



Tesis Doctoral

Gestión y evaluación de la seguridad en el trabajo mediante técnicas de visión artificial

Manlio Miguel Massiris Fernández

Programa de Doctorado en Tecnologías Informáticas

Director: Juan Álvaro Fernández Muñoz

Esta tesis cuenta con la autorización del director de la misma y de la Comisión Académica del programa. Dichas autorizaciones constan en el Servicio de la Escuela Internacional de Doctorado de la Universidad de Extremadura.

2022

Dedicación

Quiero dedicar la presente tesis a mi familia,

A mi madre Luz, mi padre Manlio y a mis hermanos Miguel y Daniel, por su constante e inmensurable sustento emocional, ustedes me recordaron cada día que sí era posible lograr este sueño.

A mis Abuelos, Tíos, Tías y Primos por brindarme su apoyo en cada pequeño paso de mi camino.

A todos, muchas gracias desde el fondo de mi corazón.

Agradecimientos

A mi director, el Dr. J. Álvaro Fernández, por brindarme la posibilidad de redactar esta tesis bajo su batuta. Le agradezco especialmente su constante paciencia y el hecho de que me haya podido aportar valores, tanto en el plano científico como en el humano.

A la Universidad de Extremadura y su Escuela de Ingenierías Industriales.

Índice general

1. Introducción	1
1.1. La ergonomía y los factores humanos	1
1.2. Planteamiento del problema	6
1.3. Objetivos de la tesis	6
1.3.1. Objetivo general	6
1.3.2. Objetivos específicos	7
1.4. Estructura del documento	8
2. Métodos y Resultados	9
2.1. Detección de equipos de protección personal mediante la red neuronal convolucional YOLO	12
2.1.1. Materiales y métodos	12
2.1.2. Resultados	15
2.2. Evaluación de riesgos ergonómicos mediante visión por computador y aprendizaje automático	17
2.2.1. Metodología propuesta	17
2.2.2. Resultados de pruebas de entorno controlado	25
2.2.3. Resultados de pruebas para ángulos de hombro y cuello	27
2.2.4. Resultados de validación	30
2.3. Sistema automatizado para monitorear el uso de equipos de protección personal en la industria de la construcción	37

2.3.1.	Materiales y métodos	37
2.3.2.	Resultados experimentales	43
2.4.	Estimación de ángulos de articulaciones mediante VIBE en avatares virtuales	49
2.4.1.	Materiales y métodos	49
2.4.2.	Resultados	54
3.	Discusión	61
3.1.	Medición automatizada del riesgo de trastornos musculoesqueléticos asistido mediante CV	61
3.2.	Monitoreo automatizado de utilización de EPP asistido mediante CV	65
4.	Conclusiones y trabajo futuro	69
4.1.	Conclusiones	69
4.2.	Trabajo futuro	70

Resumen

La OMS recomienda efectuar evaluaciones de riesgos laborales (ERL) para disminuir las posibilidades de ocurrencia de accidentes y/o dolencias musculoesqueléticas relacionadas con el trabajo. Estas ERL tienen como objetivo adaptar las condiciones de trabajo a la capacidad de los trabajadores y proponer mejoras cuando se hallen riesgos para su salud. Las ERL se llevan a cabo principalmente mediante la observación *in-situ* por parte de ergonomistas. Sin embargo, es sabido que los trabajadores alteran su desempeño habitual al sentirse observados, aunado al efecto inevitable de la subjetividad intra- e inter-observador. Así, factores diferenciales como el entorno de observación, el medio de observación, la formación y la experiencia afectan a la reproducibilidad y la trazabilidad de las ERL tradicionales.

En esta tesis doctoral se propone investigar y testear nuevas tecnologías basadas en Inteligencia Artificial (IA) aplicadas en un Sistema de Visión por Computadora (CVS) para obtener ERL automatizadas que reduzcan los factores limitativos anteriores. Como resultado, se ha demostrado en el laboratorio, en simulaciones y en entornos laborales reales que estos CVS inteligentes pueden ser entrenados y luego utilizados para el monitoreo de acciones que pueden generar lesiones en los trabajadores. Concretamente, se han diseñado y probado métodos basados en IA en el diseño de CVS para la detección automática de los usos adecuados del equipo de protección personal (EPP), así como para la supervisión automatizada de posibles riesgos musculoesqueléticos relacionados con el trabajo, ambos enmarcados en los procesos de digitalización del entorno de laboral planteada por la Industria 4.0.

Palabras clave: Ergonomía, Automatización, Prevención de riesgos laborales, Equipos de protección personal, Redes neuronales, Visión por computadora, Evaluación de riesgos laborales, Salud y seguridad en el trabajo, Trastornos musculoesqueléticos.

Abstract

WHO recommends performing occupational risk assessments (ORAs) to reduce the chances of accidents and/or work-related musculoskeletal disorders. The purpose of these ORAs is to adapt working conditions to the workers' capabilities and to propose improvements when occupational health risks are found. ORAs are mainly carried out through on-site observation by ergonomists. However, it is known that workers alter their usual performance when they feel observed, added to the inevitable effect of intra- and inter-observer subjectivity. Thus, differential factors such as observation environment, observation medium, training and experience affect the reproducibility and traceability of traditional RLEs.

In this doctoral thesis we propose to investigate and test new technologies based on Artificial Intelligence (AI) applied in a Computer Vision System (CVS) to obtain automated ERLs that reduce the above limiting factors. As a result, it has been demonstrated in the laboratory, in simulations and in real work environments that these intelligent CVS can be trained and then used for monitoring actions that can generate injuries in workers. Specifically, AI-based methods have been designed and tested in the design of CVS for the automatic detection of appropriate uses of personal protective equipment (PPE), as well as for the automated monitoring of possible work-related musculoskeletal risks, both framed in the processes of digitization of the work environment raised by Industry 4.0.

Keywords: Ergonomics, Automation, Occupational risk prevention, Personal protective equipment, Neural networks, Computer vision, Ergonomic risk assessment, Occupational health and safety, Work-related musculoskeletal disorders.

Capítulo 1

Introducción

1.1. La ergonomía y los factores humanos

Según la Asociación Internacional de Ergonomía (AIE) [Int01] y las normas técnicas españolas vigentes (UNE EN-614-1:2006, UNE-EN ISO 6385:2016) [Gob04], la ergonomía (o el estudio de los factores humanos) es la rama de la ciencia que se encarga de las interacciones entre las personas y su entorno laboral, y que aplica la teoría y los principios, datos y métodos para optimizar el bienestar humano y el rendimiento general de un sistema productivo, lo cual puede lograrse ofreciendo un ambiente con factores humanos que se adecuen en términos de seguridad y comodidad a la persona [I L02].

La AIE ha identificado tres factores clave en un análisis ergonómico completo: factores físicos, cognitivos y organizativos. Los factores físicos están relacionados con las características anatómicas y fisiológicas de los trabajadores, la ergonomía cognitiva hace referencia al pensamiento del trabajador y la ergonomía organizativa se encarga de los procesos sociales llevados a cabo en un entorno laboral [Int01]. Sumado a lo anterior, la norma (UNE EN ISO 614-1: 2006) define entorno laboral como el sistema donde uno o varios trabajadores y las herramientas de trabajo interactúan para desarrollar una función en un puesto de trabajo determinado.

Por último, la Directiva Europea 89/391/CEE y la Ley española 31/1995 [Ins95], de

Prevención de Riesgos Laborales (PRL), establecen la obligación de una organización a realizar Evaluaciones de Riesgos Laborales (ERLs) por parte de profesionales competentes. La ERL se define como el proceso de identificación y valoración de aquellos riesgos a los cuales están expuestos los trabajadores [LWA14]. Por ello, se debe controlar de forma periódica las condiciones medioambientales, la estratificación en la organizaciones, los procesos de trabajo y la salud física y psicosocial de los trabajadores. De esta forma, es posible recopilar información para tomar decisiones sobre la necesidad de adoptar medidas que reduzcan o mitiguen el riesgo laboral. Entre las decisiones resultantes usuales de las ERLs destacan las medidas de reducción de riesgo en el origen, las medidas organizativas, las medidas de protección colectiva, las medidas de protección individual como la utilización de equipo de protección personal (EPP), así como la formación e información a los trabajadores sobre la prevención de enfermedades laborales [ILO01].

Los trastornos musculoesqueléticos (TMEs) se definen como las enfermedades del aparato locomotor humano que son causadas o empeoradas por un trabajo efectuado [ILO95]. Recientemente, la Agencia Europea para la Seguridad y la Salud en el Trabajo ha concluido que los TMEs son, con diferencia, el mayor problema para la salud laboral [VLS+21], resaltando que dichas enfermedades suelen ser la mayor causa de absentismo laboral. De la misma forma, el gasto ocasionado por los TMEs es elevado en los regímenes de seguridad social de España y de la mayoría de los países occidentales [LPG+14; CPDR19]. Según la última encuesta a trabajadores españoles [Ins15], es preocupante el porcentaje de trabajadores que manifiestan estar expuestos a riesgos ergonómicos que pueden causar TMEs. Específicamente, en los sectores de la agricultura, la industria, la construcción, el comercio y la hostelería, entre el 75 % y el 79 % de los trabajadores manifiesta realizar movimientos repetitivos con sus manos o brazos. Además, entre el 44 % y 73 % de los trabajadores manifiesta manipular cargas pesadas. En la industria de la construcción, esta situación es de especial interés debido a que presenta factores que aumentan el riesgo laboral de sufrir una TME, como lo son la elevada proporción de pequeñas empresas y de trabajadores autónomos, la variedad de las obras de construcción, la elevada rotación de los trabajadores, el gran número de trabajadores estacionales y migrantes, trabajadores

novatos que no están familiarizados con los procesos de construcción, la exposición a las inclemencias del tiempo, y los múltiples oficios y ocupaciones involucrados [ILO95].

Para mitigar el gran impacto de las TMEs, ergonomistas, ingenieros, médicos y diverso personal de seguridad y salud en el trabajo (SST) han sugerido protocolos enmarcados en los procesos de mejora continua usuales de las empresas, que permiten afrontar la reducción de riesgos laborales de una forma escalonada. La ERL propuesta en 1997 por el Instituto Nacional de Seguridad y Salud Ocupacional de EE.UU. (NIOSH) definió una serie de pasos a incluir obligatoriamente en los programas de ergonomía, con el fin de generar un proceso sistemático para identificar, analizar y controlar los factores de riesgo en el lugar de trabajo [LDJ19]. Como resultado, todo programa eficaz de ergonomía debe incluir una ERL basada en evidencias para reconocer y rectificar las deficiencias halladas en un amplio abanico de situaciones laborales [CGF97].

En las últimas décadas, se ha diseñado un numeroso conjunto de métodos y herramientas para realizar ERLs. Dichos métodos pueden clasificarse, según su forma de aplicación, como sigue [LDJ19; VBTS17; Gar20]:

- Autoevaluación, donde los trabajadores se evalúan a sí mismos utilizando formularios diseñados específicamente para ello.
- Observación humana, donde personal especialmente cualificado recopila una estimación de los ángulos de las articulaciones del cuerpo mediante observaciones *in situ* (en línea) o de vídeo (fuera de línea).
- Medición directa, donde se adosan al cuerpo del trabajador herramientas antropométricas o de medición inercial para recopilar datos.
- Evaluación mediante Visión por Computadora (CV), donde se determinan automáticamente modelos del cuerpo humano a partir de fotografías o vídeos, lo que proporciona una medición ergonómica basada en modelos de Industria 4.0 (I4.0), que se caracterizan por ser más escalables, sistemáticos y objetivos.

Actualmente, el campo de investigación de las ERLs se encuentra en una encrucijada [YUYA21; HR20; PCR19], debido a que los métodos más utilizados en ERLs son los dos primeros: autoevaluación y observación por especialistas. Ambos grupos adolecen de sesgos inherentes a las evaluaciones realizadas por humanos, como son la variabilidad en los resultados obtenidos por diferentes evaluadores (sesgo inter-observador), e incluso por el mismo evaluador en diferentes contextos (sesgo intra-observador) [PEZ15; RLL+16]. La variabilidad inter-observador se debe principalmente a factores subjetivos como la formación, la experiencia o los medios de observación, mientras que la variabilidad intra-observador suele estar asociada a la fatiga cognitiva y a los cambios en el entorno de observación [SAG19; DAP17; GAHR17; ISH+20]. Adicionalmente, el estado del arte señala que las autoevaluaciones pueden subestimar los peligros y requieren la cooperación general de los trabajadores, lo que hace que sus efectos sean inexactos y difíciles de medir a lo largo del tiempo [GYS17].

En [DAP17] se cuantificaron los errores más comunes en el uso de métodos observacionales de las ERLs, obteniendo respuestas de 317 expertos de 20 países iberoamericanos. Como resultado, al menos el 30% de las evaluaciones presentaron algún error. Además, dado que la principal alternativa a una ERL basada en observación *in situ* es la realizada mediante la observación de vídeos, incluso los ergónomos más capacitados cometen errores en sus ERLs debido a condiciones propias del lugar de trabajo que pueden afectar a la calidad de las grabaciones de vídeo, como son el ángulo de visualización o captura, la iluminación reducida o no homogénea, y las oclusiones provocadas por materiales y por personas, incluyendo al propio trabajador (auto-occlusiones) [PSLM17].

La I4.0, basada en la digitalización del entorno de trabajo, ofrece nuevas soluciones para solventar estos problemas de repetibilidad en las ERLs [MUF+20]. El monitoreo de riesgos en los entornos laborales es un área de investigación de creciente interés en la I4.0 [YUYA21], ya que sus resultados permiten mejorar la toma de decisiones por parte de expertos y, por ende, reducir el costo económico, social y ético generado por los accidentes laborales. En los últimos años, dicha problemática se ha abordado principalmente de dos formas: mediante el uso de medidores inerciales (medición directa), y mediante técnicas

de CV (medición indirecta) [LLRL21; ZYL18].

Hoy en día, resulta evidente que las tecnologías de la I4.0 presentan un panorama amplio para su crecimiento y aplicabilidad en sectores clave como son la industria y la construcción. Sin embargo, este es un proceso engorroso de adoptar, debido generalmente a la necesidad de personal altamente capacitado, de tiempo y recursos económicos [YUYA21; HLS14]. La reducción de TMEs es especialmente costosa en ambientes de trabajo cambiantes, típicos del sector de la construcción, en particular con el uso de métodos de ERLs de medición directa [LLRL21], ya que requieren medidores inerciales que se adosan al cuerpo del trabajador, mermando la comodidad y limitando la naturalidad del trabajo a evaluar [ZYL18]. En contraposición, para utilizar métodos de CV solo se necesitan equipos de cómputo y cámaras portátiles.

Para superar dichas limitaciones, en esta tesis se propone investigar y testear nuevas alternativas para los procesos de mejora de la seguridad laboral basados en el uso de técnicas y modelos de CV basados en Inteligencia Artificial (IA), disciplina que ha experimentado en los últimos años una notable madurez [RLL+16; LLRL21; ZYL18; VBTS17; LZLL20; YUYA21]. Así, la tecnología actual permite diseñar y entrenar un Sistema de Visión por Computadora (CVS) para ser utilizado en el monitoreo de movimientos que puedan generar un alto riesgo ergonómico para el trabajador, o para documentar de forma automatizada la trazabilidad a lo largo del tiempo de la utilización de EPP por parte de la plantilla de trabajo [LLRL21].

El hilo integrador que brinda coherencia unitaria a la presente tesis doctoral es la hipótesis fundamental de que la CV puede apoyar el proceso de toma de decisiones de la gestión de la seguridad laboral. De esta forma, enmarcadas en la hipótesis anterior, se abordan dos temáticas específicas: el monitoreo de utilización de Equipo de protección personal (EPP) asistido mediante CV y la medición automatizada de riesgo musculoesquelético utilizando CV y aprendizaje automático (ML).

1.2. Planteamiento del problema

En la gestión de la seguridad laboral se deben identificar los riesgos a los cuales los trabajadores están expuestos en su entorno de trabajo, para después aplicar, en su caso, las medidas correctivas adecuadas. Este proceso de identificación de riesgos laborales se ha realizado tradicionalmente mediante la observación de expertos o por auto-informes de situaciones riesgosas realizados por los propios trabajadores. Sin embargo, ambos métodos producen importantes imprecisiones debidas, entre otros factores, a la subjetividad humana. En este sentido, la Visión Artificial o por Computadora (CV) representa una oportunidad para la estandarización de procesos y la reducción de costos. En particular, hay oportunidades para la medición de riesgos ergonómicos y el monitoreo de la utilización de equipos de protección personal (EPP). Estos sistemas permiten una aproximación de procesamiento paralelo y en tiempo real de imágenes provenientes de distintas cámaras y sensores, que pueden utilizarse para monitorear de forma automática y precisa un entorno de trabajo. Desde el punto de vista informático, las Redes Neuronales Convolucionales (CNN) más recientemente desarrolladas representan herramientas ideales para tareas típicas de CV, como la detección de objetos, el seguimiento de personas y la identificación de la postura instantánea de un trabajador, entre otras. Además, algunas de estas CNN que forman parte del estado del arte se encuentran disponibles en código abierto, lo que posibilita el desarrollo acelerado de aplicaciones de CV de alto nivel para realizar las pruebas necesarias en la investigación.

1.3. Objetivos de la tesis

1.3.1. Objetivo general

En esta tesis doctoral se propone investigar y evaluar nuevas alternativas para los procesos de mejora en seguridad industrial basados en CV. En la actualidad, resulta posible programar y calibrar un CVS (*ad hoc* o mediante técnicas de Inteligencia Artificial) para ser utilizado en tareas de monitoreo, clasificación y detección del comportamiento de

los trabajadores, así como para determinar de forma robusta la trazabilidad de las mejoras a obtener mediante cambios en las conductas objetivo. Todo ello permitiría prevenir riesgos laborales, mejorar la capacitación, la toma de conciencia de los actores intervinientes y aumentar la productividad en general. Se tomarán como casos específicos de estudio el sector industrial y el de la construcción, dada la complejidad y heterogeneidad de sus respectivos entornos laborales y de las tareas que allí se realizan.

1.3.2. Objetivos específicos

- Estudiar las más recientes tecnologías en Prevención de Riesgos Laborales (PRL) disponibles en la actualidad y las formas en que estas pueden ser aplicadas al sector industrial y al sector de la construcción, de acuerdo a las normativas internacionales de aplicación.
- Investigar las tecnologías actuales basadas en CV que puedan impactar en la detección e identificación no supervisada de conductas de los trabajadores que impliquen un riesgo laboral.
- Analizar, en base a lo determinado en los dos objetivos anteriores, los requerimientos en *hardware* y *software* necesarios para una instalación de monitoreo laboral basado en CV.
- Determinar, desarrollar y testear técnicas viables para el desarrollo de una instalación piloto de monitoreo de entornos laborales con características parametrizables y controlables.
- Analizar las condiciones ambientales necesarias para la instalación efectiva del sistema de monitoreo en un ambiente laboral de tipo industrial o de la construcción, no controlado, tanto en interiores como en exteriores.

1.4. Estructura del documento

La presente tesis doctoral está basada en una colección de artículos científicos publicados durante la misma, que tienen una continuidad temporal y objetivos comunes. En consecuencia, en este primer Capítulo se realiza una introducción general de la tesis y del planteamiento del problema, para establecer los objetivos de la tesis. En el Capítulo 2 se presenta la metodología planteada y los resultados de cada trabajo publicado. Seguidamente, el Capítulo 3 presenta una discusión global de los resultados de los artículos que vertebran esta tesis doctoral. Por último, en el Capítulo 4 se resumen las principales aportaciones científicas de esta tesis y se establecen las vías más interesantes, a criterio de este doctorando, sobre posibles desarrollos futuros.

Capítulo 2

Métodos y Resultados

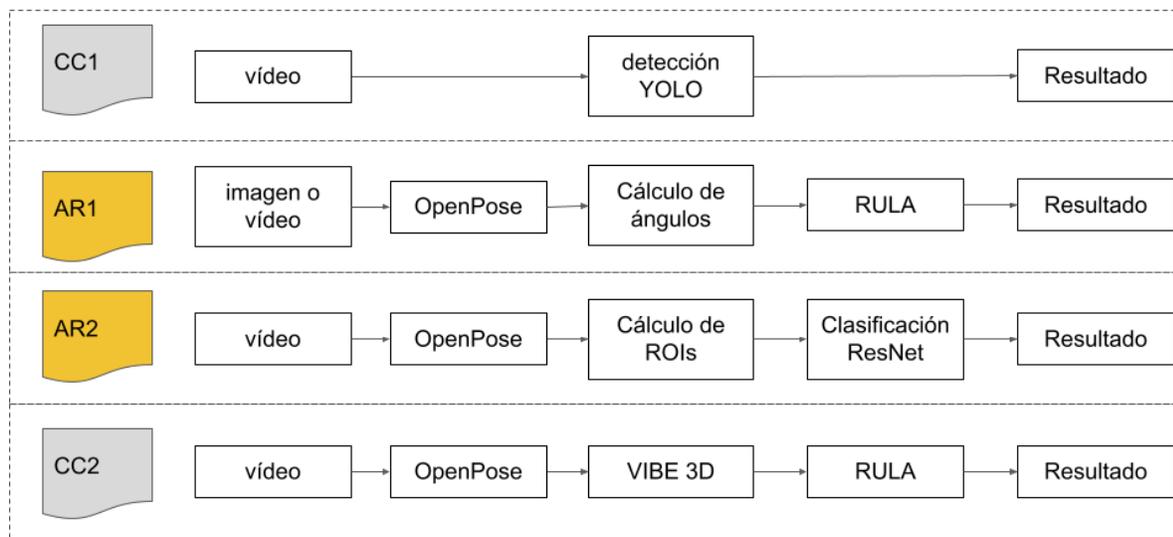


Figura 2.1: Artículos en orden cronológico.

La presente tesis se basa en la cronología de publicaciones científicas ilustrada en la Figura 2.1 y listada a continuación:

- **CC1** Manlio Massiris Fernández, Claudio A. Delrieux, J. Álvaro Fernández. 2018. “Detección de equipos de protección personal mediante red neuronal convolucional YOLO,” en Actas de las XXXIX Jornadas de Automática, 5-7 Sep 2018, Badajoz, España, pp. 1022–1029. ISBN: 978-84-09-04460-3. DOI: 10.17979/spudc.9788497497565. 1022

La comunicación [MFD18] cuenta con 23 citas acumuladas en Google Scholar hasta la fecha de redacción de la presente tesis doctoral.

