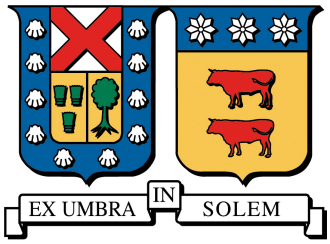
UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE ELECTRÓNICA



PLATAFORMA IOT PARA DETECCIÓN DE AGLOMERACIONES EN ESPACIOS PÚBLICOS CON ÉNFASIS EN METRO DE SANTIAGO

CELESTE ANAÍS RAMÍREZ CABEZAS

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERA CIVIL TELEMÁTICA

PROFESOR GUÍA
NICOLÁS SEBASTIÁN GÁLVEZ RAMÍREZ

PROFESOR CORREFERENTE
NICOLÁS VALENZUELA LEVI

Octubre 2026

CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): ☒Memoria o trabajo de título ☐ Tesis de Postgrado

Título del trabajo: Plataforma IoT para detección de aglomeraciones en espacios públicos con énfasis en Metro de Santiago

Nombre del candidato(a): Celeste Anaís Ramírez Cabezas

Carrera / Grado: Ingeniería Civil Telemática

Campus: San Joaquín **Departamento:** Electrónica

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, Nicolás Sebastián Gálvez Ramírez, en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente **DEJO CONSTANCIA** que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución.

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL (marcar una opción)

☒ El trabajo **NO contiene** información que amerite confidencialidad y puede ser publicado de inmediato en repositorio con acceso abierto.

☐ El trabajo **CONTIENE** información con potenciales implicancias de propiedad industrial o intelectual y requiere un periodo de confidencialidad (**embargo**) por **(marcar una opción)**:

☐ 6 meses ☐ 12 meses ☐ 2 años ☐ 3 años ☐ 5 años ☐ 10 años

Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):

_____

_____

_____

4.- FIRMAS

Profesor(a) guía o director(a) de memoria o tesis:

Fecha: 29 de octubre 2026 **Firma:** _____



Estudiante o Candidato(a):



Fecha: 29 de octubre 2026 **Firma:** _____

Este formulario debe ser insertado como página 2 de la memoria o tesis, completado y firmado por estudiante y profesor(a) antes de la entrega en portal PRISMA de Biblioteca USM.

Quiero dedicar este trabajo a mi mamá, por el amor y apoyo incondicional que me ha entregado toda la vida, y porque si llegué hasta acá, es gracias a ella.

Agradecimientos

En primer lugar, quiero agradecer a mi familia. A mi mamá, porque me dejaste ir lejos para que pudiera cumplir mis sueños, porque cuando pensé que ya no podía más, siempre creíste en mí y porque tus palabras y tu amor siempre me dieron fuerza para seguir intentándolo. A mi Abita, porque nunca me dejaste sola, porque con tu paciencia y amor estuviste apoyándome durante todos estos años de estudio. A mi hermana, mi Nanix, porque solo con tus abrazos sentía que todo estaba bien y no necesitaba nada más.

Le agradezco a mi novio, Julián. Gracias por estar en estos últimos años de universidad, gracias por tu amor, por tu compañía, por tu paciencia. Porque gracias a ti, lejos de mi casa y de mi familia, encontré un hogar. Porque en este último tiempo podía estar todo cuesta arriba, pero siempre estuviste tú y con eso es suficiente. Gracias por hacer la vida mucho más bonita.

A mis amigas de la vida. A Valentina, por todas esas noches que me acompañaste estudiando, por las veces que me escuchaste y me diste tu apoyo, porque más de alguna vez lloramos juntas. A Bárbara, por las veces que fuiste a la universidad a verme. Gracias por acompañarme y por abrirme las puertas de tu casa cuando lo necesitaba. Gracias a las dos por estar siempre conmigo.

A mis amigos, "los telemáticos", porque con buena compañía todo es mejor, y ellos me hicieron disfrutar esta carrera. Agradezco haber compartido con ustedes tareas, proyectos, laboratorios y cerritos. Los voy a llevar siempre en mi corazón.

A mis amigas "las Catas" porque entre tantos hombres, las encontré a ustedes, que han sido un apoyo dentro y fuera de la universidad, y le agradezco a Dios que las haya puesto en mi camino.

