Generación de resúmenes de trayectorias de trabajadores basado en visión por computadora

Conference Paper · October 2019

CITATIONS
0

READS
226

3 authors:

Manlio Massiris Fernandez
Universidad Nacional del Sur
16 PUBLICATIONS **58** CITATIONS
SEE PROFILE

Claudio Delrieux
Universidad Nacional del Sur
115 PUBLICATIONS **546** CITATIONS
SEE PROFILE

Juan Álvaro Fernández Muñoz
Universidad de Extremadura
45 PUBLICATIONS **106** CITATIONS
SEE PROFILE

Generación de resúmenes de trayectorias de trabajadores basado en visión por computadora

Manlio Massiris, Claudio Delrieux
Laboratorio de ciencias de las imágenes.
Dept. de Ing. Eléctrica y de Computadoras.
Universidad Nacional del Sur y CONICET.
Bahía Blanca, Argentina.
manlio.massiris@uns.edu.ar

Juan Álvaro Fernández
Dept. Ing. Eléctrica, Electrónica y Automática.
Escuela de Ingenierías Industriales.
Universidad de Extremadura.
Badajoz, España.
jalvarof@unex.es

Abstract—El seguimiento de las trayectorias del personal es una herramienta de prevención de accidentes laborales, útil en la gestión de materiales y en la optimización del espacio productivo. Los profesionales de seguridad e higiene laboral realizan el seguimiento comúnmente mediante la observación directa, con las consiguientes desventajas e inconvenientes. Por ello recientemente han surgido tecnologías de adquisición novedosas, como por ejemplo el seguimiento basado en etiquetas o la visión por computadoras. La tecnología basada en la visión por computadora ha recibido cada vez más atención debido a su funcionamiento sin etiquetas y de bajo costo.

En este trabajo evaluamos la posibilidad de generar resúmenes de trayectorias mediante visión por computadora, utilizando las redes neuronales YOLO y Deep Sort, para la detección y seguimiento de trabajadores respectivamente. La valoración se aborda primero utilizando un dataset de condiciones sencillas para el seguimiento de los trabajadores, esto para afinar todas las variables de partida de los algoritmos. Y por último, se evalúa la solución propuesta en situaciones específicas desafiantes para el seguimiento de trabajadores en la industria obteniendo resultados promisorios.

Palabras clave— seguimiento visual, prevención de riesgos laborales, redes neuronales artificiales.

I. INTRODUCCIÓN

La localización y seguimiento de los trabajadores es una herramienta que recientemente está tomando importancia para la industria, debido a que son factores utilizados para cálculos de riesgos laborales [1, 2], posicionamiento de materiales[3], distribución de los puestos de trabajo y la optimizan del espacio productivo [4]. En general un sistema de localización y seguimiento para trabajadores debe satisfacer al menos 5 criterios [4] de entre los siguientes:

- Bajo costo y mantenimiento.
- No molesto con respecto a las tareas.
- Aplicable tanto en interiores como en exteriores.
- Exacto.
- Alta frecuencia de datos.
- No intrusivo en cuestiones de privacidad para el personal.

Como respuesta han surgido las tecnologías de seguimiento basadas en etiquetas (RFID, por sus siglas en inglés) y las tecnologías basadas en visión por computador. Las tecnologías RFID presentan problemáticas al ser un procedimiento intrusivo que puede en ciertos casos limitar el libre desarrollo de la actividad productiva sin despreciar además el costo asociado a las etiquetas [5]–[7]. Por lo cual, escogemos las tecnologías basadas en visión por computadora como punto de partida para este trabajo. En la visión por computadora, es posible dividir el problema planteado en tres partes. La primera es la detección de personas, la segunda es el seguimiento de personas y la tercera es la interpretación de los datos.

Entre el abanico de técnicas para la detección de personas en visión por computadora, y en pro de en una siguiente etapa realizar un seguimiento de las mismas, podemos destacar como los descriptores de HOG (histograma de gradientes orientados en inglés) [5, 8] y la clasificación en SVM (máquina de soporte vectorial por sus siglas en inglés) [3, 9, 10], sin embargo, en los últimos años con el crecimiento general de las capacidades computacionales disponibles, las tecnologías basadas en redes neuronales han generado nuevas oportunidades, sobre todo relacionadas con la detección de peatones [11]. En la detección de personas por descriptores, suelen utilizarse características como el contorno o silueta, un punto central, un esqueleto, una elipse o rectángulo alrededor del cuerpo, además del color, y la textura [7, 11]. Sin embargo, las técnicas de detección basadas en redes neuronales ofrecen generalizaciones que pueden ser aplicables a una mayor variedad de entornos de trabajo [8].