- **AR1** Manlio Massiris Fernández, J. Álvaro Fernández, Juan M. Bajo, Claudio A. Delrieux. 2020. “Ergonomic risk assessment based on computer vision and machine learning.” *Computers and Industrial Engineering*, 149, 106816. DOI: 10.1016/j.cie.2020.106816

El artículo [MFBD20] está publicado en la revista científica *CAIE*, que se encuentra en el primer cuartil de su categoría (Q1) en el Science Journal Citation Reports (JCR). El mencionado artículo cuenta con 23 citas acumuladas en Google Scholar hasta la fecha de redacción de la presente tesis doctoral.

- **AR2** Manlio Massiris Fernández, Juan M. Bajo, J. Álvaro Fernández, Claudio A. Delrieux. 2021. “Sistema automatizado para monitorear el uso de equipos de protección personal en la industria de la construcción.” *Revista Iberoamericana de Automática e Informática Industrial*, 18(1), pp. 68-74. DOI: 10.4995/riai.2020.13243

El artículo [MFBD21] publicado en la revista científica *RIAI*, que se encuentra en el 2º cuartil de su categoría (Q2) en el Scimago Journal Ranking (SJR) y en el 4º cuartil de su categoría (Q4) en el Science Journal Citation Reports (JCR). El mencionado artículo cuenta con 1 cita acumulada en Google Scholar hasta la fecha de redacción de la presente tesis doctoral.

- **CC2** Manlio Massiris Fernández, Juan M. Bajo, J. Álvaro Fernández, Claudio A. Delrieux. 2020. “Joint angle estimation with VIBE: an evaluation using virtual avatars,” en *IEEE Congreso Bienal de Argentina (ARGENCON) 2020*. 1-4 Dic 2020, Resistencia, Argentina. ISBN: 978-1-7281-5957-7. DOI: 10.1109/ARGENCON49523.2020.9505526

La comunicación [MBFD20] no cuenta con citas en Google Scholar hasta la fecha de redacción de la presente tesis doctoral.

CAPÍTULO 2. MÉTODOS Y RESULTADOS

Estas cuatro publicaciones no son las únicas realizadas durante la tesis doctoral, aunque sí las más relevantes. Por este motivo, se han tomado como referencia para la redacción de la presente tesis doctoral.

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

En un número creciente de entornos de trabajo, el uso de EPPs se está convirtiendo en una cuestión obligatoria, ya que son la última barrera para detener posibles situaciones de riesgo físico para el trabajador. Esto significa que la supervisión periódica y fiable del cumplimiento de las normas de seguridad laboral es una tarea exigente, por lo que la supervisión automatizada representa una solución de alto impacto potencial para la seguridad. Se propone utilizar la Visión Artificial (CV) como alternativa para monitorear el uso de equipos de protección personal.

2.1.1. Materiales y métodos

2.1.1.1. Redes neuronales convolucionales (CNN)

Un caso particular de aprendizaje profundo (DL) son las Redes Neuronales Convolucionales o CNN (del inglés, *Convolutional Neural Network*) [LBOM98; Fuk88], que actualmente representan el estado del arte en múltiples ámbitos y problemas de la Visión por Computador (CV), debido a su sobresaliente rendimiento en el reconocimiento e interpretación de imágenes y vídeos [FLL+18]. Su capacidad para rendir adecuadamente en este contexto se basa en un diseño específico destinado, en primer lugar, a extraer las características de una imagen y, a continuación, utilizar dichas características para reconocer o clasificar los objetos presentes en una imagen. Los parámetros de una red neuronal convolucional (CNN) constan principalmente de una serie de coeficientes o pesos, asignados a cada capa y neurona, dentro de una estructura donde dichos pesos, aplicados sobre una entrada proveniente de una capa previa, se interpretan como filtros espaciales o de convolución 2D que se aplican sobre dicha entrada, siendo estos filtros los utilizados en el ámbito del Procesamiento Digital de Imágenes (DIP). Como ocurre también en otros tipos de Redes Neuronales, estos pesos (o filtros) pueden aprenderse o modificarse, de tal forma que se logre una optimización global de la clasificación o detección de clase deseada



Figura 2.2: Metodología propuesta: de izquierda a derecha, (1) dividir imagen en bloques de 13x13 px, (2) hacer predicciones con YOLO re-entrenado y (3) establecer umbrales para obtener solo detecciones más fiables.

respecto a una función de error [Fuk88]. Las CNN permiten que modelos informáticos compuestos por múltiples capas de procesamiento aprendan representaciones de los datos e imágenes en varios niveles de abstracción. De esta forma, una CNN correctamente entrenada es capaz de identificar modelos o clases generalizadas a partir de grandes conjuntos de datos [FLL+18].

2.1.1.2. YOLO

Esta sección utiliza la CNN conocida como You-Only-Look-Once (YOLO) [RF17] para el monitoreo de la presencia y el uso adecuado del EPP por parte de trabajadores del sector de la construcción. La topología de YOLO [RF17] está diseñada para resolver problemas de Visión Artificial con un alto rendimiento. En YOLO, la detección de objetos se aborda como un problema de regresión, donde la CNN predice simultáneamente múltiples regiones de interés (ROI) que rodean los objetos a detectar dentro de la imagen y asigna probabilidades condicionales para cada una de las clases, $p(\text{Clase}|\text{Objeto})$. Además, la red neuronal puede alcanzar una velocidad operativa de 45 fotogramas por segundo (fps) en computadoras de propósito general. Además, codifica implícitamente información contextual y modela el tamaño, forma y apariencia de los objetos [RF17] (ver Figura 2.2).

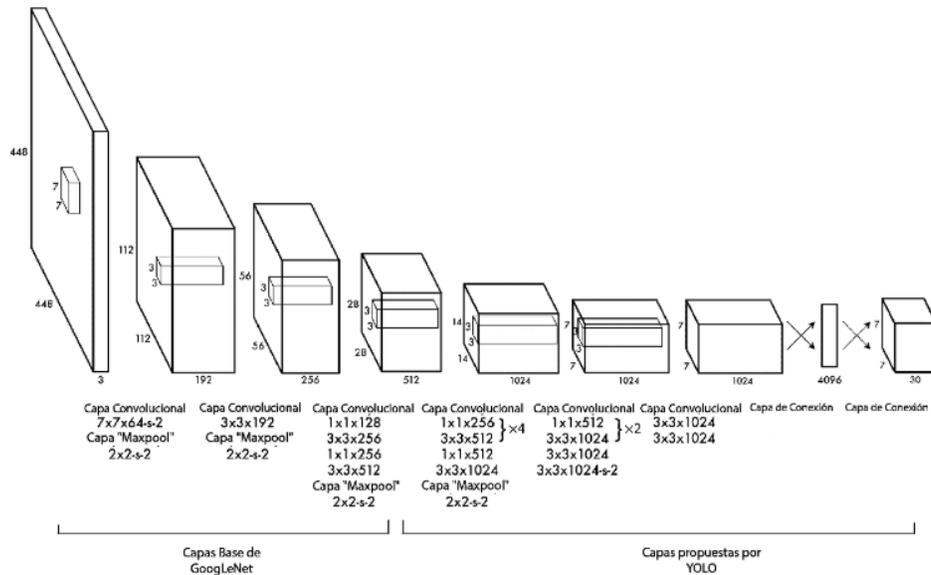


Figura 2.3: Arquitectura de la CNN YOLO [RF17]

La arquitectura de la red YOLO se inspira en el modelo GoogLeNet para la clasificación de imágenes. YOLO posee 24 capas convolucionales seguidas de 2 capas totalmente conectadas o densas (ver Figura 2.3). Las capas convolucionales iniciales de la red extraen las características de la imagen, mientras que las capas densas predicen las probabilidades de salida y las coordenadas de las ROIs de detección [RF17]. Además, en YOLO se alternan capas reductoras 1×1 con capas convolucionales 3×3 , para acortar el espacio de características entre las capas precedentes. El resultado final de la CNN es un tensor de predicciones de tamaño $7 \times 7 \times 30$ [RF17].

Las primeras capas convolucionales fueron pre-entrenadas en el conjunto de datos de competición de 1000 clases de ImageNet, hasta alcanzar una exactitud del 88% en el conjunto de validación de ImageNet 2012, comparable a los modelos de GoogLeNet en el Zoo de Modelos de Caffe [Mis17]. A continuación, se añadieron cuatro capas convolucionales y dos capas totalmente conectadas con pesos inicializados de forma aleatoria. La red resultante fue entrenada durante 135 épocas en los conjuntos de datos de entrenamiento y validación de PASCAL VOC 2007 y 2012 [HDH09]. Al realizar el testeo, también fueron incluidos los datos de testeo de ambos *datasets* [HDH09]. La capa final

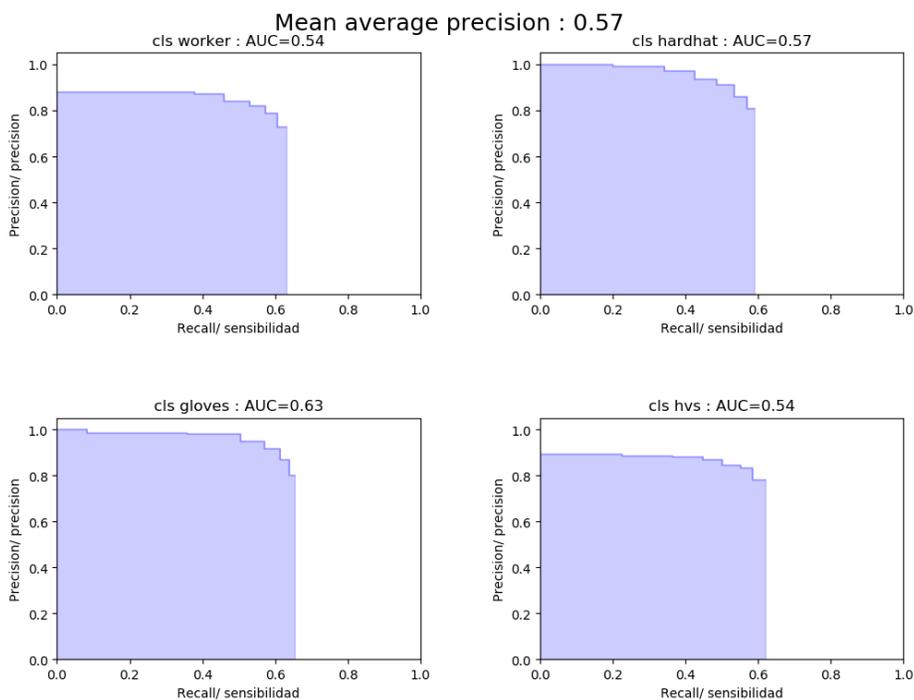


Figura 2.4: Curvas precisión-sensibilidad para las clases objetivo.

predice las probabilidades de cada clase considerada y las coordenadas de cada cuadro delimitador, cuya anchura y altura se normalizan respecto al tamaño de la imagen para obtener coordenadas entre 0 y 1 [RF17].

2.1.2. Resultados

La red neuronal YOLO original fue re-entrenada con un conjunto de datos creado a partir de videos generados mediante cámaras deportivas, con la intención de detectar diverso EPP: guantes, cascos, trajes de alta visibilidad y trabajadores. El re-entrenamiento se realizó con 1354 imágenes tomadas de 5 vídeos publicados en YouTube.com, las cuales se etiquetaron manualmente utilizando la herramienta YoloMark [Ale18]. Las imágenes etiquetadas se dividieron siguiendo las recomendaciones del artículo original en tres grupos: entrenamiento 80 %, testeo 10 % y validación 10 % (ver Figura 2.4).

Con el sistema entrenado, se presenta un análisis de caso abierto con un vídeo grabado

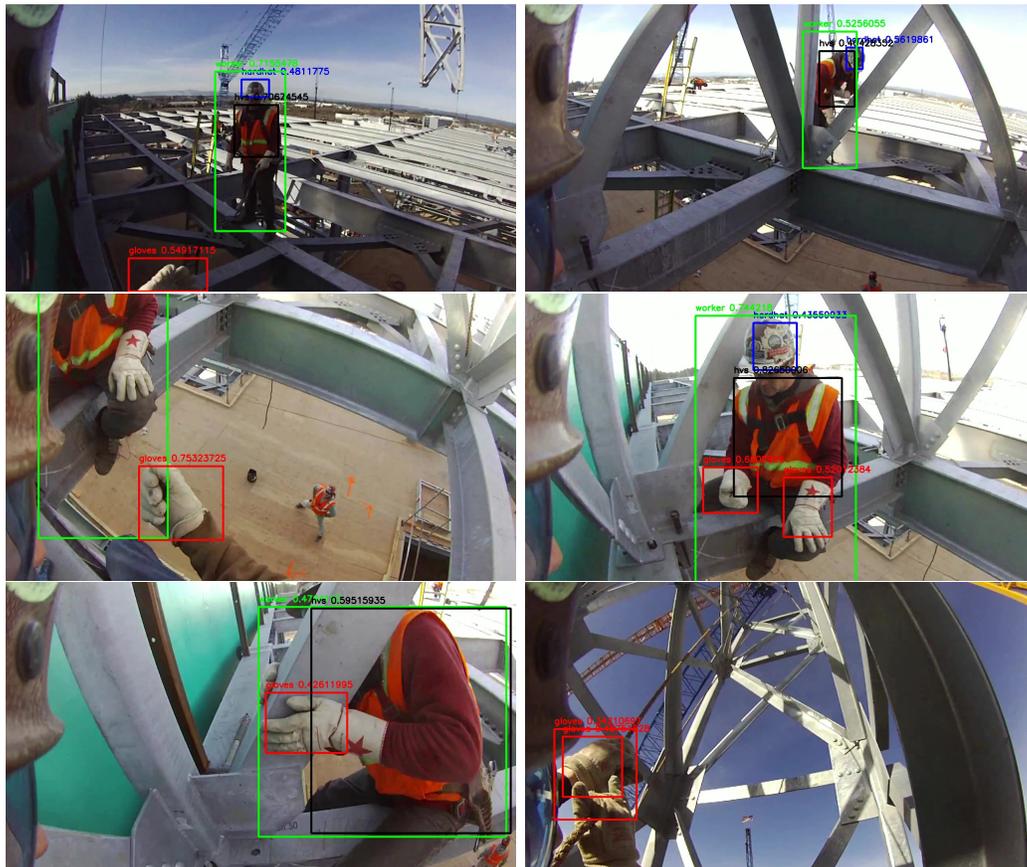


Figura 2.5: Fotogramas obtenidos a la salida de la CNN YOLO re-entrenada para detectar EPP, sobre un vídeo grabado en un entorno laboral real mediante cámara digital deportiva adosada al casco de un trabajador (ver vídeo completo en <https://youtu.be/54gwd21cjpg>).

con una cámara deportiva instalada en el casco de un trabajador metalúrgico en una obra de construcción real, no habiendo sido utilizado ningún *frame* de este vídeo para las etapas de entrenamiento o testeo. Los resultados obtenidos muestran que la estrategia propuesta es adecuada como solución implantable para estos entornos de trabajo (ver Figura 2.5).

2.2. Evaluación de riesgos ergonómicos mediante visión por computador y aprendizaje automático

Para este trabajo, se desarrolló una metodología que realiza una evaluación precisa de los riesgos ergonómicos, calculando automáticamente las puntuaciones obtenidas aplicando el método de Evaluación Rápida de las Extremidades Superiores (RULA) a partir de instantáneas o vídeos digitales, utilizando técnicas de Visión por Computador y Aprendizaje Automático. El método supera las limitaciones de otros desarrollos recientes también basados en CV, así como otros trabajos basados en el uso de sensores de medición portátiles (*wearables*), siendo capaz de realizar una evaluación no supervisada considerando múltiples trabajadores simultáneamente, incluso en condiciones de visualización no óptimas (por ejemplo, mala iluminación, oclusiones y vistas de cámara inestables). El flujo de trabajo de procesamiento utiliza redes neuronales de código abierto para detectar los esqueletos de los trabajadores, tras lo cual se infieren las posiciones y los ángulos de las articulaciones del cuerpo, de donde se calculan las puntuaciones RULA. El método se probó con conjuntos de datos de imágenes del mundo real generados por ordenador y con vídeos de libre acceso tomados en escenarios de trabajo al aire libre. Las puntuaciones RULA calculadas coincidieron con las evaluaciones de siete especialistas en la materia, alcanzando una κ de Cohen superior a 0,6 en la mayoría de los experimentos en el mundo real.

2.2.1. Metodología propuesta

En esta sección se presentan los métodos utilizados en este trabajo, que incluyen el método de ERL ergonómicos Evaluación Rápida de las Extremidades Superiores (RULA), la CNN OpenPose, la metodología propuesta para obtener puntuaciones RULA a partir de OpenPose, así como los métodos de validación empleados.

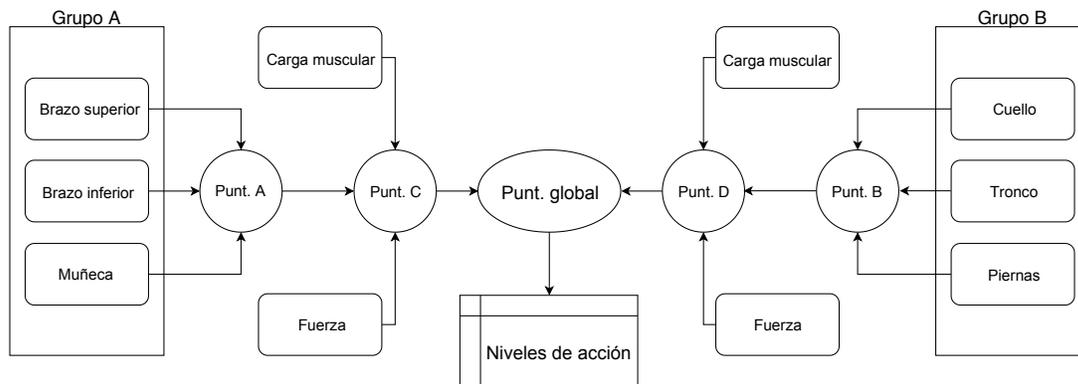


Figura 2.6: Esquema jerárquico de puntuaciones RULA.

2.2.1.1. RULA

El método RULA [MC93] fue diseñado para proporcionar una puntuación global de riesgo ergonómico conocida como *RULA Grand Score*. La puntuación global de RULA se obtiene a partir de un esquema de combinación jerárquica de puntuaciones individuales relacionadas con diversas partes del cuerpo (véase Figura 2.6), en el que cada parte del cuerpo se evalúa de forma independiente, usando principalmente para ello el valor de uno o más ángulos específicos de las articulaciones existentes en dicha parte del cuerpo, que, en función de su valor respecto a unos valores de referencia o umbral, se asocia a una puntuación o nivel [MC93].

Hoy en día, RULA es una herramienta de ERL ampliamente aceptada y utilizada por investigadores y profesionales del sector de la salud y seguridad en el trabajo (SST) [WM21; JD19; PSLM17; SBR17; DAP17]. Los principales beneficios de RULA para el profesional son su costo relativamente bajo, y la simplicidad del procedimiento, que a su vez facilita el proceso de retroalimentación generado a partir de la evaluación. Las puntuaciones de RULA se suelen obtener a partir de observación directa y también, más recientemente, a partir del análisis de fotografías y vídeos capturados *in situ* por cámaras digitales [XRC+17; Eur21]. Sin embargo, la evaluación siempre es realizada por un humano, lo cual implica inherentemente errores perceptuales que, a veces, conllevan una elevada variabilidad intra- e inter-observador [PSLM17; XRC+17].

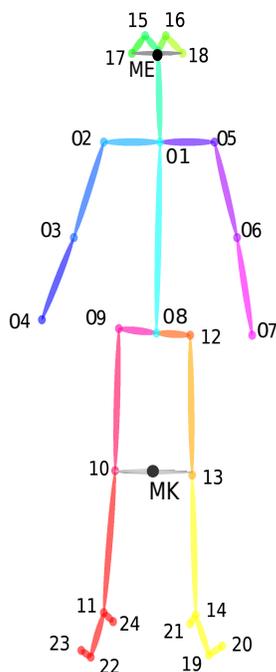


Figura 2.7: Esqueleto y articulaciones del modelo OpenPose 25 (OP25).

Para llevar a cabo el método RULA, el analista debe seleccionar las posturas corporales más críticas que bien se realizan durante una larga duración, o son repetitivas, o representan una desviación relevante de posturas seguras y neutras, o son claramente perjudiciales o peligrosas para el trabajador [CIAD12].

El método propuesto en este trabajo está diseñado para calcular automáticamente las puntuaciones RULA a partir de instantáneas individuales o secuencias de vídeo. La evaluación se basa en los siguientes supuestos: en una única foto, la imagen es representativa de una postura potencialmente crítica, mientras que en una toma de vídeo, los fotogramas se toman a intervalos regulares y son representativos en su conjunto de una única tarea o actividad laboral. Estas condiciones son similares a las especificadas en el estado del arte [CIAD12; Rom14].

2.2.1.2. OpenPose

La estimación de la postura humana mediante CV se define como la tarea de estimar la ubicación de las articulaciones del cuerpo humano en una imagen y, como segundo

paso, la búsqueda de poses específicas que coincidan con las articulaciones observadas dentro de un espacio de poses factibles. De esta forma, se facilita el modelado de la acción humana y su detección, tanto en imagen como vídeo, incluyendo situaciones en las que las articulaciones o las extremidades están parcialmente ocluidas o incluso ausentes en la imagen.

Hoy en día, herramientas de software de código abierto como la red neuronal OpenPose [CSW+21] permiten la detección de articulaciones y extremidades de personas a partir de imágenes y vídeos digitales. Centrado en la estimación simultánea de la postura de múltiples personas, OpenPose toma una imagen como entrada de una CNN de dos ramas, donde la primera rama se encarga de predecir conjuntamente mapas de confianza para la detección de articulaciones del cuerpo, y la segunda rama estima campos de afinidad para la asociación o unión física de dichas articulaciones.

