

Modelo de un sistema de alerta basado en visión artificial para la prevención de aglomeraciones en el transporte público

Model of a machine vision-based alert system for crowd prevention in public transportation

Lucia Cárdenas¹, Andrea Pipino², María Teresa Donadio³, Paulo Picota⁴

¹Universidad Tecnológica de Panamá, Facultad de Ingeniería de Sistemas Computacionales, Panamá; lucia.cardenas@utp.ac.pa; <https://orcid.org/0009-0007-2849-2720>

²Universidad Tecnológica de Panamá, Facultad de Ingeniería de Sistemas Computacionales, Panamá; andrea.pipino@utp.ac.pa; <https://orcid.org/0009-0004-4931-7145>

³Universidad Tecnológica de Panamá, Facultad de Ingeniería de Sistemas Computacionales, Panamá; maria.donadio@utp.ac.pa; <https://orcid.org/0009-0009-1628-8284>

⁴Universidad Tecnológica de Panamá, Facultad de Ingeniería de Sistemas Computacionales, Panamá; paulo.picota@utp.ac.pa; <https://orcid.org/0000-0003-2639-6798>

Fecha de recepción: 21-08-2024

Fecha de aceptación: 15-10-2024

DOI <https://doi.org/10.48204/j.vian.v8n2.a6571>

Resumen: La inteligencia artificial se ha convertido en una parte integral de muchas aplicaciones y servicios que utilizamos a diario. Este artículo presenta el diseño de un modelo que utiliza tecnologías de visión artificial para detectar e identificar la concentración de personas y generar alertas según los valores, proporcionados por un algoritmo de clasificación específico para aglomeraciones. El proyecto está enfocado en la atención al público panameño. Teniendo en cuenta que, actualmente, las empresas de transporte no cuentan con herramientas que permitan el manejo de datos para obtener información sobre las condiciones de los pasajeros lo largo de las diferentes rutas. El objetivo principal es desarrollar un modelo que pueda detectar aglomeraciones en tiempo real y lograr proporcionarles información a los operadores de transporte sobre la cantidad de usuarios en los buses, y como resultado mejorar el proceso de toma de decisiones. Se utilizan librerías aplicadas a la visión artificial, con el fin de proponer una solución eficiente y rápida que, además permite una mejora continua, facilitando la gestión y la prevención de las aglomeraciones. Como resultado, el sistema propuesto detectó las aglomeraciones en un 84.04 %, demostrando la posibilidad de la optimización del flujo de pasajeros.

Palabras clave: Visión artificial, algoritmo de clasificación, transporte público, aglomeración.

Abstract: Artificial intelligence has become an integral part. Of the many applications and services that we use on a daily basis. This paper presents the design of a model that uses artificial vision technologies to detect and identify the concentration of people and generate alerts according to the values. Provided by a specific classification algorithm for crowds. The project is focused on serving the Panamanian public. Considering that currently transportation companies do not have tools that allow the management of data to obtain information about the conditions. Of the passengers along the different routes. The main objective is to develop a model that can detect agglomerations in real time and provide information to transport operators about the number of users on the buses, and as a result, improve the decision-making process. Libraries applied to artificial vision are used in order to propose an efficient and fast solution, which also allows

continuous improvement, facilitating management. Prevention of overcrowding. As a result, the proposed system detected 84.04% of crowds, demonstrating the possibility of optimizing passenger flow.

Keywords: Artificial vision, classification models, public transportation, agglomeration.

1. Introducción

En la sociedad vemos la presencia de grandes concentraciones de personas en escenarios como el transporte público, centros comerciales y eventos, de modo que presentan una limitante para el desarrollo de diversas actividades. Este fenómeno no solo afecta el flujo de personas, sino también genera obstáculos debido al continuo aumento de la cantidad de individuos en un espacio determinado (Barbierato et al., 2020; Bilotta et al., 2018). Adicional, las aglomeraciones pueden dar lugar a un incremento en los niveles de estrés y ansiedad, disminuyendo su bienestar emocional y físico (Martínez, n.d.); de la misma manera, contribuye a la sensación de falta de seguridad. Ante este escenario surge la interrogante sobre ¿cómo podemos evitar la concentración excesiva de personas? Para abordar esta situación, se plantea que contar o detectar personas en una multitud de manera precisa y en tiempo real requiere de un enfoque actualizado y eficaz. Con este fin se podrían emplear técnicas tradicionales de conteo o control de personas; sin embargo, dada la gran cantidad de individuos involucrados, resulta complicado y nada práctico realizar un conteo manual. Por lo tanto, se necesita recurrir a nuevas tecnologías que permitan agilizar, facilitar y disminuir la carga de trabajo sobre el personal encargado de gestionar el entorno.