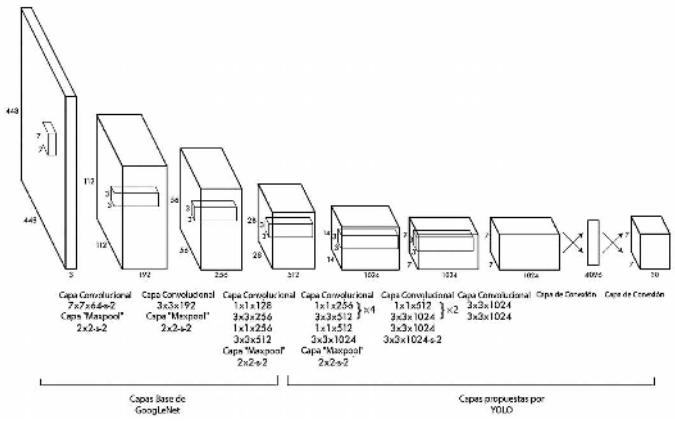

Fig. 1. Arquitectura de la red neuronal YOLO [15].

ISBN: 978-987-1648-44-3

En el seguimiento de personas existen también tecnologías que podrían llamarse clásicas, como las basadas en descriptores, objetivos puntuales, o siluetas [12, 13]. Por otro lado, recientemente también han surgido técnicas que realizan seguimiento basado en redes neuronales, técnicas que se aplican mayoritariamente para el seguimiento de peatones y deportistas, estas cumplen con 4 de 5 de los criterios mencionados anteriormente, en donde la exactitud es el único defecto [4, 14]. Las prácticas actuales de rastreo basadas en la visión por computadora fallan cuando los objetivos están parcialmente ocluidos o muy cerca unos de otros, y por lo tanto no brindan una manera eficiente de rastrear a los trabajadores en entornos laborales complejos, como por ejemplo en la industria de la construcción. Esto se debe a la naturaleza compleja de los sitios de construcción, donde a menudo en las tomas de video ocurren múltiples oclusiones de trabajadores ya sean parciales o severas. Por ello en el estado del arte del seguimiento de trayectorias de trabajadores en entornos laborales complejos, la detección se realiza utilizando la apariencia del traje de alta visibilidad del trabajador [4] y la predicción de trayectoria de movimiento con regresiones o modelos de Markov[1].

II. Materiales y métodos

El objetivo de este trabajo es la generación de resúmenes de trayectorias o seguimientos de los trabajadores en entornos laborales, planteando la novedosa utilización de las redes neuronales You Only Look Once (YOLO) [15] y Deep Sort [16], para la detección de personas y para el seguimiento de las mismas respectivamente.

A. La red YOLO

La detección de objetos en YOLO se considera un problema de regresión. Una sola red de convolución predice simultáneamente varios cuadros circundantes que enmarcan objetos en una imagen y predice la probabilidad condicional para cada clase *p(clase|objeto)* de cada cuadro [15].

Como se puede ver en la Fig. 1, la primera versión de YOLO tiene 24 capas de convolución y 2 capas totalmente conectadas, y YOLO usa la capa de reducción 1×1 y capa de convolución 3×3 en lugar del módulo inicial sugerido por GoogLeNet [15]. YOLO predice la detección en un mapa de características de 13×13, por lo cual, generalmente en un sujeto grande es suficiente, pero para encontrar un sujeto más pequeño, se deben usar funciones de partículas finas.

A diferencia de la tecnología de ventana deslizante y la tecnología de análisis de área de imagen, YOLO opera en toda la imagen a la vez cuando hace predicciones. Como resultado, se tiene en cuenta la información de contexto, el tamaño, la forma y la apariencia [15]. Utilizamos una red YOLO con pesos re-entrenados para la detección de trabajadores y equipos de protección personal. A diferencia de lo publicado previamente [17], en el presente trabajo la red YOLO re-entrenada solo es utilizada en la detección de trabajadores.

La red Deep Sort

La red neuronal Deep Sort [16] ofrece un enfoque pragmático para el seguimiento de múltiples objetos con un énfasis en algoritmos simples, efectivos, en línea y en tiempo real. Emplea el filtro de Kalman en la imagen y la asociación de datos fotograma a fotograma. La red utiliza el método húngaro con una métrica de asociación de la red que mide la superposición de los cuadros circundantes [16]. Esta simple red integra información de apariencia para mejorar el rendimiento sobre versiones anteriores [16]. Debido a esta extensión, podemos rastrear objetos a través de períodos más largos de oclusiones, reduciendo efectivamente el número de cambios de identidad.