Así, a partir de una imagen digital de entrada, esta red neuronal proporciona una lista de cuerpos humanos detectados, cada uno asociado a un esqueleto tipo con un número de articulaciones definidas previamente. En este trabajo, utilizamos el modelo de esqueleto OpenPose de 25 articulaciones (OP25), las cuales se enumeran según la Figura 2.7. Para cada articulación detectada, el modelo proporciona un vector o tupla que incluye su posición relativa (x, y) dentro de la imagen, y la confianza de estimación de dicha posición, que oscila entre valores de confianza 0 (nula) y 1 (plena) [CSW+21]. A partir de esta información, en este trabajo se ha propuesto un parámetro adicional de confianza global de detección del esqueleto, que se calcula como la media aritmética de las confianzas de estimación de todas las articulaciones consideradas, con el fin de caracterizar más fácilmente detecciones de esqueletos poco realistas.

2.2.1.3. Estimación del ángulo de las articulaciones

Los ángulos de las articulaciones de los cuerpos detectados mediante OpenPose se calculan según las fórmulas de la Tabla 2.1. Dichos ángulos son necesarios para estimar las puntuaciones RULA de forma automática. Por ejemplo, el ángulo del codo izquierdo (EL) se obtiene a partir de las posiciones observadas del hombro, el codo y la muñeca izquierdos,

CAPÍTULO 2. MÉTODOS Y RESULTADOS

Tabla 2.1: Ángulos de articulaciones para RULA a partir de esqueletos OP25.

Nombre del ángulo	Acrónimo ¹	Articulaciones implicadas ²
codo izquierdo	EL	∠04, 03, 02
codo derecho	ER	∠05, 06, 07
hombro izquierdo	SL	∠03, 02, 09
hombro derecho	SR	∠06, 05, 12
hombro izquierdo 2	SL2	∠03, 02, 01
hombro derecho 2	SR2	∠06, 05, 01
rodilla izquierda	KL	∠09, 10, 11
rodilla derecha	KR	∠12, 13, 14
pie izquierdo	FL	∠10, 11, 22
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

correspondientes a las articulaciones del esqueleto OP25 #04, 03 y 02, respectivamente, como se muestra en la Figura 2.7.

Hay dos ángulos específicos de la Tabla 2.1 que utilizan puntos fiduciales (es decir, no pertenecientes al esqueleto base) que se determinan a partir de articulaciones del modelo OP25: "*Mid-ear*" (*ME*) y "*Mid-knee*" (*MK*), que representan puntos a medio camino entre las orejas y las rodillas, respectivamente, tal y como se ilustra en la Figura 2.7. Estos puntos fiduciales permiten calcular dos ángulos que son críticos en la evaluación de RULA: la flexión del cuello (NF) y la flexión del tronco (TF), que suelen ser difíciles o imposibles de calcular de otro modo, incluyendo la observación directa, ya que las articulaciones requeridas para calcularlos suelen estar ocluidas.

¹Acrónimo dado por las respectivas iniciales en inglés

²Las articulaciones son numeradas de acuerdo con la Figura 2.7 excepto por ME y MK (véase Sec. 2.2.1.3).

2.2.1.4. Cálculo de puntuaciones RULA

RULA requiere aplicar umbrales de ángulo de las articulaciones para calcular las puntuaciones de riesgo ergonómico. Estos umbrales son claros para algunas articulaciones en RULA (por ejemplo, codos y rodillas), pero no están explícitamente delimitados para otros. Este último grupo incluye las puntuaciones de abducción de hombros, brazos fuera del cuerpo, torsión de cuello, flexión de cuello, torsión de tronco y ángulos de flexión de tronco [PSLM17; BPS14]. Por lo tanto, para obtener estas puntuaciones RULA de partes del cuerpo o locales, se definen los umbrales recogidos en la tercera columna de la Tabla 2.2.

Estos umbrales específicos, así como las fórmulas establecidas en la Tabla 2.2 sobre los ángulos de la Tabla 2.1, se derivan de múltiples experimentos, como se describe en la Sección 2.2.2. De forma resumida, las fórmulas propuestas tratan de compensar las posibles oclusiones en cada lado del cuerpo, utilizando el promedio de ángulos de las partes del cuerpo izquierda y derecha, puntos medios fiduciales o valores máximos.

Por último, el método propuesto 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. Para proporcionar una puntuación general completa de RULA, los analistas establecen manualmente las puntuaciones de torsión de muñeca y posición, la puntuación de equilibrio de piernas y las puntuaciones de uso de fuerza, siendo esta una aproximación habitual en diversas investigaciones del estado del arte [MVM+12; PSLM17; YYL+19; SFR+20].

2.2.1.5. Métodos de validación

Como se menciona en [CSW+21] y [ZYL18], la existencia de defectos habituales en la captura de imagen, como la iluminación desigual, la variación del punto de vista y

³Abreviaturas de ángulos definidas en Tabla 2.1

CAPÍTULO 2. MÉTODOS Y RESULTADOS

Tabla 2.2: Puntuaciones de RULA obtenidas a partir de los ángulos del esqueleto estimado por OpenPose.

Puntuación RULA	Fórmula ³	Valor
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 del brazo superior	$\max(SL2, SR2)$	+1 ($110^\circ, \infty$)
Posición del brazo inferior	$\overline{[EL, ER]}$	+1 ($60^\circ, 100^\circ$)
		+2 ($-\infty, 60^\circ$)
		+2 ($100^\circ, \infty$)
Brazos 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	$\text{abs}(65^\circ - \overline{[NB, NB2]})$	+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 ($95^\circ, \infty$)

las oclusiones entre varias personas y/o entre personas y objetos, pueden causar pérdidas de información relevantes en el contexto de esta investigación. Para estudiar la aplicabilidad del método propuesto, basado en CV, en función de los defectos anteriores, diseñamos varios experimentos simulados y controlados para cuantificar los efectos de la auto-oclusión, comparar la calidad de la evaluación de cada ángulo articular en tomas simuladas y reales, y de esta forma determinar umbrales de ángulo articular adecuados para RULA. Finalmente, el método se valida en situaciones reales de trabajo al aire libre, bajo la supervisión técnica de siete ergonomistas experimentados, quienes también

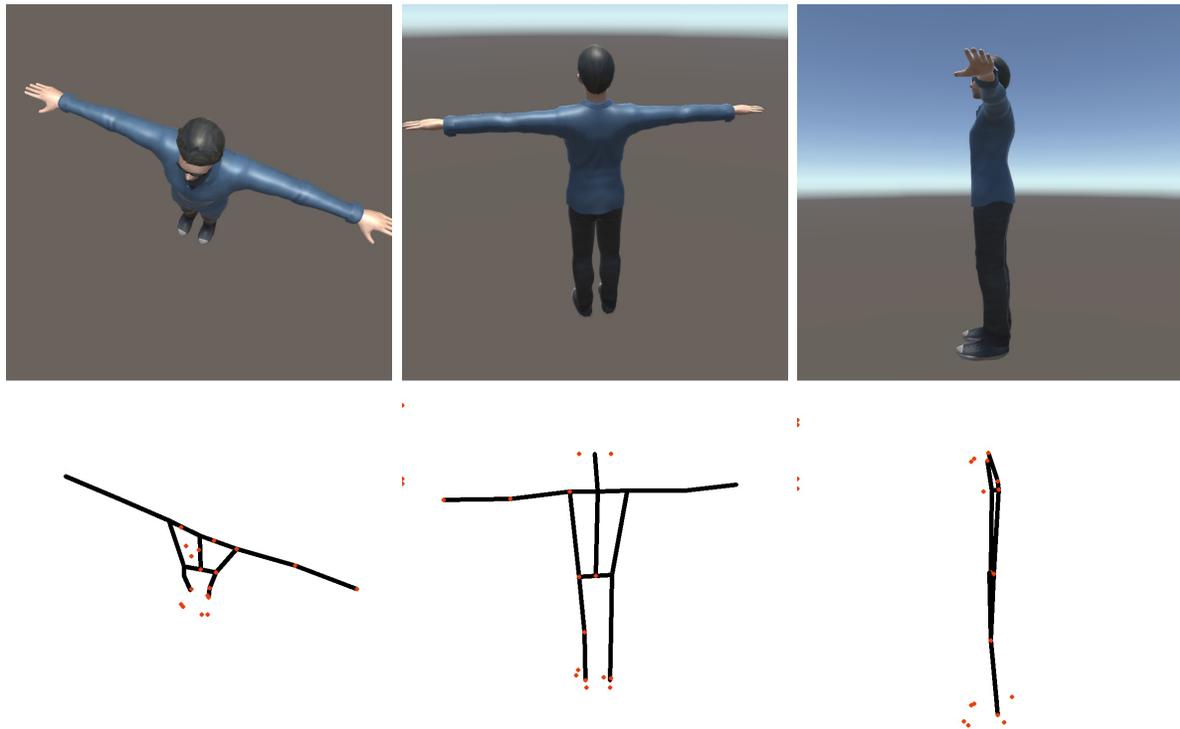


Figura 2.8: Ejemplos de procesado sobre maniquí artificial 3D.

evaluaron las puntuaciones RULA asociadas. Los métodos de validación implicaron los siguientes niveles de comparación:

1. Confianzas de detección de esqueletos y articulaciones dependiendo del punto de vista, como en [PALM15; LHG+18; GHS+15].
2. Concordancia de la puntuación RULA entre el método propuesto y las evaluaciones realizadas por ergónomos experimentados, de acuerdo con [LLGR14; LSLM17; PSLM17].

Como se detalla en los apartados 2.2.2 y 2.2.3, las pruebas iniciales de la metodología propuesta se centraron en las auto-oclusiones, es decir, oclusiones visuales de partes y articulaciones del cuerpo de una persona producidas en ausencia de ningún otro obstáculo, al ser capturado por una cámara desde un punto de vista concreto. Se utilizó un modelo

3D simulado como referencia para las articulaciones y las posiciones del cuerpo y para proporcionar puntos de vista de cámara controlados.

El conjunto de datos generado contiene 300 muestras aleatorias de puntos de vista de cámara siguiendo el método de Hammersley [CF97], que garantiza una distribución angular uniforme de las muestras. En nuestro caso, se eligió un radio de 5m para capturar en primer plano el maniquí artificial 3D desde cada punto de vista posible sobre el suelo, donde dichos puntos de vista confluyen en el centroide del maniquí (ver Figura 2.8), formando así un hemisferio muestral por encima de dicho centroide.

Finalmente, como se detalla en la sección 2.2.4, se validó el método propuesto en cinco vídeos de escenarios de trabajo reales, haciendo uso del κ de Cohen no ponderado, que mide la concordancia de las evaluaciones de riesgo RULA del método propuesto con respecto a las evaluaciones realizadas por siete ergonomistas experimentados. Como se menciona en [LHG+18], dado que RULA se realiza principalmente mediante observación manual, los expertos generalmente analizan solo algunas posturas corporales que consideran críticas. Por esta razón, el procedimiento de validación de la metodología propuesta se realizó con el asesoramiento de expertos, de acuerdo a las metodologías propuestas en [PSLM17] y [LSLM17]. Para cada vídeo, los ergonomistas seleccionaron los fotogramas a analizar según el método descrito en la sección 2.2.1.1, esto es, la selección manual de instantáneas o fotogramas tomados a intervalos regulares. Estos fotogramas fueron evaluados de forma independiente por los especialistas y mediante el método propuesto.

2.2.2. Resultados de pruebas de entorno controlado

El primer experimento se llevó a cabo para estudiar la confianza de detección de las articulaciones, p_k , en función del punto de vista de la cámara. El conjunto de datos sintéticos generado (ver Figura 2.8) fue procesado mediante OP25 para obtener un total de 300 esqueletos según la Figura 2.7. Los dos gráficos de la Figura 2.9 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 acimut (o ángulo de

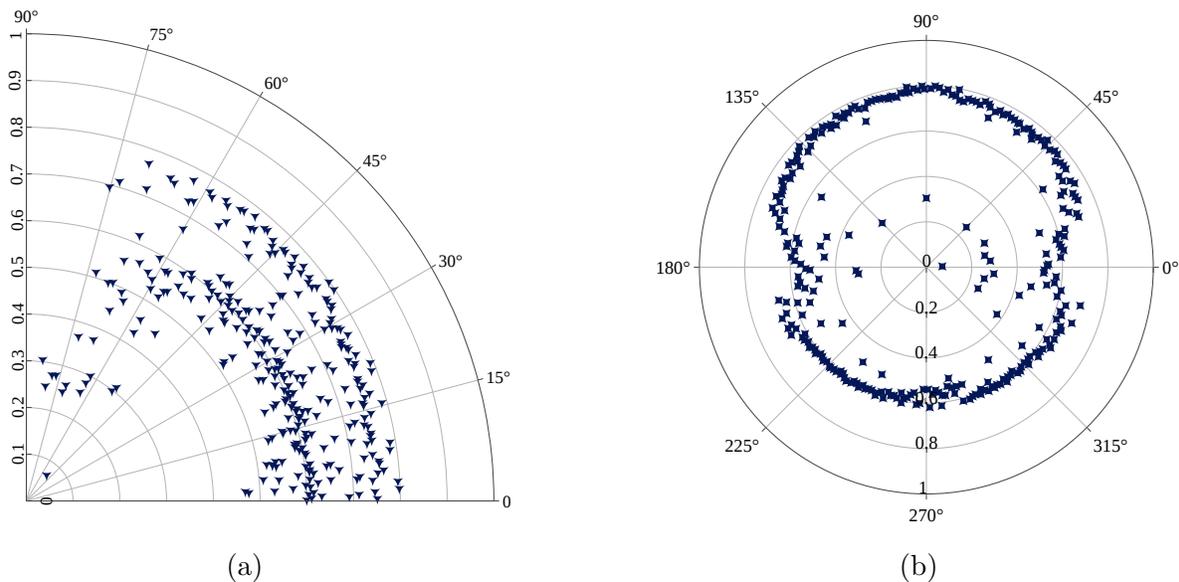


Figura 2.9: Confianza media de detección de esqueleto frente a (a) latitud y (b) acimut del punto de vista de la cámara.

Tabla 2.3: Confianza del esqueleto a partir de los rangos de latitud de los puntos de vista.

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

Tabla 2.4: Confianza del esqueleto a partir de los rangos de acimut de los puntos de vista.

rangos	$30^\circ - 150^\circ$	$210^\circ - 330^\circ$	Other
promedio	0.740	0.589	0.576
desviación estándar	0.088	0.041	0.129
min	0.250	0.374	0.070
max	0.799	0.653	0.750

visión horizontal). Como resultado, las latitudes de la cámara (Figura 2.9a) se agruparon en tres categorías: $0^\circ - 45^\circ$, $45^\circ - 75^\circ$, y $75^\circ - 90^\circ$. Los ángulos de 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 más en la última. Por lo tanto, la latitud de la cámara debe limitarse preferiblemente a 45° , o a lo sumo a menos de 75° , para evitar la pérdida de confianza debido a las auto-occlusiones.

De la misma forma, los ángulos acimutales también se agruparon en tres categorías: $30^\circ - 150^\circ$, $210^\circ - 330^\circ$, y otros ángulos (ver Figura 2.9b). La primera categoría representa los puntos de vista frontales del modelo 3D, donde las articulaciones faciales y las de los pies, tronco y extremidades casi nunca se auto-ocluyen. En la segunda categoría (puntos de vista traseros), la auto-oclusión de los dedos de los pies, la nariz y los ojos reduce el máximo \bar{p} (véase Tabla 2.4). 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 gran simetría de la postura del modelo 3D elegido (Figura 2.8), la auto-oclusión lateral afecta a aproximadamente la mitad de las articulaciones corporales consideradas. Tanto la vista frontal del sujeto como la trasera proporcionan las mejores confianzas de esqueleto requeridas para la evaluación RULA, ya que las articulaciones auto-ocuidas son pequeñas en número (por ejemplo, tobillos en vista frontal) o bien son las menos relevantes para el método (por ejemplo, ojos y nariz en vista trasera). Por lo tanto, el uso de acimuts de vista de cámara frontal ($90^\circ \pm 45^\circ$) o trasero ($270^\circ \pm 45^\circ$) debería proporcionar articulaciones corporales adecuadas para el método propuesto.

2.2.3. Resultados de pruebas para ángulos de hombro y cuello

En este experimento, se probaron varios métodos para calcular las puntuaciones RULA a partir del conjunto de datos simulado utilizado en la sección anterior. El objetivo era encontrar la mejor manera de calcular cada puntuación de RULA. De acuerdo con la Tabla 2.2, dado que tanto la parte superior del brazo como el cuello y el tronco pueden obtener en RULA una puntuación máxima (+4), fueron el objeto principal de estas pruebas. Sin embargo, RULA no incluye una descripción clara de los puntos antropométricos específicamente involucrados en su medición angular [PSLM17].

Para la puntuación del cuello, se consideró el ángulo formado por las articulaciones de la cadera media y el cuello, y tres opciones posibles para el tercer punto: la nariz, el punto medio entre las orejas y el punto medio entre los ojos. Asimismo, se seleccionaron

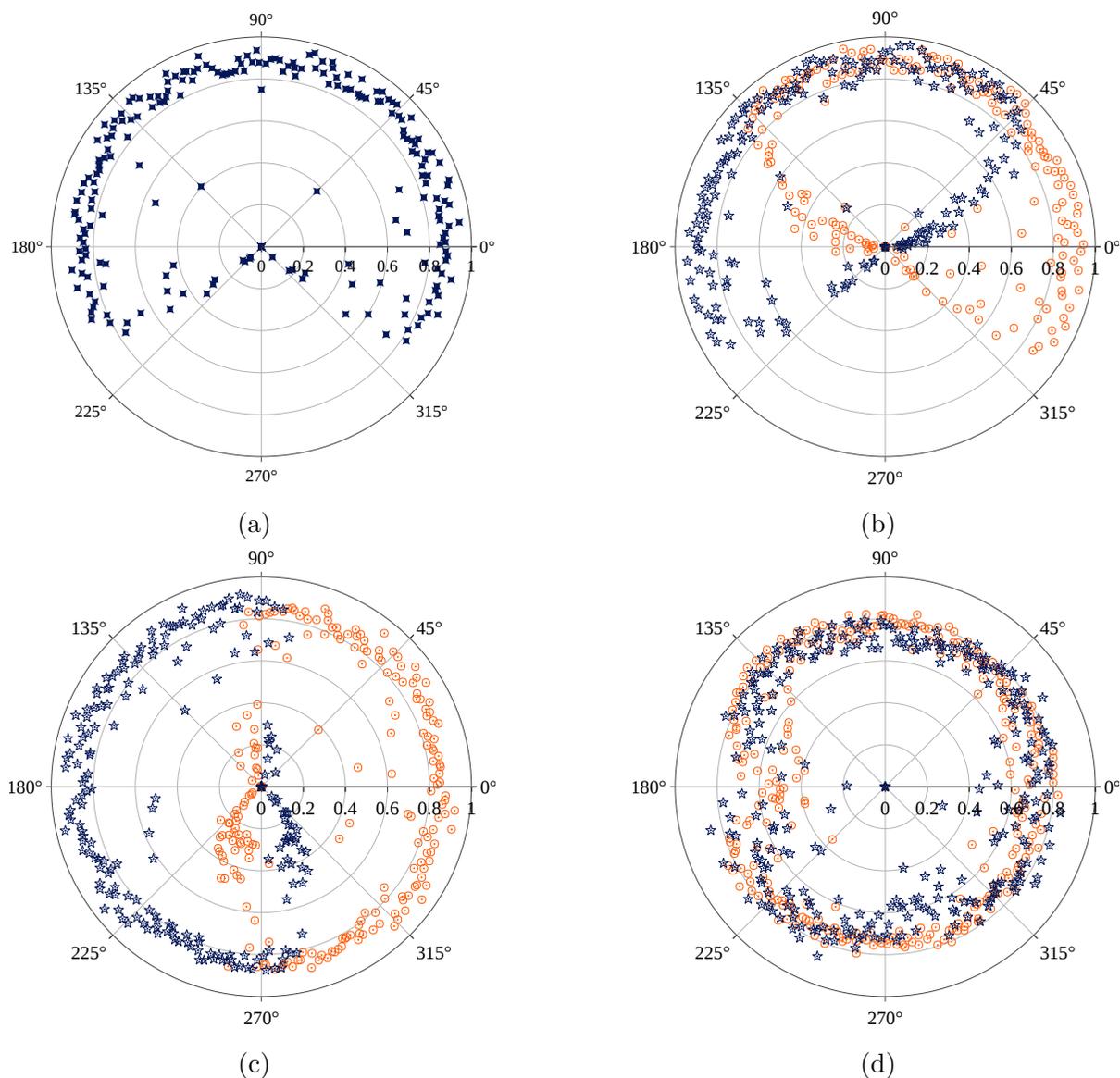


Figura 2.10: Confianza de detección de articulaciones frente a acimut del punto de vista de la cámara: (a) nariz, (b) ojos, (c) oídos y (d) hombros.

nuevamente las articulaciones de la mitad de la cadera y el cuello para calcular la puntuación del tronco, y se consideraron en el análisis los puntos medios entre rodillas, tobillos y talones como candidatos para el tercer punto. Finalmente, las puntuaciones de la parte superior del brazo se evaluaron desde los ángulos formados por las articulaciones de la cadera, el hombro y el codo de cada lado del cuerpo.

La Figura 2.10 ilustra gráficos de confianza articular individual (p_k) con acimut variable en el punto de vista para nariz, ojos, oídos y hombros. Para la nariz, la Figura 2.10a muestra confianzas de detección bastante diferentes (p_0) para las vistas frontal y trasera del sujeto. En particular, el intervalo acimutal ($225^\circ - 315^\circ$) muestra una confianza nula debido a la auto-oclusión producida por la propia cabeza.

La Figura 2.10b combina las confianzas de los ojos izquierdo y derecho (p_{16} y p_{15}). El ojo derecho se detecta correctamente en el intervalo acimutal ($315^\circ - 135^\circ$), 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 completamente con el punto ciego de la nariz.

La detección de confianza combinada de los oídos (p_{18} y p_{17}) se muestra en la Figura 2.10c. En este caso, las gráficas muestran que ambos oídos no pueden detectarse simultáneamente con suficiente confianza desde casi ningún punto de vista acimutal. La ausencia de un intervalo combinado de punto ciego para los oídos es la razón que justifica la definición del punto ME (punto medio entre orejas) para obtener la puntuación del cuello de la Tabla 2.2.