La visión artificial es una tecnología que emplea algoritmos avanzados y modelos matemáticos para analizar imágenes y videos. Estos algoritmos permiten al sistema identificar patrones y extraer información relevante de los datos visuales, lo que facilita el procesamiento automatizado. Estos algoritmos son aptos para interpretar datos visuales en términos de estructuras y conexiones, lo que le permite realizar tareas más complejas como el reconocimiento de objetos, la detección de movimiento y la identificación de características específicas en una imagen, lo que es óptimo para implementar en una solución orientada al problema, anteriormente, presentado (Sheng-Fuu Lin et al., 2001).

La gestión de aglomeraciones con inteligencia artificial ha surgido como una respuesta innovadora a los desafíos planteados por la creciente urbanización (Surasak et al., 2018;



Chaudhary et al., 2021). Tras una reunión con una empresa de transporte público en la ciudad de Panamá, se observó que, actualmente, no cuentan con un sistema para monitorear en tiempo real el estado de los pasajeros dentro de los buses, incluyendo la cantidad de personas a bordo.

Ante esta necesidad, se propone el diseño y desarrollo de un prototipo basado en visión artificial para prevenir aglomeraciones, optimizando así la gestión del flujo de pasajeros, que utiliza conceptos de visión artificial para detectar las aglomeraciones, permitiendo así un mejor manejo de recursos y seguridad con respecto a las decisiones que podría tomar la empresa al administrar las rutas basadas en datos en tiempo real (Baumann et al., 2022). En este sentido, se plantea que la implementación del diseño preliminar mejorará la precisión en la detección de pasajeros en espacios cerrados, lo que permitirá gestionar con mayor rendimiento las multitudes y optimizar la toma de decisiones operativas en tiempo real.

2. Materiales y métodos

- **Diseño**

Para el planteamiento del diseño de prototipo realizamos diversas reuniones con empresas panameñas de transporte público, quienes expusieron sus limitaciones actuales en relación con la problemática del monitoreo interno de los vehículos de transporte colectivo, la realidad de las aglomeraciones no controladas y los beneficios que tendría el obtener información en tiempo real de la cantidad de personas en un vehículo de transporte colectivo. Del mismo modo, con profesores de la Universidad Tecnológica de Panamá especializados en inteligencia artificial se desarrollaron sesiones de capacitación para familiarizarnos con las librerías aplicadas en esta área.

Basados en la información recopilada en las entrevistas y los artículos científicos revisados, se extrajeron varios métodos para el conteo de personas utilizando “*deep learning*” (Bhangale et al., 2020), entre los cuales algunos procedimientos para el análisis de características incluían la detección de cabezas, caras y/o cuerpos; con esto podemos efectuar una comparación y conocer, cuál implementar para el desarrollo del prototipo.

- **Desarrollo**

Para el desarrollo del prototipo se emplearon librerías, comúnmente, utilizadas en visión artificial como son opencv, tensorflow, numpy, math, entre otras; que facilitaron la programación para la detección de personas en tiempo real por medio de una cámara web, inicialmente, se programaron dos métodos de detección de objetos, uno aplicando opencv y el otro tensorflow, como observamos en la Figura 2 y Figura 3, con el propósito de conocer sus diferencias, facilidades y limitaciones; así pues, se gestionan observaciones y características, definiendo las librerías a emplear con base en las pruebas, la rapidez de procesamiento y la simplicidad de código.

Figura 2
Reconocimiento de objetos con opencv.