TABLE I. Arquitectura de la red neuronal DEEP SORT [16].

Nombre de la capa	Tamaño de Entrada	Tamaño de Salida
Convolucional 1	3x3/1	32x128x64
Convolucional 2	3x3/1	32x128x64
Max pool 3	3x3/2	32x64x32
Residual 4	3x3/1	32x64x32
Residual 5	3x3/1	32x64x32
Residual 6	3x3/2	64x32x16
Residual 7	3x3/1	64x32x16
Residual 8	3x3/2	128x16x8
Residual 9	3x3/2	128x16x8
Densa		128
Normalización y Lote		128

Para el presente trabajo se utilizaron los pesos publicados por los autores para el seguimiento de peatones, y se configuró la red neuronal para tener una memoria de 5 fotogramas, de tal forma que un objetivo se puede realizar un seguimiento siempre que éste se encuentre ocluido por menos de 4 fotogramas.

B. Integración en Python.

Para la integración en Python se utilizó de base una versión publicada bajo la modalidad de código abierto [18]. Dicho código fue posteriormente modificado por los autores para adaptarlo a las necesidades y resultados esperados del presente trabajo. Los módulos más destacados durante el procesamiento fueron Tensorflow, Opencv, Python Image Sequence (PIMS) y Scikit learn. Para la investigación se utilizó un computador con procesador Intel core i7, 16Gb de memoria RAM y una tarjeta gráfica GeForce GTX 1080.

C. Dataset A.

Se plantea como primer acercamiento a la solución del problema un dataset para el seguimiento de trabajadores en entornos industriales simples [16]. Inicialmente diseñado para la detección de equipos de protección personal en entornos industriales, dada su simplicidad es una oportunidad ideal para realizar las primeras pruebas. El video analizado tiene 21 segundos, 1920 x 1080 de tamaño de fotograma y 30 fotogramas por segundo.

ISBN: 978-987-1648-44-3

D. Dataset B.

Otro dataset utilizado fue el publicado en [4], elaborado para el seguimiento de trabajadores envueltos en procesos industriales complejos, listando y ofreciendo un video corto por cada una de las problemáticas más comunes a la hora de realizar el seguimiento de trabajadores, tales como:

- Movimiento abrupto.
- Combinación de problemáticas.
- Congestión.
- Oclusión por escalera.
- Oclusión por barras de metal.
- Oclusión por pared.
- Variación de posturas.
- Variación en la iluminación.

III. Resultados cualitativos y discusión

A. Experimentación exploratoria sobre el Dataset A.

La primera aproximación a la solución se realizó en videos de entornos laborales sencillos, con el objetivo de entender y profundizar en cada una de las variables de los algoritmos implementados. Dado que normalmente los movimientos de las personas no son tan rápidos como la tasa de muestreo que poseen los videos, es normal que en los algoritmos de seguimiento se tome información de forma periódica. A continuación presentamos los resultados al realizar el seguimiento con todos los fotogramas, cada 3 fotogramas y cada 5 fotogramas (Figs 2, 3 y 4, respectivamente).

En el proceso de asignación de etiquetas a cada una de las detecciones, Deep Sort asigna una nueva etiqueta a cada detección no asociada a las que tiene en memoria, sin embargo, en la modificación del código desarrollado por los autores solo se toman como verdaderas y por ende, se hacen visibles las detecciones que se realizan en al menos 3 fotogramas de los 5 fotogramas que Deep Sort tiene en memoria. Por lo cual es aceptado que el contador de entidades no siga números consecutivos.

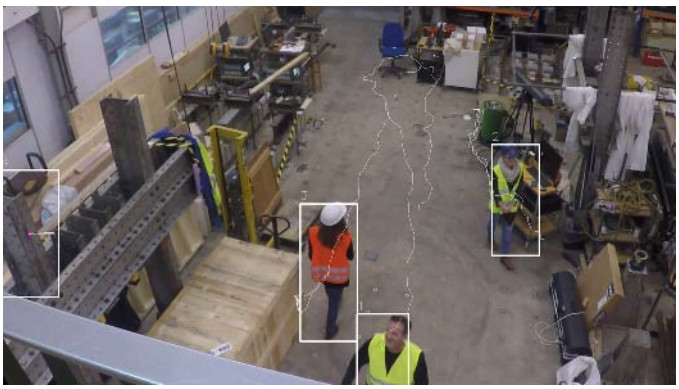

Fig. 2. Resultados de seguimiento sin muestreo.