Finalmente, las confianzas obtenidas para la articulación del hombro, p_{02} y p_{05} , muestran puntos ciegos laterales estrechos en acimut, ubicados alrededor de 0° y 180° para el hombro izquierdo y derecho, respectivamente (ver Figura 2.10d). La auto-oclusión entre hombros explica estos puntos ciegos laterales, que también se detectaron en las confianzas de las articulaciones de la rodilla (p_{10} y p_{13}). Como se presenta en la Tabla 2.2, la solución propuesta para superar este problema fue utilizar el promedio de los ángulos del hombro izquierdo y derecho (SL y SR), mientras que para calcular el ángulo de flexión del tronco (TA), se optó por definir el punto fiducial de la rodilla media (MK) a partir de la posición estimada de ambas articulaciones en la imagen.

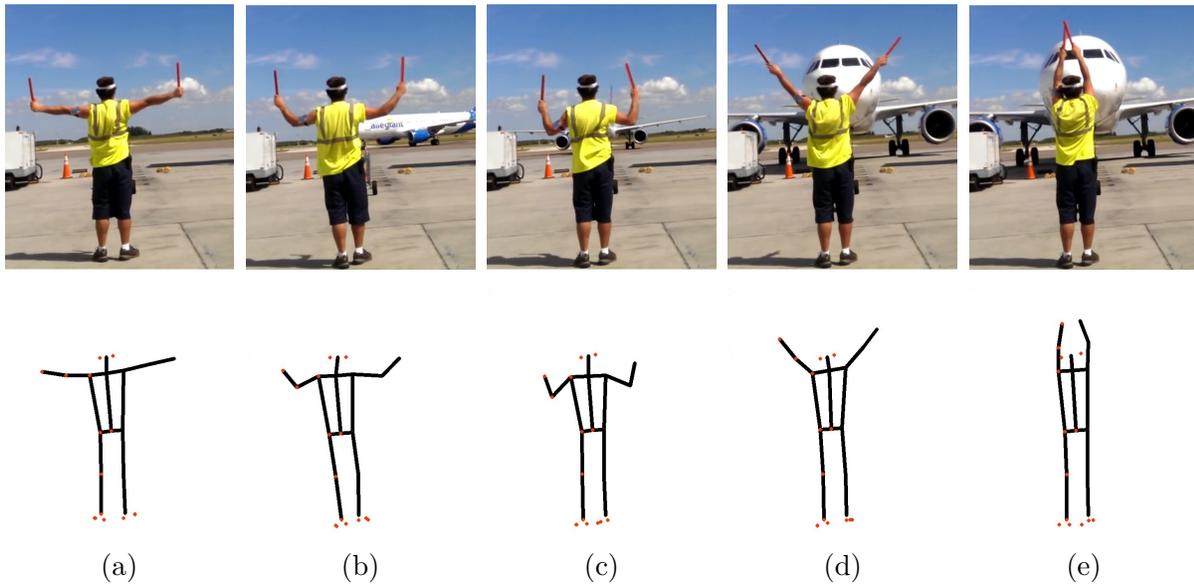


Figura 2.11: Muestras de la secuencia *Señales Marshall*.

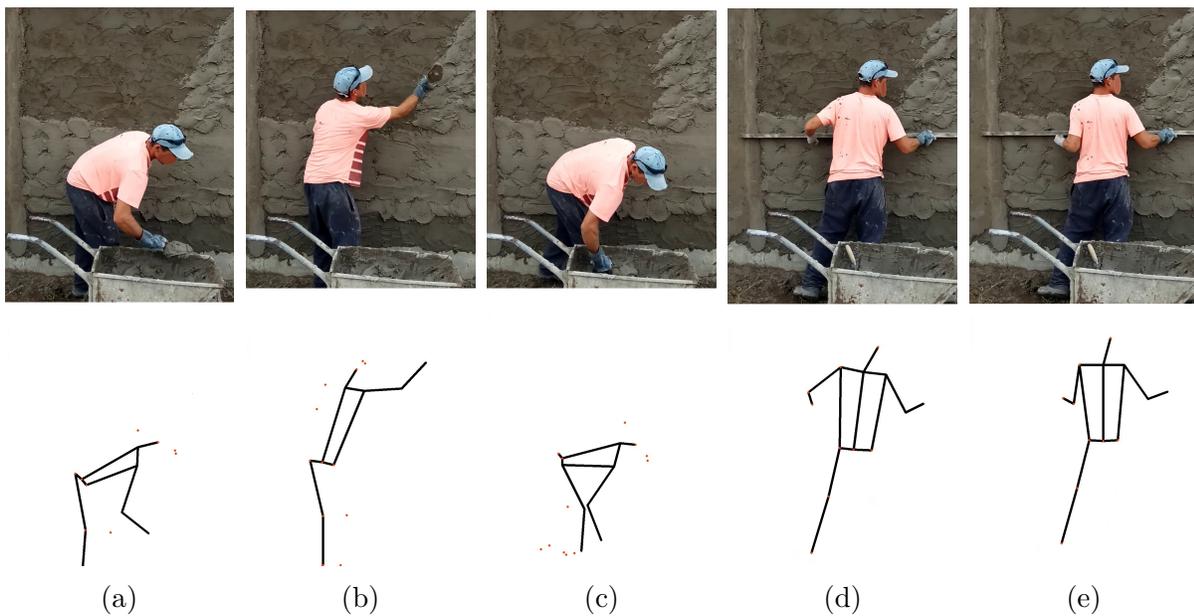


Figura 2.12: Muestras de la secuencia *Revoque de pared*.

2.2.4. Resultados de validación

Los siguientes experimentos tuvieron como objetivo validar el método propuesto mediante el procesado de vídeos de trabajadores realizando su labor en diferentes entornos de trabajo reales al aire libre, de acuerdo a lo indicado en la sección 2.2.1.5.

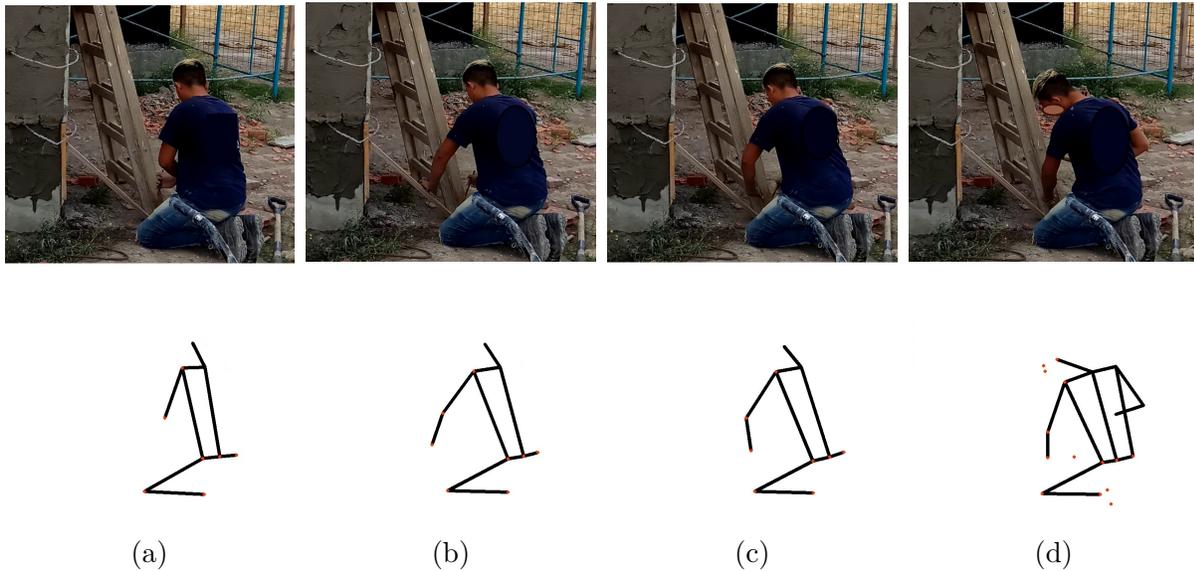


Figura 2.13: Muestras de la secuencia *Martilleado*.

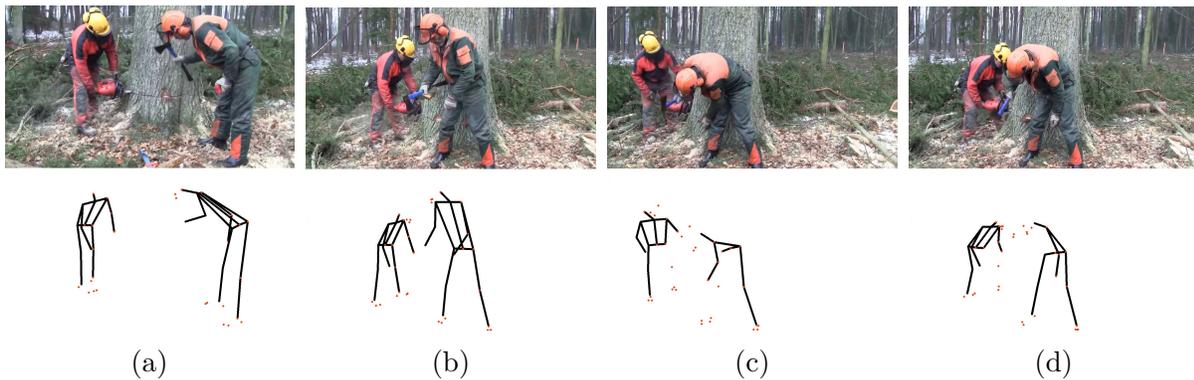


Figura 2.14: Muestras de la secuencia *Corte de árboles*.

Primero, se analizó un vídeo tomado con un *smartphone* de un operador realizando señales Marshall a un avión en pista, desde un punto de vista trasero al trabajador (véase Figura 2.11). En esta secuencia, denominada *Señales Marshall*, se tomaron fotogramas periódicamente para estudiar la precisión de la estimación del método en la evaluación de los ángulos de las extremidades superiores desde un punto de vista trasero, en ausencia de flexiones o torsiones del tronco. Este estudio fue promovido por los ergonomistas, dada la relevancia de estos ángulos en RULA. Además, se llevó a cabo un análisis temporal, que permitió la caracterización estadística (valor medio \pm desviación estándar) del ángulo de cada articulación.

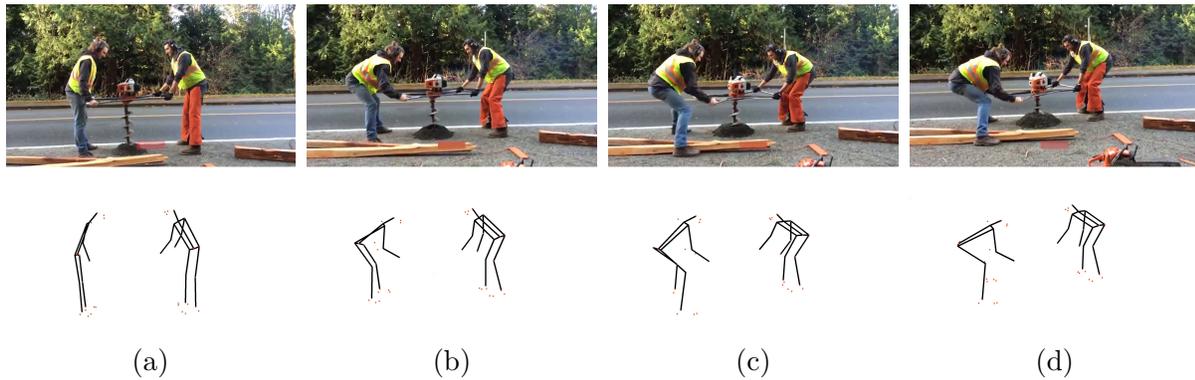
Figura 2.15: Muestras de la secuencia *Perforación*.

Tabla 2.5: Puntuaciones RULA de vídeos de trabajos al aire libre y consenso con expertos.

Secuencia de vídeo ⁴	Tronco	Cuello	Punt. A	Brazo sup.	Brazo inf.	Punt. B	Punt. global	κ
Perforación (1)	5	3	5	3	3	7	7	0,61
Perforación (2)	5	3	5	3	3	7	7	0,61
Corte de árboles (1)	4	4	5	3	2	7	7	1,0
Corte de árboles (2)	4	4	5	3	2	7	7	1,0
Revoque de pared	3	4	5	3	2	7	7	0,66
Señales Marshall	2	1	5	4	3	4	5	0,65
Martilleado	5	4	5	3	3	7	7	0,81
κ	0,44	0,68	0,66	1,0	0,61	0,58	1,0	

Utilizando las fórmulas de la Tabla 2.2, el método propuesto proporcionó ángulos promedio para los hombros (SL, SR) de $83,41^\circ \pm 26,8^\circ$, así como ángulos promedio de los codos (EL, ER) de $120,79^\circ \pm 49,17^\circ$, produciendo así puntuaciones RULA de posición de brazo superior e inferior de +3 y +2, respectivamente, para toda la secuencia.

Como se discutió anteriormente en la sección 2.2.3, el punto de vista trasero es un punto ciego para varios puntos antropométricos de OpenPose. Este fue el caso de esta secuencia, donde incluso la oreja izquierda del trabajador se auto-ocluye en algunos fotogramas (véase Figura 2.11e). Sin embargo, el uso del punto medio entre las orejas (ME) permitió en este caso obtener ángulos de cuello precisos para cada toma, con valores de $6,72^\circ \pm 4,34^\circ$, produciendo así una puntuación de cuello RULA de +1 para la mayor parte de fotogramas de esta secuencia.

A continuación, se aplicó el método propuesto en los vídeos de actividades de trabajo denominados *Revoque de pared* y *Martilleado*, ambos tomados desde teléfonos inteligentes sin trípode. En estos vídeos, los ergonomistas informaron de importantes dificultades en su evaluación, debido a la existencia de oclusiones repetidas y, sobre todo, a la inestabilidad del punto de vista de la cámara. Por ejemplo, en la secuencia de *Revoque de pared* (Figura 2.12), las piernas del albañil siempre quedan ocluidas por el equipo de construcción, mientras que las articulaciones del cuello y la mitad de la cadera se detectan con alta confianza. Sin embargo, la inferencia ocasional de las articulaciones ocluidas de las rodillas, obtenidas por OP25, permitió en este caso al método propuesto utilizar el punto medio entre las rodillas (MK) en dichas imágenes, y calcular así un ángulo de flexión del tronco (TF) promedio de $39,13^\circ \pm 30,03^\circ$ para la secuencia completa. Además, el método proporcionó promedios de ángulo de 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, a criterio de los ergonomistas, en la posición del brazo superior de +2 para la mayor parte de la secuencia, incluso en condiciones de auto-oclusión del brazo izquierdo en algunas tomas (Figuras 2.12a-2.12c).

En la secuencia de *Martilleado*, ilustrado en la Figura 2.13, el lado derecho del operario queda ocluido constantemente. Esta situación puede impedir, en los casos donde el codo derecho no es visible, que el método propuesto pueda calcular los ángulos de hombro (SR) y codo (ER) derechos implicados en la evaluación del brazo superior en RULA. Sin embargo, en el procesado de la secuencia, la articulación del hombro derecho fue detectada por OpenPose con buena confianza. Además, el método propuesto fue ocasionalmente capaz de inferir las ubicaciones del codo y muñeca derechas, lo que permitió el cálculo de la puntuación global RULA en dichos fotogramas. Por ejemplo, en la Figura 2.13d, 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 de RULA utilizando los ángulos SR y ER de la Tabla 2.1. Bajo estas condiciones subóptimas, el método propuesto generó una medición combinada del ángulo del hombro de $36,80^\circ \pm 13,12^\circ$ para esta secuencia de vídeo, produciendo así en la mayoría de los casos una puntuación RULA

de +2 en la posición del brazo superior.

A continuación, se aplicó el método propuesto en los vídeos titulados *Corte de árboles* y *Perforación*, proporcionándose de forma consistente dos esqueletos para cada fotograma de dichas secuencias. Esto permitió calcular simultáneamente las puntuaciones RULA individuales para cada trabajador. En la secuencia *Corte de árboles* (Figura 2.14), dos trabajadores realizan actividades de trabajo coordinadas alrededor de un árbol, mientras que, en el vídeo *Perforación* (Figura 2.15), dos operadores trabajan de manera sincronizada para mantener equilibrado un martillo perforador de obra.

En la secuencia *Corte de árboles*, los ergonomistas seleccionaron por consenso el fotograma clave que se muestra en la Figura 2.14b para analizar los ángulos individuales del tronco de cada trabajador. Esta elección se basó principalmente en la ausencia de auto-oclusiones (Figura 2.14a), oclusiones causadas por herramientas de trabajo (Figura 2.14c) y oclusiones ocasionadas entre trabajadores (Figura 2.14d), proporcionando así mejores condiciones para evaluar las puntuaciones de RULA. El método propuesto calculó ángulos de flexión troncal (TF) de aproximadamente 29° para el trabajador izquierdo y 18° para el derecho. Vale la pena señalar que, en esta secuencia, el método propuesto fue capaz de estimar correctamente la posición de las orejas, a pesar de estar completamente ocluidas por orejeras de seguridad, permitiendo con ello obtener puntuaciones RULA precisas en el cuello (+4) a pesar de esta dificultad.

Finalmente, en la secuencia de *Perforación* (Figura 2.15), los especialistas centraron su interés en examinar las variaciones temporales de la puntuación del tronco a lo largo de la secuencia. El método propuesto midió con precisión suficiente, a criterio de los ergonomistas, los ángulos TF de ambos trabajadores, proporcionando un rango angular promediado entre los dos trabajadores de 14° a 86° . Estos datos fueron de gran valor para los ergonomistas, ya que les permitió detectar rápidamente fotogramas con posturas críticas en la secuencia, coincidentes con el extremo superior de dicho rango.

Una vez realizadas todas las evaluaciones por los expertos, se recogieron los valores mostrados en la Tabla 2.5, que recoge las puntuaciones RULA obtenidas en las cinco secuencias de vídeo anteriores, así como el coeficiente de concordancia (κ) de Cohen entre

CAPÍTULO 2. MÉTODOS Y RESULTADOS

los resultados del método propuesto y las evaluaciones de los expertos, calculados por cada trabajador y diferenciados por cada vídeo estudiado.

2.3. Sistema automatizado para monitorear el uso de equipos de protección personal en la industria de la construcción

Este trabajo presenta un novedoso sistema de CV que genera indicadores automatizados de uso adecuado de equipos de protección personal (EPP) de gran importancia en la industria de la construcción, concretamente el uso de casco de seguridad y de chaleco de alta visibilidad. El sistema se construye sobre una arquitectura de redes neuronales que trabaja sobre imágenes digitales. Primero se utiliza la red neuronal OpenPose para la detección de puntos antropométricos de los trabajadores visualizados, los cuales se utilizan para segmentar automáticamente regiones de interés (ROI) ubicadas en la cabeza y el tronco. Sobre estas ROI, un clasificador neuronal estima la presencia o ausencia de estos dos EPP de interés. Los resultados obtenidos en vídeos tomados en movimiento por drones y *smartphones* muestran que el sistema es plenamente capaz de realizar una evaluación completa de indicadores de utilización de los dos EPP sin asistencia, con el propósito principal de prevenir incidentes potencialmente peligrosos en el lugar de trabajo.

2.3.1. Materiales y métodos

2.3.1.1. Redes neuronales utilizadas

El objetivo de este trabajo fue diseñar y probar una metodología para la detección visual de la correcta utilización de EPP en vídeos capturados con puntos de vista variable, mediante técnicas de CV y ML, para prevenir accidentes laborales.

Inicialmente, cada vídeo es tratado como un vector de imágenes, en donde cada una de ellas se procesa mediante la red neuronal OpenPose. Como se mencionó anteriormente en la sección 2.2.1.2, OpenPose es una CNN multicapa de dos etapas. En la primera etapa, se predicen los mapas de confianza de la detección de puntos antropométricos del cuerpo, mientras que, en la segunda, se asocian dichos puntos para formar posibles esqueletos.

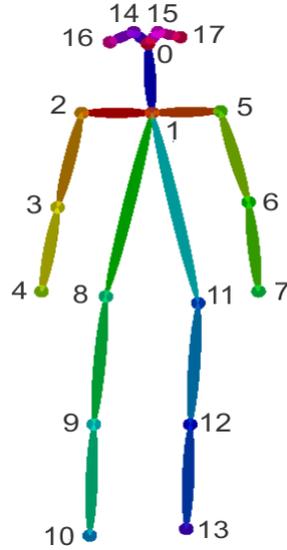


Figura 2.16: Puntos antropométricos de OpenPose18 (OP18) [CSW+21].

En esta ocasión, se seleccionó la versión disponible de OpenPose con menor cantidad de neuronas, OpenPose 18 (OP18), ya que los 18 puntos antropométricos considerados en el modelo incluyen puntos asociados a la cabeza y el tronco suficientes para nuestro propósito. OP18 obtiene a su salida un conjunto variable de tuplas $(E.x, E.y, E.c)$, asociadas a la posición estimada 2D (x, y) y la confianza de la estimación, c , de 18 puntos antropométricos agrupados en esqueletos E según la Figura 2.16.

La confianza de cada punto OP es una cantidad numérica en el rango $[0, 1]$ que depende de aspectos tales como resolución, distorsión proyectiva y oclusiones, debidas estas últimas tanto a la existencia de elementos opacos en la escena, como a la postura relativa del trabajador respecto a la cámara [XLLH18]. Para descartar detecciones de baja confianza, se definieron umbrales de tamaño y confianza mínimos para los esqueletos.

Para cada esqueleto obtenido con tamaño y confianza suficientes, se determina una ROI asociada a la cabeza, y otra asociada al tronco, donde se espera encontrar los EPP de interés. Dichas ROIs se definen a partir de puntos OP relacionados con dichas partes del cuerpo, con índices $j_{ca} = [1, 16, 17]$ para la ROI de la cabeza y $j_{tr} = [1, 2, 5, 8, 11]$ para la ROI del tronco. Esto permite definir, al igual que para los esqueletos, valores de tamaño y confianza mínimos para cada ROI, con la intención de descartar ROIs con contenido

visual de baja confianza.