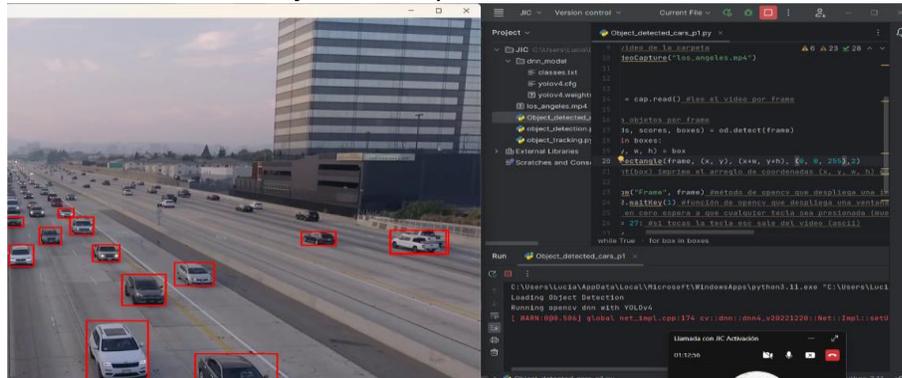
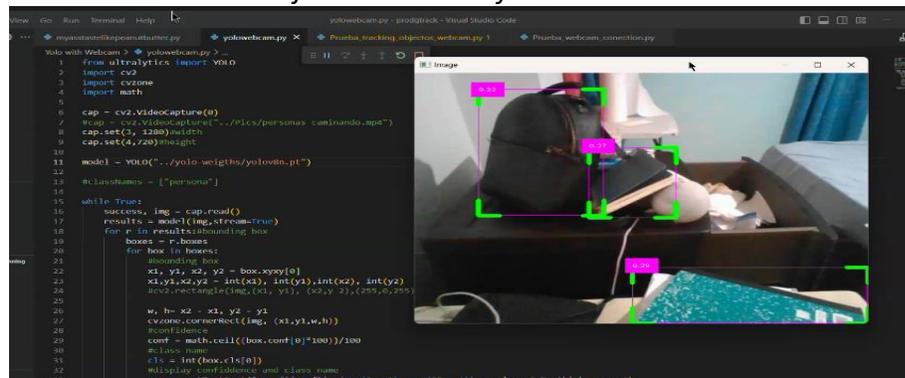


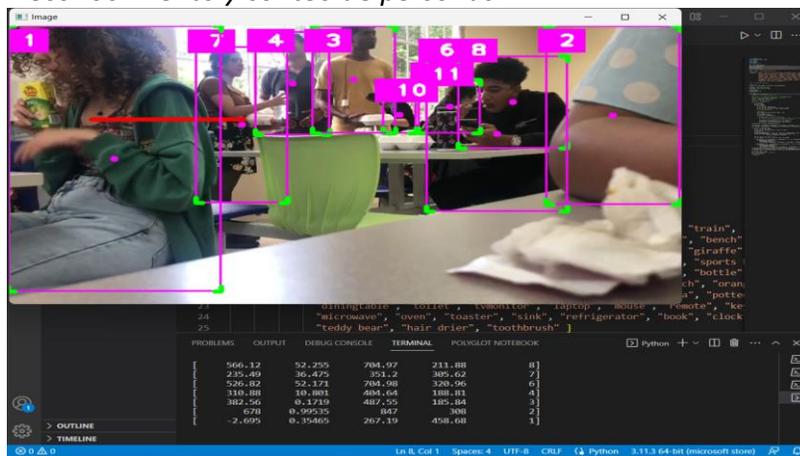
Figura 3
Reconocimiento de objetos con tensorflow.



Con el uso de ambas librerías se programó un algoritmo para la detección, rastreo y conteo de personas (Figura 4), por consiguiente, al unir el modelo de clasificación de aglomeraciones y el algoritmo de detección de objetos, se consiguen respuestas más precisas y rápidas, por otra parte, mejoramos a velocidad de procesamiento cambiando la ejecución del programa a la GPU (Unidad de Procesamiento Gráfico) de nuestros equipos.

El modelo de detección de personas y aglomeraciones fue desarrollado utilizando redes neuronales convolucionales, adaptadas al programa según las necesidades y los resultados esperados. Es importante destacar que el modelo de clasificación de aglomeraciones aún se encuentra en fase de desarrollo, ya que, fue, específicamente, diseñado y entrenado para esta investigación, lo que permite mantener un control más preciso sobre los parámetros.