Finalmente le agradezco a mi profesor guía, Nicolás Gálvez, porque es una de las razones por las que estoy acá. Porque cuando llegué a la carrera, no creí que podía seguir, pero estuvo siempre diciéndome que era capaz. Le agradezco por su vocación, por su paciencia todas las veces que le pregunté en los primeros años y, sobre todo, le agradezco por creer en mí. Voy a atesorar los recuerdos con usted, desde la primera llamada en Zoom, hasta la última reunión por Teams. Gracias porque en usted no solo encontré un profesor, sino que también encontré a un increíble ser humano.

A todos ustedes, gracias por acompañarme en esta etapa, una de las más importantes de mi vida.

Resumen

Este trabajo aborda la problemática asociada a las aglomeraciones de personas en espacios públicos de alta demanda, donde las variaciones en la afluencia dificultan la supervisión y gestión oportuna de los flujos. Los mecanismos tradicionales de monitoreo, basados principalmente en la observación de cámaras de videovigilancia, pueden presentar limitaciones para obtener de manera automática información sobre la cantidad y distribución de personas presentes en un espacio. En este contexto, el propósito de esta memoria es desarrollar e implementar un sistema de monitoreo inteligente de afluencia de personas que permita generar información para apoyar la toma de decisiones operativas. Para su desarrollo se considera Metro de Santiago como caso de estudio.

La solución implementada integra distintas etapas asociadas a la captura, procesamiento, almacenamiento y visualización de información proveniente de fuentes de video. El procesamiento se realiza mediante técnicas de visión por computador e inteligencia artificial, utilizando YOLOv8 para la detección de personas y la obtención de información espacial mediante segmentación. A partir de estos resultados, el sistema genera dos indicadores principales: el conteo de personas y un índice de aglomeración que representa la proporción del área ocupada por las personas detectadas. La información obtenida se almacena en una base de datos `MySQL` y es gestionada mediante un `backend` desarrollado con `FastAPI`. Además, se implementa una interfaz web mediante `React` que permite visualizar las fuentes de video procesadas, consultar la evolución temporal de los indicadores, gestionar cámaras y alertas, y generar reportes históricos en formato `CSV`.

Para evaluar el comportamiento del sistema se utilizan grabaciones correspondientes al caso de estudio y a una fuente de prueba local, considerando distintos escenarios de presencia y distribución de personas. Las pruebas permiten analizar el conteo de personas y el índice de aglomeración frente a las condiciones presentes en cada escena. Además, el funcionamiento del mecanismo de alertas se verifica mediante umbrales configurables, comprobando la generación, almacenamiento y posterior gestión de los eventos desde la interfaz web.

Los resultados obtenidos permiten comprobar el funcionamiento integrado de los componentes desarrollados y la generación de información que puede ser consultada y analizada por el usuario.

Por lo tanto, el trabajo desarrollado constituye una aproximación al uso de tecnologías de visión por computador e inteligencia artificial para apoyar la gestión de afluencia de personas y establece las bases para futuras implementaciones en espacios públicos de alta demanda.

Palabras Clave— Monitoreo de aglomeraciones, visión por computador, detección de personas, YOLOv8, índice de aglomeración, afluencia de personas, inteligencia artificial.

Abstract

This work addresses the problem of crowding in high-demand public spaces, where fluctuations in pedestrian flow make timely supervision and management difficult. Traditional monitoring methods, which rely mainly on operators watching video surveillance cameras, have limited ability to automatically obtain information on the number and distribution of people in a given space. In this context, this thesis aims to develop and implement an intelligent pedestrian flow monitoring system that generates information to support operational decision-making, using the Metro from Santiago, Chile, as a case study.

The solution integrates the capture, processing, storage, and visualization of information from video sources. Processing relies on computer vision and artificial intelligence techniques, using YOLOv8 for people detection and segmentation to extract spatial information. From these results, the system produces two main indicators: the people count and a crowding index, defined as the proportion of the area occupied by detected people. The data is stored in a `MySQL` database and managed through a `backend` built with `FastAPI`. In addition, a web interface developed with `React` allows users to view the processed video sources, examine how the indicators evolve over time, manage cameras and alerts, and generate historical reports in `CSV` format.

To evaluate the system, recordings from the case study and a local test source are used, covering different scenarios in terms of the presence and distribution of people. The tests analyze the people count and the crowding index under the conditions of each scene. The alert mechanism is also verified using configurable thresholds, confirming that events are generated, stored, and subsequently managed through the web interface.

