones:

1. En Massiris Fernández, Fernández, Bajo, y Delrieux (2020) y en Massiris Fernández, Bajo, y Delrieux (2019) proponemos un método para calcular automáticamente las ERE siguiendo el método RULA de McAtamney y Corlett (1993) para la investigación de trastornos en las extremidades superiores relacionados con el

trabajo, dicho método funciona con fotos o secuencias de video analizadas utilizando la red neural OpenPose (Cao y cols., 2017). OpenPose estima los esqueletos de las personas presentes en una imagen. De todos los esqueletos obtenidos filtramos los datos ruidosos. Por ello descartamos los esqueletos que se detectan con un tamaño y una confianza que están por debajo de unos umbrales definidos por el usuario. Para cada esqueleto en una imagen se calculan los ángulos de articulaciones como estipula el método RULA. Estos ángulos se utilizan posteriormente para calcular el RULA Grand Score (McAtamney y Corlett, 1993) de acuerdo con criterios establecidos.

2. En Massiris Fernández, Bajo, Martinez, Fernández, y Delrieux (2020) y en Massiris Fernández, Bajo, Fernández, y Delrieux (2020) incursionamos en una solución en cuatro dimensiones, utilizamos la red neuronal STAF (Raaj, Idrees, Hidalgo, y Sheikh, 2019) para la detección y seguimiento de trabajadores, y la red neuronal VIBE para la estimación del movimiento de las articulaciones corporales 3D a lo largo del tiempo. Adicionalmente, se utiliza el método REBA de McAtamney y Hignett (2000) para la evaluación ergonómica de todo el cuerpo.

1.3. Monitoreo de utilización de EPP en la Industria 4.0

El monitoreo de utilización de equipos de protección personal en los entornos laborales es un área de investigación de creciente interés en la temática de la industria 4.0, ya que sus resultados permiten mejorar en la toma de decisiones por parte de los expertos, y por ende, reducir el costo económico, social y ético generado por los accidentes laborales (Kelm y cols., 2013). Dicha problemática se ha abordado de dos formas, utilizando identificadores por radiofrecuencia (RFID) y aplicando visión artificial.

Identificadores RFID

Kelm y cols. (2013) han propuesto un sistema de puerto móvil para la identificación por radiofrecuencia RFID. Para verificar que solo el personal que utilice adecuadamente los EPP pueda ingresar al área protegida, se colocan antenas o lectores de RFID en la entrada del sitio de construcción, lo cual garantiza que los trabajadores que ingresan al sitio de construcción presenten el EPP. Sin embargo, esta propuesta solo asocia que el EPP marcado y la ID del trabajador se encuentran cercanos, pero no permite determinar si el EPP está usado, retenido o colocado en el suelo. Además, solo controla que se cumplan las normas en el ingreso al sitio de trabajo y no durante la ejecución de las actividades laborales.

Zhang, Yan, Li, Jin, y Fu (2019) proponen un sistema de cascos inteligentes que consta de una red de disparadores RFID a prueba de agua y recargable. En el casco monta un detector de rayos infrarrojos, un sensor térmico, una batería de litio, un receptor blue-

tooth con conectividad una aplicación para teléfono inteligente y una placa de circuitos embebidos. Además, en el controlador base propone una aplicación web y un servidor en la nube. Con todo esto el sistema es capaz de emitir una alarma en el celular y en el centro de mando si un trabajador no usa su casco por más de un segundo. El sistema es claramente complejo y de difícil implementación.

Sistemas de visión artificial

El interés por el uso de sistemas de visión artificial para prevenir riesgos laborales ha aumentado significativamente. Esto se debe, entre otras causas, al bajo costo y el uso masivo de cámaras digitales, la aparición de plataformas de código abierto que facilitan la creación rápida de prototipos de soluciones de visión por computadora, y el crecimiento de comunidades que comparten sus datos códigos y posibles soluciones a los errores (Guo, Yu, y Skitmore, 2017). En la historia del arte de la detección de trabajadores en el ambiente laboral se han utilizado características visuales para detectar equipos de construcción y trabajadores, como detección de bordes (Shrestha, Shrestha, Bajracharya, y Yfantis, 2015), el histograma de gradientes orientados (HOG), análisis en el modelo de color HSV y técnicas de aprendizaje como las maquinas de soporte vectorial (Memarzadeh, Golparvar-Fard, y Niebles, 2013; Park y Brilakis, 2012).

En video con cámara estática, los píxeles que no presentaban movimiento a lo largo del video eran filtrados con algoritmos de sustracción de fondo, para facilitar extraer las entidades en movimiento en una futura clasificación en tiempo real (Chi y Caldas, 2011). Estos procedimientos se realizaban como un primer paso en el seguimiento de las entidades que conformaban un espacio de trabajo, pues la utilización de EPP permitía distinguir a los trabajadores del resto de personas en el video (Konstantinou, Lasenby, y Brilakis, 2019; Mosberger, Andreasson, y Lilienthal, 2014; Park y Brilakis, 2016; Seong y cols., 2017). Sin embargo, no se aplicaron ampliamente en utilización de EPP, pues estaban limitadas a ciertos proyectos donde las condiciones visuales se mantienen constantes. Dichos trabajos fueron diseñados para monitorear el desempeño de los trabajadores de la construcción en la construcción en interiores y en la mayoría de sus pruebas se limitaron al laboratorio o a un espacio pequeño (Shrestha y cols., 2015). Por ello, recientemente han surgido desarrollos basados en redes neuronales, que ofrecen nuevas soluciones en la temática.

El aprendizaje profundo (DL, en inglés) permite que modelos computacionales compuestos por varias capas de procesamiento aprendan representaciones de datos en varios niveles de abstracción, y gracias a este concepto descubran representaciones precisas de forma autónoma en grandes volúmenes de datos (Nath, Behzadan, y Paal, 2020). El DL ha hecho recientemente grandes avances en el reconocimiento de imágenes y videos (Fang y cols., 2018). Un caso particular de DL es el de las redes convolucionales o redes neuronales convolucionales (CNN, en inglés) (Fukushima, 1988; LeCun, Bottou, Orr, y Müller, 1998), que actualmente constituyen el estado del arte de varios problemas de visión por computador, dado su buen desempeño, sus problemas de reconocimiento

e interpretación en imágenes y video (Fang y cols., 2018). Su capacidad para actuar de manera adecuada en estos contextos se basa en características fundamentales: conexiones locales, pesos compartidos, agrupación y uso de un gran número de capas (Fukushima, 1988).

El propósito de las CNNs es extraer una gran cantidad de características de una imagen, y luego usar esas características para detectar o clasificar los objetos en una imagen. Dependiendo de la necesidad emplean una gran cantidad de filtros, diferentes arquitecturas y parámetros que se pueden aprender en capas. El aprendizaje de una CNN, puede ser visto como un ajuste o optimización multivariada en función de los componentes de clasificación, con el objetivo final de minimizar el error (Fukushima, 1988; LeCun y cols., 1998).

La aparición de las CNN ha llevado a un rápido desarrollo en el campo de la detección de objetos. Por este motivo, destacaremos las CNN recientes en el campo del monitoreo de utilización de EPP. Fang y cols. (2018) utilizaron CNNs basadas en regiones (R-CNN) para detectar si un trabajador no llevaba casco. Mneymneh, Abbas, y Khoury (2017), en cambio, primero realizan el seguimiento de los trabajadores en video y después infieren si cada trabajador posee un casco en el área de la cabeza. Wu, Cai, Chen, Wang, y Wang (2019) propusieron un algoritmo basado en un detector de caja múltiple de disparo único (SSD) para detectar cascos, y del mismo modo, Xie, Liu, Li, y He (2018) utilizaron CNNs para detectar los cascos de los trabajadores. Sin embargo, hasta la fecha, se han realizado pocos estudios para identificar múltiples EPP, y por ende, no se aprovecha plenamente el poder de los algoritmos DL.

Soluciones propuestas

En el Capítulo 3 abordaremos dos procedimientos propuestos para el monitoreo de equipo de protección personal. Ambos se especializan en mediciones con cámara no-estática y en específico son propuestas para trabajar con videos generados en cámaras RGB sujetas al casco del trabajador o para cámaras sujetas a drones.

1. En Massiris Fernández, Fernández Muñoz, y Delrieux (2018) proponemos un sistema basado en la red neuronal YOLO Redmon y Farhadi (2017), que nos permite detectar objetos con gran eficiencia a través de imágenes, y a diferencia de los trabajos mencionados en la revisión bibliográfica, es capaz de detectar guantes, ropa de alta visibilidad, casco y también a los trabajadores, lo cual es novedoso, pues no se han encontrado evidencias de desarrollos que funcionen en visión por computador y que puedan detectar múltiples equipos de protección personal en el trabajador al mismo tiempo.
2. En Massiris Fernández, Bajo, Fernández Muñoz, y Delrieux (2020), proponemos un procesamiento en dos etapas. En la primera etapa se realiza la detección y localización del área de interés, mediante la red neuronal OpenPose (Cao y cols.,

2017), y en la segunda etapa un clasificador Resnet 152 (He, Zhang, Ren, y Sun, 2016) es re-entrenado para detectar presencia o ausencia de casco y chaleco del alta visibilidad.

1.4. Diseminación de los resultados

1.4.1. Artículos en revistas indexadas

- **Massiris Fernández, M.,** Fernández, J. Á., Bajo, J. M., y Delrieux, C. A. (2020). *Ergonomic risk assessment based on computer vision and machine learning*. Computers and Industrial Engineering, 149(November), 106816. <https://doi.org/10.1016/j.cie.2020.106816>
- **Massiris Fernández, M.,** Fernández, J. Á., Bajo, J., y Delrieux, C. (2020). *Sistema automatizado para monitorear el uso de equipos de protección personal en la industria de la construcción*. Revista Iberoamericana de Automática e Informática industrial, (Agosto), ISSN 1697-7920. <https://10.4995/riai.2020.13243>.

1.4.2. Artículos en conferencias internacionales

- **Massiris Fernández, M.,** Fernández Muñoz, J. Á., y Delrieux, C. (2018). *Detección de equipos de protección personal mediante la red neuronal convolucional YOLO*. Actas de Las XXXIX Jornadas de Automática, 1022–1029. Badajoz, España, 24 de noviembre de 2017.
- **Massiris Fernández, M.,** Bajo, J., Delrieux, C. (2019). *Aproximación a la evaluación de riesgos ergonómicos mediante visión por computadora*. XXVII Jornadas de Jovens Pesquisadores, Asociación de Universidades del Grupo Montevideo, São Carlos, Brasil, 23-25 de octubre de 2019.
- **Massiris Fernández, M.,** Bajo, J., Fernández, J. Á., Delrieux, C. (2020). *Joint Angle Estimation with VIBE: an Evaluation Using Virtual Avatars*. IEEE ARGENCON 2020, Ciudad de Resistencia, Chaco, República Argentina, 1-4 de Diciembre de 2020.
- **Massiris Fernández, M.,** Bajo, J., Martinez, S., Fernández, J. Á., Delrieux, C. (2020). *Vision-Based Ergonomic Risk Estimation: Deep-learning Strategies*. 9th International Workshop on Soft Computing Applications, SOFA 2020, Arad, Romania, 27-29th November 2020.

Capítulo 2

Estimación de riesgos Ergonómicos

2.1. Antecedentes de la problemática

La Organización Mundial de la Salud detalla que los trastornos musculoesqueléticos relacionados con el trabajo (TME) surgen cuando la tensión mecánica es mayor que la capacidad de carga de los componentes del sistema locomotor humano (Luttmann, Jager, y Griefahn, 2004). El gobierno Argentino a través de la Resolución 295/2003 (Ministerio de Trabajo, 2003), define los TME como los trastornos crónicos musculares, tendinosos y nerviosos derivados del trabajo. Este tipo de esfuerzo mecánico excesivo sobre las estructuras biológicas puede ocasionar diferentes consecuencias, como por ejemplo lesiones de músculos y tendones (distensiones o desgarros), ligamentos (esguinces o desgarros) y huesos (fracturas, microfracturas involuntarias, cambios degenerativos). También es posible que se produzcan irritaciones en el punto de inserción de los músculos y tendones y en la vaina del tendón, así como restricciones funcionales y procesos degenerativos tempranos de huesos, cartílagos, meniscos, vértebras, discos intervertebrales o articulaciones (Luttmann y cols., 2004).

Como respuesta, han surgido en las empresas los llamados los programas de evaluación de riesgos ergonómicos (ERE), cuyo fin es identificar y reducir factores de riesgo laboral. La palabra Ergonomía se deriva del griego ergon (trabajo), nomos (leyes) y el sufijo de cualidad. La Asociación Internacional del Ergonomía (IEA) define a la Ergonomía como la disciplina científica que se ocupa de la comprensión de las interacciones entre los seres humanos y otros elementos de un sistema. También puede ser definida como la profesión que aplica la teoría y los métodos para optimizar el bienestar humano y el rendimiento del sistema productivo (Wilson, 2000). La ergonomía tradicionalmente se ha trabajado desde tres perspectivas: la perspectiva física, la perspectiva cognitiva y la perspectiva organizacional. Sin embargo, es claro que para realizar el análisis ergonómico de un puesto de trabajo se debe aplicar un enfoque holístico, pues no es posible distinguir los límites entre las diferentes perspectivas (Boje, Guerriero, Kubicki, y Rezgui, 2020; David, 2005; Kadir y cols., 2019).

Los factores relacionados con los TMEs se suelen reducir aplicando la perspectiva física de la ergonomía. Dichos factores de riesgo pueden ser listados como la aplicación de fuerzas de alta intensidad, la manipulación de objetos pesados durante mucho tiempo, la manipulación frecuente y repetida de objetos, movimientos repetidos de las extremidades superiores, posición no erguida, esfuerzo muscular estático, inactividad muscular, movimientos repetitivos, vibraciones, factores relacionados con el entorno físico y las condiciones ambientales, y trabajo a altas o bajas temperaturas (Assuncao y Silva Abreu, 2017; Ha y cols., 2009; Luttmann y cols., 2004). Además, otras condiciones laborales insalubres como el ruido y los entornos cerrados, y factores psicosociales como la presión del tiempo o el apoyo social insuficiente, son factores que contribuyen a la naturaleza multicausal de estos trastornos (Assuncao y Silva Abreu, 2017; Luttmann y cols., 2004).

En la Unión Europea (UE), los TME son la principal causa (más del 50%) de las enfermedades relacionadas con el trabajo, y son responsables de más del 40% de todas las pérdidas económicas por problemas de salud y seguridad ocupacional (Bevan, 2015). En la misma línea, en los Estados Unidos de América (EE.UU.) (Gerr y cols., 2014; U.S. Bureau of Labor Statistics, 2015), estos trastornos son responsables de más del 30% de todas las enfermedades y accidentes no mortales. Por último, en Benavides y cols. (2018) se hace una comparación entre las condiciones de trabajo entre los países de América Latina, donde los países que reportaron mayor número de trabajadores expuestos a movimientos repetitivos en el trabajo (50% - 60%) son Argentina, Ecuador, Honduras y Nicaragua.

2.1.1. TMEs en La Argentina

En la Encuesta Argentina a trabajadores sobre Condiciones de Empleo, Trabajo, Salud y Seguridad (Gadea y cols., 2019), se han considerado varios atributos que definen la salud de la población trabajadora argentina. Para esto se les pidió a los trabajadores identificar áreas del cuerpo donde sienten malestar que atribuyen a posturas y/o esfuerzos derivados de su trabajo. Estas áreas de malestar se han agrupado en la Figura 2.1. En resumen el 52,4% de los trabajadores manifiestan sentir malestar en la zona de la espalda, y 3 de cada 10 sufren malestar en otras zonas del cuerpo consideradas. Finalmente, llama la atención que sólo el 28,5% no presenta ninguna manifestación de estas condiciones.

De la misma forma, en Gadea y cols. (2019) se menciona que en los trabajadores que realizan posturas corporales incómodas o manipulan cargas excesivas, la frecuencia de los TME por la sintomatología dorsal crece hasta el 71,7% (ver Figura 2.2). Asimismo, estar expuesto a al menos uno el riesgo ergonómico duplica las probabilidades de sufrir molestias en las extremidades, en la región superior y en la región dorsal, además aumenta la frecuencia (en más de 10%) de expresar malestar en las extremidades inferiores y en la región del cuello. Casi el 30% de los trabajadores que manipulan cargas y tienen posturas incómodas tienen bajo bienestar psicofísico y tienen los índices más

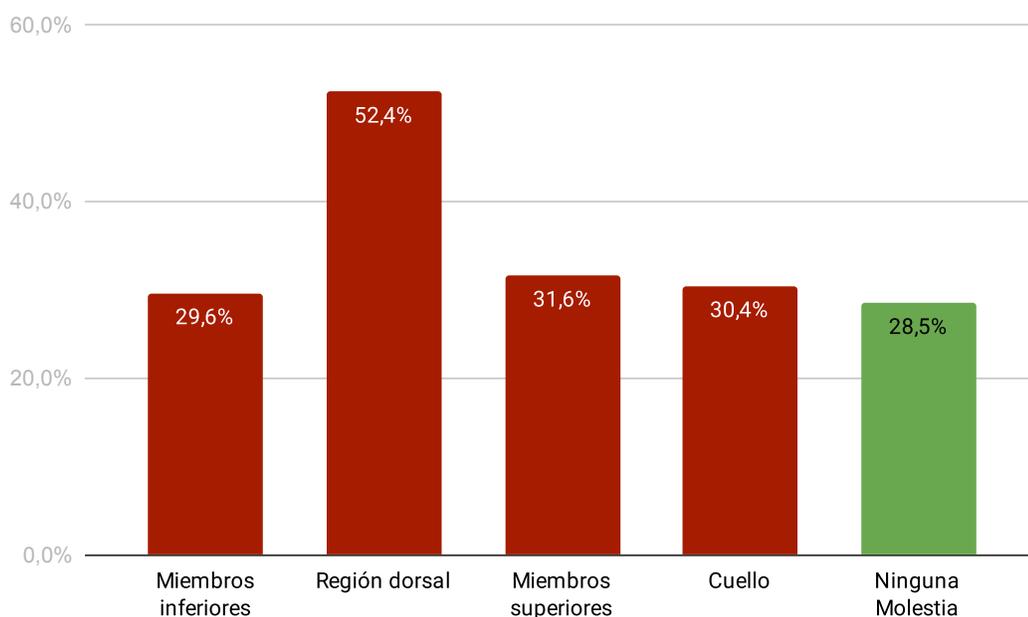


Figura 2.1: Zonas del cuerpo donde los trabajadores argentinos expresaron algún tipo de malestar relacionado con su trabajo.

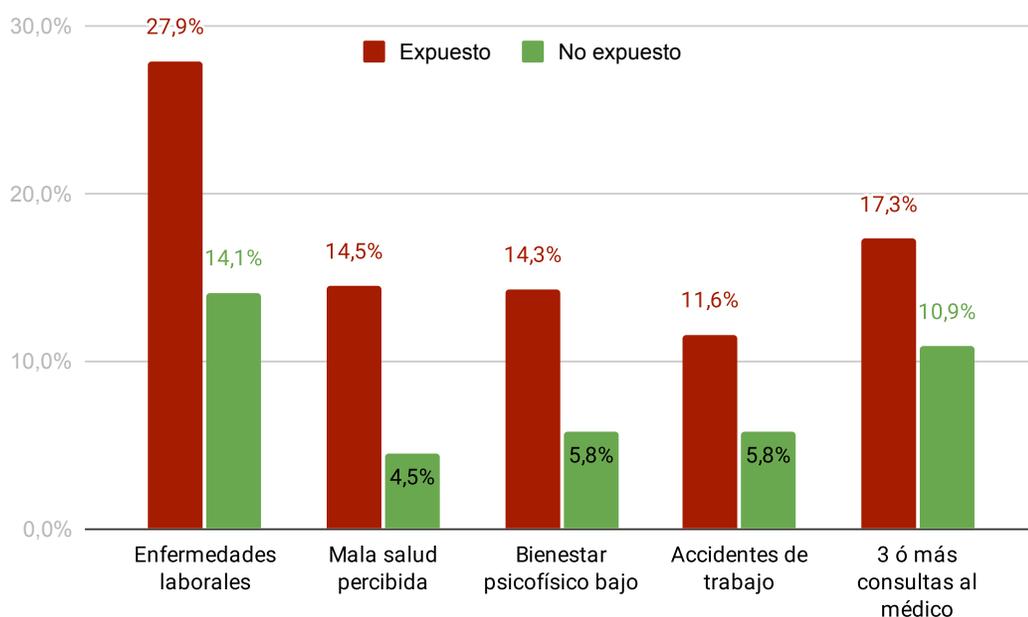


Figura 2.2: Indicadores de bienestar o situaciones nocivas según trabajadores expuestos a posturas forzadas (Gadea y cols., 2019).

altos de accidentes y enfermedades. El 50% de trabajadores expuestos a manipulación cargas, posturas corporales incómodas y Las vibraciones han demostrado que el trabajo afecta negativamente su salud.

2.2. Evaluación de riesgos ergonómicos

La evaluación de riesgos ergonómicos (ERE) fue propuesta en 1997 por el Instituto Nacional de Seguridad y Salud Ocupacional (NIOSH) de EE. UU., la cual contiene una serie de pasos que se deben incluir en cada programa de ergonomía, cuyo fin es generar un proceso sistemático para identificar, analizar y controlar los factores de riesgo en el lugar de trabajo. Como resultado, un programa de ergonomía eficaz debe incluir una ERE basada en la evidencia para reconocer y rectificar las deficiencias en una amplia gama de situaciones laborales (Cohen, Gjessing, y Fine, 1997).

Una gran cantidad de métodos y herramientas han sido diseñadas en las últimas décadas para las ERE. Dichos métodos puede ser clasificados de la siguiente manera (Battini y cols., 2014; David, 2005; Lowe, Weir, y Andrews, 2014; Vignais y cols., 2017):

- **Autoevaluación**, donde los trabajadores se evalúan a sí mismos utilizando formularios diseñados específicamente.
- **Observación humana**, donde personal especialmente calificado recopila una estimación de los ángulos de las articulaciones del cuerpo a través de observaciones *in situ* o análisis de video fuera de línea.
- **Medición directa**, donde se adosan al cuerpo del trabajador herramientas antropométricas o de medición inercial para recopilar datos.
- **Evaluación de visión por computadora**, donde los modelos del cuerpo humano se determinan automáticamente mediante CV a partir de tomas de fotos o videos, lo que proporciona una medición ergonómica basada en modelos más escalable, sistemática y objetiva.

Actualmente, los métodos más utilizados para la ERE en los lugares de trabajo industriales caen dentro de los dos primeros (autoevaluaciones y observación humana por parte de especialistas). En Diego-Mas y cols. (2015) se consigna que los métodos de ERE más comunes en Iberoamérica son RULA (79.5%), REBA (56.5%), OWAS (56.9%) y NIOSH (59%). Estos métodos han sido ampliamente descritos en la bibliografía (Chiasson, Imbeau, Aubry, y Delisle, 2012; Joshi y Deshpande, 2019; Lim, Jung, y Kong, 2011; Roman-Liu, 2014)). Por ejemplo, en Battini y cols. (2014) se realiza una comparación sencilla basada en la inclusión de factores de riesgo ergonómico como la postura, frecuencia de movimiento, fuerza aplicada, tiempo de recuperación, medio ambiente, si los métodos dinámicos o si analizan el cuerpo entero (ver Tabla 2.1).

Tabla 2.1: Comparación de los métodos de ERE (Battini y cols., 2014).

	POSTURA	FRECUENCIA	FUERZA	RECUPERACIÓN	AMBIENTE	DINÁMICO	CUERPO ENTERO
OWAS	✓	✓	X	X	X	✓	✓
OCRA	✓	✓	✓	✓	✓	X	X
RULA	✓	X	✓	X	X	X	X
REBA	✓	X	✓	X	X	X	X
NIOSH	X	✓	✓	X	✓	X	✓

Los enfoques de autoevaluación y los de observación humana aunque son prácticos de usar, se ven severamente afectados por sesgos subjetivos. El estado del arte resalta que incluso los ergonomistas experimentados a menudo se equivocan en las decisiones subjetivas, por lo cual resulta difícil comparar mediciones entre diferentes ergonomistas o comparar mediciones del mismo ergonomista espaciadas en el tiempo. Otro factor que genera error son las condiciones visuales subóptimas en el lugar de trabajo como ser la mala iluminación, las oclusiones, y ángulos de toma inadecuados en videos o imágenes (Plantard y cols., 2017). En Iberoamérica, Diego-Mas y cols. (2017) realizan un estudio donde detallan los errores más comunes cometidos por los ergónomos. Los resultados muestran que aproximadamente el 30% de las evaluaciones realizadas por los profesionales tenían errores. En el 13% de las evaluaciones, los errores fueron graves o se invalidaron por completo los resultados de la evaluación (ver Figura 2.3).

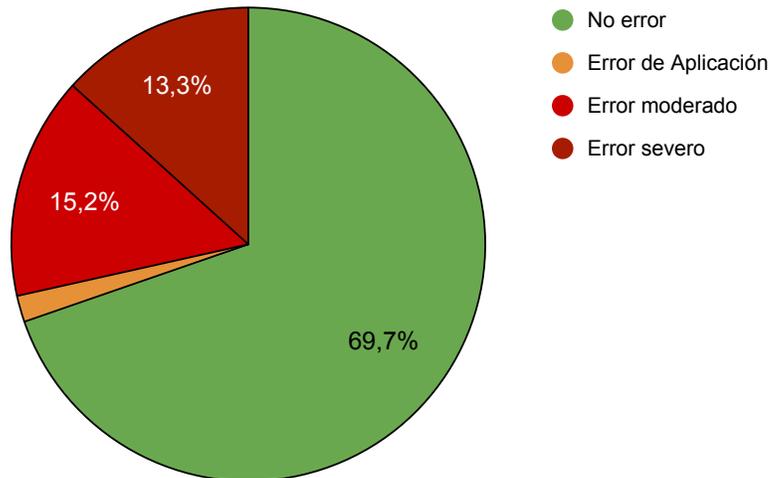


Figura 2.3: Porcentaje de valoraciones con errores y categoría de error. (Diego-Mas y cols., 2017).

2.3. Estimación de riesgos Ergonómicos mediante la red neuronal convolucional OpenPose y el método RULA

Para la primera solución exponemos en detalle la metodología y los resultados presentados en (Massiris Fernández, Fernández, y cols., 2020) y en (Massiris Fernández y cols., 2019). A continuación, describimos los fundamentos del método RULA, para luego pasar a presentar la tecnología de visión computacional adoptada para su evaluación automática. Dicha tecnología propuesta utiliza videos o fotos generados por dispositivos RGB de bajo costo para proporcionar un solución flexible que puede reproducirse fácilmente y difundirse abiertamente.

2.3.1. Descripción de la metodología RULA

El método RULA introducido por McAtamney y Corlett (1993) fue diseñado para proporcionar el llamado RULA *Grand Score* (GS), *i.e.*, un puntaje numérico global que representa el nivel de carga postural del sistema musculoesquelético de un trabajador, con respecto a las posiciones de las partes del cuerpo. Este GS se obtiene de un esquema de combinación jerárquica, donde cada parte del cuerpo se evalúa de forma independiente con respecto a los ángulos de las articulaciones por puntajes. Dichos puntajes están determinados por umbrales angulares.

Estos puntajes de partes del cuerpo se combinan luego en puntajes intermedios A y B de RULA, que resumen respectivamente el estrés de las extremidades superiores y el estrés del resto del cuerpo. Finalmente, el GS combina en un solo valor ambos puntajes intermedios, y sus factores de fuerza y estrés muscular. Por último, el GS se relaciona con cuatro niveles de acción, que indican gradualmente la necesidad de una mayor investigación o la aplicación de cambios ergonómicos en el lugar de trabajo.

El primer paso en RULA es observar las diferentes tareas que realiza el trabajador. Por este motivo, el analista debe seleccionar las posturas corporales más críticas que se realizan durante un período prolongado, que son repetitivas, que se sabe específicamente que son dañinas, o que representan una desviación crucial en comparación con las posiciones seguras y neutrales. Como regla práctica, se deben observar varios ciclos de trabajo antes de una selección de postura crítica. Por esta razón es que se establecen los supuestos expuestos al principio de esta sección. Si el ciclo de trabajo es muy largo (o no hay ciclo de trabajo), el analista a menudo opta por tomar instantáneas a intervalos de tiempo regulares (Chiasson y cols., 2012; Li, Han, Gül, Al-Hussein, y El-Rich, 2018; Roman-Liu, 2014). Esta estrategia también permite una medición rápida de la duración asociada con cada posición del cuerpo y el tiempo trabajado entre descansos (Chiasson y cols., 2012; Roman-Liu, 2014). Los principales beneficios de aplicar RULA son su nula necesidad equipamiento extra y la simplicidad del procedimiento, lo que a su vez con el resultado entendido facilita la finalización de la evaluación. Como resultado, los

especialistas adoptaron este método ERE como una herramienta de detección para diferentes escenarios de trabajo (McAtamney y Corlett, 1993).

