

Detección del uso de casco de seguridad en imágenes de trabajadores mediante el método de la bolsa de palabras visuales (BoVW)

Safety helmet wearing detection in workers images using the Bag of Visual Words (BoVW) method

Teth Azrael Cortes Aguilar*  y Adriana Tovar Arriaga 

Departamento de Sistemas Computacionales y Electrónica, Tecnológico Nacional de México: ITJMMPyH
Camino Arenero 1101, Col. El Bajío, CP. 45017. Zapopan, Jalisco, México
[*teth.cortes@zapopan.tecmm.edu.mx](mailto:teth.cortes@zapopan.tecmm.edu.mx)

PALABRAS CLAVE:

Equipo de protección personal, bolsa de palabras visuales, aprendizaje supervisado, visión por computadora

RESUMEN

El casco de seguridad es un equipo de protección personal (EPP) imprescindible para prevenir accidentes fatales en fábricas y sitios de construcción. El problema es que en muchos lugares no se lleva a cabo una inspección adecuada, particularmente cuando esta tarea queda a cargo de personas. Sin embargo, la detección automática de objetos a través de sistemas de visión por computadora con cámaras de bajo costo y algoritmos de inteligencia artificial, como el método de la bolsa de palabras visuales (BoVW) son una opción adecuada para inspeccionar en el puesto de control de acceso que los trabajadores y visitantes ocasionales usen cascos de seguridad antes de ingresar a zonas peligrosas. En este artículo reportamos que la etapa de entrenamiento se realizó con imágenes experimentales divididas en dos clases, obtenidas después de aplicar un filtro de detección de bordes y una reducción previa de su tamaño original. Como resultado, se redujo el tiempo promedio de la etapa de entrenamiento a 11.9 segundos y se logró una métrica de exactitud del 95.8%. La etapa de prueba se realizó con imágenes descargadas de internet, y se logró una exactitud del 88.3% en un tiempo promedio de 0.63 segundos. Esto muestra que el método de la bolsa de palabras visuales tiene un buen rendimiento en velocidad y precisión para la tarea de detección del casco de seguridad.

KEYWORDS:

Personal protective equipment, bag of visual words, supervised learning, computer vision

ABSTRACT

The safety helmet is an essential personal protective equipment to prevent fatal accidents in factories and building sites. The problem is that many places do not perform proper inspections, particularly when this task is done by people. However, automatic detection of objects through computer vision systems with low cost cameras and artificial intelligence algorithms such as the Bag of Visual Word (BoVW) method are a suitable option to inspect at the access control point that workers and occasional visitors wear safety helmets before entering hazardous areas. In this paper we report that the training stage was performed with experimental images arranged into two classes, obtained after applying an edge detection filter and a previous reduction of their original size. As a result, the average training time was reduced to 11.9 seconds and a 95.8% accuracy was achieved. The test stage was carried out with images downloaded from the internet, the average testing time was reduced to 0.63 seconds and an 88.3% accuracy was achieved. It shows that the bag of visual words method has a good performance on the speed and accuracy of the safety helmet detection task.

• Recibido: 11 de julio de 2023

• Aceptado: 31 de agosto 2023

• Publicado en línea: 30 de octubre de 2023

1. INTRODUCCIÓN

En México la norma NOM-017-STPS-2008 obliga a los empleadores a entregarles a sus trabajadores, el Equipo de Protección Personal (EPP) necesario, con la intención de proteger a la persona de situaciones que puedan dañar su salud e integridad física [1]. A pesar de las regulaciones, en algunos lugares de trabajo, como sitios de construcción y fábricas, una parte de los trabajadores y visitantes ocasionales ingresan en zonas peligrosas sin el equipo de protección obligatorio, como cascos contra impacto o cascos dieléctricos. Además, en algunos centros de trabajo, la supervisión sobre el uso correcto del equipo de protección personal la realiza personal de seguridad en puestos de vigilancia y control de acceso. Sin embargo, la fatiga visual y la ausencia temporal de la persona encargada de la supervisión, provoca que este sistema de vigilancia presente fallas. No obstante, en la actualidad debido al uso de nuevos sistemas de visión por computadora basados en métodos de aprendizaje automático y la reducción en el costo de las cámaras digitales, han mostrado ser una alternativa viable para resolver problemas de detección de objetos [2], [3], [4], [5].

Por otro lado, la idea de implementar un sistema de visión por computadora en un puesto de control de acceso, antes de entrar en una zona peligrosa, tiene un sentido práctico, que permite controlar muchas variables que pueden afectar los algoritmos de detección de objetos, como el ruido de la escena de fondo, la iluminación, el enfoque de la cámara, delimitar la región de interés, evitar la obstrucción del campo de visión por objetos no deseados y restringir la postura corporal del trabajador. Si el sistema de visión detecta que el trabajador usa el casco de seguridad, entonces se le autoriza el acceso a la zona de trabajo, por el contrario si el trabajador no usa casco de seguridad, entonces el sistema le negará el acceso.