Finalmente, las ROIs con tamaño y confianza suficientes son procesadas por redes ResNet [HZRS16] re-entrenadas para detectar la presencia de un EPP específico. ResNet [HZRS16] es una CNN que se caracteriza por la inclusión de conexiones residuales, que permiten extender la conectividad directa de una capa a cualquier capa posterior, lo que facilita aumentar el número de capas sin perder capacidad de entrenamiento, lo cual hace que ResNet sea, según el estado del arte, una de las mejores redes de clasificación en términos de precisión frente a coste computacional. En este trabajo, se eligió el modelo ResNet de 18 capas (RN18). Para adaptar esta red a la problemática objetivo, se re-entrenaron las tres últimas capas de cada red RN18 para identificar cada EPP objetivo.

En definitiva, las redes clasificadoras RN18^{ca} y RN18^{tr} asignan respectivamente a cada esqueleto dos etiquetas, *E.Casco* y *E.Chaleco*, ambas con valores (Sí, No). Por último, cada esqueleto procesado se almacena en una base de datos para su posterior tratamiento estadístico.

2.3.1.2. Definición de ROIs y umbrales de confianza

Para el propósito de esta investigación, se definió la ROI del tronco ($E.R^{tr}$) como el rectángulo de área mínima, con lados paralelos a los bordes de la imagen, que contiene todos los puntos antropométricos de esqueleto asociados a la cintura y los hombros (véase Figura 2.16). Por consiguiente, su tamaño puede definirse fácilmente por su altura y anchura, en función de los valores extremos de las coordenadas (x, y) de los puntos antropométricos de interés j_{tr} establecidos en el apartado anterior.

Para definir la región de la cabeza de un esqueleto, los puntos antropométricos de referencia más comunes son la base del cuello y el punto medio entre las orejas [AN00]. Sin embargo, ambas orejas solo son detectables conjuntamente en los planos frontal y posterior, respecto a la cabeza, mientras que en todos los planos laterales solo habrá una oreja visible. Por esta razón, se definió la ROI de la cabeza como un cuadrado con bordes paralelos a la imagen. con centroide en cualquiera de las orejas detectadas, o en el punto medio entre orejas, en el caso de doble detección. El lado de dicho cuadrado se define

como el doble de la distancia existente entre la base del cuello y el mencionado centroide. Para ambas ROIs, su confianza se define como el promedio de la confianza de la detección de los puntos antropométricos involucrados, j_{tr} y j_{ca} .

Según lo establecido en el apartado anterior, los umbrales de tamaño y confianza para el esqueleto, la cabeza y el tronco se utilizan para descartar esqueletos y ROIs objetivo que puedan presentar problemas en la etapa de clasificación. En particular, los umbrales de confianza evitan considerar resultados poco precisos, mientras que los umbrales de tamaño filtran pequeñas ROIs que pueden generar clasificaciones fallidas. En los experimentos del presente trabajo, se utilizó para el esqueleto, la cabeza y el tronco un umbral de confianza de 0,3. El umbral de tamaño mínimo para los esqueletos se estableció en 0,2, aunque, en realidad, es un umbral basado solo en la relación entre las alturas del esqueleto y del fotograma, sin considerar la anchura, dada su mayor variabilidad respecto a la postura. Finalmente, dado que los clasificadores RN18 están diseñados para recibir a su entrada una imagen RGB de 224×224 px, se utilizó como umbral de tamaño mínimo el valor de 112 px, considerando la longitud de la diagonal de las ROIs definidas para la cabeza y el tronco.

2.3.1.3. Dataset y entrenamiento

El *dataset* construido para este trabajo está formado por ROIs extraídas a partir de más de 30 vídeos disponibles en Internet, que recogen diversas actividades reales de trabajadores de la industria de la construcción, con distintas condiciones de iluminación y perspectiva de escena. Para acelerar el proceso de etiquetado, se seleccionaron vídeos de secuencia continua donde todos los trabajadores visualizados pueden clasificarse directamente con las mismas etiquetas de uso de los EPP objetivo en cualquier instante de tiempo.

Posteriormente, en la selección de fotogramas se maximizó la diversidad en las muestras. Así, para las clases de clasificación positiva de *Casco* y *Chaleco*, se incluyeron varios colores y modelos homologados, mientras que en la clase negativa de *Casco* se incluyeron trabajadores sin y con cabello, estos últimos con distintos cortes y colores, y trabajadores

Tabla 2.6: Matriz de confusión para validación de RN18^{ca} (*Casco de seguridad*).

		Real		Total
		<i>v</i>	<i>f</i>	
Predicho	<i>p</i>	<i>vp</i> = 4.546	<i>fp</i> = 12	4.558 Precisión = 0,997
	<i>n</i>	<i>fn</i> = 13	<i>vn</i> = 3.095	3.108 VPN = 0,995
Total		4.559 Sensibilidad = 0,997	3.107 Especificidad = 0,996	7.666 Exactitud = 0,996

con gorras o sombreros distintos de un casco de seguridad (véase Figura 2.17). Finalmente, en la clase negativa de *Chaleco* se incluyeron torsos desnudos y con camisetas y camisas de distintas tramas y colores, diferentes de un chaleco de seguridad homologado, así como variabilidad de género (véase Figura 2.18). En definitiva, el *dataset* generado en este trabajo consta de un total de 23.303 subimágenes de cabezas y 21.378 subimágenes de torsos, todas ellas obtenidas a partir de esqueletos asociados a fotogramas de vídeo con trabajadores de la industria de la construcción.

Para el re-entrenamiento del clasificador RN18^{ca}, se realizó una partición aleatoria con 15.367 subimágenes $E.R^{ca}$ para re-entrenamiento y 7.666 para validación, mientras que para el clasificador RN18^{tr}, se eligieron de igual forma 14.264 subimágenes $E.R^{tr}$ para re-entrenamiento y 7.114 para validación. Finalmente, ambos conjuntos de re-entrenamiento se dividieron en un 80 %-20 % para entrenamiento y testeo, respectivamente, proporcionando los resultados recogidos en las matrices de confusión de las Tablas 2.6 y 2.7, que contienen las definiciones usuales de parámetros de rendimiento de clasificación: sensibilidad = $vp/(vp + fn)$, precisión = $vp/(vp + fp)$, especificidad = $vn/(vn + fp)$, valor predictivo negativo (VPN) = $vn/(vn + fn)$ y exactitud = $(vp + vn)/(vp + vn + fp + fn)$.

El entrenamiento de las CNNs RN18 se realizó en la plataforma de computación en la nube Google Colab utilizando el paquete Fast.ai, mientras que OP18 se instaló de forma local en Pytorch, en un ordenador con procesador Intel Core i7 con 16 GB de RAM y tarjeta gráfica NVIDIA GeForce GTX 1080 TI. El software de evaluación se diseñó para ejecución local sobre GPU, en lenguaje Python sobre Ubuntu 16, de las tres CNN utilizadas, alcanzando una velocidad media de procesamiento de 3,57 fps.

Tabla 2.7: Matriz de confusión para validación de RN18^{tr} (*Chaleco de alta visibilidad*).

		Real		Total
		<i>v</i>	<i>f</i>	
Predicho	<i>p</i>	$vp = 2.754$	$fp = 22$	2.776 Precisión = 0,992
	<i>n</i>	$fn = 18$	$vn = 4.320$	4.338 VPN = 0,996
Total		2.772 Sensibilidad = 0,994	4.342 Especificidad = 0,995	7.114 Exactitud = 0,994

Figura 2.17: *Dataset E.Rca* con etiquetas de *Casco* (a)-(f) No y (g)-(l) Sí.



Figura 2.18: *Dataset E.R^{tr}* con etiquetas de *Chaleco* (a)-(f) No y (g)-(l) Sí.

2.3.2. Resultados experimentales

Para la verificación de la metodología propuesta, se escogieron tres vídeos adicionales, distintos de los utilizados para la generación automática del *dataset*, de labores típicas del sector de la construcción, grabados con diferentes condiciones técnicas: (1) cámara de dron volando a baja altura (Figura 2.19), (2) cámara egocéntrica fijada en la cabeza de un supervisor al atardecer (Figura 2.20), y (3) teléfono móvil con soporte manual (Figura 2.21). El vídeo (1) contiene una gran variedad de puntos de vista de un conjunto de 10 obreros trabajando en una obra de cimentación. Además, los fotogramas de las



(a)



(b)



(c)



(d)



(e)



(f)

Figura 2.19: Resultados del vídeo de evaluación 1 (dron).

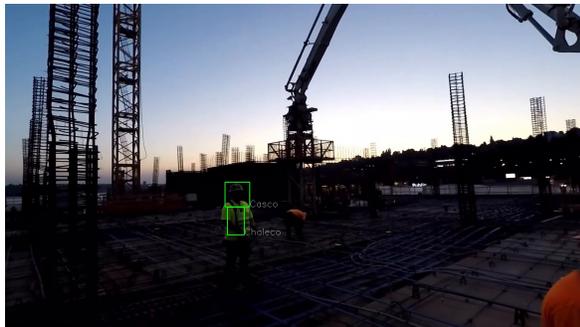
CAPÍTULO 2. MÉTODOS Y RESULTADOS



(a)



(b)



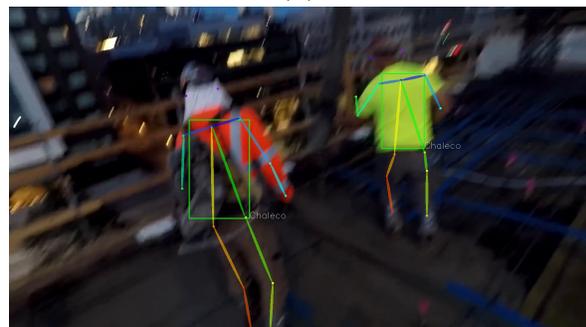
(c)



(d)

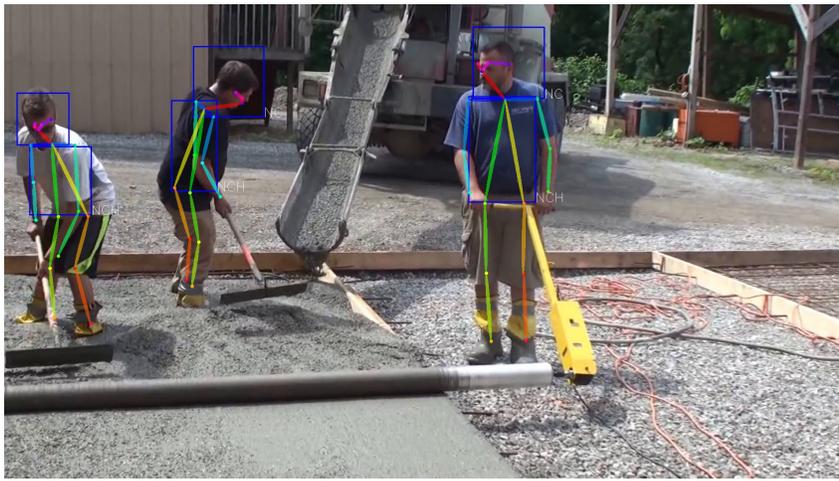


(e)

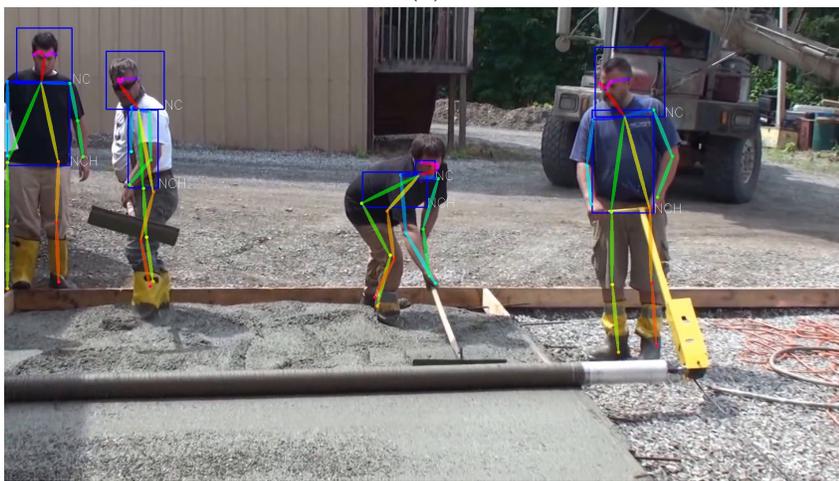


(f)

Figura 2.20: Resultados del vídeo de evaluación 2 (egocéntrica).



(a)



(b)



(c)

Figura 2.21: Resultados del vídeo de evaluación 3 (*smartphone*).

Figuras 2.19a, 2.19b y 2.19d presentan considerables oclusiones, debidas principalmente a la perspectiva de captura y a la presencia de numerosas barras de hierro, propias de un entorno de trabajo de cimentación.

La metodología propuesta permite utilizar umbrales de confianza para excluir los esqueletos de los trabajadores afectados por una oclusión parcial. Con un umbral característico de $u_{Conf}^E = 0,3$, los esqueletos de los trabajadores con alrededor de la mitad del cuerpo ocluido pueden ser descartados. Del mismo modo, el efecto de las distancias u oclusiones parciales en un objetivo puede ser controlado por su umbral de tamaño ($u_{Tam}^E = 0,2$).

OP18 se diseñó para estimar puntos antropométricos ocluidos solo a partir de otros que sí son visibles y poseen un mayor coeficiente de confianza. Esta característica de estimación permite a la metodología propuesta inferir las ROIs adecuadas para llevar a cabo la clasificación de uso de los EPPs con oclusiones parciales. Por ejemplo, un $u_{Conf}^{ca} > 0,3$ permite descartar el uso del casco del trabajador E1 en la Figura 2.19b, mientras que la inferencia obtenida con el mismo u_{Conf}^{ca} en su oreja y hombro derecho en la Figura 2.19e permite obtener un $E.R^{ca}$ adecuado para clasificar con éxito el uso del casco. Sin embargo, este valor excluye del diagnóstico a los trabajadores E2, E3 y E5 de esta figura.

La importancia de descartar las detecciones con el umbral de tamaño queda clarificada en la Figura 2.19f, donde se descartan dos falsas detecciones (D1 y D2). Sin embargo, el método falla en situaciones como la mostrada en la Figura 2.19c, donde el esqueleto del trabajador E7 es detectado de forma errónea. De forma más resumida, en la Figura 2.19a, se detectan de forma correcta el chaleco de los esqueletos E2, E3, E7 y E10, y también el casco, que no es visible en el caso de E10. En la Figura 2.19b, solo se detectan correctamente E1 (solo chaleco) y E2 (ambos EPP). En la Figura 2.19c, se detecta de forma correcta los chalecos de E7 y E1 y de forma errónea (por ubicación) el casco de E7. En la Figura 2.19d, se detectan correctamente ambos EPP en los esqueletos E1 y E5; en la Figura 2.19e, el E1 y, por último, en la Figura 2.19f, los esqueletos E1, E2, E4 y E5, donde también se descartan las detecciones erróneas D1 y D2.

Las condiciones de captura del segundo vídeo generan muchos fotogramas borrosos,

debido al movimiento de la cámara y a la pobre iluminación de la escena, que además se visualiza a una distancia considerable de la cámara (Figuras 2.20e y 2.20f). Sin embargo, la metodología propuesta es capaz de recabar información de utilidad al trabajar solamente con esqueletos y ROIs que poseen mayor confianza. Además, en este vídeo se presentan otras dificultades, como el cambio gradual de iluminación (Figura 2.20f), las oclusiones inherentes al tipo de trabajo (Figuras 2.20b y 2.20d)) y la diversidad de posturas y tamaños de los obreros (Figura 2.20c).

En las Figuras 2.20d y 2.20e se ejemplifican, en la clasificación del uso de casco, las dos dificultades más comunes halladas en la metodología propuesta. En este caso, se relacionan con la anormal utilización de una prenda extra en la cabeza, sumado al ya mencionado aspecto borroso de la captura. Sin embargo, en la Figura 2.20a, el mismo casco del trabajador situado en primer plano se clasifica de forma correcta. Finalmente, en la Figura 2.21, se presenta una secuencia de trabajo en grupo, en la que ningún trabajador utiliza los EPP objeto de este estudio. El método propuesto logra aquí discernir correctamente el no uso de EPP en trabajadores con posturas variables, ocluidas, y con distinta indumentaria laboral. Ante todos estos casos, la metodología se muestra robusta, aportando resultados fiables de muy alta precisión (por encima del 99 %).

2.4. Estimación de ángulos de articulaciones mediante VIBE en avatares virtuales

En este trabajo, se estudia una metodología alternativa a OpenPose y RULA (véase sección 2.2) para la evaluación automática de riesgos ergonómicos mediante CV y ML. Se trata de una metodología completa para la medición ergonómica del cuerpo basada en redes neuronales de reciente publicación. Como se muestra en la Figura 2.22, el método propuesto comienza con la red neuronal STAF, diseñada para la detección y seguimiento de esqueletos asociados a personas a lo largo de secuencias de vídeo. Estos resultados se envían posteriormente a VIBE (otra red neuronal de reciente aparición) para su posterior virtualización mediante avatares 3D. Por último, se evalúan los riesgos ergonómicos utilizando la herramienta REBA.

2.4.1. Materiales y métodos

2.4.1.1. STAF: campos de afinidad espacio-temporales

STAF [RIHS19] es una reciente red neuronal de código abierto para detectar y rastrear simultáneamente las poses 2D de múltiples personas en secuencias de vídeo, en contraste con desarrollos previos, como OpenPose, que trabajan solo a nivel de fotograma. STAF propone una topología temporal basada en la detección y seguimiento de las articulaciones y esqueletos de las personas detectadas en la secuencia de vídeo, capaz de trabajar con una amplia variedad de movimientos corporales. La red STAF utiliza los denominados campos de afinidad entre detecciones de articulaciones corporales en fotogramas anteriores, para estimar su ubicación en el fotograma actual.

2.4.1.2. VIBE: inferencia de vídeo para la estimación de la postura y la forma del cuerpo humano

VIBE [KAB20] es una red neuronal generativa antagónica o GAN (del inglés, *Generative Adversarial Network*), entrenada para aprender movimientos naturales del cuerpo

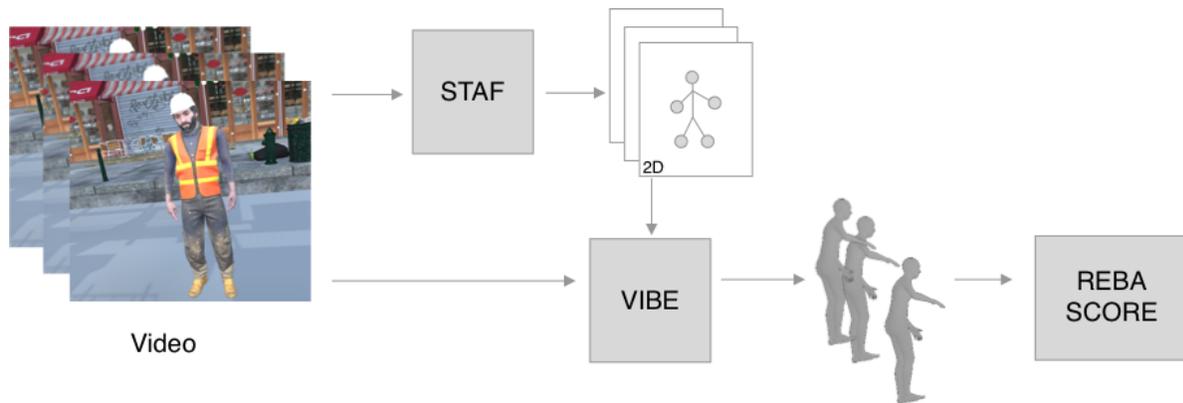


Figura 2.22: Metodología propuesta.

humano, distinguiendo los movimientos posibles entre una gama de movimientos estimados por redes temporales de regresión de postura y forma. En relación con el estado del arte, VIBE propone una novedosa arquitectura de red temporal con un mecanismo de auto-atención. Como resultado, el modelo propone una GAN entrenada que reproduce secuencias de movimiento humano verosímiles no solo en cuanto a la postura, sino también en cuanto a su cinemática.

2.4.1.3. REBA: evaluación ergonómica rápida del cuerpo humano completo

Surgido como una evolución de RULA, el método REBA [MH00] es un método de lápiz y papel destinado a generar un protocolo de examen postural sensible a los riesgos musculoesqueléticos en diversas tareas. Al igual que RULA, el método se basa en el estudio ergonómico individualizado de distintas partes del cuerpo humano, en cuanto a los ángulos observados en las distintas articulaciones relacionadas y sus planos naturales de movimiento. Dichas evaluaciones individuales se identifican con niveles de riesgo postural, los cuales se agregan de forma jerárquica hasta obtener una puntuación global (*REBA score*), que indica la necesidad y/o urgencia con que se deberían aplicar acciones correctivas. De esta manera, REBA proporciona un sistema de puntuación para la actividad muscular causada por posturas estáticas, dinámicas, repetitivas o inestables [MH00].

Si bien es cierto que RULA evalúa todo el cuerpo humano, su puntuación del riesgo musculoesquelético se centra en las extremidades superiores. En cambio, el método REBA

realiza una evaluación generalizada del riesgo del cuerpo humano, permitiendo realizar una evaluación más exhaustiva de las extremidades inferiores, estimando diferentes niveles de riesgo postural, en función del grado de flexión de las rodillas. Igualmente, permite evaluar posturas adicionales respecto al tronco, incluyendo su extensión. Sin embargo, su forma de aplicación es similar a RULA, ya que se basa en posturas individuales, no en una secuencia completa. Por ello, el ergonomista debe seleccionar previamente las posturas críticas, o bien realizar un análisis periódico de la postura del trabajador en fotogramas consecutivos de una misma secuencia temporal (ver sección 2.2.1.1).

En este trabajo, el estudio se enfocó en las desviaciones angulares de las articulaciones más importantes analizadas en REBA: codos, hombros, rodillas, tronco y cuello.

2.4.1.4. Experimentación

La motivación de este trabajo, realizado con posterioridad al de la sección 2.2, fue evaluar tecnologías alternativas de más reciente publicación, dentro del marco de la tesis (Visión por computador y Aprendizaje automático), aplicadas a la evaluación automatizada de riesgos ergonómicos de personas trabajadoras. Como ya se ha mencionado anteriormente, estas tecnologías de ERL son propensas a errores de observación, que surgen principalmente como consecuencia de las auto-oclusiones que se producen de forma natural al observar la postura de una persona. Normalmente, esto obliga al analista a capturar y analizar diferentes puntos de vista de la misma postura de interés realizada por el trabajador. La importancia real de estos errores está fuertemente influenciada por las actividades del trabajador y de la posición relativa de la cámara, así como por la capacidad de realizar múltiples evaluaciones de la misma secuencia.