Como fue mencionado en el inicio de este capítulo, RULA es una herramienta ergonómica de ERE ampliamente aceptada. Usualmente es realizada a través de la observación humana directa, y más recientemente, con ayuda de fotos o de grabaciones de video por parte de investigadores y profesionales en el campo (Joshi y Deshpande, 2019; Liu, 2019; Plantard y cols., 2017; Savino, Battini, y Riccio, 2017). Sin embargo, la inspección humana conlleva inherentemente errores de percepción y, a veces, una variabilidad sustancial intra e interobservador (Plantard y cols., 2017; Xu y McGorry, 2015). Por dicha razón, diseñamos nuestro método para calcular automáticamente las puntuaciones RULA a partir de fotos o secuencias de video.

2.3.2. Descripción de Red Neuronal OpenPose

La estimación de la pose humana es un problema de larga data en área de visión por computadora. El uso de herramientas de aprendizaje automático como las CNN ha aumentado la solidez de estos métodos, logrando en algunos casos que el modelado y la detección de la acción humana, incluidas situaciones en las que las articulaciones o las extremidades están parcialmente ocluidas o incluso ausentes en una imagen. Este proceso se define como la tarea de ubicar las articulaciones del cuerpo como las rodillas, codos, hombros) en imágenes, y como segundo paso, la búsqueda de poses específicas que coincidan con las articulaciones observadas en el espacio de posibles poses articuladas (Cao y cols., 2017).

Hoy en día, las herramientas de software de código abierto de última generación como OpenPose (Cao y cols., 2017) permiten la detección en tiempo real de articulaciones y extremidades a partir de imágenes y videos digitales. OpenPose propone un proceso donde se toma como entrada una imagen completa, aplicando una CNN de dos ramas, la primera dedicada a predecir conjuntamente mapas de confianza para la detección de articulaciones del cuerpo, y la segunda dedicada a campos de afinidad de partes para la asociación de las articulaciones (Cao y cols., 2017).

Dada una imagen de entrada, la red entrega una lista de cuerpos detectados, cada uno con un esqueleto asociado de articulaciones definidas previamente. Usamos el modelo OpenPose de 25 articulaciones (OP25), donde las articulaciones del esqueleto se enumeran siguiendo el orden específico que se muestra en la Figura 2.4. Finalmente, la articulación k -ésima perteneciente al esqueleto s -ésimo, J_k^s , está representada por OpenPose como la tupla

$$J_k^s = [x_k^s, y_k^s, p_k^s]. \quad (2.1)$$

Para cada articulación detectada, el modelo proporciona un vector con su posición relativa en la imagen y la confianza de la estimación, que varía de 0 a 1 con respecto a la confianza máxima. A partir de esta información, calculamos una confianza de detección de esqueleto general \bar{p} como el promedio de las confianzas de estimación conjunta, que

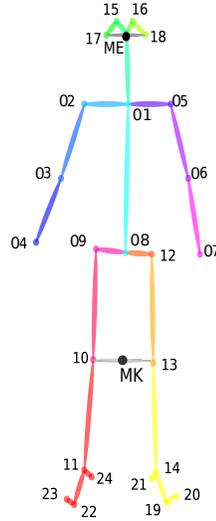


Figura 2.4: articulaciones del esqueleto de OpenPose 25.

será útil para filtrar detecciones ruidosas o espurias .

$$\bar{p} = \frac{1}{N} \sum_{k=1}^N p_k^s. \quad (2.2)$$

2.3.3. Estimación del ángulo de la articulación

Los ángulos de la articulación del cuerpo establecidos en la Tabla 2.2 se utilizan para estimar los puntajes RULA automáticamente. Por ejemplo, el ángulo de Flexión del codo izquierdo (EL) se obtiene de las posiciones observadas del hombro, el codo y la muñeca izquierdos, correspondientes respectivamente a OP25 articulaciones del esqueleto # 04, 03 y 02, como se muestra en la Figura 2.4. Dos ángulos específicos de Tabla 2.2 utilizan puntos fidúceos de referencia calculados a partir de puntos OP25 en conjunto: ME (mitad de las orejas) representa un punto a medio camino entre ambas orejas, y MK (mitad de las rodillas) denota un punto entre las rodillas (ver Algoritmo 1).

Algorithm 1 Confianza preservando media de dos articulaciones.
Articulación no detectada codificada como [0,0,0].

- 1: INPUTS: J_1, J_2
 - 2: **if** $\min(J_1, J_2) == [0, 0, 0]$ **then**
 - 3: $J_0 \leftarrow \max(J_1, J_2)$
 - 4: **else**
 - 5: $J_0 \leftarrow \frac{1}{2}(J_1 + J_2)$
 - 6: **end if**
 - 7: OUTPUT: J_0
-

Definimos estos puntos fidúceos para calcular dos ángulos que son críticos para la evaluación de RULA, es decir, flexión del cuello (NF) y flexión del tronco (TF), que co-

Tabla 2.2: Ángulos de RULA calculados a partir de OpenPose 25.

Nombre del ángulo	Acrónimo ¹	Articulaciones involucradas ²
Flexión del Codo Izquierdo	EL	∠04,03,02
Flexión del Codo Derecho	ER	∠05,06,07
Flexión del Hombro Izquierdo	SL	∠03,02,09
Flexión del Hombro Derecho	SR	∠06,05,12
Flexión del Hombro Izquierdo 2	SL2	∠03,02,01
Flexión del Hombro Derecho 2	SR2	∠06,05,01
Flexión de la Rodilla Izquierda	KL	∠09,10,11
Flexión de la Rodilla Derecha	KR	∠12,13,14
Flexión del Pie Izquierdo	FL	∠10,11,22
Flexión del Pie Derecho	FR	∠13,14,19
Torsión del Cuello	NT	∠00,01,02
Inclinación Izquierda del Cuello	NB	∠17,01,02
Inclinación Derecha del Cuello	NB2	∠18,01,05
Flexión del Cuello	NF	∠ME,01,08
Torsión Derecha del Tronco	TT	∠02,08,09
Torsión Izquierda del Tronco	TT2	∠05,08,12
Inclinación del Tronco	TB	∠09,08,01
Flexión del Tronco	TF	∠MK,08,01

¹Acrónimos de los nombres en inglés.

²Articulaciones numeradas según Figura 2.4 excepto ME y MK (ver texto).

múnmente son difíciles o imposibles de calcular de otra manera, ya que las articulaciones requeridas con frecuencia se ocluyen. Por esta razón, se calcula un punto medio a partir de dos puntos laterales que permite usar como referencia los datos no ocluidos (ver Algoritmo 1).

2.3.4. Cálculo de puntuaciones de RULA

RULA aplica umbrales a los ángulos de las articulaciones para calcular las puntuaciones. Estos umbrales están claramente definidos para algunas articulaciones, por ejemplo codos y rodillas, pero no están delimitados explícitamente para otros. Este último grupo incluye abducción del hombro, trabajo del brazo fuera del cuerpo, torsión del cuello, inclinación lateral del cuello, torsión del tronco y ángulos de inclinación lateral del tronco (Battini y cols., 2014; Plantard y cols., 2017). Por lo tanto, para obtener estas partes del cuerpo o puntajes RULA locales, definimos los umbrales que se muestran en la tercera columna de la Tabla 2.3.

Por último, se necesitan algunos parámetros adicionales para evaluar algunas puntuaciones RULA, ya que OpenPose 25 solo proporciona posiciones de articulaciones y confianzas a partir de la entrada visual. Por ello, para proporcionar una evaluación de RULA completa, los usuarios deben establecer manualmente puntuaciones de posición

Tabla 2.3: Puntuaciones RULA calculadas en base a los esqueletos detectados.

RULA	Formula ¹	Puntuaciones
Posición del brazo superior	$\overline{[SL, SR]}$	+1 ($-20^\circ, 20^\circ$)
		+2 ($-\infty, -20^\circ$)
		+2 ($20^\circ, 45^\circ$)
		+3 ($45^\circ, 90^\circ$)
		+4 ($90^\circ, \infty$)
Abducción superior del brazo	$\max(SL2, SR2)$	+1 ($110^\circ, \infty$)
Posición del brazo inferior	$\overline{[EL, ER]}$	+1 ($60^\circ, 100^\circ$)
		+2 ($-\infty, 60^\circ$)
		+2 ($100^\circ, \infty$)
Brazo trabajando hacia fuera del cuerpo	$\max(SL, SR)$	+1 ($30^\circ, \infty$)
Ángulo del cuello	NF	+1 ($0^\circ, 10^\circ$)
		+2 ($10^\circ, 20^\circ$)
		+3 ($20^\circ, \infty$)
		+4 ($-\infty, 0^\circ$)
Ángulo de Torsión del Cuello	$\text{abs}(90^\circ - NT)$	+1 ($5^\circ, \infty$)
Ángulo de Inclinación del Cuello angle	$\text{abs}(65^\circ - \overline{[NB, NB2]})$	+1 ($5^\circ, \infty$)
Ángulo de Flexión del Tronco (TFA)	TF	+1 $\approx 0^\circ$
		+2 ($1^\circ, 20^\circ$)
		+3 ($20^\circ, 60^\circ$)
		+4 ($60^\circ, \infty$)
Ángulo de Torsión del Tronco (TTA)	$\max(TT, TT2)$	+1 ($100^\circ, \infty$)
Ángulo de Inclinación del Tronco	TB	+1 ($-\infty, 85^\circ$)
		+1 ($95^\circ, \infty$)

¹Las abreviaturas de los ángulos se dan en la Tabla 2.2.

y torsión de la muñeca, puntuación de equilibrio de piernas y puntuaciones de uso de fuerza y músculo.

Además, nuestro método utiliza dos umbrales (tamaño y confianza) para descartar datos de esqueleto no deseados que pueden ser falsos positivos. El umbral de tamaño considera un área mínima del esqueleto en proporción a la resolución de la imagen, mientras que el umbral de confianza se aplica para descartar esqueletos cuyas articulaciones están ocluidas o detectadas con poca confianza. En resumen, las fórmulas proporcionadas intentan compensar las posibles oclusiones en cada lado del cuerpo, utilizando el promedio de los ángulos de las partes del cuerpo izquierda y derecha, los puntos medios fiduciales o los valores máximos. Los umbrales y fórmulas específicos se derivan de múltiples experimentos descritos en la siguiente sección.

2.3.5. Métodos de evaluación

Como se menciona en [Cao y cols. \(2017\)](#) y [Zhang y cols. \(2018\)](#), las condiciones de filmación usuales en el entorno laboral son desafiantes para las técnicas de CV, dado que es usual que se presenten situaciones como la presencia de iluminación irregular, la variación en el punto de vista y las oclusiones, todos factores que pueden provocar la pérdida de información relevante. El método propuesto se basa en los siguientes supuestos dependiendo de la información ingresada por el usuario. Si el usuario ingresa una sola foto, la imagen debe ser representativa de una postura potencialmente crítica. Si el usuario ingresa una secuencia de video, las imágenes se toman a intervalos regulares y son representativas de una sola actividad laboral bajo evaluación. En el caso de secuencias de imágenes, se obtiene una puntuación general de RULA para la actividad.

Para investigar la aplicabilidad de nuestro método basado en CV dados estos factores, diseñamos varios experimentos simulados y controlados para cuantificar los efectos de la auto-oclusión, para comparar la calidad de la evaluación del ángulo de las articulaciones y para validar los umbrales propuestos en la Tabla 2.3. Finalmente, validamos el método en una situación real de trabajo al aire libre bajo la supervisión técnica de ergonomistas experimentados, quienes también evaluaron los puntajes RULA asociados.

Los métodos de validación implicaron tres niveles de comparación:

1. Confianzas de detección de esqueletos y articulaciones dependiendo del punto de vista, como aplicaron antes [Plantard, Auvinet, Le Pierres, y Multon \(2015\)](#) y [Golabchi y cols. \(2015\)](#).
2. Comparación de ángulos calculados dependiendo de los puntos de vista controlados en el laboratorio y simulados, de la misma forma que [Li y cols. \(2018\)](#).
3. Calculo de acuerdo entre la puntuación RULA a través del método propuesto y observaciones de ergonomistas experimentados, de acuerdo con [Levanon, Lerman, Gefen, y Ratzon \(2014\)](#), [Lee y cols. \(2017\)](#) y [Plantard y cols. \(2017\)](#).

Las auto-oclusiones pueden ser definidas como las regiones de oclusión proyectadas sobre el mismo objeto de estudio. Dichas auto-oclusiones son comunes en los problemas de visión por computadora especialmente al trabajar con humanos, pues por regla general desde un punto de vista lateral, solo es posible visualizar la mitad del cuerpo ([Gupta y cols., 2006](#)).

En un primer experimento, probamos la metodología propuesta contra las auto-oclusiones. Usamos un modelo 3D simulado como referencia para las articulaciones y las posiciones del cuerpo para proporcionar vistas controladas. El conjunto de datos generado contiene 300 muestras de puntos de vista de cámara aleatorios siguiendo el método de Hammersley, que proporciona una semi-esfera de equidistribución angular entre las muestras ([Cui y Freedden, 1997](#)).

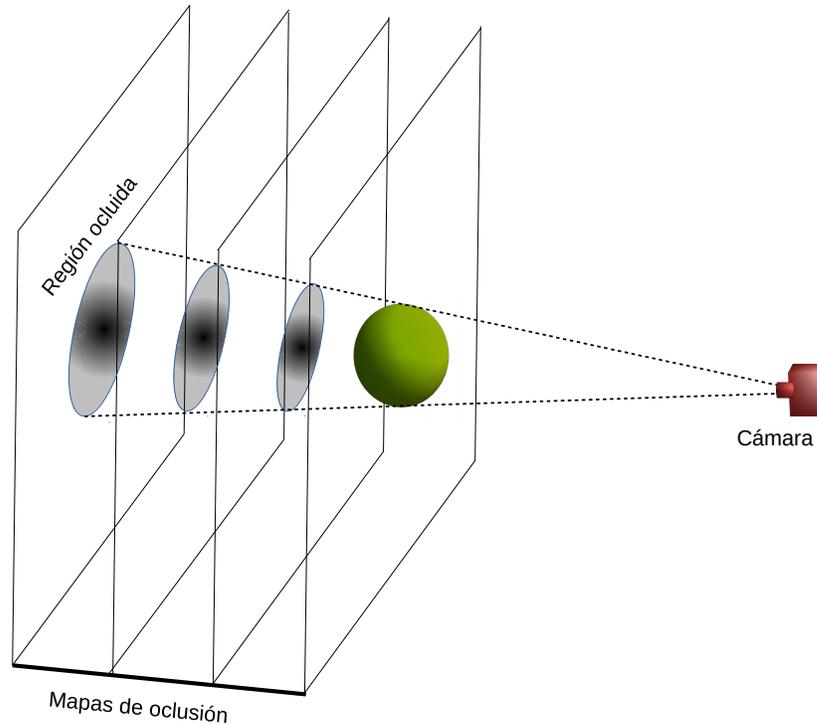


Figura 2.5: Regiones de oclusión (Gupta y cols., 2006).

En nuestro caso, un conjunto de combinaciones de poses, posiciones de articulación y ángulos de articulación fueron elegidos para capturar la confiabilidad de la detección de los puntos antropométricos en fotografías en primer plano desde todos los puntos de vista posibles, donde el centro de la semiesfera coincide con el centro de gravedad del modelo (ver Figura 2.7).

En segundo lugar, la subsección 2.3.6, muestra datos recolectados de una secuencia de imágenes controladas donde un sujeto real (con aprobación ética y sin antecedentes de lesiones) realiza acciones de torsión y flexión del tronco, respectivamente Figuras 2.18 y 2.19). Finalmente, en la sección 2.3.7, validamos nuestro método en cinco videos de escenarios de trabajo reales probando la concordancia en los datos de muestra a muestra, como se expresa mediante el kappa de Cohen no ponderado (κ) entre los puntajes de riesgo RULA del método propuesto y las evaluaciones asistidas de siete ergonomistas experimentados. Como se menciona en Li y cols. (2018), dado que RULA se realiza principalmente mediante observación manual, los expertos generalmente analizan solo algunas posturas corporales que consideran críticas.

Estos experimentos tuvieron como objetivo probar nuestro método con secuencias de tareas realizadas en diferentes lugares de trabajo reales al aire libre. Se seleccionaron cinco videos que contenían una o más de las difíciles condiciones de visualización mencionadas anteriormente (*por ejemplo*, iluminación irregular, vistas de cámara en movimiento, oclusiones). Grabamos dos trabajos diferentes (*revoque de pared* y *martillado*)

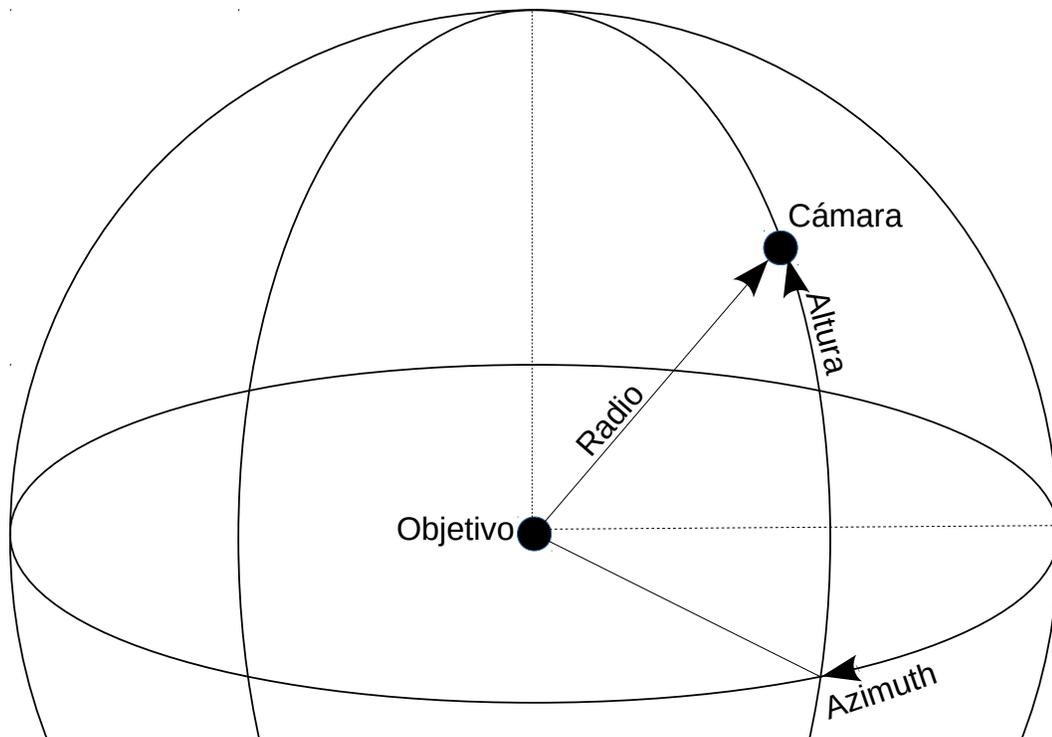


Figura 2.6: Semi-esfera 3D de Hammersley.

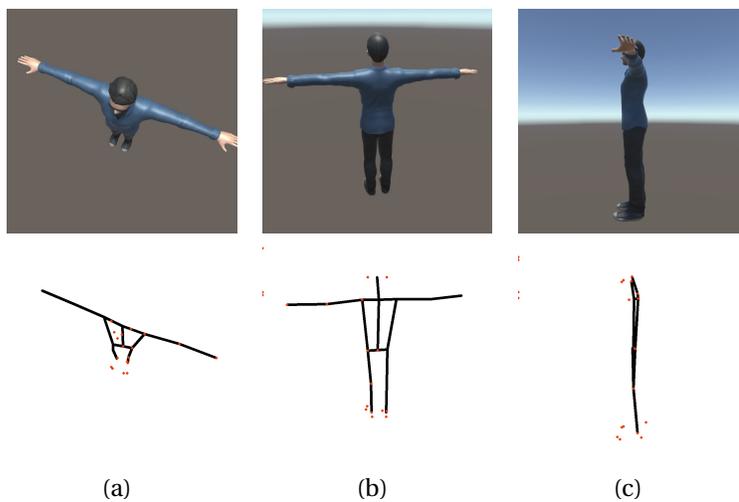


Figura 2.7: Las imágenes de la persona simulada en el dataset.

y seleccionamos tres videos de acceso público de diferente desempeño profesional real, aunque estos videos no fueron grabados específicamente para evaluaciones ergonómicas. Para cada video, los ergonomistas seleccionaron los fotogramas de acuerdo con el método descrito en la subsección 2.3.4, es decir, selección manual de imágenes o fotogramas tomados a intervalos regulares. Finalmente, el acuerdo de evaluación se calculó

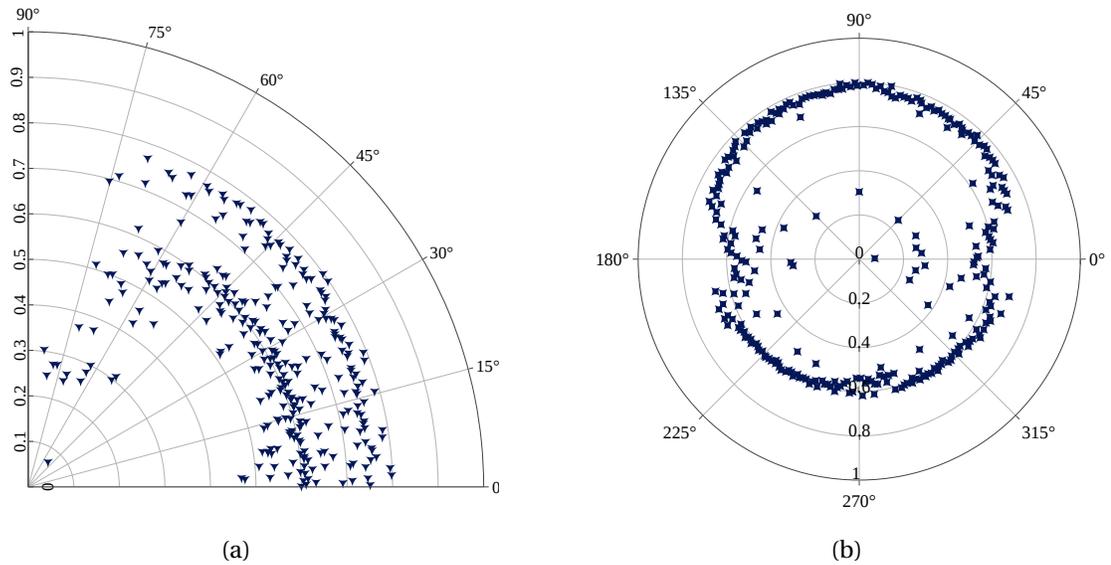


Figura 2.8: Confianza \bar{p} en la detección del esqueleto frente a la cámara vs (a) latitud y (b) azimut.

como en (Lee y cols., 2017; Plantard y cols., 2017). Los ergonomistas evaluaron conjuntamente las puntuaciones de las muñecas y las piernas, así como las de los músculos y la fuerza. Además, en estos videos se aplicaron umbrales de tamaño y confianza de 0,35 y 0,2 respectivamente.

2.3.6. Pruebas en entorno controlado

Prueba de confianza de detección de esqueletos

Primero, estábamos interesados en estudiar la confianza de detección del esqueleto \bar{p} como una función del punto de vista de la cámara. Los dos gráficos de la Figura 2.8 muestran los valores de confianza de detección promedio obtenidos, \bar{p} , para todo el conjunto de datos frente a la latitud del punto de vista (o ángulo de visión vertical) y el azimuth (o ángulo de visión horizontal). Como resultado, las latitudes de las cámaras (Figura 2.8(a)) se agruparon en tres categorías: $0^\circ - 45^\circ$, $45^\circ - 75^\circ$ y $75^\circ - 90^\circ$.

Tabla 2.4: Confianzas de la detección de esqueletos *vs.* puntos de vista en rangos de latitud.

rangos	$0^\circ - 45^\circ$	$45^\circ - 75^\circ$	$75^\circ - 90^\circ$
mean	0.660	0.588	0.360
std	0.082	0.159	0.169
min	0.469	0.070	0.250
max	0.799	0.782	0.696

Tabla 2.5: Confianzas de la detección de esqueletos *vs.* puntos de vista en rangos de azimut.

rangos	30° – 150°	210° – 330°	Otros
mean	0.740	0.589	0.576
std	0.088	0.041	0.129
min	0.250	0.374	0.070
max	0.799	0.653	0.750

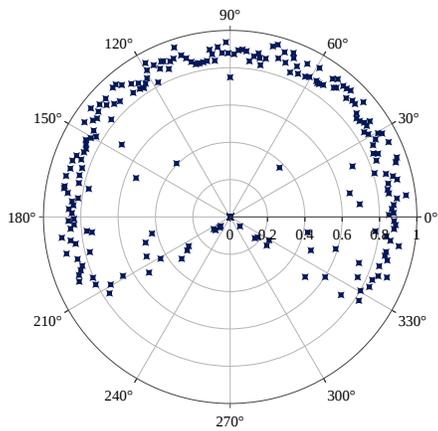
Los ángulos del punto de vista dentro de la primera categoría proporcionaron buenas confianzas de adquisición de esqueletos, que disminuyeron ligeramente en la segunda categoría y empeoraron mucho en la última. Por lo tanto, la latitud de la cámara debe limitarse preferiblemente a 45°, o como máximo a menos de 75°, para evitar la pérdida de confianza debido a las auto-oclusiones.

De manera similar, los ángulos de azimuth también se agruparon en tres categorías: 30° – 150°, 210° – 330° y otros ángulos (ver Figura 2.8(b)). La primera categoría representa puntos de vista frontales del modelo 3D, donde las articulaciones de la cara, los pies, el tronco y las extremidades casi nunca se auto-ocluyen. En la segunda categoría (puntos de vista posteriores), la auto-oclusión de los dedos de los pies, la nariz y los ojos reduce el máximo \bar{p} (ver Tabla 2.5). Finalmente, los puntos de vista laterales del sujeto se incluyen en la tercera categoría, proporcionando una alta variabilidad en las confianzas del esqueleto.

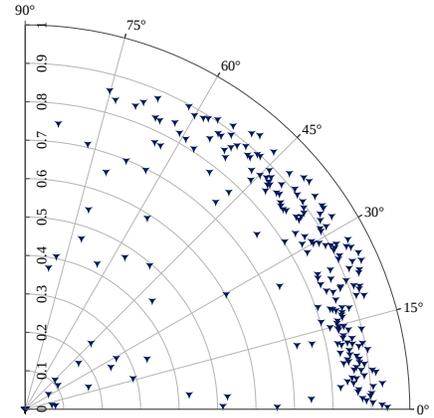
Dada la simetría de la postura precisa del modelo 3D (Figura 2.7), la auto-oclusión lateral afecta aproximadamente a la mitad de las articulaciones del cuerpo consideradas. Tanto la vista frontal como la posterior del sujeto brindan las mejores confianzas esqueléticas requeridas para la evaluación RULA, ya que las articulaciones auto-ocluídas son muy pocas, como los tobillos en vista frontal, o las menos relevantes para el método, como los ojos y nariz en vista trasera. Por lo tanto, el uso de azimuts de vista de cámara frontal (90° ± 45°) o posterior (270° ± 45°) debería proporcionar detecciones corporales adecuadas para el método propuesto.