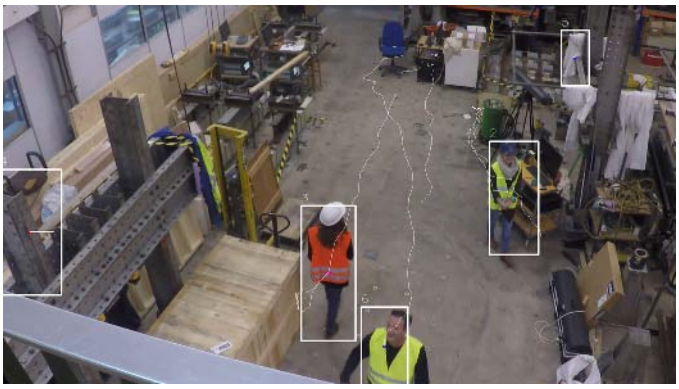

Fig. 3. Resultados de seguimiento con muestreo cada 3 fotogramas.

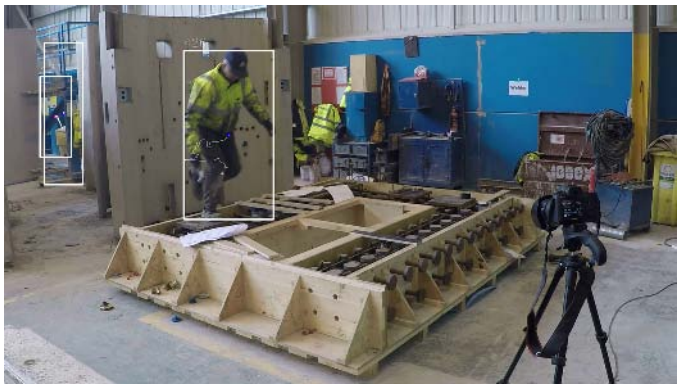

Fig. 4. Resultados de seguimiento con muestreo cada 5 fotogramas.

En los resultados de la experimentación las trayectorias se marcan en líneas blancas en el último fotograma de cada video. Para este video en específico obtenemos que el mejor muestreo es el se presenta cada 3 fotogramas, sin embargo, esto es posible dada la longitud del video analizado de 21 segundos.

El muestreo cada 3 fotogramas en este caso es ideal porque reduce los falsos positivos que no fueron posibles de eliminar, además, reduce el ruido en la trayectoria de las personas. El mencionado ruido es generado debido a que se toma el centroide del cuadro circundante como punto de referencia, el tamaño de los cuadros circundantes varía según el tamaño del sujeto en la escena, por ejemplo si el sujeto levanta un brazo este punto central se moverá en la misma dirección del levantamiento del brazo. Para trabajos futuros se plantea trabajar con funciones que suavicen la trayectoria en el pos-proceso.

B. Estudios de caso sobre el Dataset B.

La segunda aproximación se realizó en función de estudios de casos, los cuales van a ser expuestos y posteriormente discutidos. Para resultados y documentación extra referenciamos al lector al enlace [https://bit.ly/2W34dZM] en el cual se podrán descargar todos los videos resultantes del presente trabajo.

ISBN: 978-987-1648-44-3

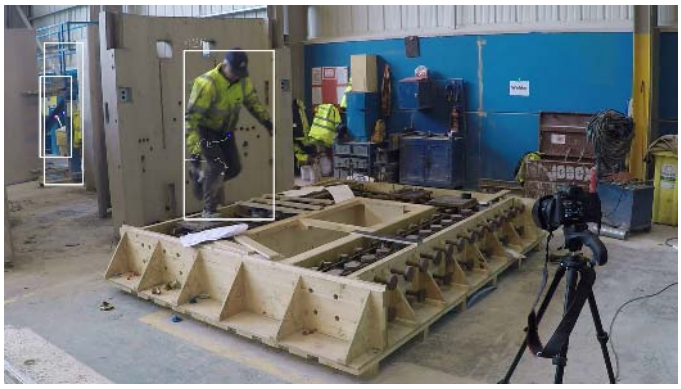

Fig. 5. Resultados de movimiento abrupto.