Figura 4
Reconocimiento y conteo de personas.



En cuanto a la programación del modelo de clasificación de aglomeraciones dividimos su desarrollo en cinco etapas:

- a) **Recopilación y preparación de datos:** Para esta etapa se buscaron datos representativos los cuales fueron etiquetados en dos clases que representan el objetivo que queremos clasificar para este proyecto. En particular, las clases etiquetadas como “Aglomerado” y “No aglomerado”. Se realiza un reajuste para normalizar los valores de píxeles en las imágenes y, finalmente, se utiliza un filtro

para eliminar los elementos de los datos que no concuerden con el formato de archivo aceptado.

- b) **Diseño de la arquitectura de la red:** El modelo construido tiene una arquitectura convolucional seguida de capas de “*MaxPooling*” para extraer y reducir las características de las imágenes. Luego, estas se aplanan y se pasan a través de capas, densamente, conectadas para realizar la clasificación final. (Sahoo et al., 2019; Wang et al., 2023; Gao et al., 2020).
- c) **Entrenamiento del modelo:** Se establece el tamaño de las muestras de entrenamiento, validación y prueba. Luego, se define la cantidad de veces que se iterará en el conjunto de datos de entrenamiento y determinar los valores de pérdida y precisión.
- d) **Evaluación:** Basándonos en los valores de pérdida y precisión registrados, determinamos si el modelo necesita modificaciones o más datos con el fin de minimizar la pérdida y aumentar la precisión de sus respuestas.
- e) **Ajuste y optimización:** Durante este proceso se incrementa, considerablemente, la cantidad de imágenes desde un lote de 800 hasta 1545, con el fin de mejorar la precisión; sin embargo, aún necesita mejoras en algunos aspectos.

Este modelo, combinado con la programación anterior y la impresión de alertas forman el sistema propuesto en el artículo.

Manejo de alertas

El manejo de alertas implica informar sobre los datos detectados por el modelo, donde utilizamos un enfoque basado en una indicación visual en pantalla. Mediante esta indicación, se muestra la cantidad de personas registradas y se determina el estado de aglomeración del área, utilizando una escala de probabilidad que va de 0 a 1. Esta visualización facilita la comprensión de los datos y permite observar los cambios a lo largo del tiempo para una mejor monitorización y evaluación al momento de analizar los resultados de las pruebas. El manejo de alerta resulta uno de los objetivos más importantes, porque establece la conexión entre el programa y el usuario, dando a conocer

los datos en tiempo real que serán utilizados por el usuario para las decisiones en sus empresas.

Pruebas de prevención

Se realizaron un total de tres pruebas en espacios controlados, esto permite un mejor manejo de las variables (cantidad de personas y estado del área) y facilita la obtención de los datos, determinando el funcionamiento del sistema donde se tomaron en cuenta los siguientes parámetros:

Cantidad de personas

- Iluminación
- Espacio lo más parecido a un bus
- Espacios vacíos en el lugar de pruebas
- Espacios vacíos entre las personas

Prueba 1: Al igual que la prueba anterior, se realizó en un salón de clases. Se seleccionó una muestra de 12 personas y se colocó la cámara a un costado para obtener una mejor visualización, como se aprecia en la Figura 5. El objetivo principal fue determinar cuándo el sistema registraba si el área estaba “No aglomerada” o “Aglomerada”.

Prueba 2: En esta prueba, se utilizó otro salón de clases, pero se colocó la cámara en la esquina opuesta respecto a la Prueba 1, para obtener un ángulo de visión diferente. Se seleccionó una muestra de cinco (5) personas. El propósito principal fue determinar cómo respondía el sistema ante diferentes niveles de iluminación.

Prueba 3: Esta prueba se llevó a cabo en un pasillo estrecho. Se seleccionó una muestra de 14 personas y se colocó la cámara en la esquina superior de la entrada, por donde ingresaban las personas. El objetivo principal fue recopilar más datos sobre las predicciones del prototipo y evaluar su precisión al controlar la cantidad de personas presentes, con el fin de alcanzar la probabilidad deseada en cada caso.