El contenido del presente artículo está dividido de la siguiente manera. En la sección 2 se describe el método de bolsa de palabras visuales utilizado para clasificar las imágenes en trabajador con casco de seguridad y trabajador sin casco de seguridad. En la sección 3 se presenta una revisión del estado del arte en la aplicación de sistemas de visión computacional y algoritmos de aprendizaje automático para el reconocimiento de equipo de protección personal. En la sección 4 se describe el método con sus etapas de desarrollo y el hardware utilizado en el proyecto, también se describe detalladamente el procedimiento aplicado a las imágenes antes de la etapa de entrenamiento. En la sección 5 se muestran los resultados de las etapas de entrenamiento del algoritmo y de prueba. En la sección 6 se presentan las conclusiones y recomendaciones para trabajo futuro.

2 BOLSA DE PALABRAS VISUALES (BoVW)

El método BoVW (*Bag of Visual Words*) es una de las herramientas utilizadas en la visión por computadora para clasificar imágenes. En el presente proyecto, una clase corresponde a las imágenes de trabajadores que no usan casco de seguridad y la otra clase corresponde a las imágenes de trabajadores que sí usan casco de seguridad. En el método BoVW una imagen se representa en función de sus características, las cuales consisten en descriptores sobre aspectos claves o puntos de interés de la imagen, que siempre estarán presentes, sin importar que la imagen se rote o se cambie de tamaño. En el método BoVW, primero se extraen los descriptores de la imagen, luego éstos se agrupan mediante algoritmos como k-means, donde el centro de la agrupación se utiliza como vocabulario del diccionario visual. Después se construyen histogramas sobre la frecuencia de los vocabularios. Los histogramas constituyen la bolsa de palabras visuales. Más adelante, el histograma se convierte en un vector de características de la imagen, ver Figura 1, al cual se le puede asociar una etiqueta, para

usarse posteriormente con un método de aprendizaje supervisado y entrenar el modelo para clasificar las imágenes mediante un algoritmo como SVM (*Support Vector Machines*) [6], [7], [8].

Sin embargo, el proceso de detección de puntos de interés puede ser afectado por distorsiones en la imagen, como sombras, texturas en el fondo y objetos irrelevantes que aparecen en la escena. Esto significa que muchos de los puntos de interés detectados, que no son informativos para la tarea de clasificación, agregan ruido a los descriptores del vocabulario, y pueden reducir el rendimiento del método de BoVW. Algunos de los algoritmos de identificación de puntos de interés intentan evitar el ruido al detectar únicamente los puntos relevantes, sin embargo, muchos de los puntos de interés no informativos son imposibles de eliminar. Por lo que algunas investigaciones han propuesto deducir una región de cobertura para cada palabra visual [7].

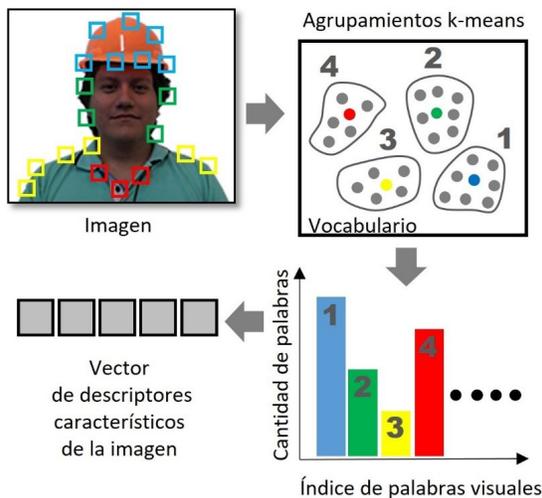


Figura 1. Descripción del método BoVW. Fuente propia.

Para medir el rendimiento de métodos de aprendizaje automático como BoVW, es común el uso de la métrica *accuracy*, la cual indica el porcentaje de casos en que el modelo ha acertado y sirve para evaluar la habilidad del algoritmo para realizar las tareas de clasificación. La métrica *accuracy* se calcula a partir de los resultados de la matriz

de confusión. En la ecuación (1), VP son los verdaderos positivos, VN son los verdaderos negativos, FN son los falsos negativos y FP son los falsos positivos. La literatura científica reporta que métodos de aprendizaje profundo como CNN logra mejores métricas de *accuracy*, de 98% contra el 95.72% de BoVW, para el caso de un conjunto de datos no balanceados [9]. Otro estudio comparativo sobre un conjunto de 1000 imágenes a color distribuidas entre 5 clases, reporta que métodos de aprendizaje profundo como CNN, logra un tiempo de cómputo menor a 2.1 horas, mientras que el tiempo de cómputo de BoVW puede variar de 0.65 hasta 26 horas de acuerdo a la cantidad de descriptores [10]. Aparentemente, estos resultados muestran que BoVW tiene un menor rendimiento que CNN, sin embargo, la diferencia en la métrica de *accuracy* y el tiempo de cómputo no es muy grande. Además BoVW se ha utilizado ampliamente con éxito en el pasado, para clasificar imágenes [8], así como para el reconocimiento en video de acciones humanas [7] y actualmente continúa como un enfoque clave en aplicaciones de aprendizaje profundo no supervisado [11].

$$accuracy = \frac{VP + VN}{VP + VN + FP + FN} \quad (1)$$

Sin embargo, trabajar con conjuntos de datos desbalanceados presenta varios desafíos; ya que la distribución de muestras en cada clase está con una proporción diferente, y existe la posibilidad de que los resultados estén sesgados hacia la clase mayoritaria, la cual, al tener un mayor número de muestras, crea un sesgo en el aprendizaje que afecta a las clases minoritarias [12]. Una forma de evitar este sesgo en los métodos de aprendizaje supervisado, es llevar a cabo el proceso de entrenamiento con un *dataset*, donde cada clase tenga la misma cantidad de muestras. Por esta razón, en el presente proyecto la etapa de entrenamiento del algoritmo BoVW se llevó a cabo con un *dataset* balanceado en dos clases.