Test detección de puntos antropométricos *vs.* punto de vista

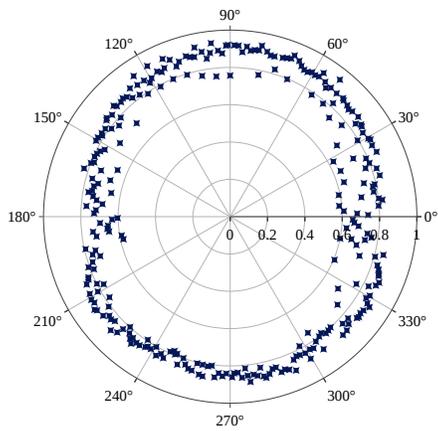
En RULA, las puntuaciones corporales de las articulaciones toman diferentes valores que se agregan mediante tablas. En este experimento, probamos varios métodos para calcular las puntuaciones de RULA del conjunto de datos simulado. Nuestro objetivo aquí es encontrar la mejor manera de calcular correctamente cada puntuación a partir de la información generada por OP25. De acuerdo con la Tabla 2.3, dado que la parte superior del brazo, el cuello y el tronco pueden obtener la puntuación máxima de la parte del cuerpo (+4), deben tratarse como las mediciones de RULA más relevantes, y por ende se debe hacer hincapié en ellas al calcular sus ángulos relacionados. Sin embargo, RULA no incluye una identificación clara de los puntos antropométricos involucrados



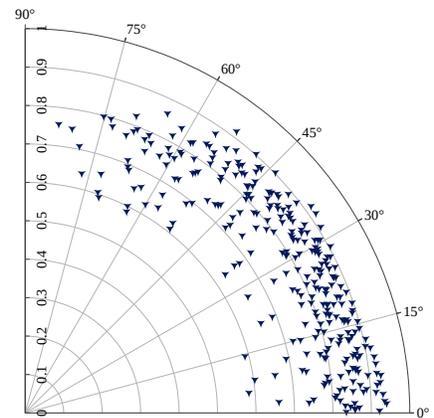
(a)



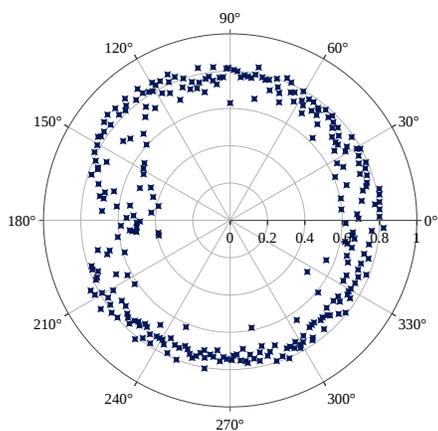
(b)



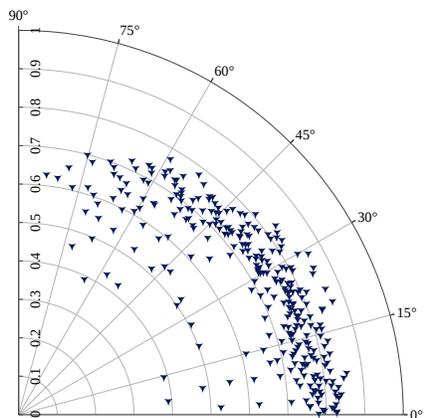
(c)



(d)

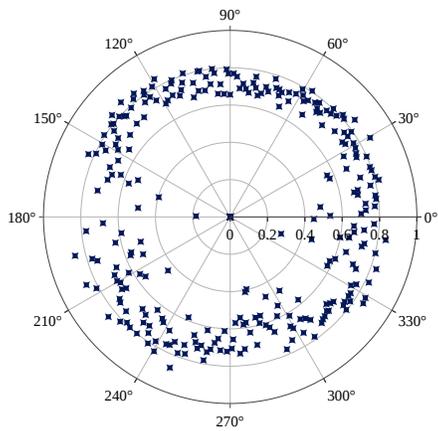


(e)

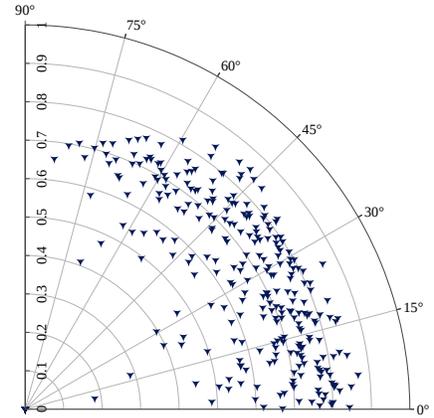


(f)

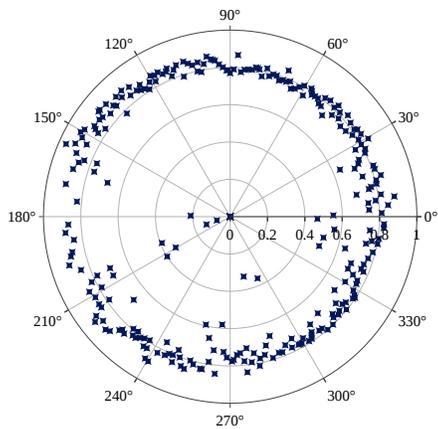
Figura 2.9: La confianza (escala en el eje radial) de la detección de punto antropométrico en función del azimuth de la cámara para (a,b) nariz, (c,d) cuello y (e,f) hombro izquierdo.



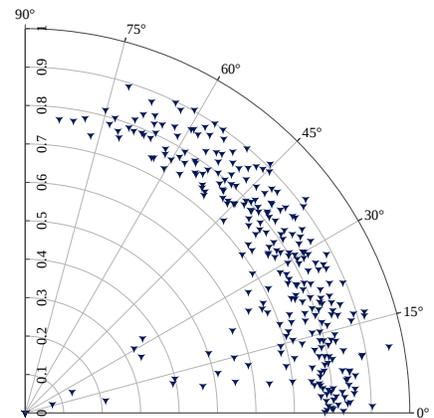
(a)



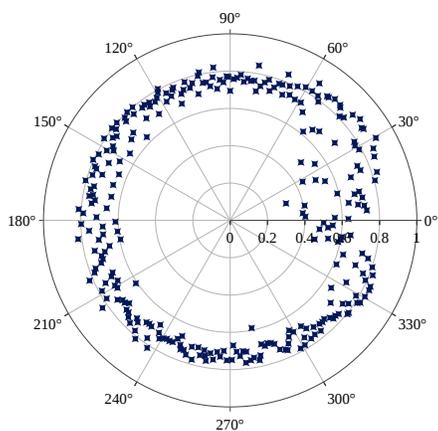
(b)



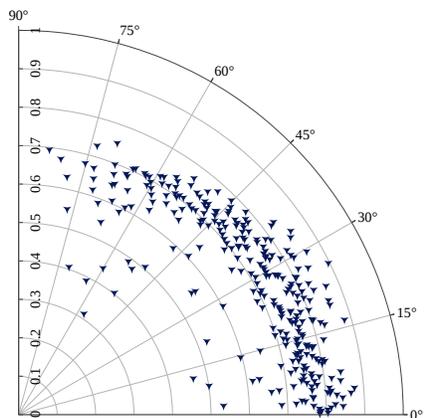
(c)



(d)

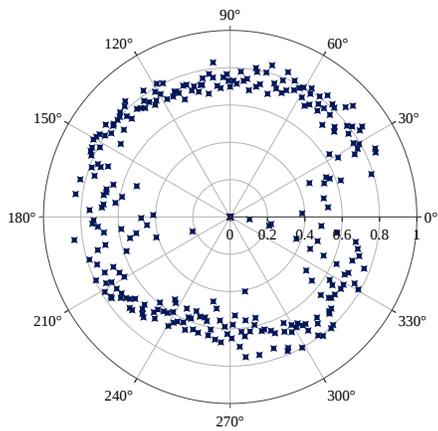


(e)

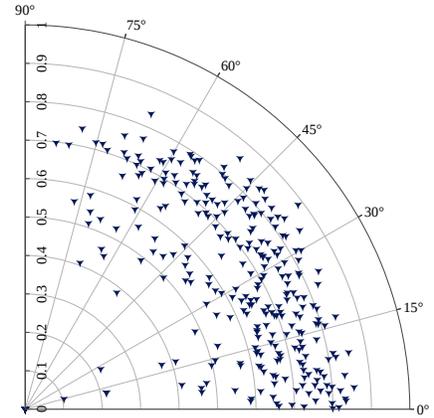


(f)

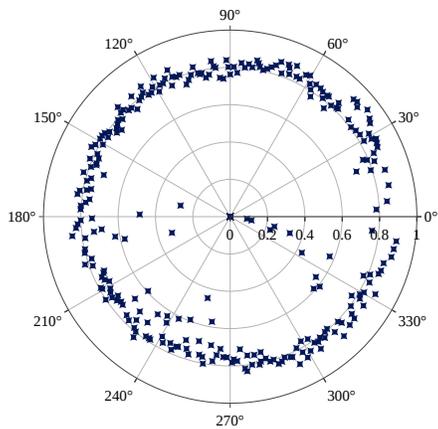
Figura 2.10: La confianza de la detección de punto antropométrico en función del azimuth de la cámara para (a,b) codo izquierdo, (c,d) mano izquierda y (e,f) hombro derecho.



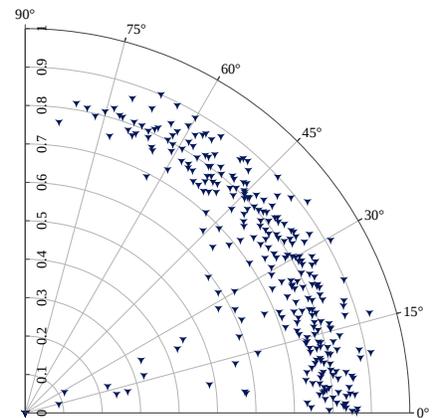
(a)



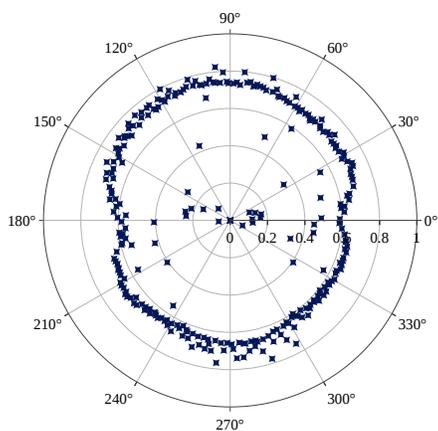
(b)



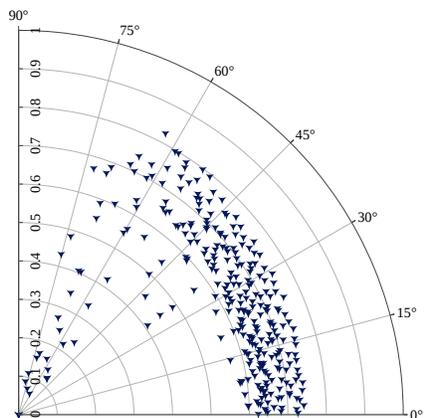
(c)



(d)

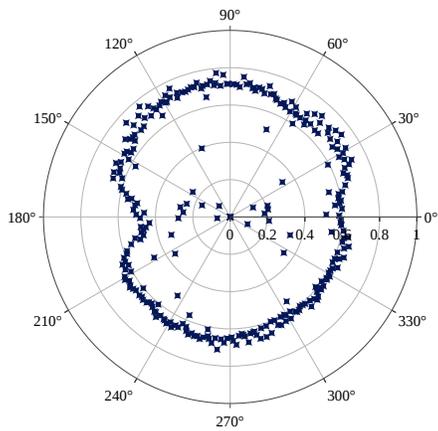


(e)

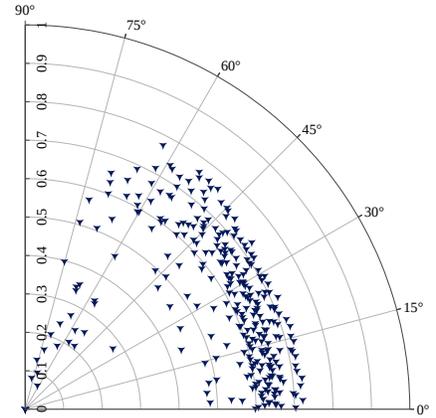


(f)

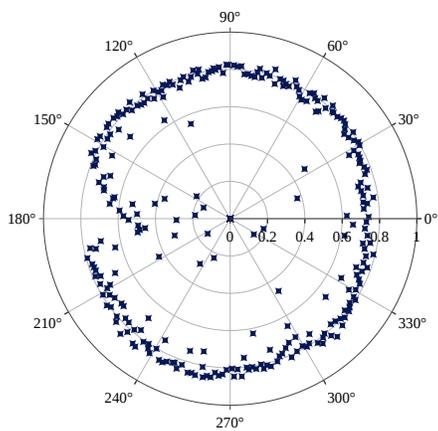
Figura 2.11: La confianza de la detección de punto antropométrico en función del azimuth de la cámara para (a,b) codo derecho, (c,d) mano derecha y (e,f) centro de cadera.



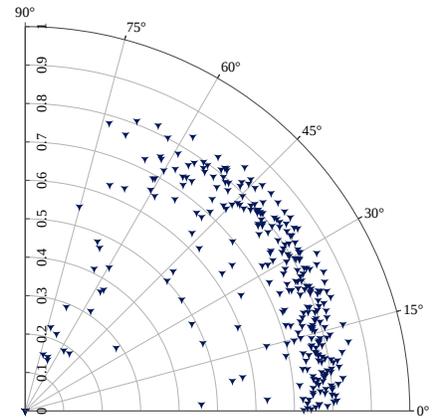
(a)



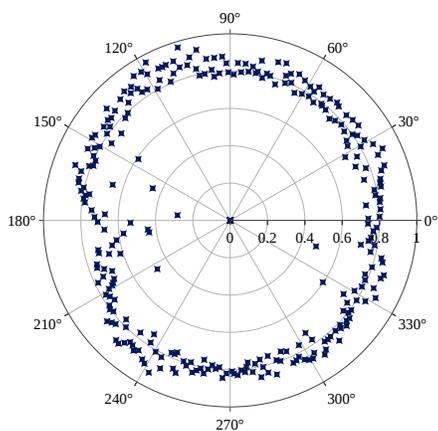
(b)



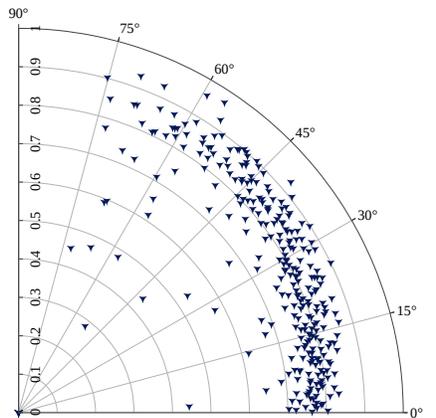
(c)



(d)

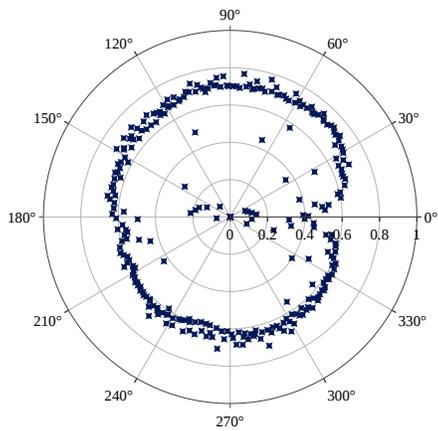


(e)

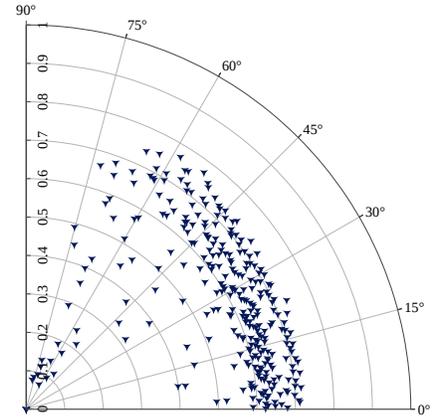


(f)

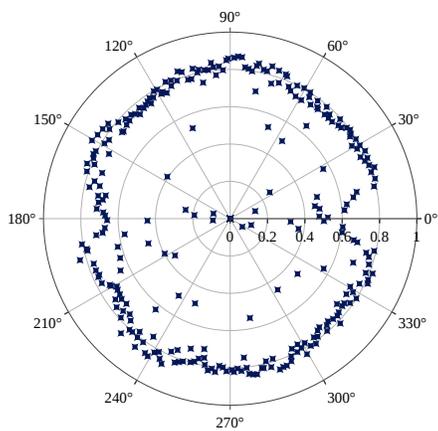
Figura 2.12: La confianza de la detección de punto antropométrico en función del azimuth de la cámara para (a,b) cadera izquierda, (c,d) rodilla izquierda y (e,f) tobillo izquierdo.



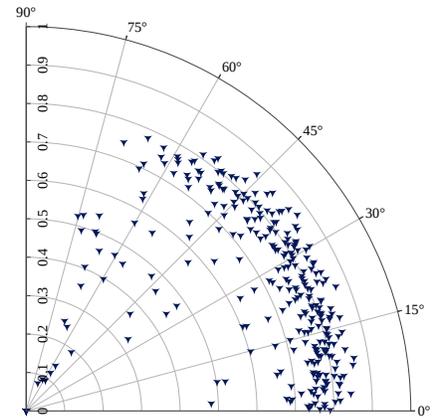
(a)



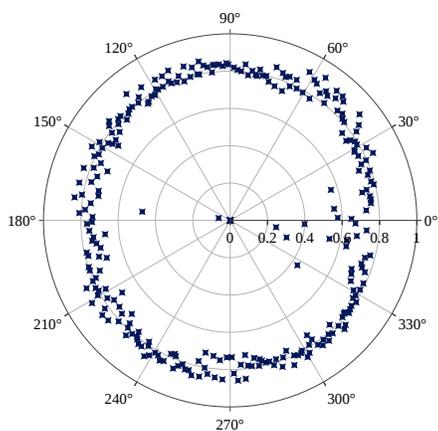
(b)



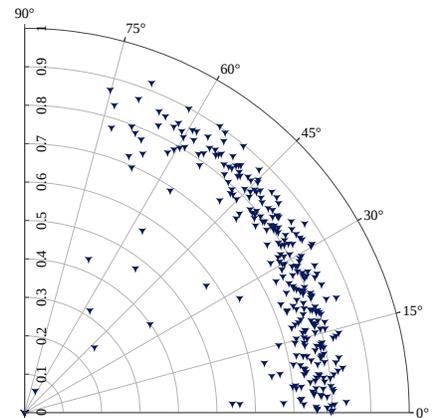
(c)



(d)

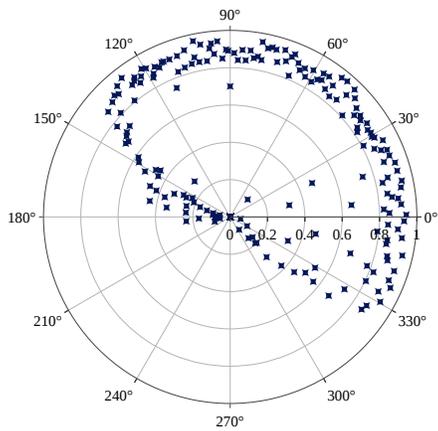


(e)

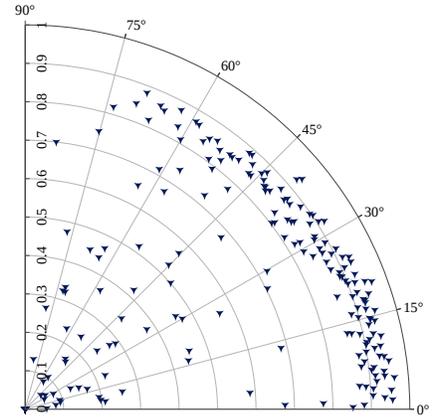


(f)

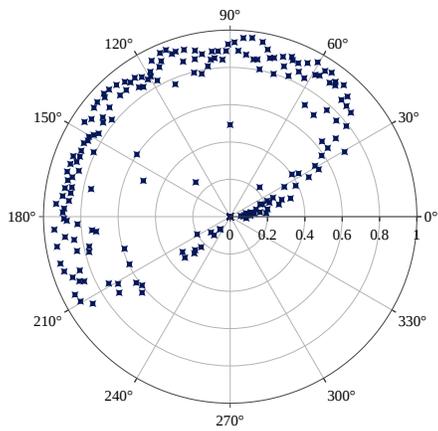
Figura 2.13: La confianza de la detección de punto antropométrico en función del azimuth de la cámara para (a,b) cadera derecha, (c,d) rodilla derecha y (e,f) pie derecho.



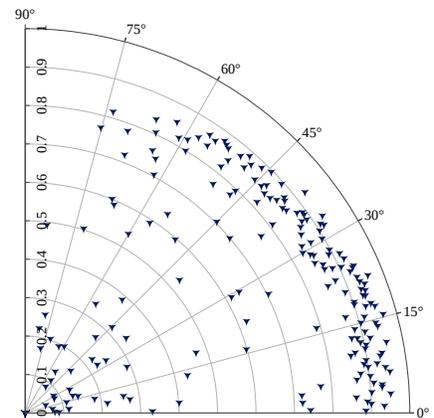
(a)



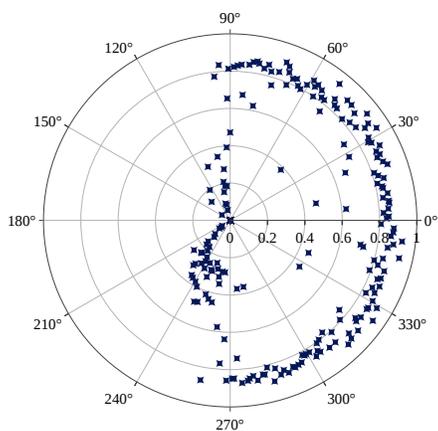
(b)



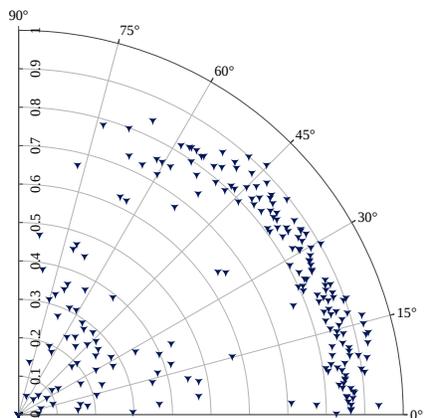
(c)



(d)

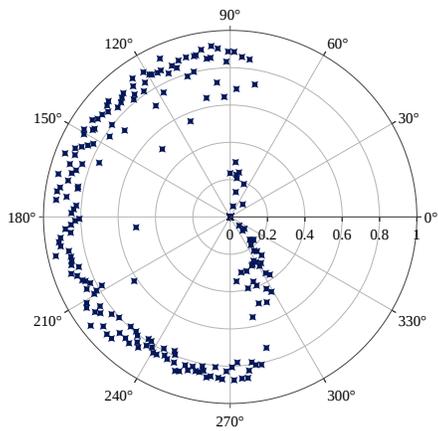


(e)

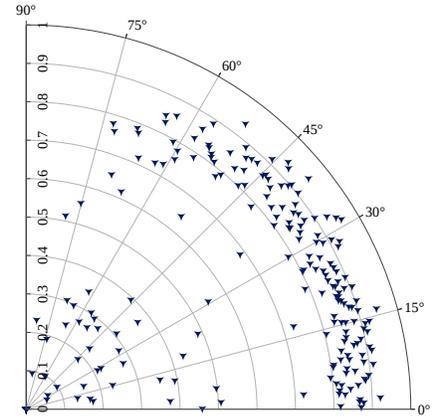


(f)

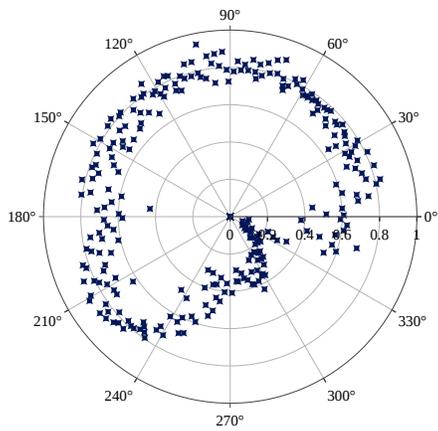
Figura 2.14: La confianza de la detección de punto antropométrico en función del azimuth de la cámara para (a,b) ojo izquierdo, (c,d) ojo derecho y (e,f) oído izquierdo.



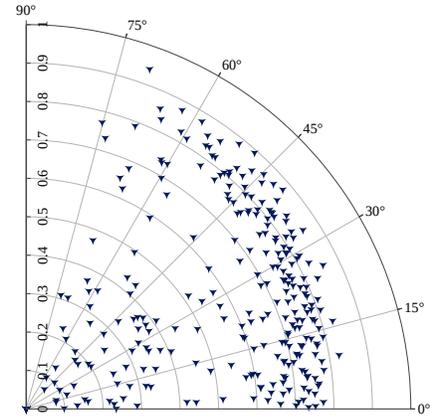
(a)



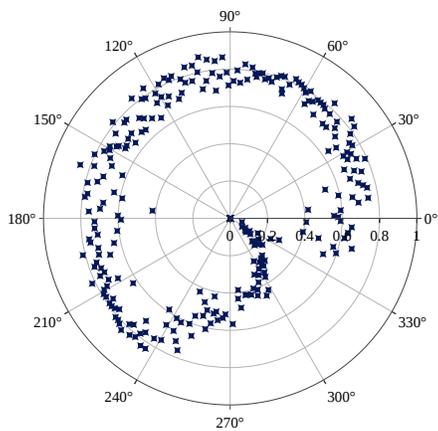
(b)



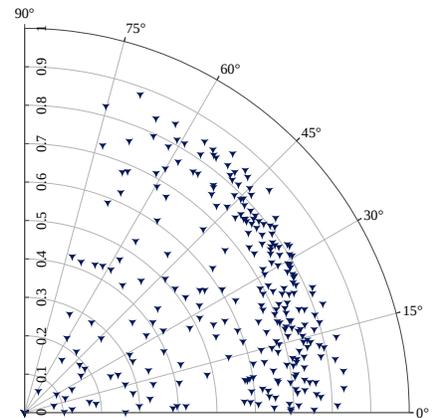
(c)



(d)

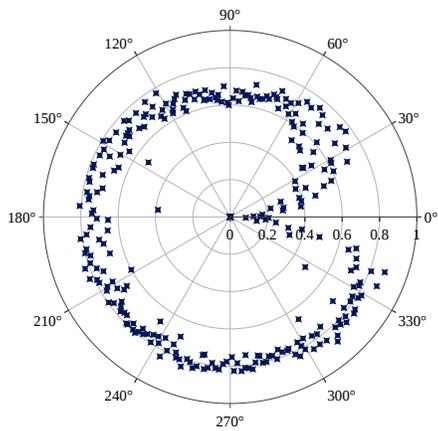


(e)

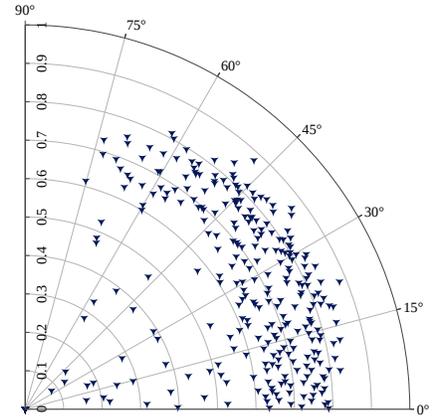


(f)

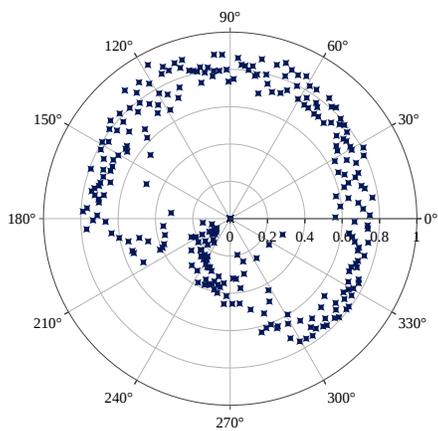
Figura 2.15: La confianza de la detección de punto antropométrico en función del azimuth de la cámara para (a,b) oído derecho, (c,d) dedo pulgar del pie derecho y (e,f) dedo meñique del pie derecho.



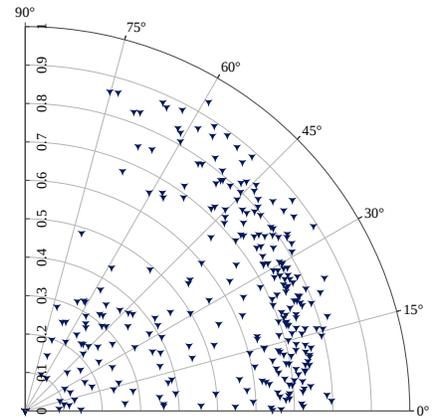
(a)



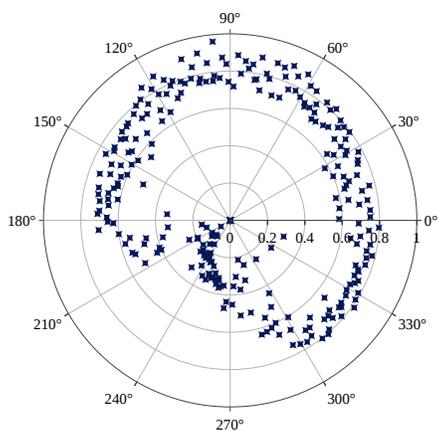
(b)



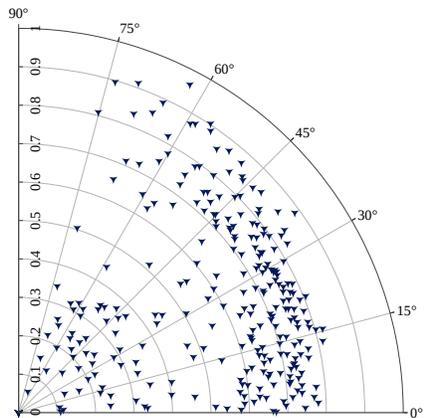
(c)



(d)



(e)



(f)

Figura 2.16: La confianza de la detección de punto antropométrico en función del azimuth de la cámara para (a,b) talón del pie derecho, (c,d) dedo pulgar del pie izquierdo y (e,f) dedo meñique del pie izquierdo.