Estas pruebas fueron diseñadas para evaluar y recopilar datos sobre el rendimiento del prototipo en distintas situaciones y condiciones. Los resultados obtenidos proporcionarán información relevante para determinar la precisión y la eficacia del modelo.

Figura 5
Prueba 1 del prototipo.



3. Resultados

Las pruebas fueron analizadas por recuadro de imagen que captaba la cámara en el video, se analizaba la probabilidad esperada y la obtenida, lo mismo respecto a la cantidad de personas. Finalmente, se capturaron los datos divididos en el entorno de las pruebas realizadas:

- Mucha luz y mucho espacio.
- Poca luz y poco espacio.
- Mucha luz y poco espacio.

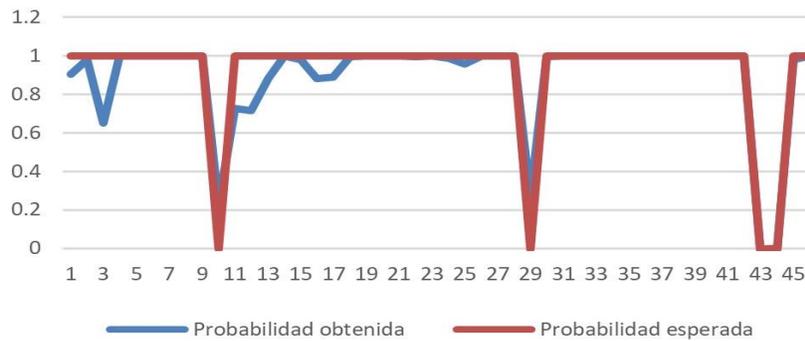
A cada prueba se le generaron gráficas y una probabilidad, conociendo la certeza del programa y sus limitaciones.

Prueba 1 (mucha luz y espacio): Fue de las más acertadas en ambos aspectos (cantidad de personas y probabilidad de aglomeración), Figuras 6 y 7, obteniendo que, cuarenta de cuarenta y siete recuadros reconocieron, correctamente, la cantidad de personas.

Figura 6
Cantidad de personas de la prueba.



Figura 7
Probabilidad de aglomeración de la prueba 1.



Prueba 2 (poca luz y espacio): Al tener un nivel de luz poco apropiado obtuvimos datos poco exactos cuando la zona estaba poco aglomerada, conseguimos que veintidós de veintinueve recuadros fueron resultados positivos, Figuras 8 y 9, es importante mencionar que estos datos positivos son para cuando había aglomeración en el área.

Figura 8
Cantidad de personas de la prueba.