The results demonstrate that the developed components work together as an integrated system and generate information that users can consult and analyze. Overall, this work offers an approach to using computer vision and artificial intelligence to support pedestrian flow management and lays the groundwork for future implementations in high-demand public spaces.

Keywords— Crowd monitoring, computer vision, people detection, YOLOv8, crowding index, people flow, artificial intelligence.

Índice general

Índice de figuras **8**

Índice de tablas **10**

1. Introducción **11**

1.1. Propósito . . . 14

1.2. Objetivos . . . 14

1.2.1. Objetivo General . . . 14

1.2.2. Objetivos Específicos . . . 14

1.2.3. Contribuciones . . . 15

1.3. Estructura de la Memoria . . . 15

2. Formalización del Problema **17**

2.1. Contexto del Monitoreo en Espacios Públicos . . . 17

2.2. Problemática de las Aglomeraciones en Entornos Urbanos . . . 18

2.2.1. Dinámica de Multitudes en Eventos Deportivos . . . 18

2.2.2. Dinámica de Multitudes en Conciertos y Festivales . . . 20

2.2.3. Dinámica de Multitudes en Manifestaciones . . . 23

2.2.4. Dinámica de Multitudes en Aeropuertos . . . 24

2.2.5. Dinámica de Multitudes en Infraestructura de Transporte . . . 26

2.2.6. Dinámica de Multitudes en Espacios Públicos Abiertos . . . 28

2.3. Limitaciones de Métodos Tradicionales de Monitoreo . . . 30

2.4. Necesidades y Desafíos Asociados a Soluciones Inteligentes para la Gestión de Aglomeraciones 34

2.4.1. Necesidades Asociadas a Soluciones Inteligentes para la Gestión de Aglomeraciones . . . 34

2.4.2. Desafíos Técnicos Asociados a la Detección Automática de Personas . . . 36

2.5. Síntesis del Problema . . . 39

3. Estado del Arte **40**

3.1. Infraestructura para el Análisis de Información mediante Video . . . 40

3.1.1. Cámaras como Fuente de Información Visual . . . 41

3.2. Transmisión y Centralización de la Información . . . 43

3.3. Procesamiento y Análisis de Información . . . 45

3.3.1. Visión por Computador Aplicada al Análisis de Multitudes . . . 46

3.3.2. Análisis de Objetos en Secuencias de Video . . . 48

3.4. Resultados del Procesamiento para el Análisis de Aglomeraciones . . . 50

3.5. Visualización de Información en Sistemas de Monitoreo . . . 56

3.6. Sistemas Inteligentes de Monitoreo de Afluencia . . . 60

4. Diseño de la Solución **66**

4.1. Requisitos del Sistema . . . 66

4.1.1. Requisitos Funcionales . . . 66

4.1.2. Requisitos No Funcionales . . . 67

4.1.3. Requisitos de Interfaces . . . 68

4.1.4. Requisitos de Ambiente . . . 69

4.1.5. Perfil de Usuario . . . 72

4.2. Arquitectura del Sistema . . . 73

4.2.1. Diagrama de Contexto . . . 73

4.2.2. Diagrama de Arquitectura . . . 74

4.3. Módulos del Sistema . . . 75

4.3.1. Capa de Captura . . . 75

4.3.2. Capa de Procesamiento . . . 76

4.3.3. Capa de Gestión de Datos y Servicios . . . 78

4.3.4. Capa de Presentación e Interacción . . . 82

4.4. Matriz de Requisitos Funcionales y Módulos . . . 83

4.5. Diseño de la Interfaz de Usuario . . . 84

4.5.1. Modelo de Navegación . . . 84

4.5.2. Vistas del Prototipo . . . 85

4.6. Transición hacia la implementación . . . 88

5. Implementación de la Solución **89**

5.1. Caso de Estudio . . . 89

5.1.1. Herramientas Utilizadas . . . 90

5.2. Análisis del Flujo de Video 93
5.2.1. Resultados del Procesamiento 95
5.3. Gestión y Almacenamiento de Datos 96
5.4. Generación de Alertas 100
5.5. Desarrollo de la Interfaz Web 101
5.5.1. Vistas Principales del Sistema 102
5.6. Verificación y Validación 107
5.6.1. Verificación 107
5.6.2. Validación 108
5.7. Análisis Crítico y Evaluación de Riesgos 108
5.8. Síntesis de la Implementación 109