3. TRABAJOS RELACIONADOS

En la actualidad, el uso masivo de cámaras digitales de bajo costo y de sistemas de visión por computadora, previenen riesgos laborales potencialmente fatales, mediante la detección del uso de equipos de protección personal en los trabajadores. Los métodos de detección de objetos se dividen en métodos tradicionales que incluyen la identificación de colores, texturas, formas, así como métodos de mayor complejidad basados en características semánticas y métodos que se basan en el uso de redes neuronales convolucionales CNN, como YOLO (*You Only Look Once*) que se han utilizado con éxito para detectar que los trabajadores porten cascos de seguridad, guantes y ropa de alta visibilidad [3], [4], [5]. Sin embargo, YOLO tiene dificultades para detectar objetos pequeños y distorsionados por su orientación respecto a la lente, además etiquetas pegadas al casco de seguridad complican su reconocimiento por color y forma. Estos problemas se pueden resolver, mejorando la resolución de las cámaras y aumentando los datos de entrenamiento, pero esto elevaría la inversión económica, también incrementaría el costo computacional y el tiempo para la adquisición del *dataset*.

El uso de EPP como cascos de seguridad, en ciertos entornos de trabajo es indispensable para evitar accidentes y lesiones en los trabajadores. Y la detección automática de cascos de seguridad, es básicamente un problema de reconocimiento de objetos, que se puede resolver mediante enfoques basados en aprendizaje profundo y visión artificial. Las investigaciones centradas en aprendizaje profundo suelen utilizar *datasets* públicos como MakeML [3], COCO [5], [13], SHD [14] o extraer imágenes de videos en internet [2], [4] para la etapa de entrenamiento de la red, donde las imágenes se clasifican como personas con casco y personas sin casco.

Los métodos tradicionales, basados en coincidencias de características de objetos, solo pueden lograr resultados confiables

sobre un fondo específico y con una sola escena. En muchos casos, la característica más obvia del casco de seguridad es su color, por esta razón, la mayor parte de los trabajos de investigación orientados a la solución de este problema, se enfocan en el uso de la información de color para la tarea de detección [15]. No obstante, las condiciones de iluminación del entorno y el fondo de la escena pueden provocar errores de predicción.

Los algoritmos de detección de objetos se dividen en dos tipos, los de una etapa, como YOLO y SSD (*Single-Shot Detector*) que utilizan un enfoque totalmente convolucional, en el que la red puede encontrar todos los objetos dentro de una imagen en un solo corrimiento y los de dos etapas, donde primero se escala la imagen a un tamaño más pequeño debido a limitaciones en la velocidad de procesamiento; después se ubican cuadros delimitadores en regiones de la imagen y se hace una evaluación de la probabilidad de que una determinada clase este asociada a un cuadro delimitador. Sin embargo, el resultado final del algoritmo de convolución es afectado, si la elección de la región no es la apropiada, además el tiempo de entrenamiento se incrementa proporcionalmente a la cantidad de cuadros delimitadores analizados en cada imagen [16].

A partir de la revisión del estado del arte en este tema de investigación, identificamos la necesidad de proponer otros métodos, que muestren resultados confiables y con un bajo costo computacional en las etapas de entrenamiento y prueba. En este sentido, el reconocimiento de EPP basado en descriptores como el método BoVW, funciona bien en situaciones controladas, pero se debe tomar en cuenta que estos descriptores pueden presentar alteraciones relevantes, para imágenes tomadas en exteriores, debido a factores inevitables como la postura del trabajador respecto al punto de observación, la iluminación del entorno y obstrucciones visuales a causa de otros trabajadores y sus herramientas de trabajo. Asimismo,

investigaciones recientes proponen la detección de puntos antropométricos de los trabajadores visualizados, para segmentar regiones de interés como la cabeza del trabajador, antes de aplicar algún método de aprendizaje profundo para detectar automáticamente EPP en entornos reales [4], [16].

En el presente proyecto se implementó el método de aprendizaje automático denominado bolsa de palabras visuales (BoVW), el cual fue capaz de clasificar correctamente la imagen de un trabajador para reconocer si usa o no usa el casco de seguridad, considerando este elemento como el más importante de los equipos de protección personal para prevenir accidentes fatales. No obstante, que BoVW es un método más simple que otros métodos más sofisticados de aprendizaje profundo, la aplicación del método BoVW logra métricas aceptables de rendimiento con un menor costo computacional. Una contribución de este trabajo es la metodología implementada, donde las imágenes usadas en la etapa de entrenamiento, fueron matrices binarias, obtenidas después de aplicarles un filtro de detección de bordes con una reducción previa del tamaño original de la imagen. Esto permitió reducir los tiempos de entrenamiento hasta 11.9 segundos en promedio y lograr una métrica de rendimiento aceptable del 95.8%.