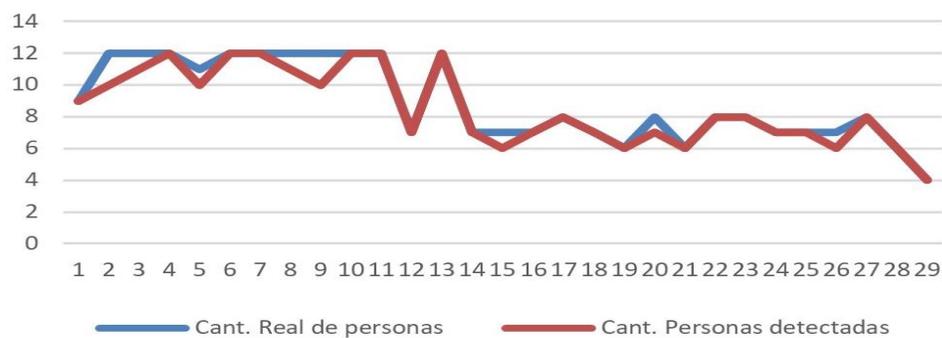
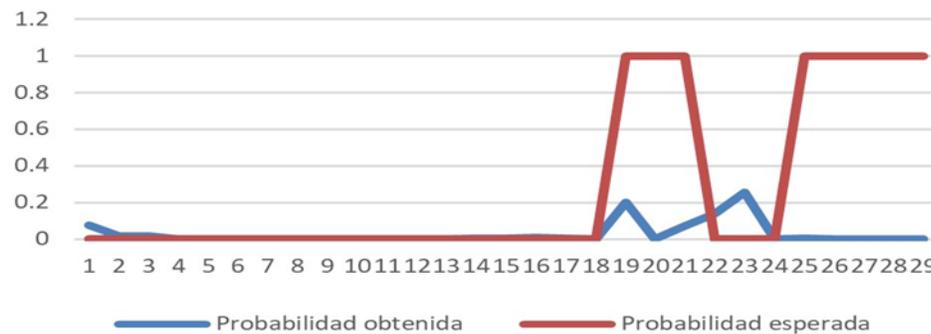
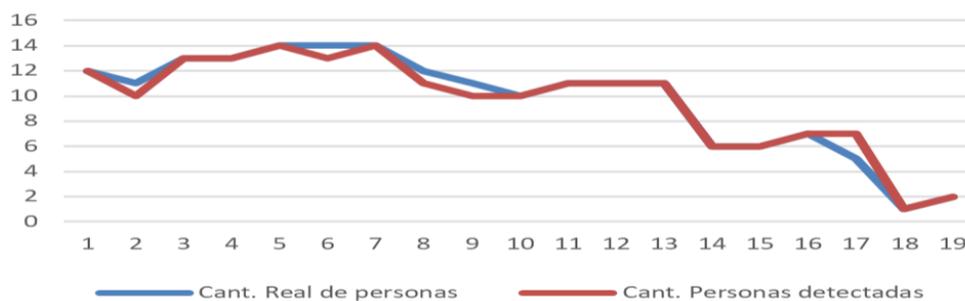


Figura 9
Probabilidad de aglomeración de la prueba 2.



Prueba 3 (mucha luz y poco espacio): En este caso la probabilidad de no aglomeración fue baja, la presencia de muchos muebles provocó que fuera menor comparada con la prueba 1, pero no incorrecta al momento de identificar que el lugar no estaba aglomerado Figuras 10 y 11.

Figura 10
Cantidad de personas de la prueba 3.

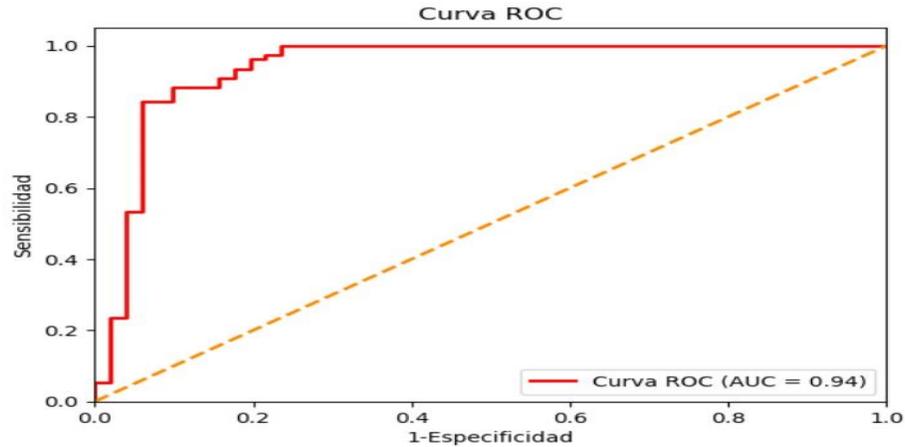


Según los resultados obtenidos y considerando cada escenario, se obtuvo un 84.04% de precisión en la detección de aglomeraciones.

La Figura 12 muestra la curva ROC que describe la precisión con la que el modelo realiza las predicciones sobre las clases de aglomerado y no aglomerado. El área bajo la curva (AUC) de esta gráfica muestra que el modelo posee un alto rendimiento para las tareas de clasificación que estamos ejecutando, luego del entrenamiento y las pruebas de laboratorio realizadas.

Figura 12

Curva de características de operador receptor (ROC) del modelo entrenado y área bajo la curva (AUC).