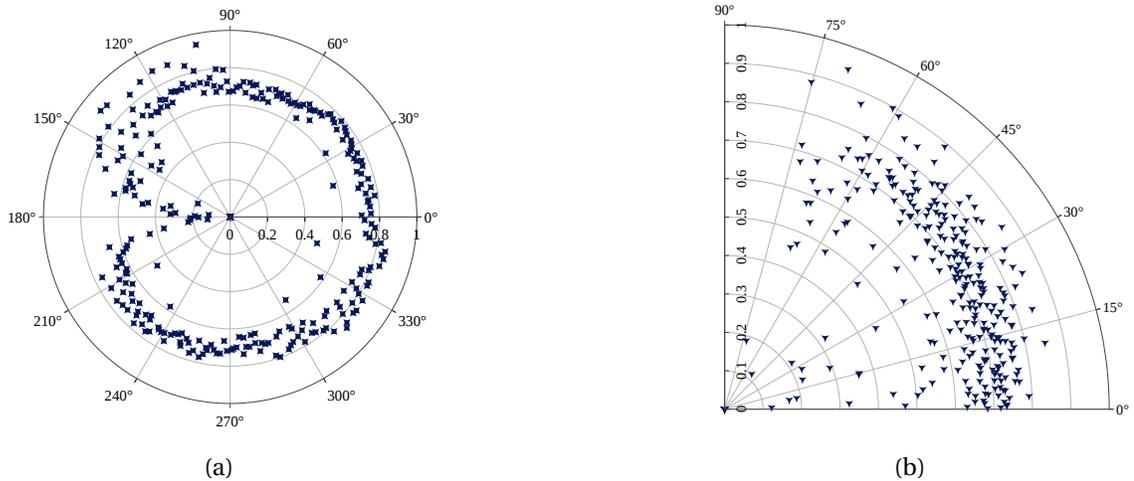


Figura 2.17: La confianza de la detección de punto antropométrico en función del azimuth de la cámara para (a,b) talón del pie izquierdo

en su medida angular (Plantard y cols., 2017). Para la puntuación del cuello, se consideró el ángulo formado por la articulación media de la cadera y el cuello, y tres posibles opciones para el tercer punto: la nariz, el punto medio entre las orejas y el punto medio entre los ojos. Asimismo, seleccionamos nuevamente las articulaciones de la mitad de la cadera y el cuello para calcular la puntuación del tronco y evaluamos los puntos medios entre las rodillas, los tobillos y los talones como candidatos para el tercer punto. Las puntuaciones de la parte superior del brazo se evaluaron a partir de los ángulos formados por las articulaciones de la cadera, el hombro y el codo de cada lado del cuerpo.

Para la nariz, la Figura 2.9(a) muestra confianzas de detección bastante diferentes (p_0) para las vistas frontal y posterior del sujeto. En particular, el intervalo de azimuth ($225^\circ - 315^\circ$) muestra una confianza nula debido a la auto-oclusión de la cabeza misma. Las Figuras 2.14(a) y 2.14(c) muestran las confianzas del ojo derecho e izquierdo respectivamente (p_{16} y p_{15}). El ojo derecho se detecta correctamente en el intervalo de azimuth ($315^\circ - 135^\circ$), mientras que el ojo izquierdo presenta un comportamiento similar en el rango ($45^\circ - 225^\circ$). Por lo tanto, no hay posibilidad de inferir la posición de ningún ojo en el rango ($225^\circ - 315^\circ$), que coincide por completo con el punto ciego posterior de la nariz.

La detección de confianza de los oídos (p_{18} y p_{17}) se muestra en la Figura 2.15(a) y 2.14(e). En este caso, las oclusiones laterales son responsables de otros puntos ciegos, *es decir*, ambos oídos no pueden detectarse simultáneamente con suficiente confianza desde casi cualquier punto de vista, excepto por los azimuths frontal (90°) y trasero (270°). La ausencia de un intervalo combinado de punto ciego para los oídos es la razón que nos llevó a definir un punto ME (medio del oído), dicho punto es utilizado para obtener la puntuación del cuello de la Tabla 2.3 aplicando el algoritmo 1 sobre ambos oídos.

Finalmente, las confianzas de la articulación del hombro obtenidas, p_{02} y p_{05} , muestran puntos ciegos laterales estrechos ubicados alrededor de 0° y 180° para el hombro izquierdo y derecho, respectivamente (ver Figuras 2.9(e) y 2.10(e)). La auto-oclusión explica estos puntos ciegos laterales, que también se detectaron en las confianzas de las articulaciones de la rodilla p_{10} y p_{13} (ver Figuras 2.12(c) y 2.13(c)). Como se presenta en la Tabla 2.3, solucionamos este problema para los ángulos de los hombros usando el promedio de los ángulos de los hombros izquierdo y derecho (SL y SR). Mientras que para calcular el ángulo de flexión del tronco (TA), se utiliza el punto medio entre las rodillas MK, dicho punto se define a partir de ambas articulaciones de la rodilla mediante el algoritmo 1.

Pruebas de umbral de flexión y torsión del tronco

Como se menciona en la Subsección 2.3.4, RULA aplica umbrales para puntuar el riesgo ergonómico. Aunque estos umbrales son claros en algunos ángulos, no están definidos explícitamente para otros. Por lo tanto, en la Tabla 2.3 proponemos umbrales basados en varias pruebas de laboratorio controladas. Un ejemplo de estos experimentos es el análisis de variación en el tiempo del ángulo del tronco (TTA) y ángulo de flexión del tronco (TFA) en tomas de video controladas, para establecer umbrales adecuados para sus puntuaciones RULA asociadas (ver Figuras 2.18 y 2.19). El gráfico de postura de torsión del tronco de la Figura 2.18(d) línea TFA (naranja) y TTA (azul) frente al tiempo. Notablemente, TTA muestra dos picos que indican que el movimiento de giro se ejecutó dos veces (en este caso, primero a la izquierda y luego a la derecha). Además, las posturas críticas de esta acción fueron identificadas por unanimidad por los especialistas en estos cuadros de pico. Por otro lado, TFA experimentó una variación casi nula durante esta acción, como se esperaba. Dado que RULA califica las posturas de torsión del tronco con una puntuación de +1, seleccionamos un umbral TTA de 100° para calcular su puntuación RULA (ver Tabla 2.3).

En la secuencia de acción de flexión del tronco (desde una posición de pie hasta agacharse y luego hacia atrás), el lado derecho del cuerpo del sujeto estaba ocluido (ver Figura 2.19(a)). Sin embargo, nuestro método fue consistentemente capaz de inferir las articulaciones ocluidas requeridas para calcular las puntuaciones locales RULA con suficiente precisión. El gráfico de tiempo de la Figura 2.19(d) permitió una forma rápida de seleccionar una pose crítica de la banda central TFA (naranja). Además, la pose crítica representada en la Figura 2.19(a), para la cual TFA toma el valor mínimo en toda la secuencia, fue considerada representativa de la acción por todos los ergonomistas para calcular sus puntajes RULA (TFA + 4, TTA 0). Este gráfico también muestra que tanto TFA como TTA disminuyeron durante la acción de flexión del tronco. Por un lado, un análisis fotograma a fotograma de este gráfico nos permitió establecer los umbrales necesarios para calcular los cuatro niveles de puntuación RULA de flexión del tronco, que también se muestran en el gráfico. Por otro lado, el umbral previamente seleccionado para TTA impidió que el método obtuviera una puntuación incorrecta de torsión del tronco RULA en cualquier imagen la secuencia.

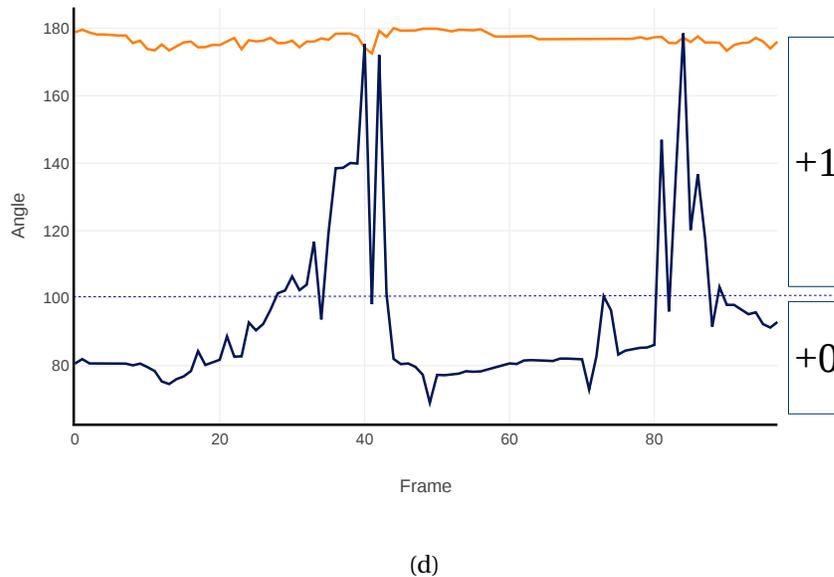
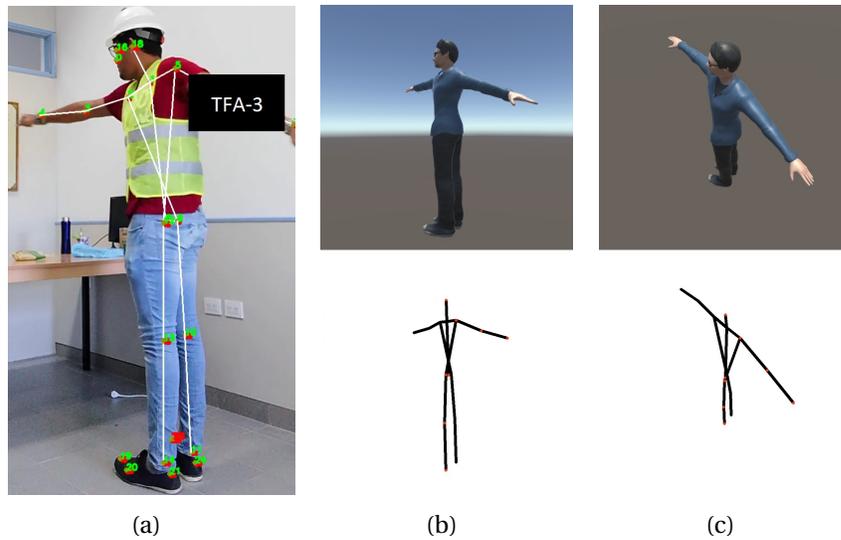
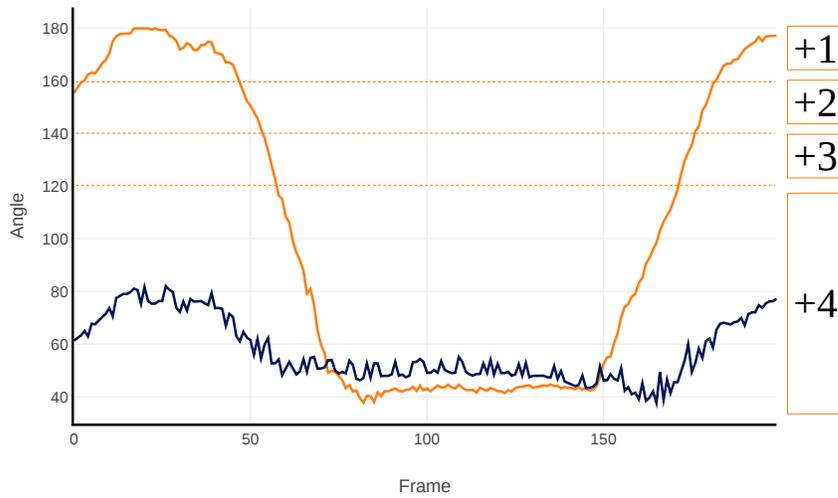
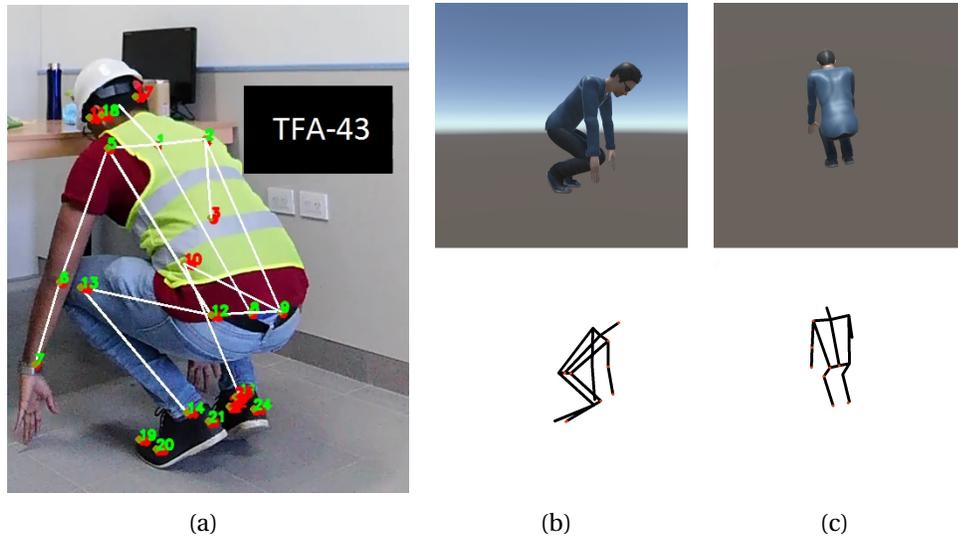


Figura 2.18: Prueba de laboratorio de torsión del tronco: (a) esqueleto OpenPose 25 superpuesto en una persona, (b,c) muestras de la simulación del conjunto de datos renderizados en 3D y esqueletos computados y (d) gráfico de ángulos del tronco inferidos TFA (naranja) y TTA (azul) vs fotograma, la acción de torsión ejecutada una vez para cada lado del cuerpo.

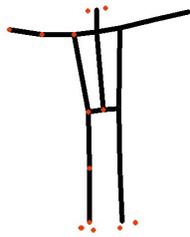
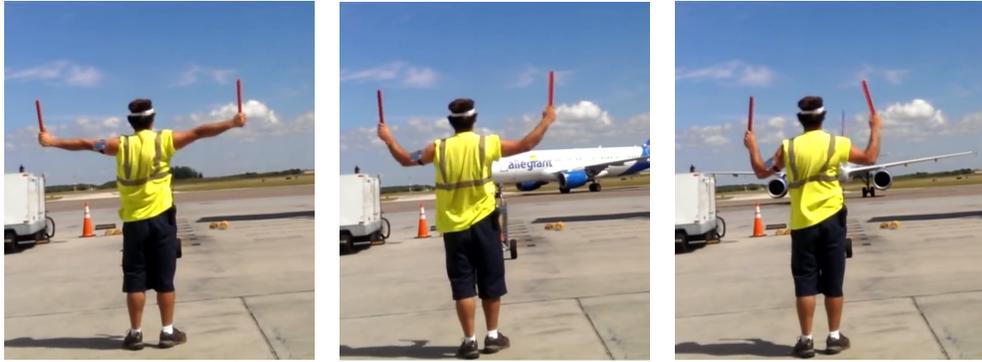
Finalmente, para verificar nuestro procedimiento, mostramos las posturas críticas previas de torsión y flexión del tronco a nuestro personaje 3D artificial para generar dos conjuntos de datos sintéticos adicionales (ver Figuras 2.18(b) y 2.18(c), y las Figu-



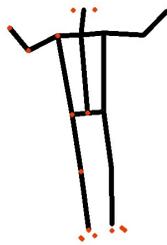
(d)

Figura 2.19: Prueba de laboratorio de flexión del tronco: (a) esqueleto OpenPose 25 superpuesto en una persona y (b,c) muestras de la simulación del conjunto de datos renderizados en 3D y esqueletos computados y (d) gráfico de ángulos del tronco inferidos TFA (naranja) y TTA (azul) vs fotograma, la acción de torsión ejecutada una vez para cada lado del cuerpo.

ras 2.19(c) y 2.19(b)), cada una desde 300 puntos de vista aleatorios. A partir de estos conjuntos de datos, pudimos corroborar la validez de nuestros resultados.



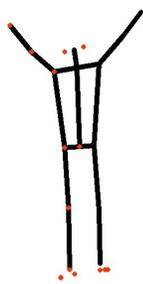
(a)



(b)



(c)



(d)



(e)

Figura 2.20: Muestras de la secuencia de trabajos con signos de Marshall.

2.3.7. Resultados en videos de secuencias de trabajo

Los parámetros de trabajo se establecieron como se indica en la Tabla 2.3. Se incluyeron los umbrales de confianza y los umbrales de tamaño vertical para minimizar los errores de detección. En las puntuaciones de RULA que no tienen un umbral claramente establecido, se asignaron manualmente las puntuaciones (ver Subsección 2.3.4). Los siguientes experimentos tenían como objetivo probar nuestro método con videos de trabajadores reales completando trabajos en diferentes escenarios de trabajo al aire libre.

Primero, analizamos un video de un operador que realiza señales de Marshall a un avión, desde un punto de vista fijo ubicado en su espalda (ver Figura 2.20). En esta secuencia, se tomaron fotogramas periódicamente para estudiar la precisión de nuestro método para evaluar los ángulos de las extremidades superiores, específicamente desde un punto de vista posterior en ausencia de posturas de flexión y torsión del tronco. Este video fue propuesto debido a la relevancia de estos ángulos en RULA. Además, un análisis temporal similar al de la Subsección 2.3.6 permitió la caracterización estadística de los ángulos entre el cuerpo y la articulación, es decir, media \pm desviación estándar. Según la Tabla 2.3, nuestro método proporcionó ángulos de hombro promediados (SL, SR) de $83,41^\circ \pm 26,8^\circ$, y ángulos de codo promediados (EL, ER) de $120,79^\circ \pm 49,17^\circ$, lo que arroja un puntaje promedio de posición del brazo superior e inferior RULA de +3 y +2, respectivamente. Como se discutió en la Subsección 2.3.6, el punto de vista posterior es un punto ciego para varias articulaciones OpenPose, que se podrían usar para calcular el ángulo del cuello en RULA. Este fue el caso de esta secuencia, donde incluso la oreja izquierda del trabajador se auto ocluyó en algunas imágenes, por ejemplo, en la Figura 2.20(e). Sin embargo, el uso del punto medio del oído (ME) nos permitió obtener ángulos de cuello precisos para cada imagen, de $6,72^\circ \pm 4,34^\circ$, lo que arroja una puntuación de cuello RULA de +1 para la mayor parte de esta secuencia.

A continuación, aplicamos nuestro método en los videos de actividades laborales de revoque de pared y martilleo, ambos tomados a mano con un teléfono celular. En estos casos, los ergonomistas tuvieron grandes dificultades de evaluación debido a las oclusiones repetidas y la visión de la cámara no estática. Por ejemplo, las piernas del yesero en la Figura 2.21 siempre están ocluidas por una carretilla, mientras que las articulaciones del cuello y la mitad de la cadera nunca se detectan con poca confianza.

Sin embargo, la inferencia ocasional de las articulaciones de la rodilla ocluidas obtenidas de OpenPose 25 permitió que nuestro método usara el punto de la rodilla media (MK) y que calcule un ángulo de flexión del tronco (TF) de $39,13^\circ \pm 30,03^\circ$ para toda la secuencia. Además, el método proporcionó promedios del ángulo del hombro izquierdo (SL) y derecho (SR) de $43,46^\circ \pm 24,86^\circ$ y $27,85^\circ \pm 12,02^\circ$, respectivamente. Estos resultados produjeron una puntuación RULA adecuada de la posición del brazo superior de +2 para la mayor parte de la secuencia, incluso bajo la influencia de la auto-oclusión del brazo izquierdo en algunas tomas (ver Figuras 2.21(a)-2.21(c)).

Tabla 2.6: Puntajes RULA y consenso con evaluaciones de expertos obtenidas en videos de trabajos al aire libre (en las secuencias de perforación y martilleo se evaluó a dos trabajadores, obteniendo los mismos puntajes).

Video	Tronco	Cuello	Puntaje A	Brazo	Antebrazo	Puntaje B	Puntaje RULA	κ de Cohen
Señales de Marshall	2	1	5	4	3	4	5	0.65
Revoque de pared	3	4	5	3	2	7	7	0.66
Martilleo	5	4	5	3	3	7	7	0.81
Tala de árboles 1	4	4	5	3	2	7	7	1.0
Tala de árboles 2	4	4	5	3	2	7	7	1.0
Perforación 1	5	3	5	3	3	7	7	0.61
Perforación 2	5	3	5	3	3	7	7	0.61
κ de Cohen	0.44	0.68	0.66	1.0	0.61	0.58	1.0	

Del mismo modo, en el video del trabajo de martilleo, el lado derecho del sujeto casi siempre está ocluido (ver Figura 2.22). La articulación del hombro derecho se detecta en todos los fotogramas. Sin embargo, cuando el codo derecho no es visible ni inferido, el método no puede calcular SR o ER. Sin embargo, nuestro método fue capaz ocasionalmente de inferir las ubicaciones de estas articulaciones ocluidas, permitiendo así el cálculo de la puntuación de RULA en estos pocos fotogramas. Por ejemplo, en la imagen que se muestra en la Figura 2.22(d), el codo derecho del trabajador es claramente visible y la muñeca derecha puede inferirse con buena confianza. Esto permitió el cálculo de las puntuaciones locales RULA utilizando los ángulos SR y ER de la Tabla 2.2. En estas condiciones subóptimas, nuestro método generó una medición combinada del ángulo del hombro de $36.80^\circ \pm 13.12^\circ$ para este video, por lo que, en la mayoría de los casos, obtuvo una puntuación RULA de la posición del brazo superior de +2. Dicha puntuación fue considerada una buena estimación por los ergonomistas. Otro ejemplo, es la medida del ángulo del cuello $42.69^\circ \pm 12.79^\circ$, el punto ME resultó como el punto medio de las orejas en las Figuras 2.22(a), 2.22(b), 2.22(c) y con la oreja izquierda en la Figura 2.22(d).

A continuación, aplicamos nuestro método en los videos de trabajo conjunto tala de árboles y perforación. El objetivo principal de estos experimentos fue corroborar que nuestro método evalúa correctamente la puntuación del tronco de varios sujetos al mismo tiempo. OpenPose 25 proporcionó consistentemente dos esqueletos para cada fotograma de cada secuencia. Esto nos permitió calcular simultáneamente puntajes RULA individuales para cada trabajador.

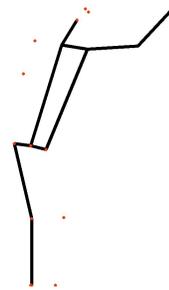
En la secuencia de tala de árboles (ver Figuras 2.23), dos trabajadores realizan actividades de trabajo coordinadas alrededor de un árbol, mientras que en el video de perforación (ver Figuras 2.24), dos operadores realizan la misma actividad de forma sincronizada para mantener equilibrado un taladro pesado. En la secuencia mencionada, los ergonomistas seleccionaron por consenso el fotograma clave que se muestra en la Figura 2.23(b) para analizar los ángulos del tronco individuales para cada trabajador. Esta elección se basó principalmente en la ausencia de auto-oclusiones (Figura 2.23(a)), oclusiones de las herramientas de trabajo (Figura 2.23(c)) y oclusiones del otro trabajador (Figura 2.23(d)), proporcionando así mejores condiciones para evaluar las pun-

tuaciones de RULA. En el fotograma seleccionado, el método propuesto calculó ángulos de flexión del tronco (TF) de aproximadamente 29° y 18° para el trabajador izquierdo y derecho, respectivamente. Vale la pena señalar que el método también fue capaz de manejar oídos completamente ocluidos (inferidos en las orejeras de seguridad), proporcionando puntajes de cuello RULA precisos (+4) de los puntos ME a pesar de esta situación desafiante.

Finalmente, en la secuencia de trabajo de perforación (Figuras 2.24), los especialistas estaban más interesados en examinar las variaciones de la puntuación del tronco RULA a lo largo de la secuencia. Nuestro método midió con precisión los TF promediados por los trabajadores desde aproximadamente 14° a 86°, lo que proporciona datos valiosos para que los ergonomistas detecten rápidamente imágenes con posturas críticas. La tabla 2.6 muestra los puntajes RULA obtenidos en los cinco videos, y la estadística κ de Cohen entre los resultados del método propuesto y las evaluaciones de los expertos, calculados por puntaje individual y por video individual.



(a)



(b)



(c)



(d)

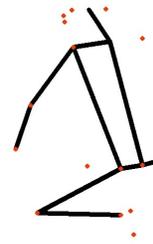


(e)

Figura 2.21: Muestras de la secuencia de trabajos de revoque de paredes.



(a)



(b)



(c)

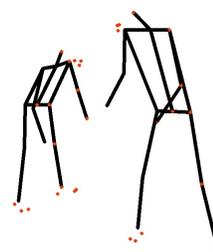


(d)

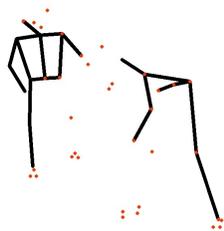
Figura 2.22: Muestras de secuencia de trabajo de martilleo.



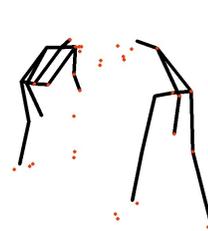
(a)



(b)



(c)

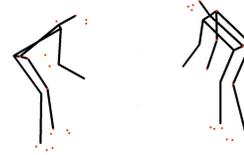


(d)

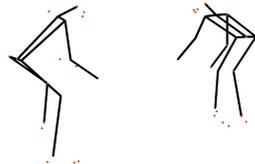
Figura 2.23: Muestras de secuencia de trabajo de tala de árboles.



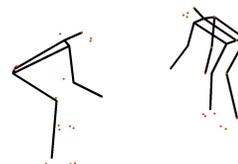
(a)



(b)



(c)



(d)

Figura 2.24: Muestras de la secuencia del trabajo de perforación.

2.4. Estimación de riesgos Ergonómicos mediante la red neuronal convolucional VIBE y el método REBA

En Massiris Fernández, Bajo, Fernández, y Delrieux (2020) y Massiris Fernández, Bajo, Martínez, y cols. (2020) proponemos un flujo de trabajo usando STAF (Raaj y cols., 2019) y VIBE (Kocabas, Athanasiou, y Black, 2020) para la extracción de articulaciones corporales y estimación de movimiento en 4D, con lo cual se computa el método de evaluación de riesgo ergonómico REBA. El modelo se evalúa utilizando escenarios virtuales diseñados específicamente en Unity3D. Finalmente, estudiamos el efecto de la posición de la cámara en la calidad de los resultados.

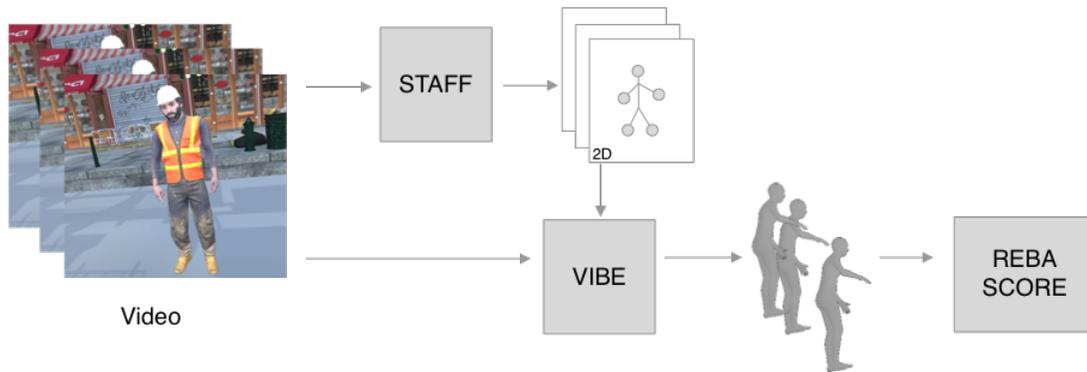


Figura 2.25: Metodología propuesta.

2.4.1. Descripción de Red Neuronal STAF

Raaj y cols. (2019) presentan una nueva metodología con un enfoque integral para el seguimiento de poses 2D de varias personas al mismo tiempo, que combina dos pasos en una sola solución. En un primer paso, detecta y agrupa partes del cuerpo en 2D, esos grupos son llamados esqueletos. En el segundo paso, STAF asocia los movimientos de los esqueletos a lo largo del tiempo utilizando una red de campos de afinidad espacio-temporal recurrente. Con una cámara estática, la red puede manejar a varias personas en movimiento (Raaj y cols., 2019). Como resultado de esta etapa, se deben obtener anotaciones de esqueletos 2D para describir los movimientos de las personas a lo largo del video (Raaj y cols., 2019).

2.4.2. Descripción de Red Neuronal VIBE

VIBE utiliza una Red Adversaria Generativa (GAN) entrenada y un módulo temporal para predecir los parámetros de pose y forma para cada fotograma de un video (Kocabas y cols., 2020). Más explícitamente, dados los esqueletos que describen el seguimiento de una persona, se utiliza un modelo temporal para predecir el modelo SMPL (*skinned*

multi-person linear model) para cada fotograma. Al mismo tiempo, un discriminador de movimiento intenta distinguir entre secuencias ostensibles y estimadas. Al hacerlo, se anima a la red a producir poses que representen movimientos plausibles. En el estado de la técnica, se ha demostrado que el método funciona para secuencias de video de longitud arbitraria y que VIBE supera a todos los métodos en conjuntos de datos 3DPW, MPI-INF-3DHP y Human3.6M (Kocabas y cols., 2020). Sin embargo, aún no se ha probado en las complejas condiciones de los entornos de trabajo.