4. METODOLOGÍA

La metodología propuesta en esta investigación se presenta en la Figura 2 y se dividió en seis etapas. La detección del EPP se logra cuando el resultado del algoritmo entrenado de bolsa de palabras visuales BoVW logra clasificar correctamente si un trabajador porta el casco de seguridad, en este caso se le autorizaría el ingreso a la zona de trabajo, por el contrario si la imagen se clasifica como trabajador sin casco, se le negaría el acceso.

- **Etapa 1.** En esta etapa se tomaron las imágenes mediante una cámara digital, en un entorno controlado, como el que se puede esperar en un puesto de control de acceso a una zona de trabajo que requiere el uso obligatorio de EPP.
- **Etapa2.** En esta etapa se realizó un procesamiento previo de las imágenes con el propósito de reducir el costo computacional en la etapa de entrenamiento, mediante una reducción en el tamaño de la imagen, su conversión a escala de grises y posterior transformación como imagen binaria.
- **Etapa 3.** En esta etapa se conformó el *dataset* de entrenamiento con imágenes agrupadas en dos clases: Trabajador CON CASCO y Trabajador SIN CASCO. También se adquirieron imágenes de internet para el *dataset* de prueba.
- **Etapa 4.** En esta etapa se realizó el entrenamiento del modelo bajo el método BoVW.
- **Etapa 5.** En esta etapa se realizaron las pruebas del modelo BoVW previamente entrenado.
- **Etapa 6.** En esta etapa de analizaron los resultados de las etapas de entrenamiento y prueba para estimar su rendimiento mediante la métrica de *accuracy* y el tiempo de ejecución que requirió cada etapa.

El método BoVW tiene la ventaja de reducir el costo computacional en la etapa de entrenamiento para clasificar correctamente dos imágenes distintas reconociendo puntos de interés que tienen en común imágenes de una misma clase. No obstante, para garantizar la exactitud de los resultados del algoritmo entrenado, es importante que en la medida de lo posible, los descriptores de estos puntos de interés entre las dos clases de imágenes sean diferentes. Por esta razón, fue necesario que las fotografías del *dataset* de entrenamiento se adquirieran de acuerdo a las siguientes cinco restricciones:

- 1 Bajo las mismas condiciones iluminación.
- 2 Usando la misma distancia focal en la cámara.

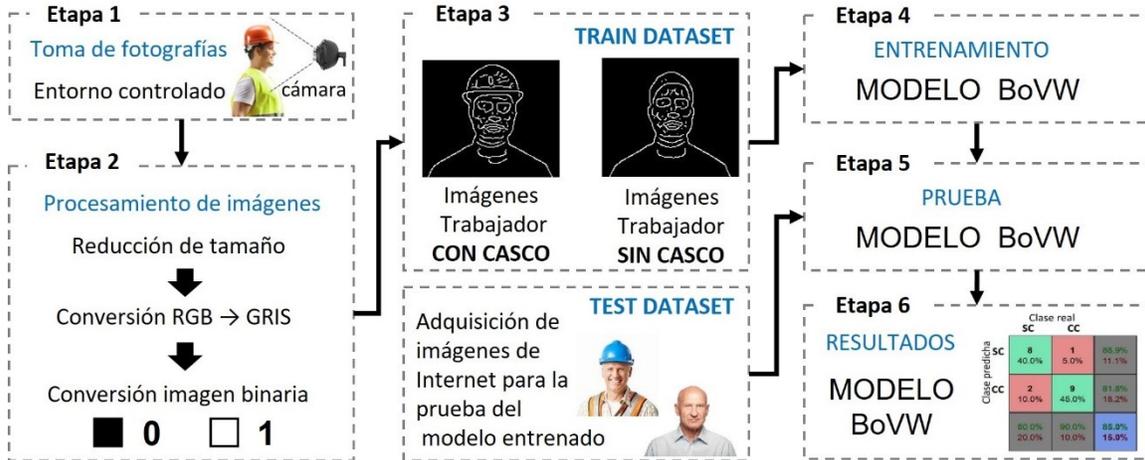


Figura 2. Metodología propuesta en el proyecto de investigación. Fuente propia.

- 3 Evitando la presencia de objetos ajenos a la escena.
- 4 Restringiendo la postura corporal de la persona, mirando de frente a la cámara.
- 5 Delimitando el campo de visión de la cámara al rostro y la parte superior del torso de la persona.

La aplicación de estas cinco restricciones en la etapa 3 de la metodología de investigación permitió reducir significativamente el ruido de la escena de fondo. Que en caso de estar presentes afectan la etapa de entrenamiento, agregando más puntos de interés que no aportan información para la tarea de clasificación.