6. Conclusiones y Trabajo Futuro **110**
6.1. Cumplimiento de Objetivos 110
6.2. Limitaciones y Dificultades 111
6.3. Trabajo Futuro 112
6.4. Reflexión Final 113

Índice de figuras

2.1. Representación de un cuello de botella [15]. 19
2.2. Concierto de Shakira en Copacabana, 2026 [22]. 22
2.3. Manifestaciones realizadas en Plaza Baquedano, Santiago de Chile. 23
2.4. Diagrama del flujo de pasajeros en un aeropuerto. 25
2.5. Etapas de pasajeros en el aeropuerto Arturo Merino Benítez. 26
2.6. Ejemplos de medidas operacionales implementadas por Metro de Santiago en eventos de alta convocatoria [39], [40] 28
2.7. Mapa del Parque Metropolitano de Santiago [46]. 30
2.8. Recorridos de CicloRecreoVía en Santiago y Viña del Mar. 30
2.9. Operadores supervisando múltiples transmisiones de cámaras CCTV [50]. 32
2.10. Ejemplo visual de dificultades en el fondo/entorno [58]. 37

3.1. Flujo general de captura y procesamiento de información. 41
3.2. Flujo general para el análisis de objetos en secuencias de video. 49
3.3. Ejemplos de oclusión inter-clase e intra-clase. 50
3.4. Campo visual y delimitación del área de análisis en la calle del Miguelete, Valencia [95]. . . . 53
3.5. Representación de las distancias interpersonales utilizadas para establecer los niveles de ocupación [95]. 54
3.6. Representación en perspectiva y vista ortogonal del área de análisis en Puri, India [96]. . . . 55
3.7. Visualización de información asociada al conteo, densidad y velocidad en un sistema de monitoreo peatonal [97]. 57
3.8. Visualización de trayectorias y direcciones de desplazamiento peatonal en un cruce urbano [98]. 58
3.9. Formas de visualización de información en un sistema de monitoreo vehicular [99]. 59
3.10. Flujo de monitoreo de congestión [100]. 61
3.11. Flujo de información para el monitoreo de afluencia mediante Wi-Fi [101]. 62

3.12. Flujo utilizado para la detección y gestión de condiciones del tráfico [99] 63
3.13. Panel principal de StreetFlow [102]. 64

4.1. Diagrama de contexto del sistema. 74
4.2. Arquitectura general del sistema. 74
4.3. Diagrama de módulos del sistema. 75
4.4. Modelo de navegación del sistema. 85
4.5. Vista inicio de sesión. 86
4.6. Vista Panel Principal. 86
4.7. Vista monitoreo de cámaras. 87
4.8. Vista de métricas. 87
4.9. Vista gestión de cámaras. 88
4.10. Visualización de alertas. 88

5.1. Estructuración de componentes del sistema. 91
5.2. Procesamiento del flujo de video para la detección de personas. 94
5.3. Esquema relacional de la base de datos del sistema. 97
5.4. Notificaciones emergentes de alertas durante el monitoreo. 100
5.5. Generación y gestión de una alerta. 101
5.6. Vista implementada de inicio de sesión. 102
5.7. Vista implementada del panel principal. 103
5.8. Vista implementada para el monitoreo de las cámaras registradas. 104
5.9. Visualización de los indicadores de monitoreo para Metro Plaza Egaña. 105
5.10. Vista implementada para la gestión de cámaras. 106
5.11. Vista implementada para la consulta y gestión de alertas. 107

Índice de tablas

2.1. Comparación de flujos y densidades según tipo de evento en un estadio [16]. 20
2.2. Emociones en distintos eventos según estado del manifestante 24
2.3. Comportamientos observados en espacios verdes urbanos públicos según periodo del día [42]. 29
2.4. Comparación de métodos tradicionales utilizados para el monitoreo de multitudes 33
2.5. Resumen de necesidades y desafíos asociados a la gestión de aglomeraciones 38