2.4.3. Descripción de la metodología REBA.

McAtamney y Hignett (2000) presentan REBA, un método de evaluación de riesgo ergonómico diseñado para generar una estructura de examen postural sensible a los riesgos musculoesqueléticos en diversas tareas. El método se basa en la separación del cuerpo en secciones articulares para evaluarlas de forma independiente, en términos de planos de movimiento. REBA proporciona un sistema de puntuación para la actividad muscular causada por posturas estáticas, dinámicas, que cambian rápidamente o inestables. El resultado de REBA es un nivel de acción con una indicación de la gravedad de la evaluación. En este capítulo, nos centraremos en las desviaciones cruciales de los ángulos articulares que establece REBA: los ángulos de los codos, los hombros, las rodillas, el torso y el cuello.

2.4.4. Dataset

Al igual que Plantard y cols. (2015), evaluamos nuestra solución mediante un experimento sintético en Unity3D que cuantifica los efectos de auto-oclusión. Diseñamos un conjunto de datos sintéticos para analizar en un entorno controlado los posibles ángulos de cámara que pueden aparecer en aplicaciones industriales reales, como cámaras montadas en la cabeza de un trabajador, cámaras estabilizadas en drones o incluso cámaras de vigilancia. El Trabajador sintético simula el levantamiento de un objeto con la mano derecha, lo que agrega una variación de los ángulos de las articulaciones con respecto al lado izquierdo del cuerpo. El software realiza los siguientes pasos:

- Calcula una distancia aleatoria de entre 2 y 5 metros desde el modelo y la cámara.
- Define un conjunto de ubicaciones de la cámara aleatorias en un hemisferio superior de radio R . Usamos la distribución de Hammersley (ver Figura 2.6) para obtener un conjunto distribuido uniformemente de posiciones de cámara Ben-Arie (1990); Yan y cols. (2017).
- Reproduce la animación del esqueleto a una velocidad de fotogramas determinada (30fps), guardando los fotogramas del renderizado como partes de un video para cada una de las posiciones del conjunto de posiciones de cámara calculado (ver Figura 2.26).



Figura 2.26: Fotogramas de un ejemplo de simulación tomados con una posición de cámara aleatoria.

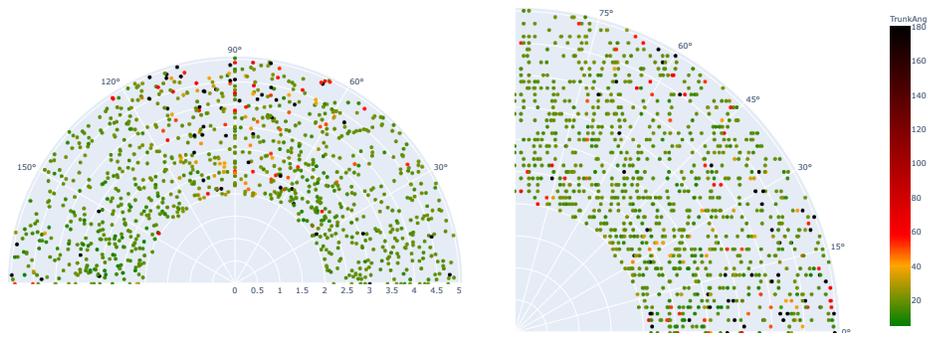


Figura 2.27: Error del ángulo del tronco vs altura y azimuth de la cámara.

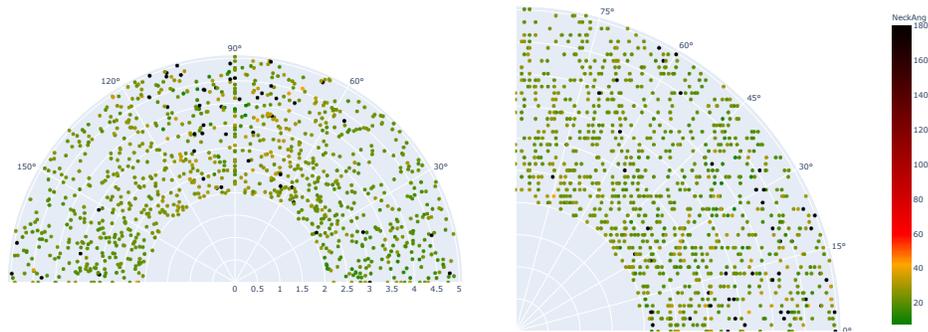


Figura 2.28: Error del ángulo del cuello vs altura y azimuth de la cámara.

- Con cada una de estas adquisiciones la posición 3D de las articulaciones especificadas por el usuario se registra en un archivo de texto CSV.

Mediante este procedimiento se realizaron un total de 1.073 tomas de video simuladas. La estimación real para cada error absoluto del ángulo de la articulación se calcula como la resta del ángulo de la articulación real (inferido de la animación del esqueleto) y el ángulo de la articulación obtenido por el modelo VIBE. Para todos los videos generados, se aplicó la metodología propuesta y luego se calculó la puntuación de precisión. El conjunto de datos está disponible en <https://shorturl.at/jow68>. El método presentado se desarrolló utilizando Python 2.7, Tensorflow, Numpy, Pims, Pandas, OpenCV y Plotly. Desarrollamos el código en una computadora con una tarjeta gráfica GeForce GTX 1080 ti, un procesador Intel Core i7, 16 Gb de RAM y bajo Ubuntu 18.04.

2.5. Resultados

En esta sección, presentamos los resultados de la evaluación del error entre el ángulo articular real en el avatar y el calculado en el flujo de trabajo de visión artificial que proponemos como alternativa a la medición ergonómica asistida por humanos. La animación consistió en un movimiento de levantamiento, que es una de las tareas más

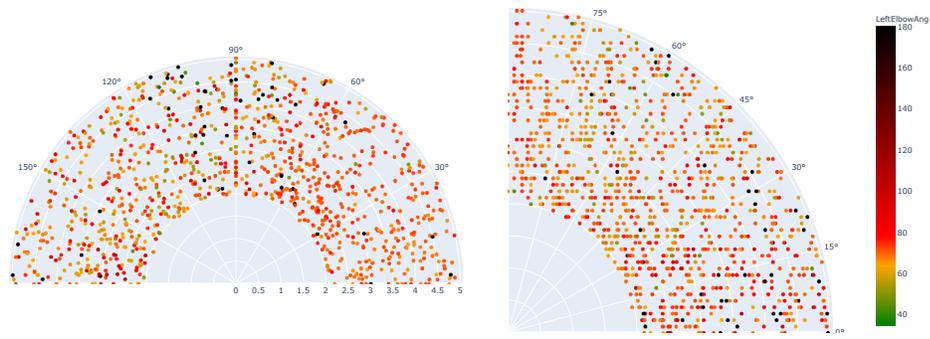


Figura 2.29: Error del ángulo del codo izquierdo vs altura y azimuth de la cámara.

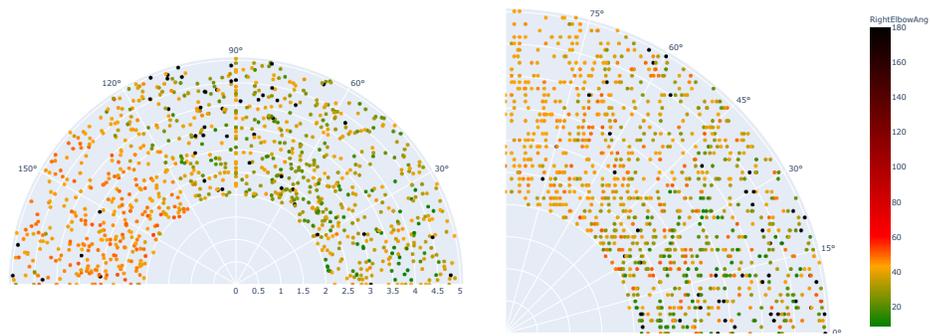


Figura 2.30: Error del ángulo del codo derecho vs altura y azimuth de la cámara.

comunes en el lugar de trabajo. El algoritmo se configuró para calcular todos los ángulos de articulación disponibles. Cuando VIBE no detecta a ninguna persona desde un punto de vista, todos los errores de ángulo de articulación se establecen en un valor máximo de 180°.

En el estado del arte de los métodos observacionales para la evaluación ergonómica, la práctica muestra que aproximadamente el 30% de las evaluaciones realizadas por los practicantes presentaron errores (Diego-Mas y cols., 2017). Por lo tanto, en la siguiente presentación, describimos el error en cada posición de la cámara usando una escala de

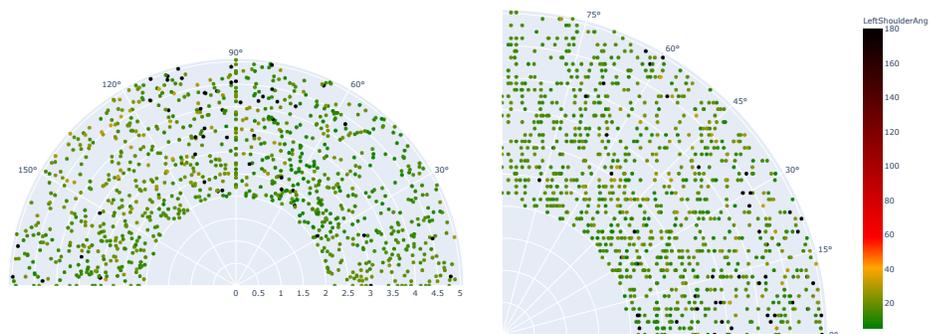


Figura 2.31: Error del ángulo del hombro izquierdo vs altura y azimuth de la cámara.

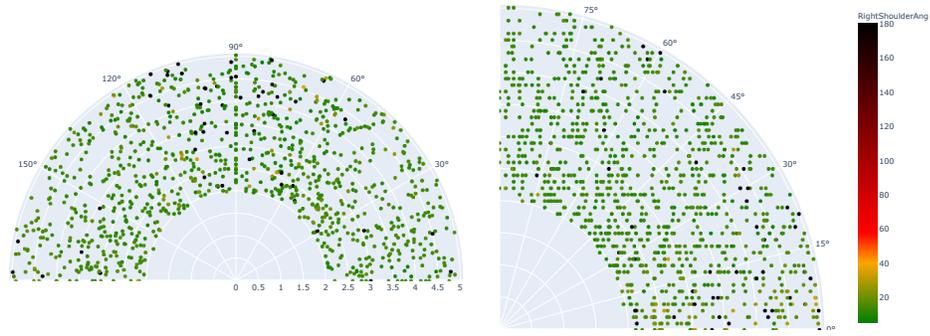


Figura 2.32: Error del ángulo del hombro derecho vs altura y azimuth de la cámara.

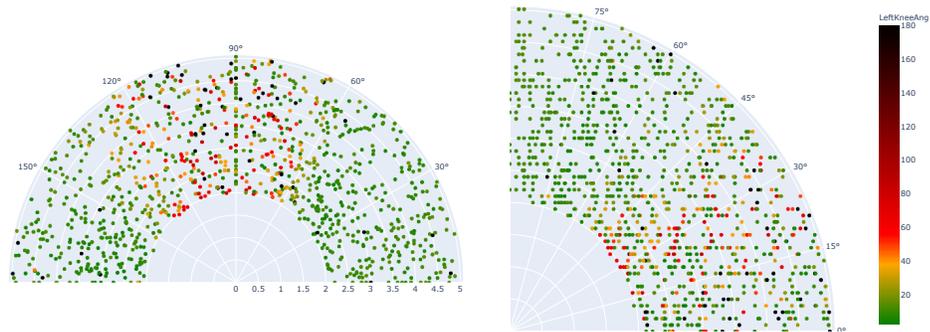


Figura 2.33: Error del ángulo de la rodilla izquierda vs. altura y azimuth de la cámara.

color que refleja errores mayores o menores que la evaluación asistida por humanos (verde menos de 30°, naranja 40°, rojo 50° y negro más de 90°).

Los errores del ángulo del tronco (Figura 2.27) y del ángulo del cuello (Figura 2.28) tienen una media similar de 34,93° y 33,38°, como el resto de ángulos los errores aumentan cuando la cámara se coloca en un punto de vista cenital. Sin embargo, los errores son razonables en el resto de puntos de vista.

Como puede verse en la Figura 2.29, el ángulo del codo izquierdo fue el que tuvo el peor error (media 74,42°, std 24,86°). Esto se puede atribuir a que esta fue la articulación

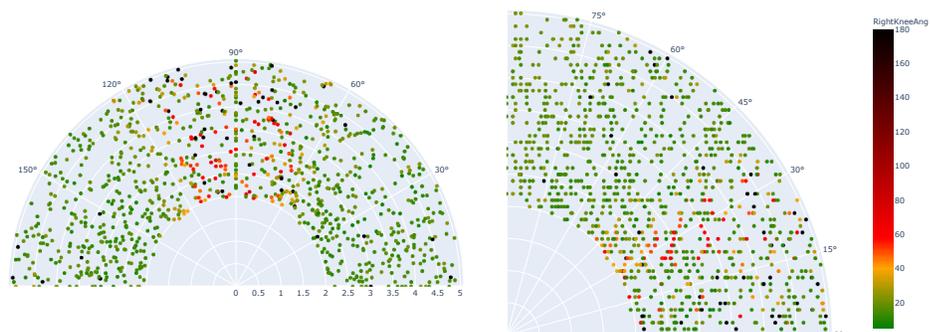


Figura 2.34: Error del ángulo de la rodilla derecha vs. altura y azimuth de la cámara.

en la toma de video simulada que tuvo el mayor movimiento, En algunas ocasiones la articulación se confundía con el fondo, lo cual generaba falsas detecciones. De manera similar, el error en el codo derecho (Figura 2.30) también cambió significativamente en la animación y por lo tanto presentó un error alto (media de $43,25^\circ$, std $31,72^\circ$).

El hombro izquierdo (Figura 2.31) y el hombro derecho (Figura 2.32) fueron estimados durante el experimento con errores menores (media $23,75^\circ$, std 21.43° a la izquierda y 34.92° , std 35.27° a la derecha). Finalmente, las rodillas generaron un error promedio de 25.0° (std 38.12°) y de 27.31° (std 35.72°) respectivamente para la rodilla izquierda y la rodilla derecha. Además, como era de esperar, también sufren autoclusiones cuando la cámara está en una perspectiva cenital.

Capítulo 3

Detección de equipos de protección personal

Los equipos de protección personal (EPP) son fundamentales para prevenir accidentes laborales y enfermedades profesionales ante la presencia de riesgos específicos que no se pueden aislar, ni eliminar (ILO-OSH, 2001). En este Capítulo nos proponemos presentar dos soluciones para el monitoreo de la utilización equipo de protección personal desde la perspectiva de la Industria 4.0.

3.1. Antecedentes de la problemática

En la Argentina la ley 19.587 de 1972 establece que el empleador debe adoptar e implementar medidas de higiene y seguridad adecuadas para proteger la vida e integridad de los trabajadores, dentro de lo cual se incluye el suministro y mantenimiento de EPPs (Gobierno de Argentina, 1972). Dichos EPPs son reglamentados por la Resolución 299 de 2011. Para una determinada actividad económica se pueden necesitar diferentes EPPs, como ser guantes, gafas, calzado de seguridad, tapones para los oídos, orejeras, cascos, respiradores, overoles, chalecos de alta visibilidad, o trajes de cuerpo entero, entre otros (Gobierno de Argentina, 2011).

Sin embargo, en la práctica debido a las limitaciones socio-económicas no siempre es posible cumplir la ley. La utilización de EPP debería ser preventiva, es decir, hasta que el riesgo pueda reducirse o anularse. Sin embargo, en los sectores económicos que trabajan con maquinaria pesada, y en particular en la industria de la construcción, la ausencia o el uso inadecuado de los EPP por parte de los trabajadores es una de las principales causas de accidentes y lesiones evitables (Mneymneh y cols., 2017). Por ejemplo, en Argentina la Encuesta Nacional a trabajadores sobre Condiciones de Empleo, Trabajo, Salud y Seguridad (Gadea y cols., 2019) de todos los trabajadores que consideran que el tipo de tarea que realizan requiere la utilización de EPPs, el 50,6% declaran dispo-

ner de todos los elementos necesarios, y el 32,3% menciona que cuenta con alguno y el 17,1% señala que no los tiene (ver Figura 3.1).

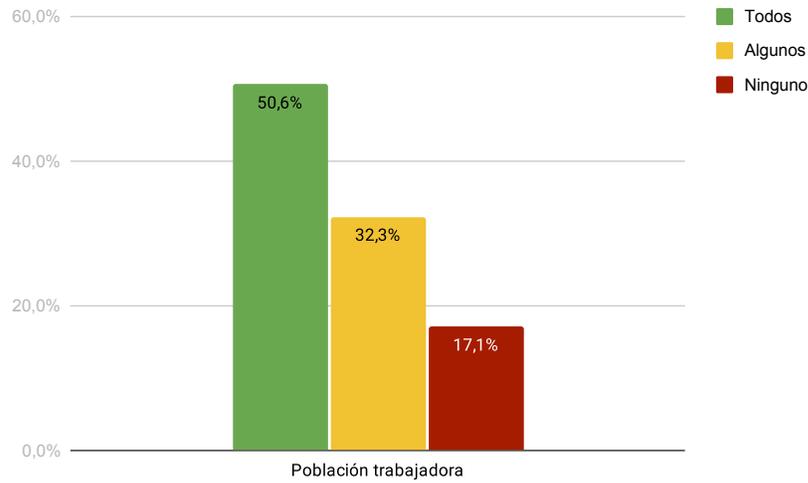


Figura 3.1: Utilización de EPP de los trabajadores que expresan necesitarlos (Gadea y cols., 2019).

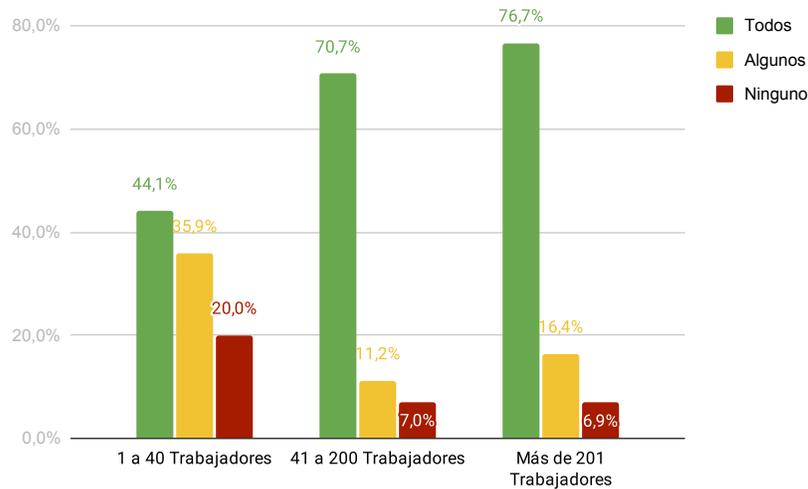


Figura 3.2: Utilización de EPP de los trabajadores que expresan necesitarlos según tamaño de empresa (Gadea y cols., 2019).

Como se puede ver en la Figura 3.2, el porcentaje de personas que reciben EPP aumenta a medida que aumenta el grado de riesgo de la tarea y el tamaño del establecimiento. Además, es mayor la utilización de EPP en trabajadores registrados en los sectores primario y secundario de la economía (ver Figura 3.3) (Gadea y cols., 2019).

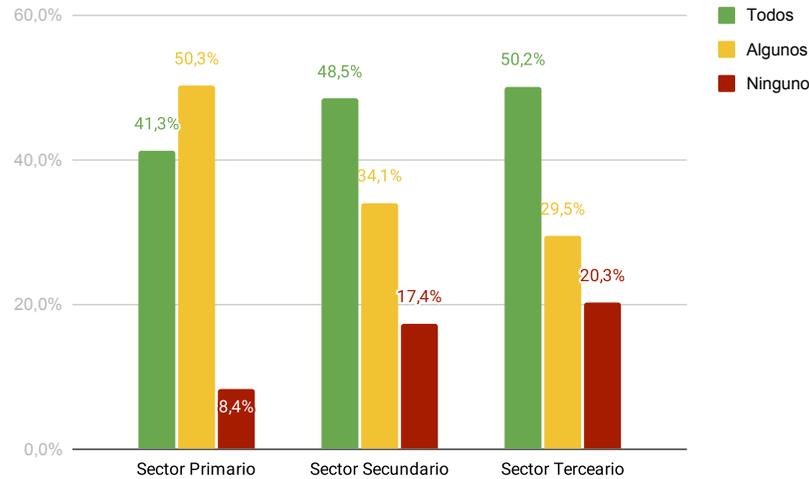


Figura 3.3: Utilización de EPP de los trabajadores que expresan necesitarlos según sector productivo.

Kelm y cols. (2013) mencionan algunos factores por los cuales un trabajador escoge no utilizar EPP, como por ejemplo, les resulta incómodo de usar, el EPP puede afectar su productividad, la exposición a un peligro es breve y por lo tanto, no creen que es necesario, la educación y capacitación son inadecuadas o no disponibles específicamente para el adecuado uso de EPP, y no usan EPP porque simplemente se olvidan. Por todas las razones mencionadas, es común que los encargados de HST utilicen metodologías para el control de utilización de equipo de protección personal.

Dichas metodologías pueden ser divididas en dos. La primera es reglamentada por el Gobierno Argentino a través de un formulario, que debe ser llenado por el personal de HST o por la ART según corresponda, y firmado por el empleado (Gobierno de Argentina, 2011). Sin embargo, la historia menciona que este tipo de informes pueden subestimar los peligros y requieren la colaboración en general de los trabajadores en la medición del uso de EPP, lo cual genera que sus efectos se tornan difíciles de medir e imprecisos a lo largo del tiempo (Guo y cols., 2017). La segunda es la metodología más utilizada en proceso de monitoreo del uso de EPP, se realiza a través de la observación in-situ por parte de expertos, lo cual resulta en una limitante y además es costosa debido a la necesidad de contratar personal especializado para realizar esta tarea (Mosberger y cols., 2014).

El estado del arte indica que estas dos metodologías tienen varias deficiencias asociadas, y en general dependen en gran medida de la intervención humana (Igelmo y cols., 2020; Yu, Guo, Ding, Li, y Skitmore, 2017), pues es común encontrar una notable variabilidad en los resultados obtenidos por diferentes evaluadores, o incluso por un mismo evaluador en diferentes contextos. La variabilidad inter-observador se debe principalmente a factores subjetivos como ser el entrenamiento, la experiencia o los medios de observación, mientras que la variabilidad intra-observador normalmente se



Figura 3.4: Proceso Propuesto, de izquierda a derecha: (1) dividir en bloques de 13x13, (2) hacer predicciones con YOLO re-entrenado y (3) establecer umbrales para obtener solo las detecciones más confiables.

asocia con fatiga cognitiva y cambios en la observación del entorno (Park, Elsafty, y Zhu, 2015). Como último paso en las metodologías expuestas, se pueden implementar correcciones utilizando realimentación visual o auditiva para prevenir o mitigar el peligro (Yu y cols., 2017), y también se pueden implementar incentivos o se pueden focalizar esfuerzos durante el entrenamiento del personal (Kelm y cols., 2013). Por las razones mencionadas anteriormente, proponemos el monitoreo de utilización de equipos de protección personal en los entornos laborales a través de las metodologías ofrecidas por la industria 4.0.

3.2. Detección de equipos de protección personal mediante la red neuronal convolucional YOLO

Esta sección se centra en el monitoreo de la presencia y el uso adecuado del EPP por parte de trabajadores de la industria metalúrgica. Para ello, se utiliza la red neuronal convolucional (CNN) denominada “You Only Look Once” (YOLO) (Redmon y Farhadi, 2017), cuya topología está específicamente diseñada para resolver problemas de visión artificial con un alto rendimiento. En YOLO, la detección de objetos se toma como un problema de regresión, la red convolucional predice simultáneamente múltiples cuadros delimitadores que rodean los objetos en la imagen y predice probabilidades condicionales para cada clase $p(\text{Clase}|\text{Objeto})$. Además, la red neuronal puede alcanzar una velocidad operativa de 45 cuadros por segundo (fps) en computadoras de propósito general. Además, codifica implícitamente información contextual, y modela el tamaño y la forma de los objetos, así como su apariencia (Redmon y Farhadi, 2017) (ver Figura 3.4).

Dataset de trabajadores metalúrgicos

Debido a las regulaciones de protección de datos de la Unión Europea y Argentina, se tomaron como muestra 5 videos obtenidos de YouTube.com, los cuales fueron grabados con cámaras deportivas sujetas al casco de un trabajador, destacando que dicha colocación de la cámara no impide el desarrollo normal de la actividad productiva y es posible grabar las tareas colaborativas de estos trabajadores metalúrgicos. Posteriormente, cada video fue dividido en imágenes (1354 en total). Finalmente cada imagen fue etiquetada manualmente utilizando la herramienta YOLO-mark (AlexeyAB, 2018), la cual fue diseñada por los mismos autores de YOLO. Las anotaciones se guardaron en archivos txt siguiendo el formato de YOLO. Por lo demás, siguiendo las recomendaciones de YOLO, se aleatorizaron las imágenes y se dividieron en grupos, el 80% de las imágenes fueron destinadas a entrenamiento, el 10% a pruebas y, finalmente, el 10% restante para validación. Por último, se ajustó la red a la cantidad de clases a entrenar y se validaron los tamaños de referencia de los objetos (*anchors*, en inglés) en relación con el mapa de características de YOLO.

Entrenamiento

Para el proceso de entrenamiento se utilizó la distribución original YOLOv2, vía Darknet (Redmon y Farhadi, 2017), empenado una computadora con 16 GB de memoria, una tarjeta gráfica NVIDIA GeForce GTX 1080 TI y un procesador Intel Core i5. Debido a la gran cantidad de herramientas y la flexibilidad que ofrece Python, se utilizó para la puesta en producción una versión de YOLO2 desarrollada bajo Keras con un backend de Tensorflow y OpenCV (Experiencor, s.f.). El proceso de entrenamiento fue relativamente sencillo, se genera un archivo con los pesos de la red cada 1000 épocas, y tras 22.000 épocas se compararon los resultados entre sí, y se seleccionaron los coeficientes de la red neuronal generados en 9.000 épocas, pues estos presentaban las mejores estadísticas.

Métricas de Precisión

Las medidas de precisión fueron seleccionadas en base al objetivo principal del método de detección de cascos, ropa de alta visibilidad, guantes y clases de trabajadores de forma individual. Para aclarar el significado de las métricas, primero debemos definir el significado de VP (verdadero positivo), FP (falso positivo) y FN (falso negativo). Más precisamente, VP es el número de clases detectadas con éxito. Se considera que una clase se detectó correctamente cuando el área predicha por la red y el área esperada se superponen al menos en un 0,7 (IOU). FP es el número de objetos detectados que no coinciden con los etiquetados manualmente (Yu y cols., 2017). FN es el número de objetos de cada clase que la red neuronal no detectó. VN, el número de clases no existentes que no fueron detectadas, no es tenido en cuenta en este tipo de experimentos dado

Categoría	Referencia	Estimado por la red
VP	Sí	Sí
FP	No	Sí
FN	No	No
VN	Sí	No

que emplearlo trivializa la validez de los resultados. La Tabla 1 presenta las definiciones de VP, FP y FN. En la sección 5.3 se muestran ejemplos de VP, FP y FN.

Precisión o valor predictivo positivo.

La primera métrica es la precisión, que es muy popular para evaluar el reconocimiento de imágenes. La precisión se define como una relación $VP / (VP + FP)$ y mide la confiabilidad de la detección o proporción de casos verdaderamente positivos entre los casos positivos detectados por la prueba, teniendo en cuenta que $VP + FP$ es el número de objetos detectados en cada clase en función de la red neuronal entrenada (Everingham, Van Gool, Williams, Winn, y Zisserman, 2010; Yu y cols., 2017).

Sensibilidad

La sensibilidad se establece como la relación $VP / (VP + FN)$, haciendo hincapié en que $VP + FN$ representa el número real de objetos de cada clase (Everingham y cols., 2010; Yu y cols., 2017).

La curva de precisión-sensibilidad

Esta curva muestra el equilibrio entre precisión y sensibilidad. Una alta precisión está relacionada con una baja tasa de falsos positivos mientras que una alta sensibilidad está relacionada con una baja tasa de falsos negativos. Un buen clasificador o detector podría encontrar un punto de equilibrio con resultados tanto precisos como de alta sensibilidad (Everingham y cols., 2010; Yu y cols., 2017).

Precisión media

Esta es una aproximación del área bajo la curva de precisión-sensibilidad (AP) que representa una generalización con respecto al umbral para cada clase (Everingham y cols., 2010).