Fig. 6. Resultados de fondo complejo y movimiento abrupto.

Fig. 7. Resultados de combinación de factores.

Fig. 8. Resultados de congestión.

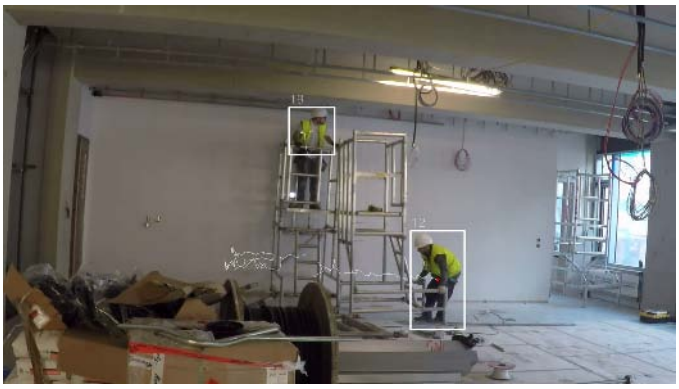

Fig. 9. Resultados de oclusión por escalera.

Fig. 10. Resultados de oclusión por barra de metal A.

Fig. 11. Resultados de oclusión por muro.

Fig. 12. Resultados de variación de posturas.

ISBN: 978-987-1648-44-3

Fig. 13. Resultados variación de iluminación exterior A.

Fig. 14. Resultados variación de iluminación exterior B.

En las Figs. 5 y 6, podemos apreciar que el algoritmo de resumen de video propuesto es capaz de lidiar con los movimientos abruptos inherentes a los puestos de trabajo complejos, sin embargo, se presentan detecciones de falsos verdaderos debido a que la red neuronal aprende generalizaciones de los objetos, por lo tanto, inferimos que la red ha aprendido que los trabajadores están relacionados con las vestimentas de alta visibilidad, en este caso ubicadas en el fondo del video.

En las Figs. 7 y 8 de combinación de factores y congestión, los trabajadores se cruzan durante su andar, el algoritmo propuesto demuestra una respuesta aceptable ante esto indicando cada una de las trayectorias de los trabajadores implicados en la escena, además tenemos trabajadores que se encuentran ocluidos y que ocupan una pequeña proporción del fotograma, lo cual dificulta su detección y seguimiento de los mismos. En estos casos tenemos resultados ambiguos, por ejemplo, en la Fig. 8 en la esquina superior izquierda vemos que es capaz de detectar los pies como parte de algún trabajador, sin embargo, no es capaz de mantener la detección de otros trabajadores ocluidos en la parte superior derecha de la Fig. 7.

En las Figs. 9, 10 y 11, nos encontramos con diferentes tipos de oclusión. En el video asociado a la Fig. 9 el algoritmo no es capaz de detectar a la persona cuando esta se encuentra ocluida por la escalera. En las Figs. 10 y 11 y en sus videos asociados podemos encontrar que el algoritmo es capaz de detectar y realizar seguimiento a los trabajadores que se encuentran parcialmente ocluidos, no obstante, el algoritmo no es capaz de realizar el seguimiento de los trabajadores cuando estos se encuentran ocluidos totalmente por las barras de metal o por el muro.

En la Fig. 12, podemos apreciar que la detección en posturas variadas se realiza bien, asimismo, es posible ver que el sujeto 429 es detectado cuando tiene una gran oclusión en la mitad del cuerpo. En las Figs. 12 y 7, se hace evidente que en un trabajo futuro es necesario un supresor de no máximos, a pesar de que, en situaciones como la de las Figs. 6 ,13 y 14 se nota la complejidad de diseñar el mismo, puesto que podemos perder dos detecciones de trabajadores que se encuentran sobrepuestos.

Por último, analizamos las Figs. 13 y 14, en las cuales se presentan escenas similares que poseen diferentes niveles de iluminación. Esto no presenta una gran dificultad para el algoritmo diseñado. Si bien es cierto, que es un error esperado, ambas figuras muestran que el algoritmo falla en la detección parcial del trabajador agachado. A lo largo del experimento en el *Dataset B* la velocidad de procesamiento de las dos redes neuronales en paralelo fue de 16.7 fps.