El procesamiento previo de imágenes, como el uso de filtros, máscaras, ajuste del brillo y reducción en el tamaño de la imagen, así como la conversión del formato de color a gris o binario, es algo común en los sistemas de visión computacional, para eliminar elementos no deseados que puedan afectar el algoritmo y también para reducir el costo computacional de analizar archivos muy grandes [17], [18]. En el presente proyecto se aplicó este procedimiento; primero se capturaron imágenes RGB de 720x1280px usando una webcam USB genérica y de bajo costo, ajustando manualmente el foco a 50 cm para establecer una misma distancia entre la lente de la cámara y el rostro del trabajador. Después se redujo el tamaño de la

imagen a 144x256px, antes de convertir la imagen en escala de grises y posteriormente aplicar un filtro de detección de bordes y obtener una imagen binaria. En la Figura 3 los píxeles oscuros del fondo se asocian al 0 lógico y los píxeles claros con los bordes detectados se asocian al 1 lógico. Este proceso, permitió eliminar la detección de puntos no informativos, debido a sombras y texturas en el fondo de la imagen, así como reducir un 91% el tamaño de los archivos de 68kB a 6kB; y en consecuencia disminuir la incidencia de predicciones erróneas y reducir significativamente el tiempo de entrenamiento.

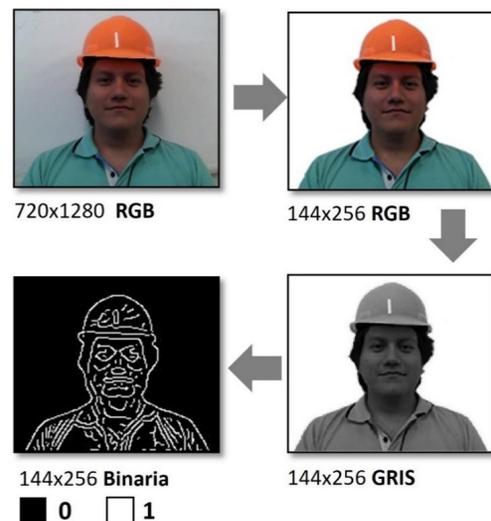


Figura 3. Procesamiento de imágenes anterior a la etapa de entrenamiento. Fuente propia.

Para el procesamiento de las imágenes del *dataset* y el entrenamiento del algoritmo BoVW, se utilizó MATLAB, que es una herramienta muy popular para el desarrollo de proyectos en entornos de educativos [19]. Para la obtención de los descriptores de imagen se utilizó la función *bagOfFeatures()*, para la etapa de entrenamiento se usó la función *trainImageCategoryClassifier()* y para la etapa de prueba se utilizó la función *predict()*.

Se construyó un conjunto balanceado de 24 imágenes de entrenamiento, distribuidas en dos clases, identificadas como trabajador CON CASCO y trabajador SIN CASCO, Figura 4. La adquisición de las imágenes del *dataset* de entrenamiento, se realizó con la colaboración de estudiantes del Tecnológico Superior de Jalisco, en un rango de edad de 20 a 27 años. El algoritmo BoVW se ejecutó en una computadora portátil con procesador INTEL core i7, 8GB de RAM y procesador gráfico NVIDIA GEFORCE 940MX.



Figura 4. Imágenes del Dataset para entrenamiento clasificadas como, trabajadores SIN CASCO y trabajadores CON CASCO. Fuente propia.

Para evitar que las imágenes de prueba fueran muy semejantes a las del *dataset* de entrenamiento. Para la etapa de prueba del algoritmo BoVW, se seleccionaron 60 imágenes de internet [20], que no se usaron durante la etapa de entrenamiento del algoritmo; 30 de las cuales corresponden a trabajadores portando casco de seguridad y otras 30 imágenes corresponden a personas

que no usan casco de seguridad, Figura 5. Estas imágenes se editaron previamente, para seleccionar la misma escena de acuerdo a las imágenes de entrenamiento, además se sobrepuso un fondo blanco, para eliminar objetos no deseados durante la detección de puntos de interés, pero no se cambió el tamaño de la imagen, dejando la más pequeña en 144x144 px y la más grande en 612x408 px. En la etapa de prueba, al igual que en la etapa de entrenamiento, el algoritmo transforma la imagen de color en escala de grises y aplica el filtro *canny* para la detección de bordes, que funciona buscando los máximos locales del gradiente de la imagen. Este método usa dos umbrales para detectar bordes intensos y débiles, e incluye los bordes débiles en la salida si están conectados a bordes intensos. Dado que emplea dos umbrales distintos, el método de *canny* es menos propenso a errores causados por el ruido que otros métodos y, al mismo tiempo, tiene más probabilidades de detectar bordes realmente débiles [21].

5. RESULTADO

En la Figura 6 se muestra la matriz de confusión para los resultados en la etapa de entrenamiento, donde 12 imágenes fueron clasificadas correctamente como trabajador sin casco SC, esto corresponde al 50% de las 24 imágenes del *dataset*. De forma similar, 11 imágenes fueron clasificadas correctamente como trabajador con casco CC. Pero, 1 imagen de trabajador con casco CC fue clasificada incorrectamente como trabajador sin casco SC. La métrica de *accuracy* calculada con (1) es del 95.8% para las imágenes de entrenamiento.

En la Figura 7 se muestra la matriz de confusión para los resultados en la etapa de prueba, donde 24 imágenes fueron clasificadas correctamente como trabajador sin casco SC, esto corresponde al 40% de total de las 60 imágenes de prueba. De manera similar, 29 imágenes fueron clasificadas de forma correcta como

trabajador con casco CC. Pero 6 imágenes de trabajadores sin casco SC fueron clasificadas erróneamente como trabajadores con casco CC y 1 imagen de un trabajador con casco CC también fue clasificada erróneamente como trabajador sin casco SC. Las imágenes clasificadas erróneamente se muestran enmarcadas en la Figura 5. La métrica de *accuracy* calculada con (1) es del 88.3% para las imágenes de prueba.