4. Discusión

En este estudio se desarrolló un sistema que permite detectar aglomeraciones y personas utilizando redes neuronales convolucionales, con el objetivo de proporcionarles información a los operadores de transporte sobre la cantidad de pasajeros en los buses a lo largo de las rutas. El uso de estas redes con un entrenamiento prolongado y un “dataset” extenso se pueden obtener valores de precisión cada vez más altos, como sucede en el estudio (Nogueira et al., 2019).

La alta precisión alcanzada por el modelo puede atribuirse a la arquitectura utilizada y a las técnicas de procesamiento de datos aplicadas, como la normalización y el aumento de datos; estas técnicas permitieron al modelo distinguir mejor sobre las variaciones en las imágenes, como diferentes iluminaciones y posiciones de las personas. Estos resultados también se ven en el trabajo de (Baumann et al., 2022) donde mencionan que el uso de CNN para el conteo de personas en el transporte público es, computacionalmente, eficiente y proporciona resultados precisos con respecto a otras técnicas.

Finalmente, podemos mencionar que existen otros métodos para el reconocimiento y detección de objetos en imágenes, como la segmentación espaciotemporal de imágenes y la clasificación de características utilizando algoritmos de “clustering”, como se presenta en el estudio de (Galic & Loncaric, n.d.). Además, en aplicaciones similares en el área del

transporte público, se han planteado diferentes enfoques, como la utilización de redes WiFi para el conteo de personas, procedimiento que ha demostrado ser muy ventajoso para la obtención de datos a largo plazo (Hidayat et al., 2018). Sin embargo, el enfoque aplicado en este proyecto resulta ser el más apropiado para las condiciones y el área geográfica en la que se encuentra ubicado, aprovechando las ventajas de las técnicas de visión artificial disponibles.

5. Conclusiones

Se ha demostrado la relevancia significativa del modelo propuesto para la detección de aglomeraciones en espacios cerrados. Las pruebas realizadas han evidenciado que el uso de librerías de visión artificial para contar personas presenta un gran potencial al reducir el tiempo y aumentar la capacidad de detección en un periodo determinado. Además, facilita la identificación de áreas amontonadas de manera instantánea.

El modelo presentado cuenta con ventajas inherentes, como un margen y potencial de mejora sucesiva debido a su relación con tecnologías de inteligencia artificial. Esto permite que pueda progresar, constantemente, para aumentar su precisión y obtener resultados óptimos. Si bien la implementación de esta tecnología puede ser, relativamente, sencilla debido a los avances actuales. Es importante considerar las limitaciones relacionadas con la calidad del hardware utilizado para capturar y procesar las imágenes. Aunque no es imprescindible visualizar en tiempo real el estado del área de interés mientras se generan las alertas, contar con una GPU acelera, significativamente, la velocidad de procesamiento y análisis de imágenes.

En última instancia, es importante destacar que, aunque se han realizado trabajos similares en otros países, esta investigación proporciona un enfoque novedoso para el sistema de transporte panameño. Además, resalta la versatilidad y utilidad de las herramientas de visión artificial para abordar problemas de la vida cotidiana y los desafíos que enfrentan las organizaciones responsables del transporte en su rutina operativa.

Con base en lo expuesto, se pueden proponer algunas recomendaciones para trabajos futuros. Por ejemplo, se puede mejorar el sistema de reconocimiento entrenando el

modelo en tiempo real a medida que se realizan las pruebas, lo cual permitiría que aprenda a clasificar y diferenciar los estados del área de estudio de manera más precisa. Además, en el contexto del transporte en Panamá, se podría desarrollar un sistema que utilice las cámaras instaladas en los autobuses para gestionar de manera centralizada las alertas y optimizar la toma de decisiones en el centro de logística en tiempo real, aplicando también un sistema de manejo eficiente.