Por tanto, el objetivo principal de este estudio fue evaluar la precisión de la estimación de los ángulos del cuerpo en una amplia variedad de configuraciones de adquisición con el uso de estas técnicas alternativas. Siguiendo el procedimiento propuesto por [PALM15], utilizado también en el estudio de la sección 2.2, se utilizó un avatar virtual para la recolección de datos, empleando en este caso un modelo 3D sintético en el motor de juegos Unity, que realiza movimientos corporales específicos (asignados como estándar), y



Figura 2.23: Entorno de simulación con trabajador virtual 3D. Cada columna muestra distintos puntos de vista de la misma postura.

con posiciones de cámara aleatorias, con el fin comparar las métricas adquiridas y evaluar el efecto de la auto-oclusión y la posición de la cámara en la calidad de la adquisición (ver Figura 2.23). En este sentido, el *dataset* de simulación obtenido es una clara mejora en cuanto a capacidades, respecto al *dataset* correspondiente desarrollado en el trabajo de la sección 2.2 (véase Figura 2.8). Esta mejora se puede caracterizar en tres aspectos:

- Postura variable, frente a postura única, en específico el maniquí virtual realiza un movimiento similar a las sentadillas para realizar un levantamiento.
- Nivel de detalle del escenario, que en este caso representa un entorno laboral urbano, con múltiples objetos en la escena.
- Nivel de realismo del maniquí 3D, en concreto su capacidad de movimientos y la indumentaria laboral, que en este caso conlleva el uso del EPP propio de un trabajador de la industria de la construcción.

Para ello, se desarrolló una herramienta *software* específica conectada a la plataforma Unity, para registrar las articulaciones reales del cuerpo en el sistema de coordenadas 3D del entorno virtual elegido. Las entradas del *software* son un modelo de malla poligonal (esqueleto del trabajador) y una animación de dicho esqueleto, en este caso, un movimiento de levantamiento desde el suelo. El mencionado esqueleto utilizado en Unity se corresponde con el obtenido a partir de VIBE. Además, el usuario puede escoger las articulaciones de interés antes de la ejecución. Con estas entradas, el programa desarrollado realiza los siguientes procesos:

1. Establecer una distancia aleatoria de captura R , entre 2 y 5 m, desde el centroide del maniquí hasta la cámara virtual.
2. Definir un conjunto aleatorio de posiciones de cámara, que apuntan al maniquí desde una semiesfera de radio R centrada en su centroide. Se eligió utilizar para este fin la distribución uniforme de Hammersley [CF97].

3. Generar cuadro a cuadro la animación del esqueleto asociado al maniquí 3D, en función de la velocidad de reproducción de vídeo (en fps), guardando cada fotograma renderizado como imagen, en cada una de las posiciones de cámara.
4. Registrar en cada instante de tiempo la posición 3D de las articulaciones seleccionadas por el usuario, en coordenadas propias del entorno de simulación 3D. Dichas coordenadas se almacenan en un archivo de texto tipo CSV.

Mediante el procedimiento anterior, se produjeron un total de 1073 vídeos del entorno de simulación (con 52 fotogramas cada uno) para todas las articulaciones de interés en REBA, acompañados de las mediciones angulares correspondientes en formato CSV. El dataset se encuentra disponible en la siguiente dirección: <https://shorturl.at/jow68>.

Para determinar el error del método propuesto, cada esqueleto coordinado de referencia, asociado a un tiempo de secuencia, podría compararse directamente con los datos posicionales correspondientes del esqueleto estimado por la etapa final de la metodología propuesta, es decir, por la red neuronal VIBE. Sin embargo, dado que la entrada del método de evaluación de riesgos elegida opera con ángulos de entrada, se eligió como métrica de error la diferencia en valor absoluto entre la estimación VIBE de cada ángulo de articulación considerada, con el ángulo de referencia correspondiente (calculado a partir de las articulaciones del esqueleto renderizadas en el entorno de simulación).

La metodología propuesta fue desarrollada utilizando Python 2.7 y los paquetes TensorFlow, NumPy, PIMS, Pandas, OpenCV, STAF [RIHS19] y VIBE [KAB20]. El código resultante se ejecutó íntegramente sobre Ubuntu 18 en un ordenador tipo PC equipado con tarjeta gráfica NVIDIA GeForce GTX 1080 TI y procesador Intel Core i7 con 16 GB de RAM.

2.4.2. Resultados

En esta sección, se presentan los resultados de evaluación del error absoluto entre el ángulo real de la articulación en el maniquí simulado 3D y el calculado por la metodología

propuesta, como alternativa al análisis ergonómico realizado por expertos, de acuerdo a lo especificado en la sección 2.4.1.4.

El algoritmo se configuró para calcular todos los ángulos de las articulaciones disponibles. Es importante señalar que, en los casos en los que VIBE no pudo detectar en este estudio ningún esqueleto en una imagen renderizada desde un punto de vista concreto, se decidió por asignar todos los ángulos de las articulaciones bajo estudio al valor máximo de error (180°). En el estado del arte de los métodos de observación para la evaluación ergonómica, la práctica demuestra que aproximadamente el 30 % de las evaluaciones realizadas por los profesionales presentan errores [DAP17].

Para mostrar los resultados obtenidos, se generaron diagramas de dispersión polares respecto al acimut y a la elevación (espejada) asociadas al punto de vista de renderización de la imagen de referencia, donde cada punto del diagrama representa el error absoluto cometido en su estimación por VIBE mediante un mapa de color. Este mapa de color representa, de forma aproximada, errores absolutos de estimación de ángulos menores de 30° con color verde, menores de 40° con color naranja, menores de 50° con color rojo y mayores de 90° con color negro.

Estos diagramas se representan en figuras individuales para los ángulos articulares del tronco (Figura 2.24) y el cuello (Figura 2.26) y conjunta para ambos codos (Figura 2.25), hombros (Figura 2.27) y rodillas (Figura 2.28). De esta forma, puede observarse a simple vista que, en general, el método propuesto no genera errores de importancia en la estimación de los ángulos asociados al tronco, cuello y ambos hombros, mientras que su comportamiento empeora en la estimación de los ángulos de las rodillas y, especialmente, en los ángulos de los codos.

Más en detalle, los errores angulares de tronco y cuello presentan valores similares en promedio, de $34,93^\circ$ y $33,38^\circ$, respectivamente. Al igual que ocurre con el resto de ángulos bajo estudio, los errores aumentan cuando la cámara se aproxima a un punto de vista cenital. Sin embargo, los errores son razonablemente pequeños en el resto de puntos de vista.

Por otra parte, como se puede apreciar en la Figura 2.25a, el ángulo del codo izquierdo

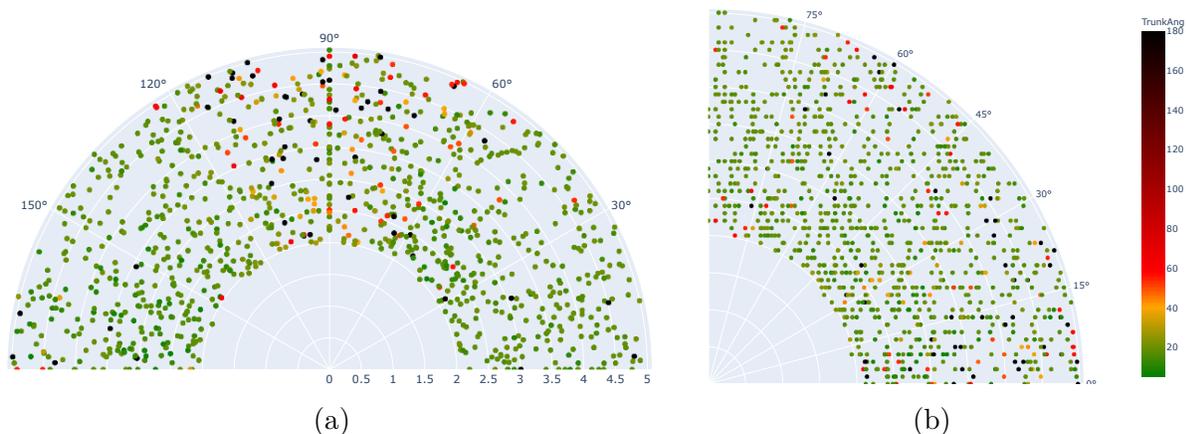


Figura 2.24: Error de estimación del ángulo del **tronco** frente a distancia de captura y (a) acimut y (b) elevación del punto de vista de la cámara virtual.

es el que produjo el peor error: $74,42^\circ \pm 24,86^\circ$ (valor medio \pm desviación estándar). Este resultado se puede atribuir al hecho de que esta articulación, en la actividad simulada (levantamiento desde el suelo), es la que posee una mayor cantidad de desplazamiento posicional relativo al esqueleto, seguida muy de cerca por el otro codo. Así, se justifica también el error del codo derecho: $43,25^\circ \pm 31,72^\circ$ (Figura 2.25c).

En cuanto a los hombros (Figura 2.27), los errores medidos fueron significativamente menores que en los codos: $23,75^\circ \pm 21,43^\circ$ (izquierdo) y $34,92^\circ \pm 35,27^\circ$ (derecho).

Por último, las rodillas (Figura 2.28) generaron errores angulares bastante bajos, de $25,00^\circ \pm 38,12^\circ$ (izquierda) y $27,31^\circ \pm 35,72^\circ$ (derecha). Sin embargo, se percibe un área problemática de puntos de captura, próximo a 90° en acimut y a 30° en elevación para ambas rodillas. Esto se justifica, en este caso, al tipo de actividad, donde se ocluyen temporalmente las rodillas a medida que el trabajador eleva los brazos y el tronco desde el suelo. Además, es la parte inicial del movimiento donde la auto-oclusión resulta todavía mayor, ya que las rodillas son ocluidas también por la cabeza y el tronco del trabajador.

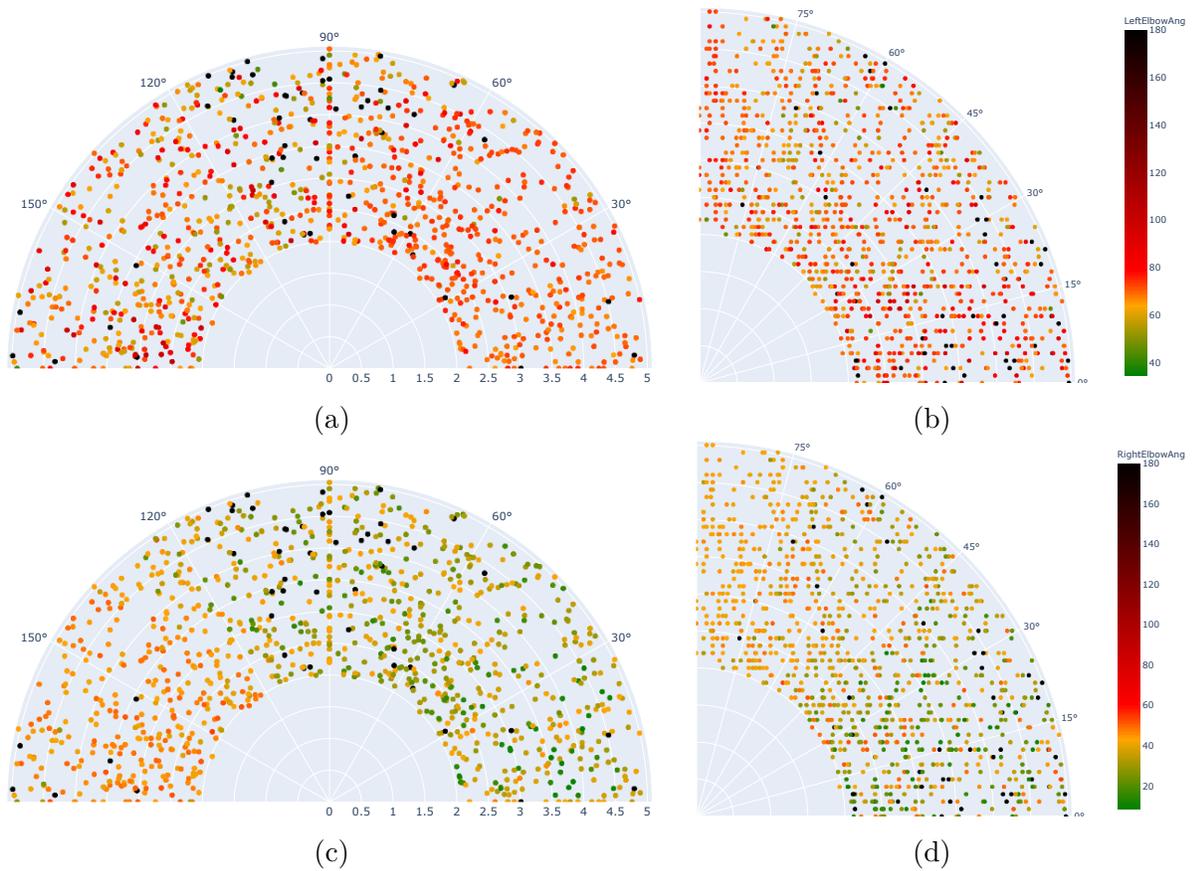


Figura 2.25: Error de estimación de los ángulos de los **codos** izquierdo y derecho, frente a distancia de captura y (a,c) acimut y (b,d) elevación del punto de vista de la cámara virtual.

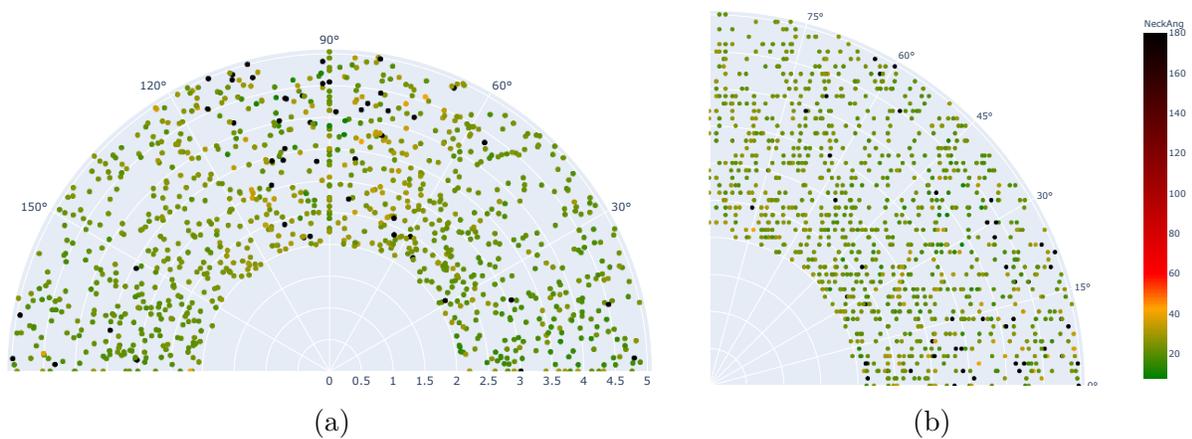


Figura 2.26: Error de estimación del ángulo del **cuello** frente a distancia de captura y (a) acimut y (b) elevación del punto de vista de la cámara virtual.

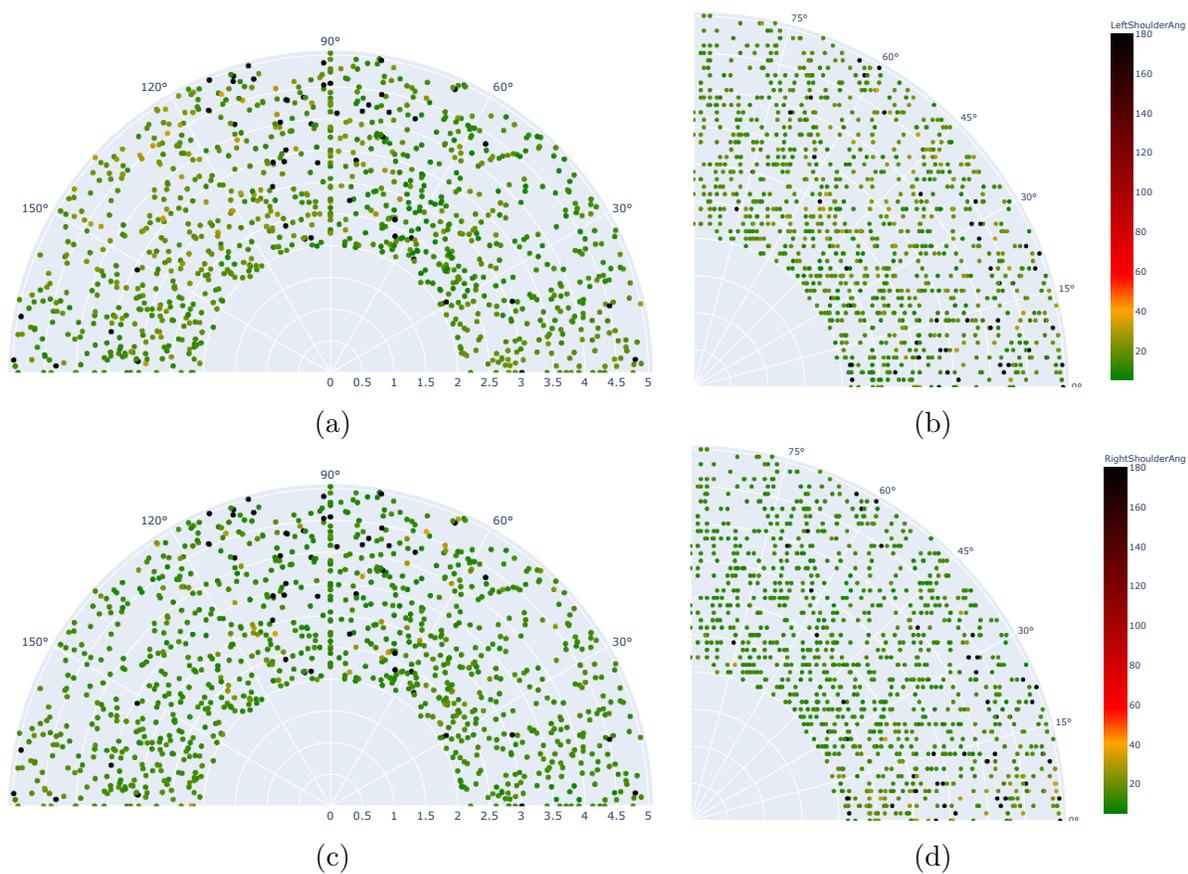


Figura 2.27: Error de estimación de los ángulos de los **hombros** izquierdo y derecho, frente a distancia de captura y (a,c) acimut y (b,d) elevación del punto de vista de la cámara virtual.

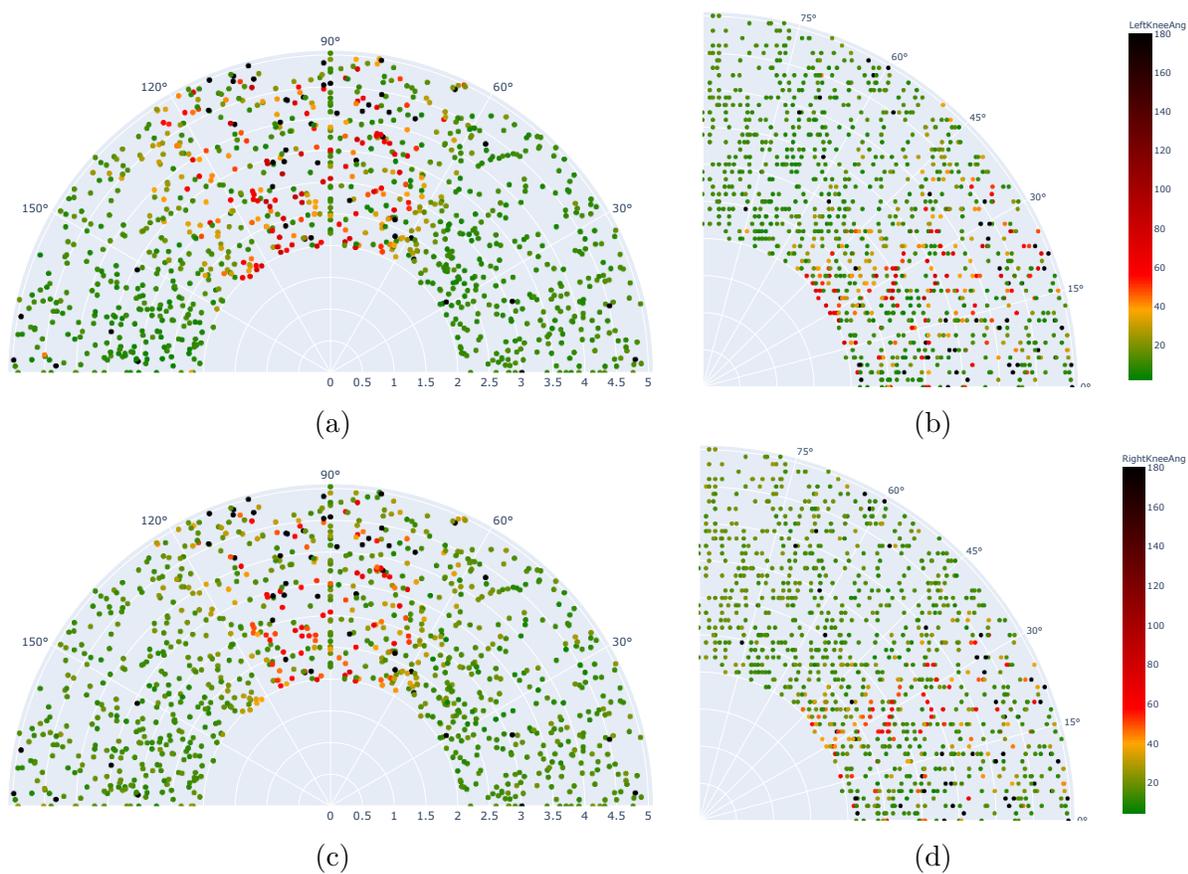


Figura 2.28: Error de estimación de ángulos de las **rodillas** izquierda y derecha, frente a distancia de captura y (a,c) acimut y (b,d) elevación del punto de vista de la cámara virtual.