Los datos estadísticos de la Figura 8 muestran que la metodología implementada logra reducir el costo computacional de implementar el algoritmo BoVW sin comprometer su rendimiento.

Por otro lado, el diagrama de caja de la Figura 9, muestra una variación en el tiempo de predicción para una única imagen de prueba, entre 0.4618 y 1.1056 segundos con una mediana de 0.6059 segundos. La media



Figura 5. Imágenes de la etapa de prueba [20]. Los errores de clasificación se muestran enmarcados.

		Clase real		
		SC	CC	
Clase predicha	SC	12 50.0%	1 4.2%	92.3% 7.7%
	CC	0 0.0%	11 45.8%	100% 0.0%
		100% 0.0%	91.7% 8.3%	95.8% 4.2%

Figura 6. Matriz de confusión para los resultados en la etapa de entrenamiento. Fuente propia

En el diagrama de caja de la Figura 8, se muestra el tiempo que requirió el algoritmo BoVW para completar la etapa de entrenamiento con una variación de 8.8586 a 15.4131 segundos y una mediana de 10.9823 segundos. La media del tiempo de entrenamiento fue de 11.9051 segundos con una desviación estándar de 2.2812 segundos.

del tiempo de predicción fue de 0.63049 segundos con una desviación estándar de 0.12942 segundos. Los datos estadísticos de la Figura 9 muestran que para una aplicación de control de acceso por visión computacional, el tiempo que requiere el algoritmo BoVW para hacer una predicción sobre si el trabajador porta el casco de seguridad, es rápido y confiable.

		Clase real		
		SC	CC	
Clase predicha	SC	24 40.0%	1 1.7%	96.0% 4.0%
	CC	6 10.0%	29 48.3%	82.9% 17.1%
		80.0% 20.0%	96.7% 3.3%	88.3% 11.7%

Figura 7. Matriz de confusión para los resultados en la etapa de prueba. Fuente propia

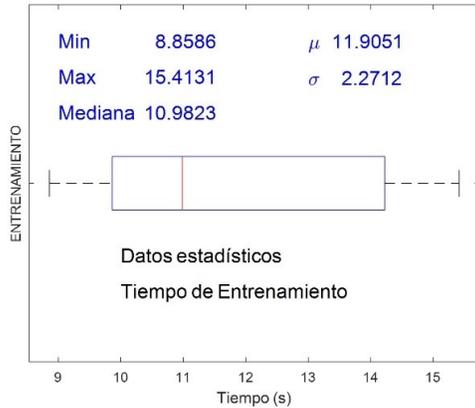


Figura 8. Datos estadísticos para el tiempo de entrenamiento del algoritmo BoVW. Fuente propia

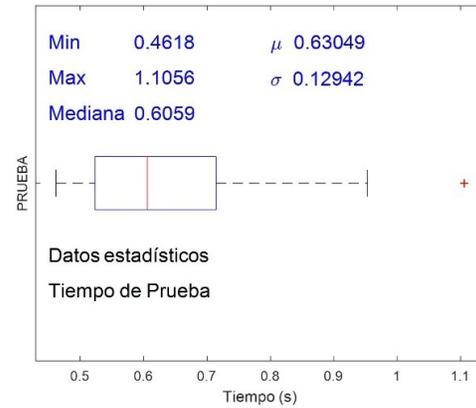


Figura 9. Datos estadísticos para el tiempo de prueba del algoritmo BoVW. Fuente propia

6. CONCLUSIONES

En el presente artículo se presentó un sistema de visión por computadora, que se puede implementar en un puesto de control de acceso para detectar si los trabajadores utilizan casco de seguridad, considerando este equipo como imprescindible de entre los demás equipos de protección personal. Que de acuerdo a normas nacionales e internacionales es un accesorio obligatorio para ayudar a reducir accidentes de trabajo. Los resultados experimentales muestran que es posible aplicar algoritmos simples como el método BoVW, para clasificar imágenes en dos clases, trabajador CON CASCO y trabajador SIN CASCO; con un menor costo computacional, pero bajo condiciones controladas en el ruido de la escena de fondo, la iluminación, el enfoque de la cámara, limitando la región de interés, evitando la obstrucción del campo de visión por objetos extraños y restringiendo la postura corporal del trabajador durante la captura de la imagen. Estas restricciones en las condiciones necesarias para la adquisición de las imágenes pueden ser difíciles de cumplir dentro del área de trabajo, pero son fáciles de implementar en un puesto de control de acceso, previo al ingreso del trabajador en la zona de riesgo.