IV. Conclusiones y trabajo futuro.

En conclusión podemos destacar que el algoritmo propuesto basado en YOLO para la detección de los trabajadores y Deep Sort para el seguimiento de los mismos fue capaz de realizar un resumen del seguimiento de trabajadores en las diferentes escenas planteadas. No obstante, se espera seguir desarrollando trabajo futuro para mejorar el seguimiento de trabajadores totalmente ocluidos ampliando el horizonte de la memoria de 5 fotogramas utilizada en el presente trabajo, así mismo, se trabajará en el diseño de un algoritmo de supresión de no máximos.

Después de presentar los resultados obtenidos a ingenieros especializados en la prevención de riesgos laborales, estos sugirieron para futuros trabajos, que es posible generar resúmenes de videos en los cuales solo se segmente individualmente cada trabajador en una secuencia, esto en pro de generar calificaciones ergonómicas y de riesgos laborales individualizadas para cada uno de los puestos de trabajo.

Agradecimientos

Este trabajo ha sido elaborado con el apoyo del Consejo Nacional de Investigaciones Científicas y Técnicas de Argentina (CONICET) y de la Junta de Extremadura (España) a través del Fondo Europeo de Desarrollo Regional (FEDER, proyecto GR18135).

Referencias

[1] K. M. Rashid and A. H. Behzadan, “Risk Behavior-Based Trajectory Prediction for Construction Site Safety Monitoring,” J. Constr. Eng. Manag., vol. 144, no. 2, p. 04017106, 2018.

[2] W. Wu, H. Yang, D. A. S. Chew, S. Yang, A. G. F. Gibb, and Q. Li, “Towards an autonomous real-time tracking system of near-miss accidents on construction sites,” Autom. Constr., vol. 19, no. 2, pp. 134–141, 2010.

[3] M. Memarzadeh, M. Golparvar-Fard, and J. C. Niebles, “Automated 2D detection of construction equipment and workers from site video streams

ISBN: 978-987-1648-44-3

using histograms of oriented gradients and colors," Autom. Constr., vol. 32, pp. 24–37, Jul. 2013.

[4] E. Konstantinou, J. Lasenby, and I. Brilakis, "Adaptive computer vision-based 2D tracking of workers in complex environments," Automation in Construction, vol. 103. Elsevier B.V., pp. 168–184, 01-Jul-2019.

[5] M. W. Park and I. Brilakis, "Construction worker detection in video frames for initializing vision trackers," Autom. Constr., vol. 28, pp. 15–25, 2012.

[6] A. Montaser and O. Moselhi, "RFID indoor location identification for construction projects," Autom. Constr., vol. 39, pp. 167–179, 2014.

[7] I. Brilakis, M. W. Park, and G. Jog, "Automated vision tracking of project related entities," Adv. Eng. Informatics, vol. 25, no. 4, pp. 713–724, 2011.

[8] Y. J. Lee and M. W. Park, "3D tracking of multiple onsite workers based on stereo vision," Autom. Constr., vol. 98, pp. 146–159, 2019.

[9] L. Zhang, B. Wu, and R. Nevatia, "Pedestrian detection in infrared images based on local shape features," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2007.

[10] V. Escorcia, M. A. Dávila, M. Golparvar-Fard, and J. C. Niebles, "Automated vision-based recognition of construction worker actions for building interior construction operations using RGBD cameras," in Construction Research Congress 2012: Construction Challenges in a Flat World, 2012, pp. 879–888.

[11] P. Dollár, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: A benchmark," in IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009, pp. 304–311.

[12] G. Welch and E. Foxlin, "Motion tracking survey," IEEE Comput. Graph. Appl., vol. 22, no. 6, pp. 24–38, 2002.

[13] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," Acm Comput. Surv., vol. 38, no. 4, p. 13, 2006.

[14] A. W. M. Smeulders, D. M. Chu, R. Cucchiara, S. Calderara, A. Dehghan, and M. Shah, "Visual tracking: An experimental survey," IEEE Trans. Pattern Anal. Mach. Intell., vol. 36, no. 7, pp. 1442–1468, 2013.

[15] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016, vol. 2016-December, pp. 779–788.

[16] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in Proceedings - International Conference on Image Processing, ICIP, 2018, vol. 2017-September, pp. 3645–3649.

[17] Manlio Massiris, Claudio Delrieux y J. Álvaro Fernández, "Detección de Equipos de Protección Personal Mediante Red Neuronal Convolucional YOLO", Actas XXXIX Jornadas de Automática, Badajoz (España), septiembre 2018, pp. 1022-1029.

[18] bendidi, "Tracking with darkflow." 2018. [Online]. Available: https://github.com/bendidi/Tracking-with-darkflow

ISBN: 978-987-1648-44-3