Referencias Bibliográficas

- Afonso, E., Thulliez, P., Pontier, D. y Gilot-Fromont, E. (2007). Toxoplasmosis in prey species and consequences for prevalence in feral cats: Not all prey species are equal. *Parasitology*, 134(14), 1963–1971. <https://doi.org/10.1017/S0031182007003320>
- Abbas, A., Jayaprakash, P. O., Anitha, M., & Jaini, X. V. (2017). Crowd detection and management using cascade classifier on ARMv8 and OpenCVPython. *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, 1–6. <https://doi.org/10.1109/ICIIECS.2017.8275988>
- Barbierato, E., Gribaudo, M., Iacono, M., & Levis, A. H. (2020). Evaluating the Safety of Crowds in Enclosed Spaces by Markovian Agents. *Electronic Notes in Theoretical Computer Science*, 353, 61–75. <https://doi.org/10.1016/j.entcs.2020.09.019>
- Baumann, D., Sommer, M., Schrempp, Y., & Sax, E. (2022). Use of Deep Learning Methods for People Counting in Public Transport. *2022 International Conference on Connected Vehicle and Expo (ICCVE)*, 1–6. <https://doi.org/10.1109/ICCVE52871.2022.9742924>
- Bhangale, U., Patil, S., Vishwanath, V., Thakker, P., Bansode, A., & Navandhar, D. (2020). Near Realtime Crowd Counting using Deep Learning Approach. *Third International Conference on Computing and Network Communications (CoCoNet'19)*, 171, 770–779. <https://doi.org/10.1016/j.procs.2020.04.084>
- Bilotta, E., Vaid, U., & Evans, G. W. (2018). Environmental Stress. En L. Steg & J. I. M. De Groot (Eds.), *Environmental Psychology* (pp. 36–44). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781119241072.ch4>
- Chaudhary, D., Kumar, S., & Dhaka, V. (2021). Estimating crowd size for public place surveillance using deep learning. *Deep learning and big data for intelligent systems*, 175–197. https://doi.org/10.1007/978-3-030-65661-4_9
- Correa, J. J. (2019). Identificación de factores de riesgo en eventos con aglomeración de personas en la ciudad de Bogotá. <http://hdl.handle.net/10654/32326>
- Dichev, C., & Dicheva, D. (2017). Towards Data Science Literacy. *International Conference on Computational Science, ICCS 2017, 1214 June 2017, Zurich, Switzerland*, 108, 2151–2160. <https://doi.org/10.1016/j.procs.2017.05.240>
- Galic, S., & Loncaric, S. (2000). Spatio-temporal image segmentation using optical flow and clustering algorithm. 63–68. <https://doi.org/10.1109/ISPA.2000.914892>

- Gao, G., Gao, J., Liu, Q., Wang, Q., & Wang, Y. (2020). *CNN-based Density Estimation and Crowd Counting: A Survey* (arXiv:2003.12783). arXiv. <http://arxiv.org/abs/2003.12783>
- Hidayat, A., Terabe, S., & Yaginuma, H. (2018). WiFi Scanner Technologies for Obtaining Travel Data about Circulator Bus Passengers: Case Study in Obuse, Nagano Prefecture, Japan. *Transportation Research Record: Journal of the Transportation Research Board*, 2672(45), 45–54. <https://doi.org/10.1177/0361198118776153>
- Lin, S.-F., Chen, J.-Y., & Chao, H.-X. (2001). Estimation of number of people in crowded scenes using perspective transformation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 31(6), 645–654. <https://doi.org/10.1109/3468.983420>
- Nogueira, V., Oliveira, H., Silva, A., Vieira, T., & Oliveira, K. (2019). RetailNet: A deep learning approach for people counting and hot spots detection in retail stores. 155–162. <https://doi.org/10.1109/SIBGRAPI.2019.00029>
- Sahoo, K., Samal, A. K., Pramanik, J., & Pani, S. K. (2019). Exploratory Data Analysis using Python. *International Journal of Innovative Technology and Exploring Engineering*, 8(12), 4727–4735. <https://doi.org/10.35940/ijitee.I3591.1081219>
- Surasak, T., Takahiro, I., Cheng, C., Wang, C., & Sheng, P. (2018). Histogram of oriented gradients for human detection in video. *2018 5th International Conference on Business and Industrial Research (ICBIR)*, 2018, 172–176. <https://doi.org/10.1109/ICBIR.2018.8391187>
- Wang, R., Tan, F., Yang, K., Hao, Y., Li, F., & Yu, X. (2023). LCNNet: Light-weight convolutional neural networks for understanding the highly congested scenes. *J. Intell. Fuzzy Syst.*, 45, 2. <https://doi.org/10.3233/JIFS-224081>