Capítulo 3

Discusión

A continuación, se plantea la discusión sobre los trabajos analizados en el Capítulo 2, agrupados por temáticas, de tal forma que primero se debatirán los resultados obtenidos en las publicaciones relacionadas con la medición del riesgo musculoesquelético (AR1 y CC2) y, en una segunda instancia, se analizarán los estudios relacionados con la detección automatizada de EPP (CC1 y AR2) (véase Figura 2.1).

3.1. Medición automatizada del riesgo de trastornos musculoesqueléticos asistido mediante CV

Se proponen las soluciones detalladas en las secciones 2.2 (AR1) y 2.4 (CC2) para la primera temática, esto es, el uso de técnicas de CV para la evaluación automatizada del riesgo ergonómico relativo a trastornos musculoesqueléticos (TME). El método de AR1 (sección 2.2.1) implica el uso de la CNN OpenPose para detectar los puntos antropométricos de los trabajadores. A continuación, se calculan los ángulos de las articulaciones, para obtener finalmente puntuaciones de riesgo según el método RULA. Con el fin de validar el método propuesto, se realizaron evaluaciones ergonómicas basadas en vídeos reales de trabajos al aire libre, bajo la supervisión de ergónomos profesionales (ver sección 2.2.4). Los principales resultados de validación de esta metodología sobre vídeo fueron los siguientes:

- El método proporcionó evaluaciones RULA para más de un trabajador de forma simultánea.
- El método se adaptó a vídeos con diversas condiciones de iluminación, oclusiones y ángulos de cámara variables.
- Las pruebas exploratorias realizadas en entorno controlado indicaron que los distintos puntos de vista de la cámara no influyen particularmente en los resultados, siempre y cuando no se trabaje en condiciones de elevación superiores a 30°.
- El método logró un notable consenso con las evaluaciones de un grupo de evaluadores expertos.

Otros estudios publicados con posterioridad reafirman estas conclusiones. En [KSS+21], se propone un sistema basado en OpenPose para calcular los ángulos de las articulaciones y las calificaciones RULA y REBA. La validación se realizó con un sistema de captura de movimiento (*MoCap*), y su rendimiento se comparó con un sistema basado en el sensor Kinect, obteniendo mejores resultados con OpenPose. En el trabajo de [LCCL22], se da un paso más allá, implementando un árbol de decisión para seleccionar el método de evaluación más adecuado según los ángulos de las articulaciones derivados del proceso de imágenes de OpenPose para distintas técnicas de ERL. En [SFR+20], se realizan pruebas dinámicas en las que se compara OpenPose con información filtrada, obtenida de giroscopios adosados al cuerpo de seis trabajadores. Como resultado, OpenPose presentó mejores resultados para los movimientos del torso y tuvo un peor rendimiento para los ejercicios de brazos y antebrazos. Finalmente, en el trabajo de [LZLL20] se propone un sistema llamado *Quick Capture*, que se validó con un grupo de tres ergonomos mediante REBA. Como resultado, se obtuvo una concordancia κ de 0,738 con las puntuaciones de los expertos.

En relación con los resultados mostrados en la Tabla 2.5, la mejor concordancia entre los expertos y la metodología AR1 se dio en las puntuaciones de la parte superior del brazo. Esta puntuación es la más influyente en la evaluación RULA, al pertenecer a los miembros

superiores, por lo que esta concordancia proporcionó una aprobación crucial para este diseño. Por el contrario, el menor consenso obtenido fue para las puntuaciones del cuello, obteniendo un coeficiente κ de Cohen de 0,44, que se interpreta en la literatura [LK77] como un acuerdo moderado. No obstante, este coeficiente es comparable en magnitud al consenso entre los propios especialistas que participaron en el experimento. Por tanto, este dato puede atribuirse, al menos parcialmente, a las discrepancias subjetivas de los expertos [RAD+09]. Por último, se obtuvo un consenso pleno en la puntuación global de RULA. Estos resultados son indicativos de la utilidad del método propuesto para llevar a cabo ERLs a partir de vídeos o imágenes tomados al aire libre o en condiciones de visualización no óptimas.

En el trabajo de [PSLM17], se menciona que las pequeñas discrepancias en las puntuaciones locales de RULA suelen ser poco significativas con respecto al RULA Grand Score. Esto se debe al diseño de agregación jerárquica utilizado en la evaluación de RULA, donde las puntuaciones individuales, asociadas a distintas partes del cuerpo, se agregan de forma ponderada para disminuir el impacto de posibles fallos de medida individuales, en especial para la mitad superior del cuerpo humano. Esta característica resulta útil para las metodologías desarrolladas, ya que cualquier error cometido en la estimación de la ubicación de las articulaciones del cuerpo, además del error debido a la distorsión proyectiva, pueden asimilarse como ruido de medición.

Esta circunstancia puede extrapolarse también al método REBA, al ser también un método jerárquico, derivado de RULA. Sin embargo, debe tenerse en cuenta que, a diferencia de RULA, REBA reduce el énfasis en las extremidades, centrándose más en el análisis postural general ofreciendo estimaciones de riesgo balanceadas entre las entre los miembros inferiores, superiores y la columna.

En el trabajo desarrollado en la sección 2.4 (CC2), se propone un flujo de trabajo alternativo, con el fin de procesar las mediciones ergonómicas automatizadas en un modelo de esqueleto 3D y a lo largo de una secuencia temporal (véase Figura 2.22). Para ello, se hace uso de una combinación de dos redes neuronales diseñadas específicamente para este

fin. En primera instancia, se utiliza la CNN STAF para generar un vector de puntos antropométricos en 2D a lo largo del tiempo, donde cada esqueleto, asociado a un fotograma de vídeo, se estima no solo en función del contenido de dicho fotograma (labor que realiza OpenPose en AR1), sino que, además, se tiene en cuenta en la estimación el contenido de los fotogramas anteriormente procesados. En segunda instancia, dicho vector de esqueletos 2D es enviado a la CNN VIBE, que realiza una estimación de la postura y la forma del cuerpo humano, asociados a movimientos naturales ya aprendidos, en un modelo de esqueleto 3D. Por último, se utilizan estos datos posturales en 3D de salida de VIBE para generar mediciones REBA a lo largo del tiempo. Al igual que en [PALM15; LHG+18], se validaron las estimaciones realizadas utilizando un entorno de trabajo simulado.

En el estado del arte [MVM+12; PSLM17; YYL+19; SFR+20], se ha argumentado que los procedimientos de ERL ergonómicos basados en CV no suministran toda la información necesaria para calcular las puntuaciones de RULA y REBA, ya que sigue habiendo una dependencia de información a definir por los expertos, como son las puntuaciones de torsión y posición de la muñeca, la puntuación de equilibrio de piernas y las puntuaciones del uso de fuerza, que dependen del tipo de actividad laboral analizada y no de los distintos ángulos articulares medidos. Sin embargo, esto también ocurre al utilizar otras estrategias de adquisición, como es el uso de sensores vestibulares (por ejemplo, acelerómetros e inclinómetros [Pér11; Yan19]), dado que algunos están ideados para medir articulaciones específicas (codos, hombros, cadera o rodillas). Además, dichas tecnologías adosan al cuerpo instrumentos especiales que resultan intrusivos y alteran el examen real, ya que los empleados son conscientes de su uso, lo que puede llevarles a ejecutar sus tareas de forma sesgada [ZYL18; KLB19; YYL+19; YLW+17; FLL+18; PSLM17].

Por otro lado, las señales captadas por estos instrumentos son más complejas de gestionar, debido al ruido y las interferencias típicas asociadas a la tecnología de sensores [SFR+20; PSLM17], requiriendo el uso de hardware específico que ha de conectarse por cable o radio al sensor. Finalmente, otras tecnologías basadas en CV, como son las cámaras RGB-D (por ejemplo, Microsoft Kinect), que capturan, además de imágenes, información de profundidad de cada píxel, no pueden trabajar con luz solar directa y tienen un alcance

operativo limitado (entre 1,2 y 3,5 m) [YUYA21].

3.2. Monitoreo automatizado de utilización de EPP asistido mediante CV

Como segunda temática, la presente tesis propone el uso de CV para la medición cuantitativa del uso de EPP mediante los estudios detallados en las secciones 2.1 y 2.3, correspondientes a las publicaciones AR2 y CC1. El método descrito en CC1 aplica una CNN YOLO v.2 re-entrenada para detectar personas, guantes, chaleco de alta visibilidad y cascos de seguridad. Se logró demostrar la viabilidad de la metodología propuesta en un entorno desafiante (sitio de construcción en exteriores), utilizando un vídeo egocéntrico, con cámara inestable y oclusiones, en el cual se logró detectar las EPP propuestas con una exactitud del 62%. Este artículo fue el primero publicado durante la tesis, en 2018, contando con una aceptación muy buena tanto de los revisores del congreso como de la comunidad científica en general, como refleja su elevado número de citas.

Antes de la disrupción del ML, la detección de EPP se había abordado en la literatura como un medio basado en CV para detectar y posteriormente distinguir a los trabajadores con respecto a las demás personas presentes en una imagen o vídeo, razón por la cual la detección de EPP no era visto como un fin para llevar la trazabilidad en la seguridad laboral [MGN13; CC11; Kon17]. Las tecnologías del estado del arte sobre detección de estos EPP al momento de la publicación de CC1 estaban basadas en la información de gradiente (HOG + SVM) [PEZ15; PB16], el color (LAB + SVM) [SCC+17] y el dominio transformado (DCT + SVM) [RTK+17]. Dichas tecnologías presentaron una gran exactitud en interiores. Sin embargo, como se menciona en [KKHB18; SCSK19], estas tecnologías presentan deficiencias de funcionamiento en exteriores y, en general, cuando se presentan variaciones en la iluminación, el color y la pose. Estas fueron precisamente las condiciones en las que se testeó el desarrollo propuesto para este trabajo.

Por otra parte, la solución publicada en AR2 se basa en el uso de la red neuronal

OpenPose. Como resultado de las investigaciones que dieron lugar a AR1, se observó que OpenPose presenta un buen desempeño para caracterizar los puntos anatómicos de interés utilizados para definir de forma automatizada las regiones (ROI) ocupadas por la cabeza y el tronco en una imagen. En este trabajo, se definieron de forma explícita conjuntos de puntos de OpenPose para establecer dichas ROI de la cabeza y tronco del trabajador. Si el resultado de la detección de estos ROI supera un umbral de confianza, cada ROI es enviada a un clasificador neuronal binario tipo ResNet, específicamente entrenado para cada EPP (casco de seguridad y chaleco de alta visibilidad).

En otros trabajos publicados recientemente, también se utilizan metodologías de dos etapas similares a la propuesta en AR2:

- En [SDGI20], se utiliza una combinación de detectores de personas, un estimador de la pose del cuerpo y clasificadores para el etiquetado de un *dataset*. Posteriormente, se modifica la arquitectura neuronal de cada clasificador añadiendo un mecanismo de atención espacial que, al re-entrenar la red, reemplaza al estimador de pose. Evidentemente, esta arquitectura de procesamiento es más compleja que la propuesta en AR2.
- En [NBP20], se comparan tres modelos de ML basados en YOLO-v3. El primer enfoque usa esta red junto a un árbol de decisión. El segundo enfoque es similar a CC1, pues el algoritmo detecta simultáneamente a todos los trabajadores visibles en cada fotograma y verifica la utilización de EPP, sin hacer uso de ninguna otra red aparte de YOLO. El tercer enfoque es similar a AR2, porque inicialmente detecta a los trabajadores, luego genera las ROIs de interés y, por último, estima la presencia o ausencia de EPP mediante otros clasificadores neuronales más complejos (VGG-16, ResNet-50 y Xception). Como resultado, el segundo enfoque consiguió el mejor rendimiento, con un 72,3% de precisión media (mAP).
- En el trabajo de [HLH19], se emplea tecnología basada en YOLO-v3, pero solo fue testada en ambientes controlados de laboratorio para detectar 6 tipos de EPP de seguridad (casco, camisa, guantes, cinturón, pantalones y zapatos).

- En [WAK+21], se propone una combinación de técnicas de DL para la detección de objetos y la posterior localización del punto clave de la cabeza, con un razonamiento simple basado en reglas, con el fin de determinar la presencia del casco de seguridad.

La principal dificultad de los métodos basados únicamente en detectores de objetos [FLL+18; WXY+20; WCC+19; HLH19] es la de captar objetos pequeños como, por ejemplo, cascos o botas de seguridad, desde una distancia lejana. Como alternativa, los métodos de dos etapas detectan primero a las personas, que son proporcionalmente más grandes en la imagen y, por tanto, sufren menos problemas de resolución en la detección. Por esta razón, estos últimos enfoques poseen mayores índices de precisión en la detección de EPP [SDGI20; WAK+21].

Para superar las limitaciones de tamaño en las detecciones, en [Kur20] se propone un *dataset* etiquetado de forma semi-automática mediante OpenPose, en el cual destacan distintas herramientas según su posición en la imagen con respecto al cuerpo. Esta estrategia coincide con la utilizada en AR2 para etiquetar el *dataset* a partir de vídeos de Internet.

Capítulo 4

Conclusiones y trabajo futuro

4.1. Conclusiones

La premisa básica de esta tesis es que tecnologías como la Inteligencia Artificial (IA) y la Visión por Computador (CV) pueden apoyar el proceso de Evaluación de Riesgos Laborales (ERL) en un entorno de trabajo real. Esta hipótesis implica el uso de mecanismos automatizados para ERL, junto con otras fuentes de información, cuyo fin es mejorar el proceso de toma decisiones y de medidas preventivas que permitan mitigar los riesgos laborales.

Las principales contribuciones de esta tesis son las siguientes:

- Se estudió el estado del arte de las nuevas tecnologías de la Industria 4.0 (I4.0) en Prevención de Riesgos Laborales (PRL).
- Se investigaron las tecnologías actuales en CV que puedan impactar en la detección e identificación no supervisada de riesgos laborales.
- Se analizó la factibilidad de las tecnologías estudiadas en el sector de la construcción de acuerdo con la normativa nacional e internacional.
- Se establecieron los requerimientos en *hardware* y *software* necesarios para una instalación de monitoreo basada en CV, en base a lo determinado en los resultados

anteriores.

- Se desarrollaron dos (2) técnicas de CV, basadas en RULA y REBA, evaluadas en instalaciones experimentales, simulaciones y en ambientes de trabajo real.
- Se propusieron dos (2) métodos basados en CV, que permiten la medición automática de utilización de EPP, la primera de ellas basada en YOLO y, la segunda, en clasificadores ResNet de baja profundidad.
- Se evaluaron y validaron los resultados obtenidos en diversos entornos laborales, tanto en interiores como exteriores, con distintos medios de adquisición de vídeo: teléfonos móviles, drones y cámaras deportivas egocéntricas.
- Las soluciones obtenidas se han mostrado robustas y fiables en condiciones de iluminación dispar, captura de vídeo en movimiento, tanto del punto de vista como del trabajador, así como oclusiones con objetos y otros trabajadores.

Las pruebas presentadas en la presente tesis se centran en dispositivos de captura modernos como drones, cámaras deportivas, vídeos egocéntricos o teléfonos inteligentes, que en general suelen ser dispositivos con condiciones de captura inestables desde el punto de vista de un sistema de análisis automatizado de imágenes. En consecuencia, las metodologías mencionadas se presentan como opciones de solución practicables y oportunas, utilizables en situaciones reales de trabajo y apropiadas para la monitorización continua y no supervisada de entornos laborales.

4.2. Trabajo futuro

Los resultados obtenidos en la presente tesis refuerzan la posibilidad de enfocar ciertos problemas de ingeniería a las capacidades de sistemas automatizados basados en CV. Como resultado, pueden considerarse varias vías de trabajo futuro. La más destacada es incorporar otros métodos de ERL al análisis propuesto en los trabajos AR1 y CC2. La finalidad de esta línea de trabajo futuro es complementar el análisis ergonómico con otras

herramientas utilizadas ampliamente en la literatura para realizar ERL, destacando entre ellos el método OWAS [Ova09]. Algunos ejemplos recientes de este tipo de aplicaciones pueden hallarse en [HKZX20; KSS+21; LCCL22].

Además, se planea a corto plazo poner a disposición de la comunidad científica un repositorio web en línea, que permita la instalación y prueba de los códigos y algoritmos desarrollados a lo largo de esta tesis. Esto podría facilitar futuras actividades de investigación cooperativa a gran escala en ergonomía, a través de la recopilación y el montaje de conjuntos de datos más amplios y debidamente anonimizados, con los que debería facilitarse la realización de análisis estadísticos de mayor alcance. Para este fin, se han iniciado conversaciones con la Red Académica Colombiana de Ergonomía (RACE), y también con otros expertos en Argentina y España.

Por otro lado, una de las vías de trabajo futuro más prometedoras es la caracterización del ciclo de evaluación ergonómica. Se pretende aquí seguir la metodología de [Sta06] para generar una taxonomía de evaluaciones ergonómicas, basada en el análisis de tareas, que integre las preguntas típicas que mejor describen el proceso de análisis ergonómico, de acuerdo con el conocimiento de las actividades de los expertos en la materia.

Otro objetivo de interés es el desarrollo de técnicas de visualización útiles para el manejo de datos de series temporales de información ergonómica [CNC+19; SFR+20], como son las puntuaciones RULA y REBA utilizadas en esta tesis. y que, a partir del análisis de la propia secuencia de vídeo, puedan facilitar a los ergónomos aquellas tareas de evaluación que requieren localizar de forma rápida y efectiva situaciones de riesgo, tareas repetitivas y otras condiciones insalubres que pueden surgir en los lugares de trabajo. En esta línea de trabajo, resulta también de interés el diseño efectivo de interfaces adecuadas al usuario especialista, en aras de optimizar su experiencia de usuario, así como la producción automática de resúmenes trasladables a documentos de registro. Se estima que todo ello podría facilitar la utilización de los métodos desarrollados en esta tesis por usuarios expertos ajenos a las técnicas aquí desarrolladas, facilitando su labor de observación de entornos laborales para la mejora de la seguridad y salud de los trabajadores.

Bibliografía

- [Ale18] AlexeyAB, *Yolo_mark*, 2018. dirección: https://github.com/AlexeyAB/Yolo%7B%5C_%7Dmark.
- [AN00] D. R. Ankrum y K. J. Nemeth, «Head and neck posture at computer workstations – what is neutral?» En *Proceedings of the 14th Triennial Congress of the International Ergonomics Association*, vol. 5, 2000, pp. 565-568. DOI: 10.5100/jje.33.261.
- [BPS14] D. Battini, A. Persona y F. Sgarbossa, «Innovative real-time system to integrate ergonomic evaluations into warehouse design and management,» *Computers and Industrial Engineering*, 2014. DOI: 10.1016/j.cie.2014.08.018.
- [CC11] S. Chi y C. H. Caldas, «Automated Object Identification Using Optical Video Cameras on Construction Sites,» *Computer-Aided Civil and Infrastructure Engineering*, vol. 26, n.º 5, pp. 368-380, 2011. DOI: 10.1111/j.1467-8667.2010.00690.x.
- [CF97] J. Cui y W. Freeden, «Equidistribution on the Sphere,» *SIAM Journal on Scientific Computing*, vol. 18, n.º 2, pp. 595-609, 1997.
- [CGF97] A. L. Cohen, C. C. Gjessing y L. J. Fine, «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*, n.º 97-117, pp. 16-30, 1997.