3.1. Comparación de dispositivos comerciales para la captura de información visual, ordenados de menor a mayor precio referencial. 43
3.2. Comparación de configuraciones para la transmisión y centralización de información visual. . 45
3.3. Comparación de herramientas utilizadas para el procesamiento de imágenes y visión por computador. 48
3.4. Características utilizadas para el análisis de multitudes. 56
3.5. Relación entre módulos del flujo general y los sistemas inteligentes de monitoreo revisados. . . 64

4.1. Matriz de requisitos funcionales y no funcionales y módulos del sistema. 84

5.1. Herramientas y tecnologías utilizadas en la implementación de la solución. 90
5.2. Comparación de rendimiento entre configuraciones de YOLOv8 [116]. 92
5.3. Resultados obtenidos durante el procesamiento de los flujos de video. 96
5.4. Estructura de datos utilizada para el almacenamiento de cámaras. 98
5.5. Estructura de datos utilizada para el almacenamiento del historial de monitoreo. 98
5.6. Estructura de datos utilizada para el almacenamiento y gestión de alertas. 99
5.7. Estructura de datos utilizada para el almacenamiento de usuarios. 99

Capítulo 1

Introducción

El crecimiento continuo de la población mundial y la expansión urbana han transformado de manera significativa la dinámica de los espacios públicos. En la actualidad, las ciudades concentran grandes volúmenes de personas en centros comerciales, estaciones de transporte, recintos deportivos y espacios de uso cotidiano, generando desafíos importantes para quienes se encargan de administrar estos espacios y garantizar su seguridad y orden. En centros comerciales, los operadores enfrentan la problemática de atraer un flujo importante de visitantes, sin generar una sensación de hacinamiento, pues esto afecta negativamente la experiencia del cliente [1]. Por otro lado, en recintos deportivos, detectar y contabilizar la multitud resulta crucial para apoyar la toma de decisiones y la gestión de seguridad y control de flujo de los asistentes [2]. En el caso de las estaciones de metro, las aglomeraciones en andenes representan un desafío creciente que lleva su infraestructura al límite en momentos de mayor afluencia, con repercusiones directas en la percepción de seguridad y el comportamiento de riesgo de los pasajeros [3]. En este contexto, la gestión de multitudes surge como una disciplina de creciente importancia, que abarca situaciones donde la concentración de personas puede alcanzar niveles críticos en períodos acotados. Las características del espacio físico, el perfil de los usuarios y la naturaleza del evento generan condiciones diversas que exigen una integración de conocimientos en seguridad pública, gestión de emergencias y comunicación, para anticipar y responder ante situaciones de riesgo. Con el avance de la tecnología, enfoques más sofisticados permiten abordar este desafío de manera más eficiente, abriendo paso a soluciones que combinan análisis de datos, automatización y monitoreo en tiempo real [4].

Ante este escenario, planificar la gestión de la afluencia resulta indispensable en cualquier estrategia de control de flujos. La clave no es reaccionar a una aglomeración cuando ya se produce, sino que está en anticiparla. La planificación permite identificar y mitigar riesgos de forma proactiva, asignar los recursos

de manera eficiente, establecer canales de comunicación claros, desarrollar protocolos de respuesta ante emergencias y coordinar de manera efectiva a los distintos involucrados en la gestión de un espacio o evento. Estructurar y ejecutar adecuadamente un plan, mejora de manera significativa la seguridad, el orden y la gestión eficaz de multitudes en diversos contextos. Sin embargo, planificar no es suficiente cuando los flujos de personas son a gran escala y no son predecibles. En esos casos, es importante contar con herramientas que monitoreen el comportamiento de la multitud, además de generar información accionable para quienes deban tomar decisiones en el momento adecuado. Una gestión eficiente de multitudes no solo aumenta la seguridad, sino que también fortalece la confianza de los usuarios en la administración del espacio gestionado y permite utilizar los recursos disponibles de mejor manera [4].

El transporte público representa uno de los escenarios donde estas problemáticas se vuelven urgentes. En latinoamérica, éste cumple un rol central para la movilidad urbana, pero presenta dificultades asociadas a la cobertura, calidad del servicio, seguridad y accesibilidad. Entre los principales problemas identificados, se encuentra el hacinamiento en sistemas de buses de tránsito rápido, la pérdida de regularidad, la evasión del pago, la inseguridad y la falta de mantenimiento. Estas condiciones afectan la experiencia del usuario y la confianza en el sistema. Además, luego de la pandemia, aumentó la sensibilidad de los pasajeros frente a las multitudes, lo que conlleva a que la percepción de seguridad sea un factor aún más relevante para la gestión del transporte público [5].