Se llevó a cabo un procedimiento previo en las imágenes, anterior a la etapa de entrenamiento del método BoVW, que consistió en una reducción en el tamaño de las imágenes, así como su conversión a imágenes binarias mediante la aplicación de un filtro de detección de bordes. Esto permitió reducir el tamaño de los archivos un 91%, lo cual contribuyó a un menor tiempo de entrenamiento de 11.9051 segundos, de acuerdo a la media de los datos estadísticos y lograr una métrica de rendimiento aceptable del 95.8%, usando imágenes reales adquiridas mediante una webcam genérica de bajo costo. En la etapa de prueba se usaron imágenes descargadas de internet, que no se usaron en la etapa de entrenamiento. El método BoVW fracasó en clasificar correctamente 6 trabajadores sin casco de entre 30 valores reales, alcanzado una exactitud del 80%. Pero la exactitud para detectar una persona con casco fue del 96.7%, fallando solo en 1 predicción de entre 30 valores reales. Sin embargo, para la aplicación que proponemos, de usar este sistema en un puesto un control de acceso, antes del ingreso a una zona de riesgo, consideramos que este resultado se puede mejorar si el *dataset* se construye a partir de las imágenes de los trabajadores autorizados, como se hizo en la etapa de entrenamiento, y las posibles fallas ocurrirán únicamente con visitantes ocasionales que desearan ingresar. Bajo este argumento, recomendamos que la supervisión de los

trabajadores regulares de la fábrica o zona de trabajo quede a cargo del sistema de visión por computadora basado en el método BoVW, mientras que la supervisión de visitantes ocasionales se verifique a través del personal de vigilancia.

Otra recomendación para extender los alcances del proyecto como trabajo futuro, es la detección de otros elementos del equipo de protección personal del trabajador, como el calzado de seguridad, chaleco reflejante, lentes de protección y guantes, entre otros. El método BoVW tiene la ventaja de clasificar dos o más imágenes distintas detectando puntos de interés que tienen en común imágenes de una misma clase. Sin embargo, BoVW tiene la desventaja de que no es capaz de identificar elementos presentes en una imagen. Para solucionar este tipo de problema, los algoritmos de aprendizaje profundo de segmentación semántica serían más apropiados, porque pueden reconocer un conjunto de píxeles vinculados a un determinado elemento del EPP a partir del análisis de imágenes de trabajadores de cuerpo entero.

AGRADECIMIENTOS

Agradecemos al Tecnológico Nacional de México, al Instituto Tecnológico José Mario Molina Pasquel y Henríquez, Tecnológico Superior de Jalisco. Y al estudiante de ingeniería electrónica Leonel Luna Pérez por su valiosa contribución para la realización del presente proyecto.

REFERENCIAS

[1] Equipo de Protección Personal, Selección, uso y manejo en los centros de trabajo, NOM-017-STPS-2008, Secretaria de Trabajo y Previsión Social, México. Diciembre 2008. Recuperado el 19 de junio de 2023 de: www.dof.gob.mx.

[2] Massiris, M., Delrieux C., Fernández J.A. Detección de equipos de protección personal mediante red neuronal convolucional YOLO, en *Actas de las XXXIX Jornadas de Automática*, Badajoz, España, 1022-1029, Sep. 2018, doi: [10.17979/spudc.9788497497565.1022](https://doi.org/10.17979/spudc.9788497497565.1022).

[3] Hayat A., Morgado-Dias, F. Deep Learning-Based Automatic Safety Helmet Detection System for Construction Safety, *Applied Sciences*. 2022, 12(16), doi: [10.3390/app12168268](https://doi.org/10.3390/app12168268)

[4] Massiris, M., Fernández, J.A., Bajo, J., Delrieux, C. Sistema automatizado para monitorear el uso de equipos de protección personal en la industria de la construcción, *Revista Iberoamericana de Automática e Informática industrial*, 2020. 18(1), 68-74, doi: [10.4995/riai.2020.13243](https://doi.org/10.4995/riai.2020.13243).

[5] Zhou, F., Zhao H., Nie Z. Safety Helmet Detection Based on YOLOv5, in *2021 IEEE Int. Conf. on Power Electronics, Computer Applications (ICPECA)*, Shenyang, China, 2021, 6-11, doi: [10.1109/ICPECA51329.2021.9362711](https://doi.org/10.1109/ICPECA51329.2021.9362711).

[6] Karim A.A.A., Sameer R.A. Image Classification Using Bag of Visual Words (BoVW), *Al-Nahrain Journal of Science*. 2018, 21(4), 76-82, doi: [10.22401/ANJS.21.4.11](https://doi.org/10.22401/ANJS.21.4.11).

[7] Cortés, X., Conte D., Cardot, H. A new bag of visual words encoding method for human action recognition, in *24th Int. Conf. on Pattern Recognition (ICPR)*, Beijing, China, 2018, 2480-2485, doi: [10.1109/ICPR.2018.8545886](https://doi.org/10.1109/ICPR.2018.8545886).

[8] Molefe M., Tapamo, J.R. Classification of Rail Welding Defects Based on the Bag of Visual Words Approach, in *13th Mexican Conf. of Pattern Recognition (MCPR-2021)*, E. Roman-Rangel et al. (Eds.) Jun. 2021, CDMX, México, 207-218, doi: [10.1007/978-3-030-77004-4_20](https://doi.org/10.1007/978-3-030-77004-4_20).

[9] Saini M., Susan S. Comparison of Deep Learning, Data Augmentation and Bag of-Visual-Words for Classification of Imbalanced Image Datasets, in *Recent Trends in Image Processing and Pattern Recognition*, K. Santosh, R. Hegadi (Eds.) Communications in Computer and Information Science, Vol. 1035. Springer, Singapore, Jul. 2019, doi: [10.1007/978-981-13-9181-1_49](https://doi.org/10.1007/978-981-13-9181-1_49).