-
- [CIAD12] M. è. Chiasson, D. Imbeau, K. Aubry y A. Delisle, «Comparing the results of eight methods used to evaluate risk factors associated with musculoskeletal disorders,» *International Journal of Industrial Ergonomics*, 2012. DOI: 10.1016/j.ergon.2012.07.003.
- [CNC+19] G. Y.-Y. Chan, L. G. Nonato, A. Chu, P. Raghavan, V. Aluru y C. T. Silva, «Motion Browser: Visualizing and Understanding Complex Upper Limb Movement Under Obstetrical Brachial Plexus Injuries,» *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, n.º 1, pp. 981-990, 2019. DOI: 10.1109/tvcg.2019.2934280. arXiv: 1907.09146.
- [CPDR19] J. A. Carrillo-Castrillo, V. Pérez-Mira, M. Del Carmen Pardo-Ferreira y J. C. Rubio-Romero, «Analysis of required investigations of work-related musculoskeletal disorders in Spain,» *International Journal of Environmental Research and Public Health*, vol. 16, n.º 10, 2019. DOI: 10.3390/ijerph16101682.
- [CSW+21] Z. Cao y col., «Realtime multi-person 2d pose estimation using part affinity fields,» en *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 43, 2021, pp. 172-186. DOI: 10.1109/TPAMI.2019.2929257. arXiv: 1812.08008.
- [DAP17] J. A. Diego-Mas, J. Alcaide-Marzal y R. Poveda-Bautista, «Errors Using Observational Methods for Ergonomics Assessment in Real Practice,» *Human Factors*, vol. 59, n.º 8, pp. 1173-1187, 2017. DOI: 10.1177/0018720817723496.
- [Eur21] European agency for safety and health at Work, *Reducing musculoskeletal disorders of viticulture workers through a participatory approach involving video analysis*, 2021. dirección: <https://osha.europa.eu/en/publications/reducing-musculoskeletal-disorders-viticulture-workers-through-participatory-approach-involving-video-analysis>.
- [FLL+18] Q. Fang y col., «Detecting non-hardhat-use by a deep learning method from far-field surveillance videos,» *Automation in Construction*, vol. 85, pp. 1-9, 2018. DOI: 10.1016/j.autcon.2017.09.018.
-

- [Fuk88] K. Fukushima, «Neocognitron: A hierarchical neural network capable of visual pattern recognition,» *Neural networks*, vol. 1, n.º 2, pp. 119-130, 1988.
- [GAHR17] R. L. Greene, D. P. Azari, Y. H. Hu y R. G. Radwin, «Visualizing stressful aspects of repetitive motion tasks and opportunities for ergonomic improvements using computer vision,» *Applied Ergonomics*, vol. 65, pp. 461-472, 2017. DOI: 10.1016/j.apergo.2017.02.020.
- [Gar20] D. C. Garzon Leal, «Nuevas tecnologías aplicadas a la ergonomía ocupacional. Empleo de sensores RGB-D y Eye-Tracking en la mejora ergonómica de puestos de trabajo,» Tesis doctoral, universitat politecnica de valencia, 2020.
- [GHS+15] A. Golabchi, S. Han, J. Seo, S. Han, S. Lee y M. Al-Hussein, «An Automated Biomechanical Simulation Approach to Ergonomic Job Analysis for Workplace Design,» *Journal of Construction Engineering and Management*, vol. 141, n.º 8, pp. 1-12, 2015. DOI: 10.1061/(ASCE)CE.1943-7862.0000998.
- [Gob04] Gob. de España, Ministerio de Empleo y Seguridad Social., *UNE-EN ISO 6385 de 2004 Principios ergonómicos para el diseño de sistemas de trabajo*. 2004. dirección: <https://www.une.org/encuentra-tu-norma/busca-tu-norma/norma?c=N0057285>.
- [GYS17] H. Guo, Y. Yu y M. Skitmore, «Visualization technology-based construction safety management: A review,» *Automation in Construction*, vol. 73, pp. 135-144, 2017. DOI: 10.1016/j.autcon.2016.10.004.
- [HDH09] D. Hoiem, S. K. Divvala y J. H. Hays, «Pascal VOC 2008 challenge,» *World Literature Today*, vol. 24, 2009.
- [HKZX20] C. Huang, W. Kim, Y. Zhang y S. Xiong, «Development and validation of a wearable inertial sensors-based automated system for assessing work-related musculoskeletal disorders in the workspace,» *International Journal*

- of Environmental Research and Public Health*, vol. 17, n.º 17, pp. 1-15, 2020. DOI: 10.3390/ijerph17176050.
- [HLH19] H. M. Hung, L. T. Lan y H. S. Hong, «A deep learning based method for real-time personal protective equipment detection,» *Journal of Science and Technique- Le Quy Don Technical University*, vol. 199, n.º 199, pp. 23-34, 2019.
- [HLS14] S. Hermawati, G. Lawson y A. P. Sutarto, «Mapping ergonomics application to improve SMEs working condition in industrially developing countries: a critical review,» *Ergonomics*, vol. 57, n.º 12, pp. 1771-1794, 2014. DOI: 10.1080/00140139.2014.953213.
- [HR20] J. Hegde y B. Rokseth, «Applications of machine learning methods for engineering risk assessment – A review,» *Safety Science*, vol. 122, n.º September 2019, p. 104 492, 2020. DOI: 10.1016/j.ssci.2019.09.015.
- [HZRS16] K. He, X. Zhang, S. Ren y J. Sun, «Deep residual learning for image recognition,» *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-Decem, pp. 770-778, 2016. DOI: 10.1109/CVPR.2016.90. arXiv: 1512.03385.
- [I L02] L. I. Leirós, «Historia de la Ergonomía, o de cómo la Ciencia del Trabajo se basa en verdades tomadas de la Psicología,» *Revista de Historia de la Psicología*, vol. 30, n.º 4, pp. 35-53, 2002.
- [ILO01] ILO-OSH, *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, 2001, ISBN: 92-2-311634-1.
- [ILO95] ILO, *ILO Safety, health and welfare on construction sites A training manual*. 1995, p. 134, ISBN: 9221091821.

BIBLIOGRAFÍA

- [Ins15] I. Instituto Nacional de Seguridad e Higiene en el Trabajo, *Encuesta Nacional de Condiciones de Trabajo EWCS2015*, 6.^a ed., I. N. d. S. e. H. en el Trabajo, ed. Madrid, 2015.
- [Ins95] Instituto Nacional de Seguridad e Higiene en el Trabajo, INSHT, *LEY 31 de 1995 de Prevención de Riesgos Laborales*, 1995.
- [Int01] International Ergonomics Association, IEA, «Definitions of Ergonomics,» en *International Encyclopedia of Ergonomics and Human Factors*, W. Karwowski, ed., vol. 1, London: Taylor & Francis, 2001.
- [ISH+20] V. Igelmo, A. Syberfeldt, D. Högberg, F. García Rivera y E. Pérez Luque, «Aiding observational ergonomic evaluation methods using MOCAP systems supported by AI-based posture recognition,» *Advances in Transdisciplinary Engineering*, vol. 11, pp. 419-429, 2020. DOI: 10.3233/ATDE200050.
- [JD19] M. Joshi y V. Deshpande, «A systematic review of comparative studies on ergonomic assessment techniques,» *International Journal of Industrial Ergonomics*, vol. 74, n.º October, p. 102865, 2019. DOI: 10.1016/j.ergon.2019.102865.
- [KAB20] M. Kocabas, N. Athanasiou y M. J. Black, «VIBE: Video Inference for Human Body Pose and Shape Estimation,» en *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 5253-5263. arXiv: 1912.05656. dirección: <http://arxiv.org/abs/1912.05656>.
- [KKHB18] H. Kim, H. Kim, Y. W. Hong y H. Byun, «Detecting Construction Equipment Using a Region-Based Fully Convolutional Network and Transfer Learning,» *Journal of Computing in Civil Engineering*, vol. 32, n.º 2, 2018. DOI: 10.1061/(ASCE)CP.1943-5487.0000731.

- [KLB19] E. Konstantinou, J. Lasenby y I. Brilakis, «Adaptive computer vision-based 2D tracking of workers in complex environments,» *Automation in Construction*, vol. 103, n.º june, pp. 168-184, 2019. DOI: 10.1016/j.autcon.2019.01.018.
- [Kon17] E. Konstantinou, «Vision-based construction worker task productivity monitoring,» Tesis doctoral, University of Cambridge, 2017, p. 146. DOI: 10.17863/CAM.20613.
- [KSS+21] W. Kim, J. Sung, D. Saakes, C. Huang y S. Xiong, «Ergonomic postural assessment using a new open-source human pose estimation technology (OpenPose),» *International Journal of Industrial Ergonomics*, vol. 84, n.º February, p. 103 164, 2021. DOI: 10.1016/j.ergon.2021.103164.
- [Kur20] F. C. Kurnaz, «ALET (automated labeling of equipment and tools: a dataset for tool detection and humanworker safety detection,» Tesis doctoral, Middle East Technical University, 2020, p. 55.
- [LBOM98] Y. LeCun, L. Bottou, G. B. Orr y K.-R. Müller, «Efficient backprop,» en *Neural networks: Tricks of the trade*, Springer, 1998, pp. 9-50. DOI: 10.1007/978-3-642-35289-8_3.
- [LCCL22] P. C. Lin, Y. J. Chen, W. S. Chen y Y. J. Lee, «Automatic real-time occupational posture evaluation and select corresponding ergonomic assessments,» *Scientific Reports*, vol. 12, n.º 1, pp. 1-9, 2022. DOI: 10.1038/s41598-022-05812-9.
- [LDJ19] B. D. Lowe, P. G. Dempsey y E. M. Jones, «Ergonomics assessment methods used by ergonomics professionals,» *Applied Ergonomics*, vol. 81, n.º June, p. 102 882, 2019. DOI: 10.1016/j.apergo.2019.102882.
- [LHG+18] X. Li, S. Han, M. Gül, M. Al-Hussein y M. El-Rich, «3D Visualization-Based Ergonomic Risk Assessment and Work Modification Framework and Its Validation for a Lifting Task,» *Journal of Construction Engineering and*

- Management*, vol. 144, n.º 1, p. 04017093, 2018. DOI: 10.1061/(asce)co.1943-7862.0001412.
- [LK77] J. R. Landis y G. G. Koch, «The Measurement of Observer Agreement for Categorical Data,» *Biometrics*, vol. 33, n.º 1, p. 159, 1977. DOI: 10.2307/2529310.
- [LLGR14] Y. Levanon, Y. Lerman, A. Gefen y N. Z. Ratzon, «Validity of the modified RULA for computer workers and reliability of one observation compared to six,» *Ergonomics*, vol. 57, n.º 12, pp. 1856-1863, 2014. DOI: 10.1080/00140139.2014.952350.
- [LLRL21] S. Lee, L. Liu, R. Radwin y J. Li, «Machine Learning in Manufacturing Ergonomics: Recent Advances, Challenges, and Opportunities,» *IEEE Robotics and Automation Letters*, vol. 6, n.º 3, pp. 5745-5752, 2021. DOI: 10.1109/LRA.2021.3084881.
- [LPG+14] P. Lázaro, E. Parody, R. García-Vicuña, G. Gabriele, J. Á. Jover y J. Sevilla, «Coste de la incapacidad temporal debida a enfermedades musculoesqueléticas en España,» *Reumatología Clínica*, vol. 10, n.º 2, pp. 109-112, 2014. DOI: 10.1016/j.reuma.2013.07.001.
- [LSLM17] W. Lee, E. Seto, K. Y. Lin y G. C. Migliaccio, «An evaluation of wearable sensors and their placements for analyzing construction worker's trunk posture in laboratory conditions,» *Applied Ergonomics*, vol. 65, pp. 424-436, 2017. DOI: 10.1016/j.apergo.2017.03.016.
- [LWA14] B. D. Lowe, P. Weir y D. Andrews, «Observation-based posture assessment: review of current practice and recommendations for improvement,» NIOSH, National Institute for Occupational Safety y Health. CDC, Canadian Centre of Research Expertise for the prevention of Musculoskeletal Disorders., inf. téc., 2014. dirección: <https://stacks.cdc.gov/view/cdc/24085>.

- [LZLL20] Z. Li, R. Zhang, C. H. Lee y Y. C. Lee, «An evaluation of posture recognition based on intelligent rapid entire body assessment system for determining musculoskeletal disorders,» *Sensors (Switzerland)*, vol. 20, n.º 16, pp. 1-21, 2020. DOI: 10.3390/s20164414.
- [MBFD20] M. Massiris Fernández, J. M. Bajo, J. Á. Fernández y C. A. Delrieux, «Joint Angle Estimation with VIBE : an Evaluation Using Virtual Avatars,» en *IEEE ARGENCON 2020*, Resistencia, Argentina: IEEE Computer Society, 2020, pp. 1-4, ISBN: 9781728159577.
- [MC93] L. McAtamney y N. Corlett, «RULA: A survey method for the investigation of work-related upper limb disorders,» *Applied Ergonomics*, vol. 24, n.º 2, pp. 91-99, 1993.
- [MFBD20] M. Massiris Fernández, J. Á. Fernández, J. M. Bajo y C. A. Delrieux, «Ergonomic risk assessment based on computer vision and machine learning,» *Computers and Industrial Engineering*, vol. 149, n.º November, p. 106816, 2020. DOI: 10.1016/j.cie.2020.106816.
- [MFBD21] M. Massiris, J. A. Fernández, J. Bajo y C. Delrieux, «Sistema automatizado para monitorear el uso de equipos de protección personal en la industria de la construcción,» *Revista Iberoamericana de Automática e Informática industrial*, vol. 18, pp. 68-74, 2021. DOI: 10.4995/riai.2020.13243.
- [MFD18] M. Massiris Fernández, J. Á. Fernández Muñoz y C. Delrieux, «Detección de equipos de protección personal mediante la red neuronal convolucional YOLO,» en *Actas de las XXXIX Jornadas de Automática*, Badajoz, España, 2018, pp. 1022-1029. dirección: <https://www.researchgate.net/publication/327449170>.
- [MGN13] M. Memarzadeh, M. Golparvar-Fard y J. C. Niebles, «Automated 2D detection of construction equipment and workers from site video streams using histograms of oriented gradients and colors,» *Automation in Construction*, vol. 32, pp. 24-37, 2013. DOI: 10.1016/J.AUTCON.2012.12.002.

BIBLIOGRAFÍA

- [MH00] L. McAtamney y S. Hignett, «Rapid Entire Body Assessment REBA,» *Applied Ergonomics*, vol. 31, n.º 2, pp. 201-205, 2000. DOI: 10.1016/S0003-6870(99)00039-3.
- [Mis17] D. Mishkin, *Models accuracy on imagenet 2012 val*, 2017. dirección: <https://github.com/BVLC/caffe/wiki/Models-accuracy-on-ImageNet-2012-val>.
- [MUF+20] V. M. Manghisi, A. E. Uva, M. Fiorentino, M. Gattullo, A. Boccaccio y A. Evangelista, «Automatic ergonomic postural risk monitoring on the factory shopfloor -The Ergosentinel tool,» *Procedia Manufacturing*, vol. 42, n.º 2019, pp. 97-103, 2020.
- [MVM+12] K. Mura y col., «Innovative system for real-time ergonomic feedback in industrial manufacturing,» *Applied Ergonomics*, vol. 44, n.º 4, pp. 566-574, 2012. DOI: 10.1016/j.apergo.2012.11.008.
- [NBP20] N. D. Nath, A. H. Behzadan y S. G. Paal, «Deep learning for site safety: Real-time detection of personal protective equipment,» *Automation in Construction*, vol. 112, n.º January, p. 103085, 2020. DOI: 10.1016/j.autcon.2020.103085.
- [Ova09] O. Ovako, «OWAS (Ovako Working posture Assessment System),» *Finnish Institute of Occupational Health*, vol. 1, n.º June, pp. 1-6, 2009. DOI: 10.1016/0003-6870(96)00009-9.
- [PALM15] P. Plantard, E. Auvinet, A. S. Le Pierres y F. Multon, «Pose estimation with a kinect for ergonomic studies: Evaluation of the accuracy using a virtual mannequin,» *Sensors (Switzerland)*, vol. 15, n.º 1, pp. 1785-1803, 2015. DOI: 10.3390/s150101785.
- [PB16] M. W. Park y I. Brilakis, «Continuous localization of construction workers via integration of detection and tracking,» *Automation in Construction*, vol. 72, pp. 129-142, 2016. DOI: 10.1016/j.autcon.2016.08.039.

-
- [PCR19] N. Paltrinieri, L. Comfort y G. Reniers, «Learning about risk: Machine learning for risk assessment,» *Safety Science*, vol. 118, n.º July 2018, pp. 475-486, 2019. DOI: 10.1016/j.ssci.2019.06.001.
- [Pér11] F. J. Pérez Duarte, «análisis global de la postura adoptada por los cirujanos en cirugía laparoscópica influencia de la formación en ergonomía,» Tesis Doctoral, Universidad de Extremadura, 2011.
- [PEZ15] M.-W. Park, N. Elsafty y Z. Zhu, «Hardhat-wearing detection for enhancing on-site safety of construction workers,» *Journal of Construction Engineering and Management*, vol. 141, n.º 9, p. 4015024, 2015. DOI: 10.1061/(ASCE)CE.1943-7862.0000974.
- [PSLM17] P. Plantard, H. P. Shum, A. S. Le Pierres y F. Multon, «Validation of an ergonomic assessment method using Kinect data in real workplace conditions,» *Applied Ergonomics*, vol. 65, pp. 562-569, 2017. DOI: 10.1016/j.apergo.2016.10.015.
- [RAD+09] M. Robertson y col., «The effects of an office ergonomics training and chair intervention on worker knowledge, behavior and musculoskeletal risk,» *Applied Ergonomics*, vol. 40, n.º 1, pp. 124-135, 2009. DOI: 10.1016/j.apergo.2007.12.009.
- [RF17] J. Redmon y A. Farhadi, «YOLO9000: better, faster, stronger,» *arXiv preprint*, 2017. arXiv: /arxiv.org/abs/1612.08242 [https:].
- [RIHS19] Y. Raaj, H. Idrees, G. Hidalgo y Y. Sheikh, «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, 2019, pp. 4615-4623, ISBN: 9781728132938. DOI: 10.1109/CVPR.2019.00475. arXiv: arXiv:1811.11975v3.
- [RLL+16] R. G. Radwin, S. Lee, K. Li, M. Lieblich y B. K. D. Park, «Discussion panel on computer vision and occupational ergonomics,» en *Proceedings of the*
-

BIBLIOGRAFÍA

- Human Factors and Ergonomics Society*, 2016, pp. 957-959. DOI: 10.1177/1541931213601220.
- [Rom14] D. Roman-Liu, «Comparison of concepts in easy-to-use methods for MSD risk assessment,» *Applied Ergonomics*, vol. 45, n.º 3, pp. 420-427, 2014. DOI: 10.1016/j.apergo.2013.05.010.
- [RTK+17] A. H. Rubaiyat y col., «Automatic detection of helmet uses for construction safety,» en *IEEE/WIC/ACM International Conference on Web Intelligence Workshops, WIW 2016*, Institute of Electrical y Electronics Engineers Inc., 2017, pp. 135-142, ISBN: 9781509047710. DOI: 10.1109/WIW.2016.10.
- [SAG19] A. H. Schwartz, T. J. Albin y S. G. Gerberich, «Intra-rater and inter-rater reliability of the rapid entire body assessment (REBA) tool,» *International Journal of Industrial Ergonomics*, vol. 71, n.º March 2018, pp. 111-116, 2019. DOI: 10.1016/j.ergon.2019.02.010.
- [SBR17] M. M. Savino, D. Battini y C. Riccio, «Visual management and artificial intelligence integrated in a new fuzzy-based full body postural assessment,» *Computers and Industrial Engineering*, vol. 111, pp. 596-608, 2017. DOI: 10.1016/j.cie.2017.06.011.
- [SCC+17] H. Seong, H. Choi, H. Cho, S. Lee, H. Son y C. Kim, «Vision-based safety vest detection in a construction scene,» en *ISARC 2017 - Proceedings of the 34th International Symposium on Automation and Robotics in Construction*, 2017, pp. 288-293. DOI: 10.22260/isarc2017/0039.
- [SCSK19] H. Son, H. Choi, H. Seong y C. Kim, «Detection of construction workers under varying poses and changing background in image sequences via very deep residual networks,» *Automation in Construction*, vol. 99, n.º October 2018, pp. 27-38, 2019. DOI: 10.1016/j.autcon.2018.11.033.
- [SDGI20] A. Sandru, G.-E. Duta, M.-I. Georgescu y R. T. Ionescu, *SuPEr-SAM: Using the Supervision Signal from a Pose Estimator to Train a Spatial Attention*

- Module for Personal Protective Equipment Recognition*, 2020. arXiv: 2009.12339. dirección: <http://arxiv.org/abs/2009.12339>.
- [SFR+20] S. Santos, D. Folgado, J. Rodrigues, N. Mollaei, C. Fajão y H. Gamboa, «Explaining the ergonomic assessment of human movement in industrial contexts,» *BIOSIGNALS 2020 - 13th International Conference on Bio-Inspired Systems and Signal Processing, Proceedings; Part of 13th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOS-TEC 2020*, n.º March, pp. 79-88, 2020. DOI: 10.5220/0008953800790088.
- [Sta06] N. A. Stanton, «Hierarchical task analysis: Developments, applications, and extensions,» *Applied Ergonomics*, vol. 37, n.º 1 SPEC. ISS. pp. 55-79, 2006. DOI: 10.1016/j.apergo.2005.06.003.
- [VBTS17] N. Vignais, F. Bernard, G. Touvenot y J. C. Sagot, «Physical risk factors identification based on body sensor network combined to videotaping,» *Applied Ergonomics*, vol. 65, pp. 410-417, 2017. DOI: 10.1016/j.apergo.2017.05.003.
- [VLS+21] S. Vandekerckhove y col., «Musculoskeletal disorders and psychosocial risk factors in the workplace: statistical analysis of EU-wide survey data Executive,» European Agency for Safety y Health at Work, Luxembourg, inf. téc., 2021. DOI: 10.2802/39948.
- [WAK+21] B. Wójcik, M. Arski, K. Ksiieik, J. A. Mischczak y M. J. Skibniewski, «Hard hat wearing detection based on head keypoint localization,» n.º June, 2021. arXiv: 2106.10944v1.
- [WCC+19] J. Wu, N. Cai, W. Chen, H. Wang y G. Wang, «Automatic detection of hardhats worn by construction personnel: A deep learning approach and benchmark dataset,» *Automation in Construction*, vol. 106, n.º February, p. 102894, 2019. DOI: 10.1016/j.autcon.2019.102894.

- [WM21] A. H. Wibowo y A. Mawadati, «The Analysis of Employees' Work Posture by using Rapid Entire Body Assessment (REBA) and Rapid Upper Limb Assessment (RULA),» *IOP Conference Series: Earth and Environmental Science*, vol. 704, n.º 1, p. 012022, 2021. DOI: 10.1088/1755-1315/704/1/012022.
- [WXY+20] L. Wang, L. Xie, P. Yang, Q. Deng, S. Du y L. Xu, «Hardhat-wearing detection based on a lightweight convolutional neural network with multi-scale features and a top-down module,» *Sensors (Switzerland)*, vol. 20, n.º 7, pp. 3-7, 2020. DOI: 10.3390/s20071868.
- [XLLH18] Z. Xie, H. Liu, Z. Li y Y. He, «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*, pp. 430-434, 2018. DOI: 10.1109/PIC.2018.8706269.
- [XRC+17] X. Xu, M. Robertson, K. B. Chen, J. hua Lin y R. W. McGorry, «Using the Microsoft Kinect™ to assess 3-D shoulder kinematics during computer use,» *Applied Ergonomics*, vol. 65, pp. 418-423, 2017. DOI: 10.1016/j.apergo.2017.04.004.
- [Yan19] L. Yang, «Ergonomic risk assessment and intervention through smart workwear systems,» Tesis doctoral, KTH Royal Institute of Technology, 2019, p. 77, ISBN: 9789178733798.
- [YLW+17] X. Yan y col., «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*, vol. 34, n.º November, pp. 152-163, 2017. DOI: 10.1016/j.aei.2017.11.001.
- [YUYA21] Y. Yu, W. Umer, X. Yang y M. F. Antwi-Afari, «Posture-related data collection methods for construction workers: A review,» *Automation in Construction*, vol. 124, n.º January, 2021. DOI: 10.1016/j.autcon.2020.103538.

- [YYL+19] Y. Yu, X. Yang, H. Li, X. Luo, H. Guo y Q. Fang, «Joint-level vision-based ergonomic assessment tool for construction workers,» *Journal of Construction Engineering and Management*, vol. 145, n.º 5, p. 04019025, 2019. DOI: 10.1061/(ASCE)CE.1943-7862.0001647.
- [ZYL18] H. Zhang, X. Yan y H. Li, «Ergonomic posture recognition using 3D view-invariant features from single ordinary camera,» *Automation in Construction*, vol. 94, pp. 1-10, 2018. DOI: 10.1016/j.autcon.2018.05.033.