Media de precisión media

La precisión media de cada clase es equivalente al AP, tomando este valor se calcula la media. En la práctica, un valor de precisión media (mAP) más alto indica un mejor rendimiento de la red neuronal (Everingham y cols., 2010).

3.2.1. Resultados

El rendimiento de la red YOLO re-entrenada se midió en función de su precisión y velocidad. En este apartado se analizan 14,23 minutos de video. Dicho video es independiente de los que se utilizan para entrenar y probar la red neuronal, pero cumple con las características necesarias para dicho propósito. Fue grabado con una cámara deportiva sujeta al casco de un trabajador metalúrgico en una construcción, por lo que esta es una oportunidad para medir los resultados en condiciones reales, y ahondar en las oportunidades de mejora. En el video, los trabajadores detectados están etiquetados en verde, la ropa de alta visibilidad detectada en negro, los guantes detectados en rojo y los cascos detectados en azul (ver <https://youtu.be/54gwd2lcjmg>). Como punto de equilibrio al analizar las curvas de precisión-sensibilidad para cada una de las clases, se eligió 0.55 como el umbral de confianza para el estudio.

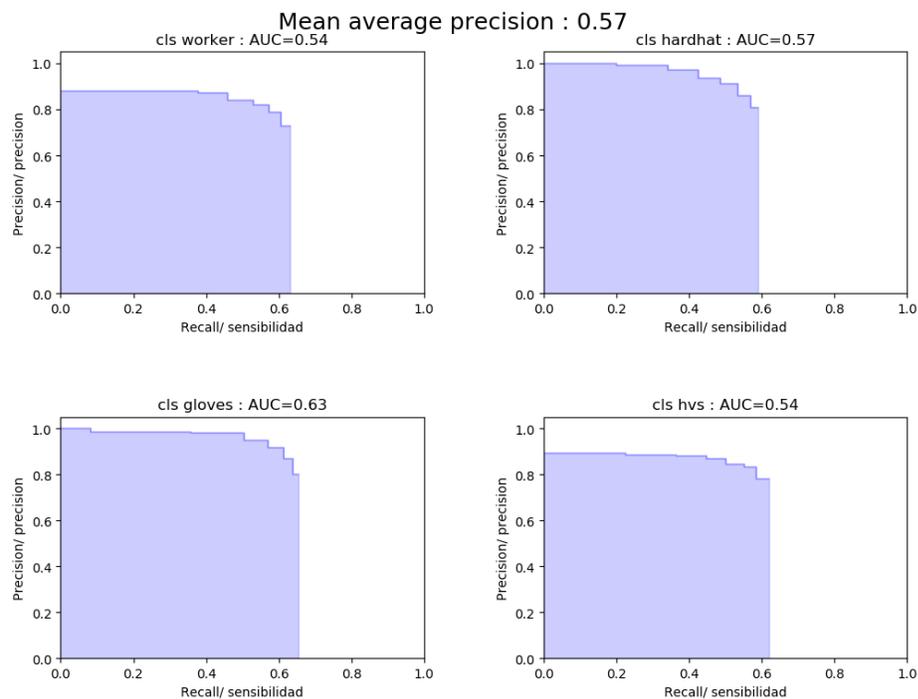


Figura 3.5: Curvas precisión-sensibilidad para cada una de las clases.



Figura 3.6: Ejemplo de verdadero Positivo en el fotograma 17956.

Un ejemplo de verdadero positivo se muestra en la Figura 3.6 Destacamos la dificultad de reconocer el casco que tiene diferentes adhesivos que dificultan el reconocimiento por color o forma, además de reconocer los guantes en dos posiciones distintas y al trabajador sentado de forma poco convencional. En el fotograma de la Figura 3.7 se observa que la red neuronal es capaz de detectar parcialmente al trabajador, sin embargo, se clasifica como falso negativo al guante del trabajador y el trabajador que está en el suelo. En unas pocas imágenes los objetos son etiquetados erróneamente causando falso positivos, en el caso de la Figura 3.8 se etiqueta de trabajador a un tubo vertical, por lo cual se cuenta para cada una de las clases como FP.

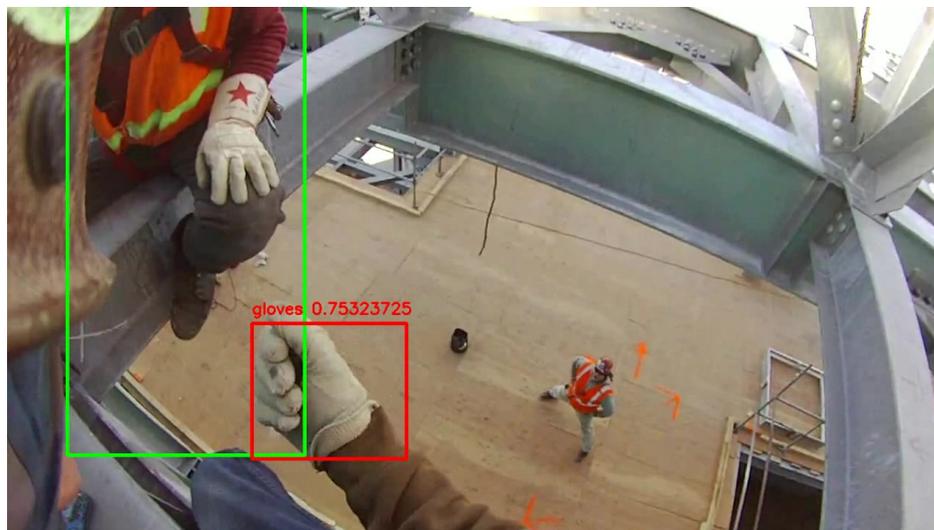


Figura 3.7: Ejemplo de falso negativo en el fotograma 17871.



Figura 3.8: Ejemplo de falso positivo en el fotograma 02376.



Figura 3.9: Ejemplo de etiquetado múltiple en el fotograma 17786.

Como se puede ver en la Figuras 3.11 y 3.10, para la eliminación de cuadros delimitadores superpuestos es necesario implementar en futuro un supresor, de tal forma que, si dos detecciones comparten un área determinada, el supresor debe eliminar la detección que presenta menor confianza. El objetivo de dicho supresor debe ser que un determinado objeto no sea reconocido más de una vez.

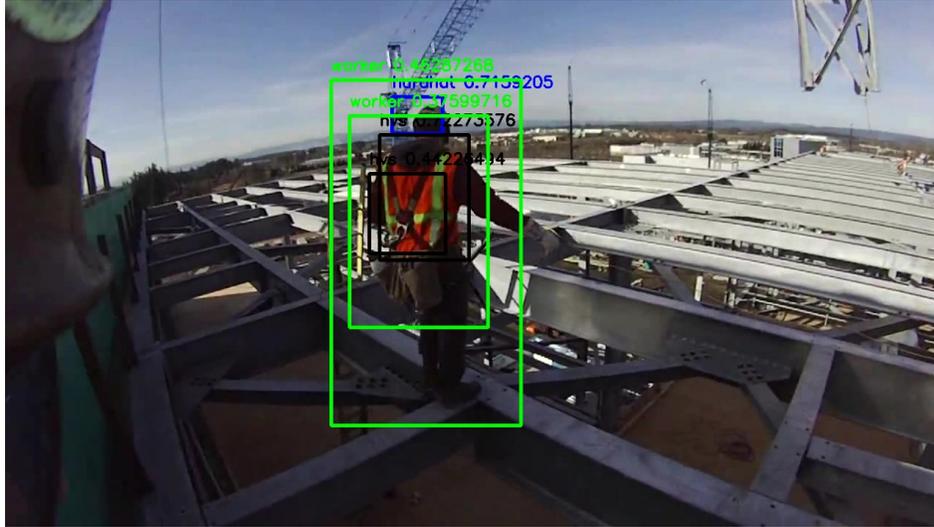


Figura 3.10: Ejemplo de etiquetado múltiple en el fotograma 00015.



Figura 3.11: Ejemplo de etiquetado múltiple en el fotograma 20448.

Ejemplo de impacto del rango visual

Como la trayectoria de los trabajadores es estocástica (Yu y cols., 2017), los trabajadores fueron capturados en diferentes tamaños en los videos, señalando que si los trabajadores están cerca de la cámara, se capturan con un tamaño más grande y tienen características de imagen más ricas. Por el contrario, los trabajadores lejos de la cámara son capturados con un tamaño más pequeño y exhiben características de pérdida de detalles y desenfoque. Esto se amplifica cuando se habla de un EPP específico, ya que son más pequeños que cualquier trabajador. Si bien es cierto que, como se puede ver en el video, existe una dificultad en detectar los guantes cuando el trabajador está lejos,

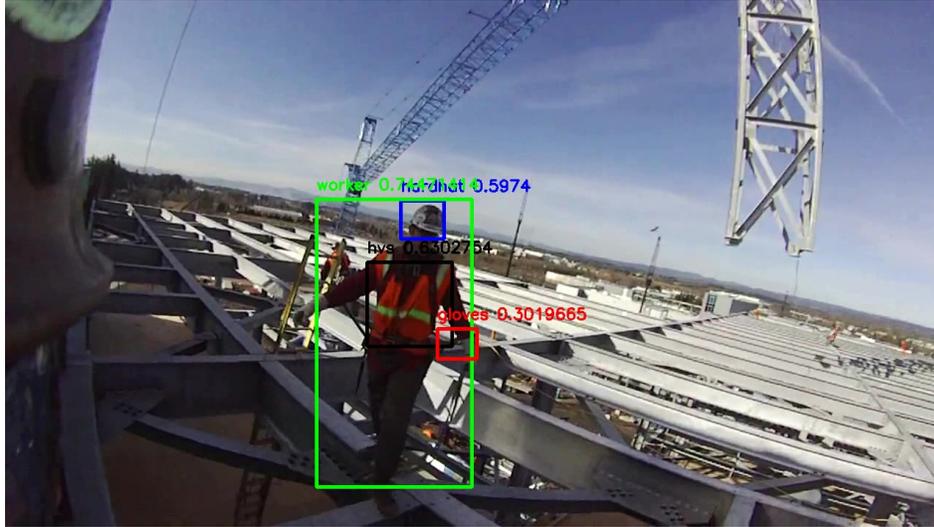


Figura 3.12: Ejemplo de impacto del rango visual 00187.



Figura 3.13: Ejemplo de impacto del rango visual 00095 .

este es uno de los inconvenientes esperados al usar YOLOv2, pues en su entrenamiento de base es difícil de encontrar artículos pequeños en comparación con con el tamaño de la imagen (Redmon y Farhadi, 2017).

Ejemplo de impacto de las oclusiones

La red neuronal demuestra que es capaz de reconocer los guantes cuando están cerca de la cámara. Esto es importante porque nos permite detectar si el trabajador que lleva la cámara está usando guantes, como el que se muestra en la Figura 3.16. La red

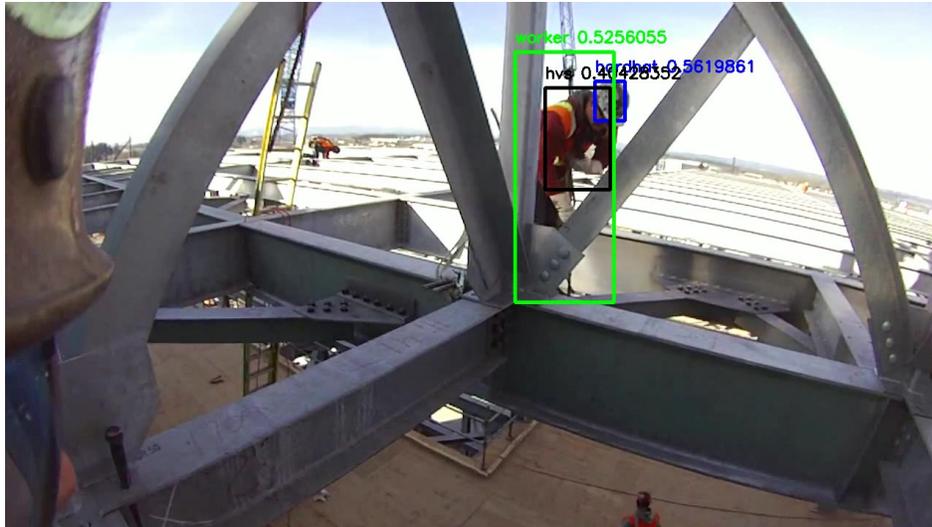


Figura 3.14: Ejemplo de impacto de las oclusiones 15950.



Figura 3.15: Ejemplo de impacto de las oclusiones 19532.

neuronal también funciona en la detección parcial de trabajadores, como se muestra en la Figura 3.17 .

Ejemplo de impacto de la postura individual y de la distorsión generada por el tipo de lente

La red Yolo está diseñada para captar las generalizaciones de las clases entrenadas, por lo cual es menos probable que falle cuando se expone a las diferentes posturas que puede mantener un trabajador. Como se puede apreciar en la Figura 3.19 la distorsión



Figura 3.16: Ejemplo de impacto de las oclusiones 19560.

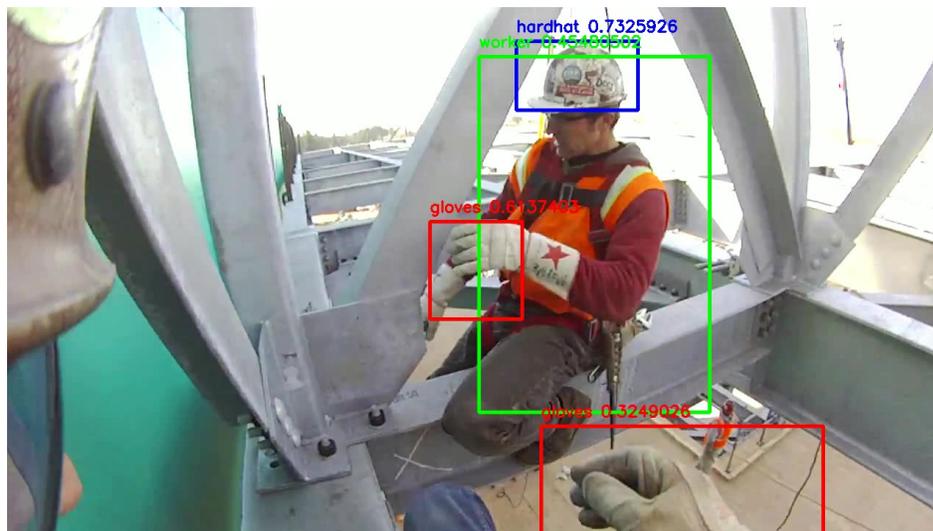


Figura 3.17: Ejemplo de impacto de la postura individual y de la distorsión generada por el tipo de lente 19764.

en la lente genera dificultad para reconocer las líneas rectas, sin embargo, la red neuronal entrenada mantiene la capacidad de superar las dificultades que esto genera e interpretar correctamente la escena.

Velocidad

Si bien es cierto que este tipo de análisis no necesita realizarse siempre en tiempo real, la velocidad es una métrica importante porque uno de los objetivos de este trabajo

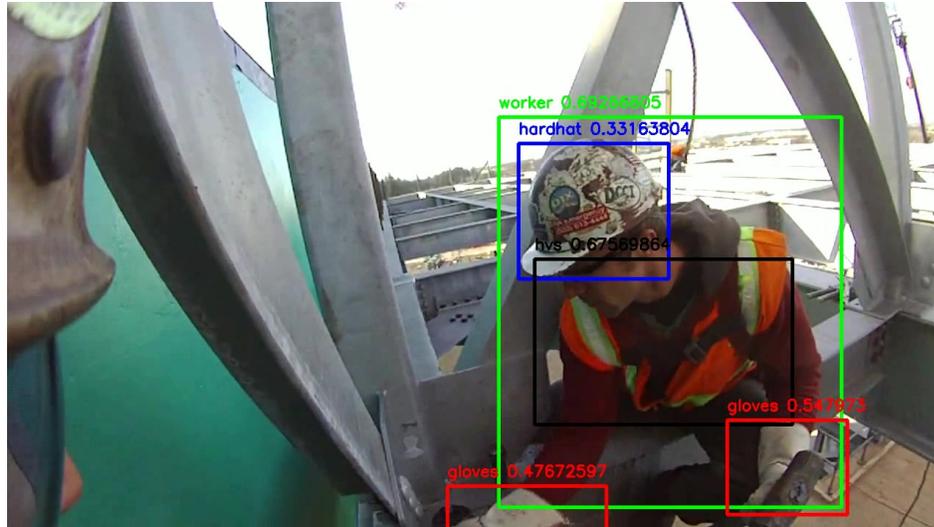


Figura 3.18: impacto de la postura individual y de la distorsión generada por el tipo de lente 18141.

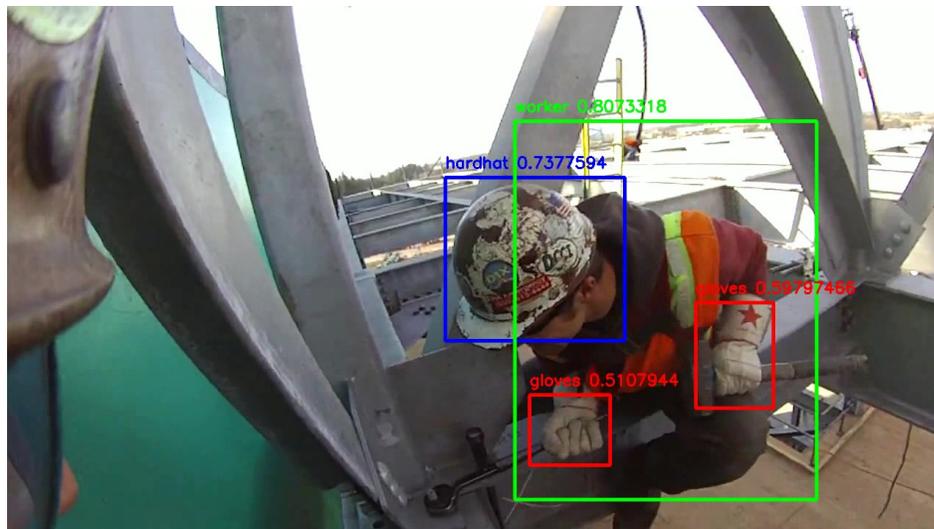


Figura 3.19: impacto de la postura individual y de la distorsión generada por el tipo de lente 18380.

es minimizar el tiempo que dedican los ingenieros de seguridad a analizar este tipo de escenarios. La velocidad de la red neuronal entrenada YOLOv2 se refiere al tiempo que consume dicha red neuronal al detectar todas las clases entrenadas en el video de 14,25 minutos, con un formato de 1200 x 720 píxeles y 30 cuadros por segundo. Para el cálculo de la velocidad, se utiliza la misma tarjeta gráfica GTX 1080 utilizada durante el entrenamiento. Como resultado, se obtiene la red neuronal para utilizar 14,23 minutos para analizar el video mencionado.

3.3. Detección de equipos de protección personal mediante las redes neuronales convolucionales OpenPose y Resnet

Nuestro objetivo en esta sección es diseñar y probar una nueva metodología para detectar el uso correcto de EPP en videos sin cámara fija, utilizando técnicas de CV y aprendizaje automático, con el fin de prevenir accidentes laborales.

3.3.1. Descripción de Red Neuronal OpenPose

Como ya mencionáramos, OpenPose (OP) se destaca de otras CNN de última generación, porque tiene una alta tasa de detección de personas al identificar los puntos antropométricos de un esqueleto (Cao y cols., 2017). A diferencia del objetivo establecido en el Capítulo anterior, para la detección de EPP se requiere la versión más pequeña de OP. La red OP18 fue seleccionada porque es más liviana y rápida, y sus esqueletos contienen 18 puntos antropométricos, suficientes para nuestro propósito. OP18 toma una imagen digital RGB como entrada de un CNN multicapa de dos etapas. En el primer paso se predicen los mapas de confianza de los 18 puntos antropométricos del cuerpo, mientras que en la segunda etapa, estos puntos se asocian con todos los esqueletos E posibles. La red obtiene como salida un conjunto variable de tuplas $(E.x, E.y, E.c)$, asociado a la posición estimada 2D (x, y) y con la confiabilidad de la estimación, c , de 18 puntos antropométricos agrupados en E esqueletos de acuerdo con la figura 3.20.

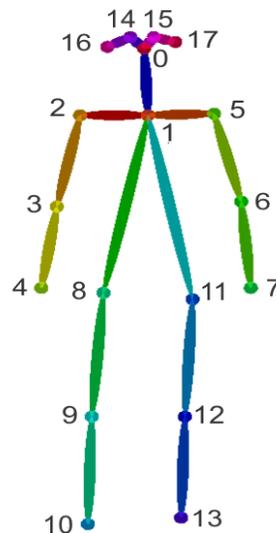


Figura 3.20: Puntos antropométricos en OP (Cao y cols., 2017).

3.3.2. Descripción de Red Neuronal Resnet

ResNet (*Residual Network*) es una CNN caracterizada por la inclusión de conexiones residuales, que permiten extender la conectividad directa de una capa a cualquier capa posterior, facilitando aumentar el número de capas sin pérdida de la capacidad de generalización (He y cols., 2016). Esto la convierte en una de las mejores redes de clasificación en términos de precisión frente al costo computacional. En este trabajo se eligió el modelo ResNet de 18 capas (RN18). Para adaptar esta red a nuestro problema, se han re-entrenado las últimas tres capas de dos RN18, la primera para clasificar cascos y la segunda para clasificar chalecos de alta visibilidad.

3.3.3. Implementación de la metodología

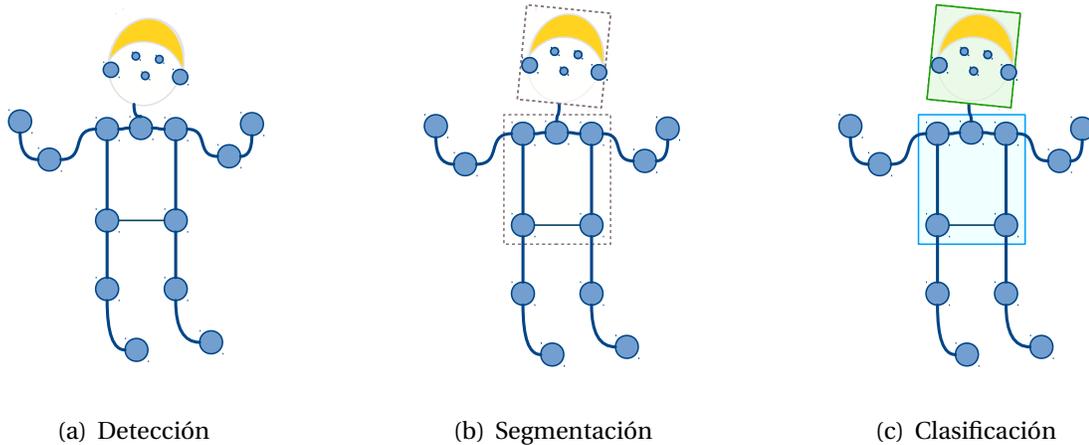


Figura 3.21: Determinación de las regiones de interés.

Algorithm 2 Detección antropométrica de EPP

Input: $Vid, u_{Conf}^e, u_{Tam}^e, u_{Conf}^{ca}, u_{Tam}^c, u_{Conf}^t, u_{Tam}^t$ **forall** j **in** Vid **do**
 $img \leftarrow Vid[j]$
 $esqueletos \leftarrow OP18(img)$
 forall k **in** $esqueletos$ **do**
 $E \leftarrow esqueletos[k]$
 $E.Casco \leftarrow NC$
 $E.Chaleco \leftarrow NC$
 if $E_{Conf} > u_{Conf}^E$ **and** $E_{Tam} > u_{Tam}^E$ **then**
 $E.Rca \leftarrow segmentarCabeza(E)$
 if $E.Rca_{Conf} > u_{Conf}^{ca}$ **and** $E.Rca_{Tam} > u_{Tam}^{ca}$ **then**
 $E.Casco \leftarrow RN18^{ca}(E.Rca)$
 end
 $E.Rtr \leftarrow segmentarTronco(E)$
 if $E.Rtr_{Conf} > u_{Conf}^{tr}$ **and** $E.Rtr_{Tam} > u_{Tam}^{tr}$ **then**
 $E.Chaleco \leftarrow RN18^{tr}(E.Rtr)$
 end
 end
 end
end

end
Result: $E.Casco, E.Chaleco$

El algoritmo 2 y la Figura 3.21 presentan una descripción general de la metodología adoptada. Cada fotograma del video de entrada Vid se analiza inicialmente con la red OP18. Por cada esqueleto E obtenido con suficiente tamaño E_{Tam} y confiabilidad E_{Conf} , se determina una región de interés (ROI) asociada con la cabeza ($E.Rca$), y otra asociada con el tronco ($E.Rtr$). En dichas regiones se espera encontrar cada EPP de interés. Como se muestra en la Figura 3.21, estos ROI se definen a partir de los puntos OP18 relacionados con las partes del cuerpo con índices $j_{ca} = [1, 16, 17]$ para $E.Rca$ y $j_{tr} = [1, 2, 5, 8, 11]$ para $E.Rtr$. Esta segmentación permite definir para cada E , valores mínimos de tamaño y confiabilidad para cada ROI, llamados $u_{Conf}^{ca}, u_{Tam}^{ca}, u_{Conf}^{tr}$ y u_{Tam}^{tr} en el algoritmo 2.

Finalmente, las sub-imágenes obtenidas de $E.Rca$ y $E.Rtr$ con tamaño y confiabilidad suficientes son procesadas por RN18 entrenadas para detectar la presencia de un EPP específico. Por lo tanto, las redes de clasificación $RN18^{ca}$ y $RN18^{tr}$ asignan respectivamente a cada E dos etiquetas, $E.Casco$ y $E.Chaleco$, ambas con valores (Sí, No). Finalmente, cada E procesado se almacena en una base de datos para su posterior procesamiento estadístico.

3.3.4. Umbrales y ROI

Para los propósitos de esta investigación, la función *segmentarTronco()* del algoritmo 2 define *ERtr* como el rectángulo de área mínima, con lados paralelos a los bordes de la imagen, que contiene todos los puntos antropométricos de *E* asociados con la cintura y los hombros. Para definir la región de la cabeza, los puntos de referencia antropométricos más comunes son la base del cuello y el punto medio entre ambas orejas (Ankrum y Nemeth, 2000). Sin embargo, las dos orejas solo son detectables en los planos frontal y posterior, mientras que en los planos laterales, solo una oreja es visible. Debido a la compleja geometría involucrada en esta región, en este trabajo la función *SegmentarCabeza()* del algoritmo 2 delimita *E.Rca* como un cuadrado con lados paralelos a los bordes de la imagen, cuyo centroide se estima como el punto medio de los oídos en el plano frontal, o como el punto antropométrico asociado al oído visible en un plano lateral, y cuyo lado es el doble de la distancia desde la base del cuello hasta dicho centroide (véase Figura 3.22).

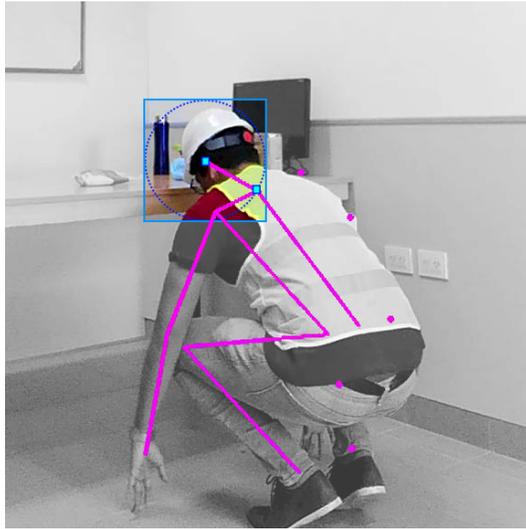


Figura 3.22: Puntos antropométricos de *E.Rca*.

Como se estableció en la sección anterior, el tamaño del esqueleto y los umbrales de confiabilidad (u_{Tam}^E y u_{Conf}^E), la cabeza (u_{Tam}^{ca} y u_{Conf}^{ca}) y el tronco (u_{Tam}^{tr} y u_{Conf}^{tr}) se utilizan para eliminar esqueletos y objetivos de ROI que pueden presentar problemas en la etapa de clasificación. En particular, los umbrales de confiabilidad evitan considerar resultados inexactos, mientras que los umbrales de tamaño filtran pequeños ROI que pueden generar clasificaciones falibles. En los experimentos de este trabajo, se utilizan los siguientes umbrales de confianza: $u_{Conf}^E = 0.35$, $u_{Conf}^{ca} = 0.3$ y $u_{Conf}^{tr} = 0.3$. El umbral $u_{Tam}^E = 0.2$, se define como el valor mínimo de la relación entre la altura del esqueleto E_{Tam} y la altura de la imagen. Finalmente, dado que los clasificadores RN18 están diseñados para recibir una imagen RGB de 224×224 px en su entrada, se usa un umbral mínimo a lo largo de la diagonal de cada ROI de interés $u_{Tam}^{ca} = 112$ px y $u_{Tam}^{tr} = 112$ px.