[10] Okafor E. Pawara P., Karaaba F., Surinta O., Codreanu V., Schomaker L. Wiering M. Comparative study between deep learning and bag of visual words for wild-animal recognition, in *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, Athens, Greece, 2016, 1-8, doi: [10.1109/SSCI.2016.7850111](https://doi.org/10.1109/SSCI.2016.7850111).

[11] Gidaris, S., Bursuc, A. Puy, G. Komodakis, N. Cord M. Pérez P. OBoW: Online Bag-of-Visual-Words Generation for Self-Supervised Learning, in *2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, 2021, 6826-6836, doi: [10.1109/CVPR46437.2021.00676](https://doi.org/10.1109/CVPR46437.2021.00676).

[12] Saini M. Susan S. Bag-of-Visual-Words codebook generation using deep features for effective classification of imbalanced multi-class image datasets, *Multimedia Tools and Applications*. 2021, 80, 20821-20847, doi: [10.1007/s11042-021-10612-w](https://doi.org/10.1007/s11042-021-10612-w).

[13] Di Benedetto, M., Carrara, F., Meloni, E., Amato, G., Falchi, F., Gennaro, C. Learning accurate personal protective equipment detection from virtual worlds. *Multimedia Tools and Applications*, 2021, 80, 23241-23253, doi: [10.1007/s11042-020-09597-9](https://doi.org/10.1007/s11042-020-09597-9).

[14] Otgonbold, M.E. Gochoo, M. Alnajjar, F. Ali, L. Tan, T.H. Hsieh, J.W., Chen, P.Y. SHEL5K: An Extended Dataset and Benchmarking for Safety Helmet

- Detection. *Sensors*, 2022, 22(6), doi: [10.3390/s22062315](https://doi.org/10.3390/s22062315).
- [15] Huang, L., Fu, Q., He, M., Jiang, D., Hao, Z. Detection algorithm of safety helmet wearing based on deep learning. *Concurrency and Computation*. 2021, 33(13), doi: [10.1002/cpe.6234](https://doi.org/10.1002/cpe.6234).
- [16] Son, H., Choi, H., Seong, H., Kim, C. Detection of construction workers under varying poses and changing background in image sequences via very deep residual networks, *Automation in Construction*. 2019, 99, 27-38, doi: [10.1016/j.autcon.2018.11.033](https://doi.org/10.1016/j.autcon.2018.11.033).
- [17] Organista, V.D., Montero, J.A., Martínez M., Cárdenas, E. Extracción y selección de características en imágenes digitales mostrando lesión de piel, *Programación Matemática y Software*. 2021, 13(3), 91-104, doi: [10.30973/progmat/2021.13.3/7](https://doi.org/10.30973/progmat/2021.13.3/7).
- [18] D.L. Hernández y J.P. Sánchez. Redes neuronales convolucionales para el reconocimiento de imágenes con presencia de cenicilla polvorienta en cultivos de tomate, *Programación Matemática y Software*. 2022, 14(3), 21-28, doi: [10.30973/progmat/2022.14.3/3](https://doi.org/10.30973/progmat/2022.14.3/3).
- [19] Echeverría, D., Mosso, A., Dzul, R.I., Lória, J.E., Pech, G.I.A., Decena C.A., González, R.J. Plataforma didáctica para robótica articulada por medio de una interfaz gráfica, *Programación Matemática y Software*. 2022, 14(3), 37-46, doi: [10.30973/progmat/2022.14.3/5](https://doi.org/10.30973/progmat/2022.14.3/5).
- [20] Gettyimages. Search: worker and worker with helmet. Recuperado el 29 de mayo de 2023 de: www.gettyimages.com.mx.
- [21] Matlab. Encontrar los bordes de una imagen 2D en escala de grises. Recuperado el 12 de junio de 2023 de: la.mathworks.com/help/images/ref/edge.html.



La profesora Adriana Tovar Arriaga recibió en 1999 el título de Ingeniera en Sistemas Computacionales por el Instituto Tecnológico de Querétaro y en 2016 el grado de Maestría en Sistemas Computacionales por el Instituto Tecnológico Superior de Zapopan. Actualmente se desempeña como docente en el departamento de Ciencias Computacionales en la Universidad ITESO y en la carrera de Ingeniería en Sistemas Computacionales de la unidad académica Zapopan del Instituto Tecnológico José Mario Molina Pasquel y Henríquez. Tiene el reconocimiento de perfil deseable PRODEP por el Tecnológico Nacional de México. tovar.arriaga@zapopan.tecmm.edu.mx

ACERCA DE LOS AUTORES



El Dr. Teth Azrael Cortes Aguilar es profesor investigador de tiempo completo en la unidad académica Zapopan del Instituto Tecnológico José Mario Molina Pasquel y Henríquez, cuenta con reconocimiento de perfil deseable PRODEP del Tecnológico Nacional de México. En 2003 recibió el título de ingeniero en Comunicaciones y Electrónica por la Universidad de Guadalajara. En 2005 recibió el grado de Maestro en Ciencia en Óptica por el centro de investigación CICESE, Ensenada, Baja California. En 2022 recibió el grado de Doctor en Manufactura Avanzada por CIATEQ, unidad Jalisco. teth.cortes@zapopan.tecmm.edu.mx