3.3.5. Conjunto de datos y entrenamiento

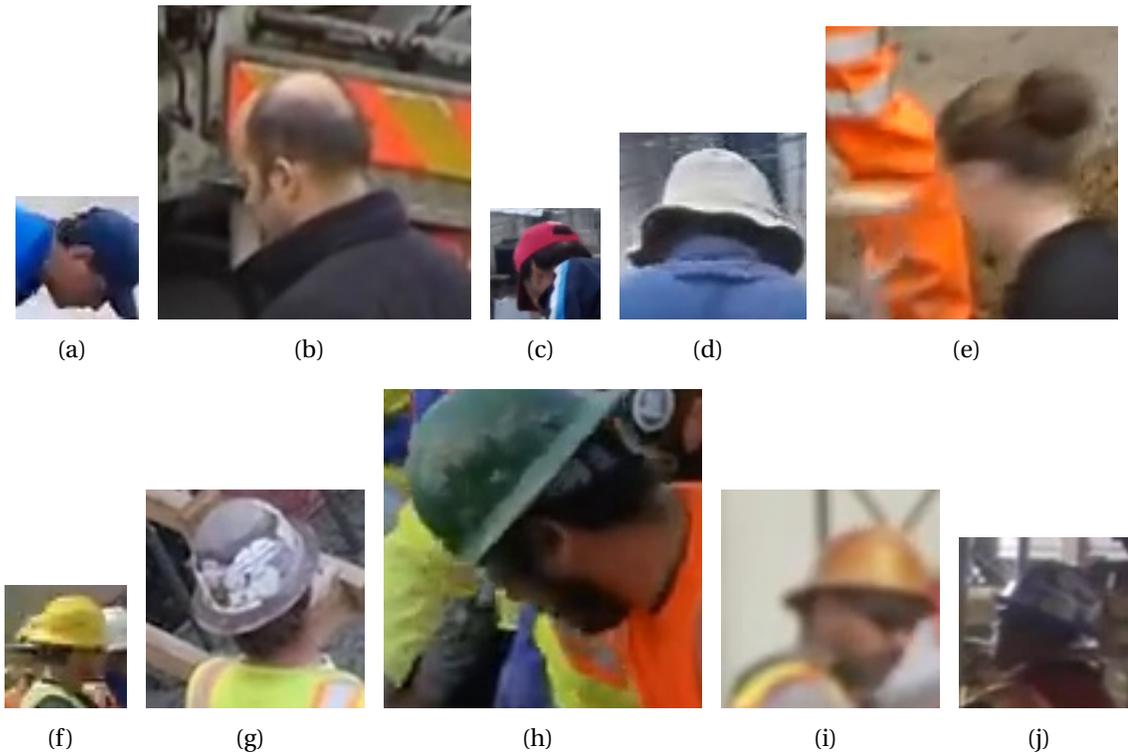


Figura 3.23: *Dataset E.R^{ca}* con etiquetas de “Casco” (a)-(e) No y (f)-(j) Sí.

El *dataset* construido para este trabajo consta de ROI extraídos de más de 30 videos disponibles en Internet como se indica en la Sec. 3.3.4. Dichos videos reúnen diversas actividades laborales reales de los trabajadores de la industria de la construcción, con diferentes condiciones de iluminación y perspectiva de la escena. Para acelerar el proceso de etiquetado, se han seleccionado videos de secuencia continua en los que todos los trabajadores que aparecen en la escena pueden ser clasificados directamente con las mismas etiquetas de EPP objetivo en todo momento.

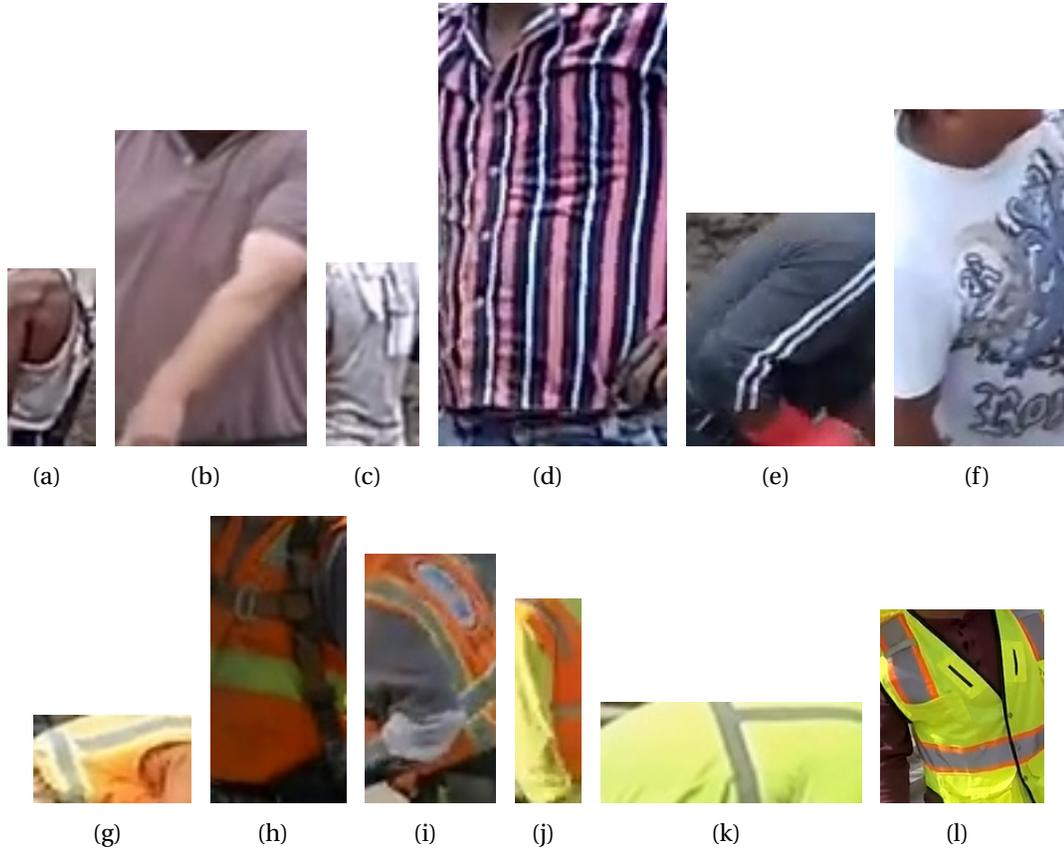


Figura 3.24: *Dataset E.R^{Tr}* con etiquetas de “Chaleco” (a)-(f) No y (g)-(l) Sí.

Posteriormente, al seleccionar los fotogramas, se maximiza la diversidad de las muestras. Por lo tanto, para las clases de notación positiva *Casco* y *Chaleco*, se han incluido varios colores y patrones. La clase negativa de *Casco* incluyó a trabajadores sin cabello, con diferentes cortes y colores de cabello, con y sin gorra o sombrero, y en general elementos visuales que no sean un casco aprobado (ver Figura 3.23). Finalmente, la clase negativa de *Chaleco* incluyó torsos desnudos o camisetas y camisas de diferentes patrones y colores, así como la variabilidad de género (ver Figura 3.24). Brevemente, el *dataset* incluye un total de 23.303 subimágenes de la cabeza y 21.378 subimágenes del torso, obtenidas de esqueletos asociados con trabajadores de la industria de la construcción.

Para el entrenamiento del clasificador $RN18^{ca}$, se realizó una partición aleatoria con 15.367 subimágenes E.Rca para entrenamiento y 7.666 para validación, mientras que para el clasificador $RN18^{Tr}$ se eligieron 14.264 subimágenes E.Rtr para el entrenamiento y 7.114 para validación. Finalmente, los dos conjuntos de re-aprendizaje se dividieron en 80%-20% para entrenamiento y prueba, respectivamente, proporcionando los resultados recopilados en las matrices de confusión de las tablas 3.1 y 3.2, que contiene las definiciones habituales de los parámetros de rendimiento de clasificación multiclase: sen-

sibilidad = $vp/(vp + fn)$, precisión = $vp/(vp + fp)$, especificidad = $vn/(vn + fp)$, valor predictivo negativo (VAN) = $vn/(vn + fn)$ y exactitud = $(vp + vn)/(vp + vn + fp + fn)$.

Tabla 3.1: Matrices de confusión para la validación de ResNet18^{ca} *Casco de seguridad*

		Real		Total
		v	f	
Predicho	p	$vp = 4.546$	$fp = 12$	4.558 Precisión = 0.997
	n	$fn = 13$	$vn = 3.095$	3.108 NPV = 0.995
Total		4.559 Sensibilidad = 0.997	3.107 Especificidad = 0.996	7.666 Exactitud = 0.996

Tabla 3.2: Matrices de confusión para la validación de ResNet18^{tr} *Chaleco de alta visibilidad*.

		Real		Total
		v	f	
Predicho	p	$vp = 2.754$	$fp = 22$	2.776 Precisión = 0.992
	n	$fn = 18$	$vn = 4.320$	4.338 NPV = 0.996
Total		2.772 Sensibilidad = 0.994	4.342 Especificidad = 0.995	7.114 Exactitud = 0.994

El entrenamiento de las redes RN18 se realizó en la plataforma de computación en la nube de Google Colab utilizando el paquete Fast.ai, mientras que OP18 se instala localmente con Pytorch en una PC con procesador Intel Core i7 con 16 GB de RAM y una tarjeta Gráfica NVIDIA GeForce GTX 1080 TI. El software de evaluación fue diseñado para ejecución local en GPU en lenguaje Python en Ubuntu, logrando una velocidad de procesamiento promedio de 3.57 *fps*.

3.3.6. Resultados experimentales

Para la validación de la metodología propuesta se plantearon tres videos de actividades laborales usuales del sector de la construcción. Dichos videos fueron escogidos porque presentan diferentes condiciones y dificultades técnicas: (1) cámara estabilizada en un dron volando a baja altura al mediodía (Figura 3.25), (2) cámara deportiva sujeta al casco de un trabajador al atardecer (Figura 3.26) y (3) *smartphone* sostenido manualmente con buena iluminación (Figura 3.27).

El primer video (Figura 3.25¹) contiene un grupo de 10 trabajadores en fundición de concreto. Las imágenes de las Figuras 3.25(a), 3.25(b) y 3.25(f) muestran oclusiones importantes debido principalmente a la perspectiva de captura y la presencia de barras de hierro verticales. La metodología permite rechazar los esqueletos de trabajadores con oclusión parcial a través de los umbrales de confiabilidad. Un umbral típico de $u_{Conf}^E = 0.35$ permite descartar esqueletos de trabajadores con aproximadamente la mitad del cuerpo ocluido, y por lo tanto, se reduce la cantidad posible de errores posibles. Asimismo, el efecto de la distancia u oclusiones parciales en un objetivo puede controlarse por su umbral de tamaño ($u_{Tam}^E = 0.2$).

La red OP18 está diseñada para inferir puntos antropométricos ocluidos a partir de otros puntos visibles, sin embargo, los puntos inferidos presentan valores bajos de confiabilidad. Esta capacidad de inferencia permite a la metodología obtener ROI adecuadas para realizar la clasificación del uso de EPP con oclusiones parciales. Por ejemplo, un umbral de $u_{Conf}^{ca} = 0.3$ permite rechazar el análisis del casco del trabajador E1 en la figura 3.25(b), mientras que la inferencia se hace con el mismo u_{Conf}^{ca} en su oreja derecha y hombro en la Figura 3.25(g), lo cual proporciona un *E.Rca* apropiado para clasificar correctamente el uso de casco. Sin embargo, este umbral excluye del análisis a los trabajadores E2, E3 y E5 en esta figura.

La importancia de rechazar las detecciones con el umbral de tamaño se aclara en la Figura 3.25(h), donde se descartan dos detecciones falsas (D1 y D2). Sin embargo, el método falla con el trabajador E7 en situaciones como la que se muestra en la Figura 3.25(c). Adicionalmente, se presenta otra falla con el mismo trabajador E7 en la Figura 3.25(e), porque el casco es poco visible y presenta calcomanías que dificultan su detección y clasificación.

En resumen, en la Figura 3.25(a) E2, E3, E7 y E10 se detectan correctamente en la Figura 3.25(b) E1 y E2, en la Figura 3.25(c) se detecta correctamente los chalecos de E7 y E1, y en la detección incorrecta el casco de E7, en la Figura 3.25(d) se detectan correctamente E1 y E2. En la Figura 3.25(e), se detectan correctamente los chalecos de E5 y E7, en los cascos se etiqueta positivamente en E5 y negativamente en E7. En las últimas tres figuras de cada video, se ejemplifican los esqueletos utilizados para la primera instancia de detección, por lo cual se aceptan como buenas detecciones en la Figura 3.25(f) E1 y

¹Ver <https://youtu.be/yuNDOfnuUhU>



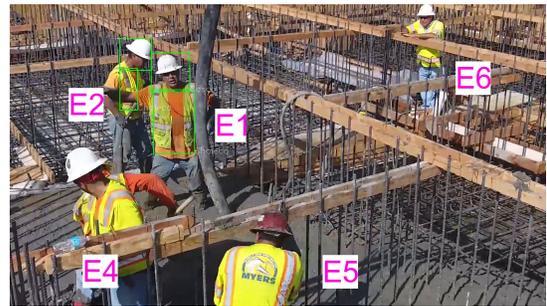
(a)



(b)



(c)



(d)



(e)



(f)

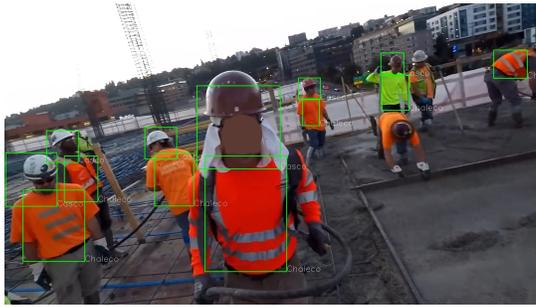


(g)

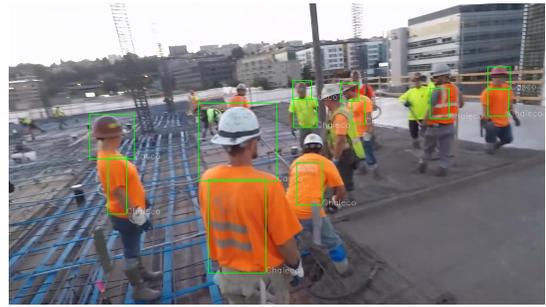


(h)

Figura 3.25: Resultados del video de evaluación 1 *Dron*.



(a)



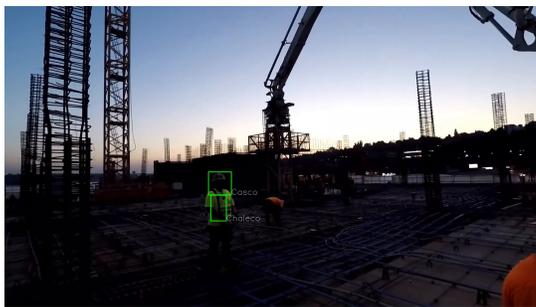
(b)



(c)



(d)



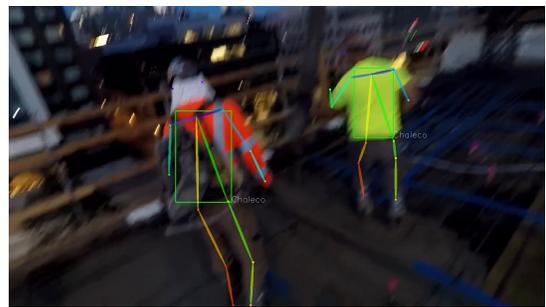
(e)



(f)



(g)



(h)

Figura 3.26: Resultados del video de evaluación 2 *Atardecer*.



(a)



(b)



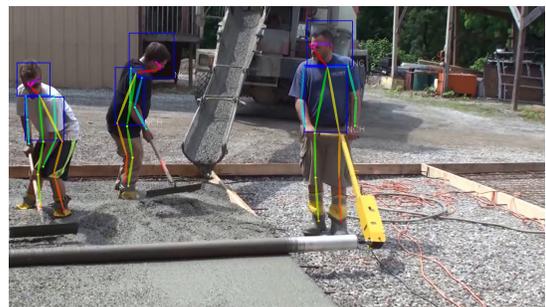
(c)



(d)



(e)



(f)



(g)



(h)

Figura 3.27: Resultados del video de evaluación 3 (*smartphone*).

E5, en la Figura 3.25(g) E1 y finalmente en la Figura 3.25(h) E1, E2, E4 y E5, así como las falsas detecciones D1 y D2, se ignoran.

Las condiciones de adquisición del segundo video generan muchas imágenes borrosas debido al movimiento de la cámara (Figuras 3.26(g) y 3.26(d)). Además, este video presenta ciertas dificultades mencionadas en el estado del arte (Konstantinou y Brilakis, 2018), como ser cambios en la iluminación (Figuras 3.26(h) y 3.26(e)), oclusiones inherentes al tipo de trabajo (Figuras 3.26(b) y 3.26(f)) y la diversidad de posturas y tamaños (Figura 3.26(e)). Sin embargo, la metodología propuesta puede recopilar información útil al trabajar solo con los esqueletos y ROI más confiables.

En las Figuras 3.26(f) y 3.26(g) se ilustran las dos dificultades más comunes encontradas en la metodología propuesta para la clasificación del casco. En este caso, se relaciona con el uso anormal de ropa adicional en la cabeza, sumado a la apariencia borrosa antes mencionada generada por el movimiento y la esperada distancia de la cámara. Sin embargo, en las Figuras 3.26(a) el mismo casco del trabajador en primer plano está correctamente clasificado, por lo cual probablemente este error pueda ser solucionado utilizando algún tipo de re-identificador.

Finalmente, la Figura 3.27 muestra una secuencia de trabajo en grupo en la que ningún trabajador está usando EPP. El método propuesto logra discernir correctamente la no utilización de EPP entre trabajadores con posturas variables, ocluidas y con diferente ropa de trabajo. En los pocos casos en los que una clasificación es errónea, como en la Figura 3.27(a), se originan por la deformación de ROIs rectangulares, que deben ser transformadas a ROIs cuadradas antes de ser ingresadas en RN18. En una vasta proporción de los casos, la metodología es robusta, proporcionando resultados fiables y de alta precisión.

Capítulo 4

Discusión, conclusiones y trabajo futuro

4.1. Aportes de esta tesis

Esta investigación comenzó con los siguientes objetivos generales de investigación:

1. Proporcionar un método económico de captura de la postura del cuerpo humano, no invasivo y preciso, que calcule los factores de riesgo para la evaluación de riesgos ergonómicos.
2. Ofrecer un método basado en CV para determinar la utilización de equipo de protección personal, enfocándose primero en la detección del trabajador y luego en la posterior detección de presencia o ausencia del EPP recomendado.

Para el primer objetivo se propuso la detección de las articulaciones de los trabajadores mediante fotos o videos, utilizando la red neuronal OP25 y el método para la evaluación ergonómica RULA. Para probar la validez del método propuesto se evaluó la confianza en la detección de las articulaciones y de los ángulos ergonómicos en función del punto de vista del observador. Además, simulamos avatares virtuales en 3D, lo que proporcionó una forma adecuada y sistemática de generar un conjunto de datos conocidos que de otra manera serían difíciles y costosos de obtener (Li y cols., 2018; Plantard y cols., 2015). Se eligieron latitudes inferiores a 45° como los puntos de vista más adecuados para mantener las auto-oclusiones corporales al mínimo, mientras que la selección del azimuth se recomienda dependiendo del interés en la evaluación de articulaciones particulares.

Los cálculos de ángulos de hombros, cuello y tronco también se probaron a fondo con conjuntos de datos construidos a partir de modelos 3D simulados y experimentos de laboratorio con personas reales. Estos ángulos de los hombros, el cuello y el tronco son de suma importancia en RULA, pues poseen la mayor influencia en la medición del riesgo ergonómico en RULA. Se propuso el uso de dos puntos fidúceos de referencia (ME

y MK) para facilitar mediciones de los ángulos del cuello y el tronco bajo las frecuentes condiciones de auto-oclusión. Los resultados experimentales también permitieron estimar los rangos y umbrales angulares de las articulaciones del cuerpo adecuados, según sea necesario para calcular las puntuaciones de RULA como es mencionado en la Tabla 2.3.

Por último, se llevaron a cabo evaluaciones ergonómicas basadas en video en escenarios de trabajo reales al aire libre y bajo la supervisión de ergonomistas experimentados. Nuestro método proporcionó puntajes RULA robustos simultáneos para más de un trabajador. Cada video evaluado presentaba diferentes condiciones de iluminación, oclusiones y puntos de vista con cámara inestable. Los resultados obtenidos fueron altamente coincidentes con las evaluaciones de los expertos (ver Tabla 2.6). Estos resultados muestran que las variaciones razonables en la vista de la cámara no influyen significativamente en los resultados en condiciones de trabajo reales, según lo predicho por las pruebas experimentales implementadas en sección 2.3.6.

Las puntuaciones de la parte superior de los brazos tuvieron la mejor concordancia entre los especialistas y nuestra propuesta. Este puntaje es el más significativo en la evaluación RULA y, por lo tanto, este acuerdo brinda un respaldo significativo a nuestro sistema. Por otro lado, las puntuaciones del tronco tuvieron el menor acuerdo (aunque un coeficiente Kappa de 0,44 se considera un acuerdo moderado en la literatura (Landis y Koch, 1977)). Sin embargo, este coeficiente es similar al acuerdo entre los propios expertos y, por lo tanto, puede atribuirse al menos parcialmente a la variación inter-subjetiva en lugar de a la variación en nuestro sistema. Finalmente, hay un acuerdo total en el GS. Estos resultados demuestran la utilidad de la CV en general, y del método propuesto en particular, para realizar una ERE a partir de fotografías tomadas al aire libre o en condiciones de observación subóptimas.

En el capítulo 3 se han propuesto dos soluciones para el segundo objetivo, dichas soluciones son basadas en visión artificial para la medición cuantitativa de la utilización de equipo de protección personal. La primera solución se basa en la red neuronal convolucional Yolo, y cuyo objetivo es la detección de cascos, guantes, ropa de alta visibilidad y los trabajadores. La segunda solución parte de lo desarrollado en la sección 2.3.2, pues utiliza OP para segmentar la cabeza y el pecho de un trabajador, por último, dichas áreas segmentadas fueron enviadas a un clasificador para detectar presencia o ausencia de casco y chaleco de alta visibilidad. La evidencia presentada en el capítulo 3 se centra en métodos de captura modernos como drones, cámaras deportivas, videos egocéntricos, o *smartphones*, que tienden a ser dispositivos de captura muy inestables. Adicionalmente, al apoyarse en el descubrimiento de esqueletos, la metodología aborda el problema de manera *injointangle/novadora*, para lo cual no existe evidencia previa en el estado del arte. Por tanto, estas metodologías se presentan como alternativas de solución viables y prácticas, aplicables en condiciones reales de trabajo y aptas para un seguimiento continuo, sin supervisión y en tiempo real.

En la primera solución propuesta para la detección de equipo de protección personal, se llevó a cabo un análisis de caso complejo y los resultados revelaron que el detec-

tor Yolo entrenado era robusto bajo varios escenarios y condiciones. Además, se comprobó que es relativamente eficiente en el tiempo. El proceso podría mejorarse agregando un supresor de no máximos para así reducir los etiquetados múltiples y ampliar el abanico evaluaciones para poder corroborar las buenas mediciones obtenidas. La segunda solución propuesta se basa en la detección de las personas con la red neuronal OP18, para segmentar el área de la cabeza y del torso, cada área de interés segmentada es enviada a una red de clasificación Resnet18 entrenada. El método propuesto permite una correcta distinción en la no utilización de equipos de protección personal en trabajadores con posiciones variables y cerradas y diferente vestimenta de trabajo.

En todos estos casos la metodología es robusta, proporcionando resultados fiables con una precisión de más del 99%, lo cual va en congruencia con otros métodos del estado del arte. Sin embargo, las demás metodologías comparadas no son evaluadas en video generado por cámaras deportivas o drones (ver tabla 4.1). En comparación con la técnica propuesta, las tecnologías basadas en color (DCT, LAB, HOG y SVM) presentaron una gran exactitud en interiores. Sin embargo, como fue mencionado en el estado del arte estas tecnologías presentan deficiencias en exteriores y en general cuando se presentan variaciones en la iluminación, el color y la pose (Sección 1.3). La Tecnología YOLO3 solo fue testada en ambientes controlados en el laboratorio (Hung, Lan, y Hong, 2019).

Tabla 4.1: Resumen comparativo de las técnicas de detección de EPP

Tecnología	Exactitud	Casco	Chaleco
YOLO 2 (Massiris Fernández y cols., 2018)	62 %	✓	✓
HOG + SVM (Park y cols., 2015)	81 %	✓	X
DCT + SVM (Rubaiyat y cols., 2017)	91 %	✓	X
Fast R-CNN (Fang y cols., 2018)	95 %	✓	X
LAB + SVM (Seong y cols., 2017)	95 %	X	✓
HOG + SVM (Park y Brilakis, 2016)	98 %	X	✓
YOLO 3 (Hung y cols., 2019)	99 %	✓	✓
OP18 + RN18 (Massiris Fernández, Bajo, y cols., 2020)	99 %	✓	✓

4.2. Limitaciones de los resultados

Dento de la solución propuesta para la estimación de RULA utilizando CV, cabe señalar dos puntos a mejorar. En algunos casos, los sesgos de detección del esqueleto generados por oclusiones o auto-clusiones pueden dar lugar a desviaciones importantes de la medición de ángulos. Por esta razón, nuestro método incorpora umbrales de confianza y de tamaño basados en el esqueleto para descartar detecciones potencialmente defectuosas. Adicionalmente, los datos de confianza de las articulaciones del cuerpo estimadas de video pueden usarse para construir estadísticas sobre las estimaciones dadas, además, dicho análisis estadístico podría tener en cuenta las tendencias centrales

y también los valores extremos. A su vez, estas estadísticas podrían habilitar métodos para obtener medidas angulares más precisas, que serán trabajadas en el futuro. En este sentido, el acopio de datos permitirá evaluar la existencia de sesgos sistemáticos y eventualmente se podrán proponer métodos para corregirlos. De todas maneras es previsible que en el futuro continúe la tendencia a aumentar la resolución y otras prestaciones de los sensores, por lo que es esperable tener capacidades robustas muy superiores a las que se tiene con el ojo desnudo.

La segunda limitación se debe al hecho de que las medidas angulares no se calculan a partir de estimaciones de articulaciones corporales en 3D, sino a partir de proyecciones en 2D. Esto claramente puede generar distorsiones proyectivas. Como se mostró, el fenómeno se puede controlar eligiendo la mayor distancia focal posible y con vistas de cámara con azimuth y elevación adecuadas en las que no se ocluyan los ángulos requeridos de la articulación del cuerpo. Nuestro método desaconseja el uso de ángulos de visión verticales mayores a 45° o vistas de cámara en las que los trabajadores se toman lateralmente. Estos dos problemas aún pueden conducir a errores y a veces sesgos sistemáticos, que pueden debilitar la importancia de una estimación de RULA a partir de una sola instantánea, dado que si bien la posición de la cámara puede controlarse hasta cierto punto, las posiciones de los trabajadores no. Sin embargo, cuando se analizan muchos fotogramas de forma conjunta, y cuando la postura de trabajo que se va a medir no es estática según se ve desde la cámara, la información acumulada de estos fotogramas se puede utilizar para reforzar tanto la importancia como la precisión de la medición. Esto requeriría desarrollar métodos de identificación y regresión a lo largo de varios cuadros conjuntos, y un método heurístico para computar los valores a partir de las estimaciones individuales.

Vale la pena notar que las pequeñas discrepancias en los puntajes locales de RULA suelen tener poca importancia con respecto al puntaje general. Esto se debe al diseño jerárquico de RULA, que utiliza la agregación de puntuaciones ponderadas como método para reducir el impacto del ruido en sus diferentes etapas (Plantard y cols., 2017). Esta característica es sin duda favorable para nuestro método, ya que los errores de localización y proyección de la articulación corporal se asimilan como ruido, pero no descartamos la posibilidad de que nuestro método se aplique a otros modelos de estimación ergonómica, en los cuales esta situación ventajosa puede no ocurrir.

También se ha afirmado que los métodos de ERE ergonómicos basados en CV no proporcionan absolutamente todos los datos requeridos para calcular las puntuaciones de RULA (Mura y cols., 2012; Plantard y cols., 2017), pues aún se depende de algunas mediciones realizadas por los expertos. Sin embargo, otros métodos de adquisición como sensores portátiles (*por ejemplo*, acelerómetros e inclinómetros) apenas miden algunas articulaciones (Nath y cols., 2017). Por lo general, requieren dispositivos de uso específicos, haciéndolos intrusivos, influyendo y en ocasiones restringiendo la evaluación real, ya que los trabajadores son conscientes de su uso, lo que puede hacer que ejecuten sus tareas de una forma sesgada (Yu y cols., 2019; Zhang y cols., 2018). Además, las señales adquiridas de estos dispositivos son mucho más difíciles de manejar que una imagen digital, debido al ruido característico y las perturbaciones inherentes a la tecnología de

detección. Dado que estos sensores solo producen datos de posición a lo largo del tiempo, los cálculos del ángulo de articulación dependen en gran medida del procesamiento de la señal y los problemas de sincronización, que están lejos de ser propensos a errores.

Por último, otras tecnologías basadas en CV, como las cámaras RGB-D, no pueden funcionar bajo la luz solar directa y tienen un rango operativo más restringido (Yu y cols., 2019). En general, a nuestro leal saber y entender, esto hace que nuestra contribución sea pionera en lograr una evaluación de puntaje RULA confiable y robusta, basada en CV y completamente automatizada, pudiendo realizar en condiciones realistas (*ie*, iluminación desigual, oclusiones y oclusiones, posición variable de la cámara y varios trabajadores).

4.3. Trabajo futuro

Se espera que la evolución de las tecnologías de aprendizaje profundo proporcione bibliotecas de CV de código abierto aún más precisas y flexibles, lo que puede mejorar las características de nuestra propuesta en un futuro próximo y mitigar los posibles inconvenientes mencionados anteriormente. Esto puede habilitar, entre otras cosas, una detección especializada de puntos de la mano y el rostro que puede fortalecer nuestro método automatizado para ERE relacionadas con la muñeca y la cabeza, lo cual podría reducir la dependencia de las mediciones realizadas por los expertos.

Como se indica en Xu y McGorry (2015), también serán necesarias más mejoras para cumplir con las tareas relacionadas con ERE más desafiantes. Por ejemplo, todavía se desconoce si el género, los somatotipos o la etnia ejercen una influencia significativa en la precisión de la detección conjunta. Por esta razón, es posible que se requiera una evaluación previa de características antropométricas para diseñar un modelo que pueda adaptarse a estas y otras condiciones particulares. Como en tantos otros escenarios, el acopio de datos de calidad es un factor determinante, aspecto al cual nos estamos dedicando. Como se mencionó arriba, además de RULA, hay otras propuestas de evaluación rápida que parecen valiosas de incorporar, como la Evaluación Rápida de Todo el Cuerpo (REBA (McAtamney y Hignett, 2000)) y la RULA modificada para trabajadores informáticos (MRULA (Levanon y cols., 2014)), ya que ambos métodos se basan en los principios de RULA.

Finalmente, en el contexto de la I4.0, la adquisición de la postura corporal en tiempo real es valiosa para realizar la fusión de datos con otras fuentes de información, específicamente en las industrias de fabricación discreta y de procesos, para extraer información latente sobre las condiciones de trabajo y la optimización logística. En este entorno, nuestro sistema es totalmente adaptable e integrable, ya que proporciona información en los formatos de intercambio de datos más utilizados, como los archivos Json. Además, estas tecnologías basadas en CV se pueden utilizar para clasificar las actividades humanas además de evaluar la seguridad de las posturas y los movimientos. Estas evaluaciones combinadas abrirán nuevos espacios ergonómicos para el análisis y

la interpretación, por ejemplo, por actividad y posiblemente por hora y lugar. Al mismo tiempo, la detección de objetos basada en CV puede articularse junto con análisis ergonómicos, para supervisar y calificar automáticamente el uso adecuado de los equipos de protección personal.

Los estudios futuros también tendrán como objetivo integrar este proceso de detección de EPP en un software completo de inspección de seguridad industrial que pueda ser utilizado por el usuario final. También se requiere explorar pruebas adicionales para evaluar la precisión de otros algoritmos de detección de objetos, por lo cual se debe ampliar el dataset tanto en entrenamiento como en nuevas pruebas. Se trabajará más para mejorar la precisión del proceso de detección y eliminar las detecciones falsas entrenando y evaluando el dataset en la nueva versión de Yolo, la cual reduce un poco la velocidad de ejecución de la red neuronal pero aumenta su tasa de aciertos. Como extensión del método basado en OP18 y RN18, el objetivo futuro es hacer escalable el sistema e incluir otros equipos de protección personal como guantes y calzado de seguridad, así como diseñar o evaluar otras redes para la detección y clasificación esquelética de los sujetos de interés.

Referencias

- AlexeyAB. (2018). *Yolo_mark*. Descargado de https://github.com/AlexeyAB/Yolo_{_}mark
- Ankrum, D. R., y Nemeth, K. J. (2000). Head and neck posture at computer workstations – what is neutral? En *Proceedings of the 14th triennial congress of the international ergonomics association* (Vol. 5, pp. 565–568). doi: 10.5100/jje.33.261
- Assuncao, A. Á., y Silva Abreu, M. N. (2017). Factor associated with self-reported work-related musculoskeletal disorders in Brazilian adults. *Revista de Saude Publica*, 51, 1S–12S. doi: 10.1590/S1518-8787.2017051000282
- Battini, D., Persona, A., y Sgarbossa, F. (2014). Innovative real-time system to integrate ergonomic evaluations into warehouse design and management. *Computers and Industrial Engineering*. doi: 10.1016/j.cie.2014.08.018
- Ben-Arie, J. (1990). The Probabilistic Peaking Effect of Viewed Angles and Distances with Application to 3-D Object Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(8), 760–774. doi: 10.1109/34.57667
- Benavides, F. G., López-Ruiz, M., Rojas, M., Silva-Peñaherrera, M., Chavarria, J., y Cornelio, C. (2018). *Brechas de desigualdad en la salud de la población trabajadora en América Latina y el Caribe* (Inf. Téc. n.º 35). Heredia, Costa Rica: SALTRA / IRET-UNA.
- Bevan, S. (2015). Economic impact of musculoskeletal disorders (MSDs) on work in Europe. *Best Practice and Research: Clinical Rheumatology*, 29(3), 356–373. Descargado de <http://dx.doi.org/10.1016/j.berh.2015.08.002> doi: 10.1016/j.berh.2015.08.002
- Boje, C., Guerriero, A., Kubicki, S., y Rezgui, Y. (2020). Towards a semantic Construction Digital Twin: Directions for future research. *Automation in Construction*, 114(November 2019), 103179. Descargado de <https://doi.org/10.1016/j.autcon.2020.103179> doi: 10.1016/j.autcon.2020.103179
- Cao, Z., Simon, T., Wei, S.-E., y Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. En *Proceedings of the iee conference on computer vision and pattern recognition* (pp. 7291–7299).
- Chi, S., y Caldas, C. H. (2011). Automated Object Identification Using Optical Video Cameras on Construction Sites. *Computer-Aided Civil and Infrastructure Engineering*, 26(5), 368–380. doi: 10.1111/j.1467-8667.2010.00690.x
- Chiasson, M. è., Imbeau, D., Aubry, K., y Delisle, A. (2012). Comparing the results of eight methods used to evaluate risk factors associated with musculoskeletal disorders. *International Journal of Industrial Ergonomics*. doi: 10.1016/j.ergon.2012.07.003

- Cohen, A. L., Gjessing, C. C., y Fine, L. J. (1997). A Primer Based on Workplace Evaluations of Musculoskeletal Disorders. *US Department of Health and Human Services, National Institute for Occupational Safety and Health, DHHS (NIOSH) Publication(97-117)*, 16–30.
- Cui, J., y Freeden, W. (1997). Equidistribution on the Sphere. *SIAM Journal on Scientific Computing*, 18(2), 595–609.
- David, G. C. (2005). *Ergonomic methods for assessing exposure to risk factors for work-related musculoskeletal disorders*. doi: 10.1093/occmed/kqi082
- Diego-Mas, J. A., Alcaide-Marzal, J., y Poveda-Bautista, R. (2017). Errors Using Observational Methods for Ergonomics Assessment in Real Practice. *Human Factors*, 59(8), 1173–1187. doi: 10.1177/0018720817723496
- Diego-Mas, J. A., Poveda-Bautista, R., y Garzon-Leal, D. C. (2015). Influences on the use of observational methods by practitioners when identifying risk factors in physical work. *Ergonomics*, 58(10), 1660–1670. doi: 10.1080/00140139.2015.1023851
- Everingham, M., Van Gool, L., Williams, C., Winn, J., y Zisserman, A. (2010). The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2), 303–338.
- Experiencor. (s.f.). *YOLOv2 in Keras and Applications*. Descargado de <https://github.com/experiencor/keras-yolo2>
- Fang, Q., Li, H., Luo, X., Ding, L., Luo, H., Rose, T. M., y An, W. (2018, jan). Detecting non-hardhat-use by a deep learning method from far-field surveillance videos. *Automation in Construction*, 85, 1–9. doi: 10.1016/j.autcon.2017.09.018
- Figlali, N., Cihan, A., Esen, H., Figlali, A., Çeşmeci, D., Güllü, M. K., y Yilmaz, M. K. (2015). Image processing-aided working posture analysis: I-OWAS. *Computers and Industrial Engineering*. doi: 10.1016/j.cie.2015.03.011
- Fukushima, K. (1988). Neocognitron: A hierarchical neural network capable of visual pattern recognition. *Neural networks*, 1(2), 119–130.
- Gadea, M., Cornelio, C., Itatí Iñiguez, M. J., Marino, J., Contartese, D., Mazonra, X., y Schachtel, L. (2019). *Encuesta Nacional Argentina a trabajadores sobre Condiciones de Empleo, Trabajo, Salud y Seguridad (ECETSS2018)*. Buenos Aires. Descargado de <http://www.trabajo.gov.ar/downloads/estadisticas/ecetss/ecetss{ }informe.pdf>
- Gerr, F., Fethke, N. B., Merlino, L., Anton, D., Rosecrance, J., Jones, M. P., ... Meyers, A. R. (2014, feb). A prospective study of musculoskeletal outcomes among manufacturing workers: I. effects of physical risk factors. *Human Factors*, 56(1), 112–130. doi: 10.1177/0018720813491114
- Gobierno de Argentina. (1972). *Ley 19.587 de Higiene y Seguridad en el Trabajo* (Vol. 79). Descargado de <http://servicios.infoleg.gob.ar/infolegInternet/anexos/15000-19999/17612/norma.htm>
- Gobierno de Argentina. (2011). *Resolución 299/2011 Adóptanse las reglamentaciones que procuren la provisión de elementos de protección personal confiables a los trabajadores*. Superintendencia de Riesgos del Trabajo. Descargado de <http://servicios.infoleg.gob.ar/infolegInternet/anexos/180000-184999/180669/norma.htm>
- Golabchi, A., Han, S., Seo, J., Han, S., Lee, S., y Al-Hussein, M. (2015). An Automated Bio-

- mechanical Simulation Approach to Ergonomic Job Analysis for Workplace Design. *Journal of Construction Engineering and Management*, 141(8), 1–12. doi: 10.1061/(ASCE)CO.1943-7862.0000998
- Guo, H., Yu, Y., y Skitmore, M. (2017). Visualization technology-based construction safety management: A review. *Automation in Construction*, 73, 135–144. Descargado de <http://dx.doi.org/10.1016/j.autcon.2016.10.004> doi: 10.1016/j.autcon.2016.10.004
- Gupta, A., Mittal, A., y Davis, L. S. (2006). Constraint integration for multiview pose estimation of humans with self-occlusions. En *Third international symposium on 3d data processing, visualization, and transmission, 3dpvt 2006* (Vol. 30, pp. 900–907). doi: 10.1109/3DPVT.2006.45
- Ha, C., Roquelaure, Y., Leclerc, A., Touranchet, A., Goldberg, M., y Imbernon, E. (2009). The French musculoskeletal disorders surveillance program: Pays de la Loire network. *Occupational and Environmental Medicine*, 66(7), 471–479. doi: 10.1136/oem.2008.042812
- He, K., Zhang, X., Ren, S., y Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016-Decem*, 770–778. doi: 10.1109/CVPR.2016.90
- Hegde, J., y Rokseth, B. (2020). Applications of machine learning methods for engineering risk assessment – A review. *Safety Science*, 122(September 2019), 104492. Descargado de <https://doi.org/10.1016/j.ssci.2019.09.015> doi: 10.1016/j.ssci.2019.09.015
- Hung, H. M., Lan, L. T., y Hong, H. S. (2019). A DEEP LEARNING-BASED METHOD FOR REAL-TIME PERSONAL PROTECTIVE EQUIPMENT DETECTION. *Journal of Science and Technique- Le Quy Don Technical University*, 199(199), 23–34.
- Igelmo, V., Syberfeldt, A., Högberg, D., García Rivera, F., y Pérez Luque, E. (2020). Aiding observational ergonomic evaluation methods using MOCAP systems supported by AI-based posture recognition. *Advances in Transdisciplinary Engineering*, 11, 419–429. doi: 10.3233/ATDE200050
- ILO-OSH. (2001). *ILO-OSH 2001. Directrices relativas a los sistemas de gestión de la seguridad y la salud en el trabajo*. Ginebra (Suiza): Oficina Internacional del Trabajo (OIT) 4,.
- Joshi, M., y Deshpande, V. (2019). A systematic review of comparative studies on ergonomic assessment techniques. *International Journal of Industrial Ergonomics*, 74(October), 102865. Descargado de <https://doi.org/10.1016/j.ergon.2019.102865> doi: 10.1016/j.ergon.2019.102865
- Kadir, B. A., Broberg, O., y Souza da Conceição, C. (2019). Current research and future perspectives on human factors and ergonomics in Industry 4.0. *Computers & Industrial Engineering*, 137, 106004. Descargado de <https://doi.org/10.1016/j.cie.2019.106004> doi: 10.1016/j.cie.2019.106004
- Kelm, A., Laußat, L., Meins-Becker, A., Platz, D., Khazaee, M. J., Costin, A. M., ... Teizer, J. (2013, dec). Mobile passive Radio Frequency Identification (RFID) portal for automated and rapid control of Personal Protective Equipment (PPE) on construction sites. *Automation in Construction*, 36, 38–52. Descargado de <https://www.sciencedirect.com/science/article/pii/S0926580513001234> doi: 10

.1016/J.AUTCON.2013.08.009

- Kocabas, M., Athanasiou, N., y Black, M. J. (2020). VIBE: Video Inference for Human Body Pose and Shape Estimation. En *Ieee conference on computer vision and pattern recognition (cvpr2020)* (pp. 5253–5263). Descargado de <http://arxiv.org/abs/1912.05656>
- Konstantinou, E., y Brilakis, I. (2018). Matching Construction Workers across Views for Automated 3D Vision Tracking On-Site. *Journal of Construction Engineering and Management*, 144(7), 1–12. doi: 10.1061/(ASCE)CO.1943-7862.0001508
- Konstantinou, E., Lasenby, J., y Brilakis, I. (2019, jul). Adaptive computer vision-based 2D tracking of workers in complex environments. *Automation in Construction*, 103(june), 168–184. doi: 10.1016/j.autcon.2019.01.018
- Krüger, J., y Nguyen, T. D. (2015). Automated vision-based live ergonomics analysis in assembly operations. *CIRP Annals - Manufacturing Technology*, 64(1), 9–12. doi: 10.1016/j.cirp.2015.04.046
- Landis, J. R., y Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159. doi: 10.2307/2529310
- LeCun, Y., Bottou, L., Orr, G. B., y Müller, K.-R. (1998). Efficient backprop. En *Neural networks: Tricks of the trade* (pp. 9–50). Springer.
- Lee, W., Seto, E., Lin, K. Y., y Migliaccio, G. C. (2017, nov). An evaluation of wearable sensors and their placements for analyzing construction worker's trunk posture in laboratory conditions. *Applied Ergonomics*, 65, 424–436. doi: 10.1016/j.apergo.2017.03.016
- Levanon, Y., Lerman, Y., Gefen, A., y Ratzon, N. Z. (2014). Validity of the modified RULA for computer workers and reliability of one observation compared to six. *Ergonomics*, 57(12), 1856–1863. doi: 10.1080/00140139.2014.952350
- Li, X., Han, S., Gül, M., Al-Hussein, M., y El-Rich, M. (2018). 3D Visualization-Based Ergonomic Risk Assessment and Work Modification Framework and Its Validation for a Lifting Task. *Journal of Construction Engineering and Management*, 144(1), 1–13. doi: 10.1061/(ASCE)CO.1943-7862.0001412
- Lim, C. M., Jung, M. C., y Kong, Y. K. (2011). Evaluation of upper-limb body postures based on the effects of back and shoulder flexion angles on subjective discomfort ratings, heart rates and muscle activities. *Ergonomics*, 54(9), 849–857. doi: 10.1080/00140139.2011.600777
- Liu, M. (2019). *Video-Based Human Motion Capture and Force Estimation for Comprehensive On-Site Ergonomic Risk Assessment* (Tesis Doctoral, University of Michigan). doi: .1037//0033-2909.I26.1.78
- Lowe, B. D., Weir, P., y Andrews, D. (2014). *Observation-based posture assessment: review of current practice and recommendations for improvement* (Inf. Téc.). NIOSH, National Institute for Occupational Safety and Health. CDC, Canadian Centre of Research Expertise for the prevention of Musculoskeletal Disorders. Descargado de <https://stacks.cdc.gov/view/cdc/24085>
- Luttmann, A., Jager, M., y Griefahn, B. (2004). Prevención de trastornos musculoesqueléticos en el lugar de trabajo. *Organización Mundial de la Salud*(5), 1 – 30. Descargado de http://www.who.int/occupational_health/publications/muscdisorders/es/

- Massiris Fernández, M., Bajo, J. M., Fernández Muñoz, J. Á., y Delrieux, C. (2020). *An automated system for monitoring the use of personal protective equipment in the construction industry* (Tesis Doctoral). doi: <https://doi.org/10.4995/riai.2020.13243>
- Massiris Fernández, M., Fernández, J. Á., Bajo, J. M., y Delrieux, C. A. (2020). Ergonomic risk assessment based on computer vision and machine learning. *Computers and Industrial Engineering*, 149(November), 106816. doi: 10.1016/j.cie.2020.106816
- Massiris Fernández, M., Fernández Muñoz, J. Á., y Delrieux, C. (2018). Detección de equipos de protección personal mediante la red neuronal convolucional YOLO. En *Actas de las xxxix jornadas de automática* (pp. 1022–1029). Badajoz, España. Descargado de <https://www.researchgate.net/publication/327449170>
- Massiris Fernández, M. M., Bajo, J. M., y Delrieux, C. A. (2019). Aproximación a la evaluación de riesgos ergonómicos mediante visión por computadora. En *Xvii jornadas de jovens pesquisadores* (pp. 1–7). São Carlos, Brasil: Asociación de Universidades del Grupo Montevideo.
- Massiris Fernández, M. M., Bajo, J. M., Fernández, J. Á., y Delrieux, C. A. (2020). Joint Angle Estimation with VIBE : an Evaluation Using Virtual Avatars. En *Ieee argencon 2020* (pp. 1–4). Resistencia, Argentina: IEEE Computer Society.
- Massiris Fernández, M. M., Bajo, J. M., Martínez, S., Fernández, J. Á., y Delrieux, C. A. (2020). Vision-Based Ergonomic Risk Estimation : Deep-learning Strategies. En *9th international workshop on soft computing applications* (pp. 1–10). Arad, Romania: Springer.
- McAtamney, L., y Corlett, E. N. (1993). RULA: a survey method for the investigation of work-related upper limb disorders. *Applied ergonomics*, 24(2), 91–99.
- McAtamney, L., y Hignett, S. (2000). Rapid Entire Body Assessment. *Applied Ergonomics*, 31, 201–205. doi: 10.1201/9780203489925.ch8
- Memarzadeh, M., Golparvar-Fard, M., y Niebles, J. C. (2013, jul). Automated 2D detection of construction equipment and workers from site video streams using histograms of oriented gradients and colors. *Automation in Construction*, 32, 24–37. Descargado de <https://www.sciencedirect.com/science/article/pii/S0926580512002403?via=ihub> doi: 10.1016/J.AUTCON.2012.12.002
- Ministerio de Trabajo, E. y. S. S. (2003, nov). *Resolución 295/2003, especificaciones técnicas sobre ergonomía y levantamiento manual de cargas*. Ministerio de Trabajo, Empleo y Seguridad Social. Descargado de <http://servicios.infoleg.gob.ar/infolegInternet/anexos/90000-94999/90396/norma.htm>
- Mnemyneh, B. E., Abbas, M., y Houry, H. (2017, jan). Automated Hardhat Detection for Construction Safety Applications. *Procedia Engineering*, 196, 895–902. Descargado de <https://www.sciencedirect.com/science/article/pii/S1877705817331430> doi: 10.1016/J.PROENG.2017.08.022
- Mosberger, R., Andreasson, H., y Lilienthal, A. J. (2014). A customized vision system for tracking humans wearing reflective safety clothing from industrial vehicles and machinery. *Sensors (Switzerland)*, 14(10), 17952–17980. doi: 10.3390/s141017952
- Mura, K., Vignais, N., Miezal, M., Gorecky, D., Marin, F., Bleser, G., ... Marin, F. (2012). Innovative system for real-time ergonomic feedback in industrial manufacturing. *Applied Ergonomics*, 44(4), 566–574. doi: 10.1016/j.apergo.2012.11.008
- Nath, N. D., Akhavian, R., y Behzadan, A. H. (2017). Ergonomic analysis of construction

- worker's body postures using wearable mobile sensors. *Applied Ergonomics*, 62, 107–117. doi: 10.1016/j.apergo.2017.02.007
- Nath, N. D., Behzadan, A. H., y Paal, S. G. (2020). Deep learning for site safety: Real-time detection of personal protective equipment. *Automation in Construction*, 112(January), 103085. Descargado de <https://doi.org/10.1016/j.autcon.2020.103085> doi: 10.1016/j.autcon.2020.103085
- Ovako, O. (2009). OWAS (Ovako Working posture Assessment System). *Finland Institute of Occupational Health*, 1(June), 1–6. Descargado de <http://www.ttl.fi/en/ergonomics/methods/workload{ }exposure{ }methods/table{ }and{ }methods/Pages/default.aspx>
- Park, M. W., y Brilakis, I. (2012, dec). Construction worker detection in video frames for initializing vision trackers. *Automation in Construction*, 28, 15–25. doi: 10.1016/j.autcon.2012.06.001
- Park, M. W., y Brilakis, I. (2016, dec). Continuous localization of construction workers via integration of detection and tracking. *Automation in Construction*, 72, 129–142. Descargado de <http://dx.doi.org/10.1016/j.autcon.2016.08.039> doi: 10.1016/j.autcon.2016.08.039
- Park, M.-W., Elsafty, N., y Zhu, Z. (2015). Hardhat-wearing detection for enhancing on-site safety of construction workers. *Journal of Construction Engineering and Management*, 141(9), 4015024. doi: 10.1061/(ASCE)CO.1943-7862.0000974
- Peppoloni, L., Filippeschi, A., Ruffaldi, E., y Avizzano, C. A. (2014, aug). (WMSDs issue) A novel wearable system for the online assessment of risk for biomechanical load in repetitive efforts. *International Journal of Industrial Ergonomics*, 52, 1–11. doi: 10.1016/j.ergon.2015.07.002
- Plantard, P., Auvinet, E., Le Pierres, A. S., y Multon, F. (2015, jan). Pose estimation with a kinect for ergonomic studies: Evaluation of the accuracy using a virtual mannequin. *Sensors (Switzerland)*, 15(1), 1785–1803. doi: 10.3390/s150101785
- Plantard, P., Shum, H. P., Le Pierres, A. S., y Multon, F. (2017). Validation of an ergonomic assessment method using Kinect data in real workplace conditions. *Applied Ergonomics*, 65, 562–569. Descargado de <http://dx.doi.org/10.1016/j.apergo.2016.10.015> doi: 10.1016/j.apergo.2016.10.015
- Raaj, Y., Idrees, H., Hidalgo, G., y Sheikh, Y. (2019). Efficient online multi-person 2D pose tracking with recurrent spatio-temporal affinity fields. En *Proceedings of the IEEE computer society conference on computer vision and pattern recognition* (Vol. June, pp. 4615–4623). doi: 10.1109/CVPR.2019.00475
- Rauch, E., Linder, C., y Dallasega, P. (2019, jan). Anthropocentric perspective of production before and within Industry 4.0. *Computers & Industrial Engineering*, 105644. Descargado de <https://www.sciencedirect.com/science/article/pii/S0360835219300233> doi: 10.1016/J.CIE.2019.01.018
- Rauch, E., Linder, C., y Dallasega, P. (2020). Anthropocentric perspective of production before and within Industry 4.0. *Computers and Industrial Engineering*, 139(January 2019), 105644. Descargado de <https://doi.org/10.1016/j.cie.2019.01.018> doi: 10.1016/j.cie.2019.01.018
- Redmon, J., y Farhadi, A. (2017). YOLO9000: better, faster, stronger. *arXiv preprint*.
- Roman-Liu, D. (2014). Comparison of concepts in easy-to-use methods for MSD risk

- assessment. *Applied Ergonomics*, 45(3), 420–427. Descargado de <http://dx.doi.org/10.1016/j.apergo.2013.05.010> doi: 10.1016/j.apergo.2013.05.010
- Rubaiyat, A. H., Toma, T. T., Kalantari-Khandani, M., Rahman, S. A., Chen, L., Ye, Y., y Pan, C. S. (2017, jan). Automatic detection of helmet uses for construction safety. En *Proceedings - 2016 IEEE/WIC/ACM International Conference on Web Intelligence Workshops, WIW 2016* (pp. 135–142). Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/WIW.2016.10
- Savino, M. M., Battini, D., y Riccio, C. (2017). Visual management and artificial intelligence integrated in a new fuzzy-based full body postural assessment. *Computers and Industrial Engineering*, 111, 596–608. Descargado de <http://dx.doi.org/10.1016/j.cie.2017.06.011> doi: 10.1016/j.cie.2017.06.011
- Seong, H., Choi, H., Cho, H., Lee, S., Son, H., y Kim, C. (2017). Vision-based safety vest detection in a construction scene. En *Isarc 2017 - proceedings of the 34th international symposium on automation and robotics in construction* (pp. 288–293). doi: 10.22260/isarc2017/0039
- Shrestha, K., Shrestha, P. P., Bajracharya, D., y Yfantis, E. A. (2015). Hard-hat detection for construction safety visualization. *Journal of Construction Engineering*, 2015.
- U.S. Bureau of Labor Statistics. (2015). Census of fatal occupational injuries. 2014 Cfoi, 20. Descargado de <http://www.bls.gov/iif/oshwc/cfoi/cfch0013.pdf>
- Vignais, N., Bernard, F., Touvenot, G., y Sagot, J. C. (2017, nov). Physical risk factors identification based on body sensor network combined to videotaping. *Applied Ergonomics*, 65, 410–417. doi: 10.1016/j.apergo.2017.05.003
- Wilson, J. R. (2000). Fundamentals of ergonomics in theory and practice. *Applied Ergonomics*, 31(6), 557–567. doi: 10.1016/S0003-6870(00)00034-X
- Wu, J., Cai, N., Chen, W., Wang, H., y Wang, G. (2019). Automatic detection of hardhats worn by construction personnel: A deep learning approach and benchmark dataset. *Automation in Construction*, 106(February), 102894. Descargado de <https://doi.org/10.1016/j.autcon.2019.102894> doi: 10.1016/j.autcon.2019.102894
- Xie, Z., Liu, H., Li, Z., y He, Y. (2018). A convolutional neural network based approach towards real-time hard hat detection. *Proceedings of the 2018 IEEE International Conference on Progress in Informatics and Computing, PIC 2018*, 430–434. doi: 10.1109/PIC.2018.8706269
- Xu, X., y McGorry, R. W. (2015, jul). The validity of the first and second generation Microsoft Kinect for identifying joint center locations during static postures. *Applied Ergonomics*, 49, 47–54. doi: 10.1016/j.apergo.2015.01.005
- Yan, X., Li, H., Wang, C., Seo, J. O., Zhang, H., y Wang, H. (2017). Development of ergonomic posture recognition technique based on 2D ordinary camera for construction hazard prevention through view-invariant features in 2D skeleton motion. *Advanced Engineering Informatics*, 34(November), 152–163. Descargado de <https://doi.org/10.1016/j.aei.2017.11.001> doi: 10.1016/j.aei.2017.11.001
- Yu, Y., Guo, H., Ding, Q., Li, H., y Skitmore, M. (2017). An experimental study of real-time identification of construction workers' unsafe behaviors. *Automation in Construction*, 82, 193–206.
- Yu, Y., Li, H., Umer, W., Dong, C., Yang, X., Skitmore, M., y Wong, A. Y. (2019). Automatic Biomechanical Workload Estimation for Construction Workers by Computer Vi-

- sion and Smart Insoles. *Journal of Computing in Civil Engineering*, 33(3), 1–13. doi: 10.1061/(ASCE)CP.1943-5487.0000827
- Zhang, H., Yan, X., y Li, H. (2018). Ergonomic posture recognition using 3D view-invariant features from single ordinary camera. *Automation in Construction*, 94, 1–10. doi: 10.1016/j.autcon.2018.05.033
- Zhang, H., Yan, X., Li, H., Jin, R., y Fu, H. F. (2019). Real-Time Alarming, Monitoring, and Locating for Non-Hard-Hat Use in Construction. *Journal of Construction Engineering and Management*, 145(3), 1–13. doi: 10.1061/(ASCE)CO.1943-7862.0001629