En el caso de Chile, esta problemática se manifiesta en distintas escalas. Las ciudades pequeñas y medianas cuentan con limitaciones importantes para la planificación del transporte, ya que existen escasos estudios detallados de la movilidad y ausencia de información detallada para definir soluciones que se ajusten a la realidad urbana. Además, los servicios de buses y taxis colectivos no funcionan de manera complementaria, sino que se superponen en sus recorridos y compiten por la misma demanda, lo que genera una operación deficiente [6]. Esto puede aumentar la congestión y reducir la calidad del servicio. Por lo tanto, el transporte público chileno además de enfrentar el desafío de movilizar grandes volúmenes de usuarios, también tiene se enfrenta a la tarea de hacerlo de forma eficiente y con capacidad de respuesta frente a condiciones de alta demanda.

En el caso de la Región Metropolitana, el sistema de buses cubre aproximadamente cinco millones de usuarios distribuidos en 680 km^2 del área urbana, registrando mas de tres millones de viajes en un día laboral y con paraderos que en horario punta, pueden concentrar hasta 5.950 usuarios a nivel regional [7]. A lo anterior, se le suman los terminales interurbanos, donde el flujo se ve intensificado especialmente en horarios punta y en períodos de vacaciones, generando congestión vehicular, saturación del espacio público y condiciones de inseguridad para peatones y habitantes del sector [8]. En el caso de los aeropuertos, donde la concentración de pasajeros ocurre en zonas de check-in y controles de seguridad, gestionar la multitud es

importante para organizar los desplazamientos, reducir tiempos de espera y garantizar un nivel de servicio adecuado [9]. Por otra parte, sistemas como el Metro de Santiago, opera con siete líneas que movilizan a diario a 2,6 millones de personas [10], donde la concentración de pasajeros en andenes, accesos y pasillos se vuelve significativamente mayor y es fácil que alcancen niveles críticos en cuestión de minutos. Además, su infraestructura es subterránea, por lo tanto, agrega una capa adicional de complejidad en la gestión de emergencias, dado que las opciones de evacuación son limitadas y el tiempo de respuesta a estos incidentes es determinante. En este tipo de espacios, la capacidad de detectar una posible aglomeración de manera oportuna y actuar antes de que se vuelva un evento mayor resulta fundamental. En esta materia, los beneficios se traducen a una reducción de costos operativos, mayor comodidad de los pasajeros y mayor viabilidad a largo plazo del negocio de transporte [4], estos son argumentos que refuerzan la pertinencia de invertir en soluciones tecnológicas orientadas a este ámbito.

Sin embargo, los métodos tradicionales de monitoreo, basados en la observación directa por parte de personal humano, presentan limitaciones claras en términos de cobertura, continuidad y capacidad de respuesta. Un operador monitorea en promedio 40 cámaras de forma simultánea, donde el 45 % de los incidentes no son identificados luego de 10 minutos de haber ocurrido, y ese porcentaje escala al 95 % después de 22 minutos. Lo anterior, refleja que el enfoque tradicional está orientado a reaccionar después de ocurrido un incidente, en lugar de prevenirlo de manera proactiva [11].

Bajo esta premisa, combinar la tecnología, el análisis de datos y el conocimiento humano surge como un componente fundamental para que la gestión de emergencias en contextos de alta afluencia sea exitosa. Esta integración permite alertar a los operadores y autoridades de forma inmediata, para que se tomen las medidas pertinentes y sea posible garantizar la seguridad de los usuarios y del público general mediante el monitoreo en tiempo real y el análisis de datos [4]. Pero entonces, ***¿cómo es posible implementar un sistema de monitoreo que supere las limitaciones de enfoques tradicionales y permita una gestión proactiva del flujo de personas en espacios públicos de alta demanda?***

El presente trabajo propone el desarrollo de una plataforma de monitoreo inteligente orientada a la detección de aglomeraciones en espacios públicos, con especial énfasis en Metro de Santiago. La solución integra técnicas de inteligencia artificial y visión por computador, para procesar videos desde cámaras instaladas en estos espacios, con el objetivo de contar personas, estimar niveles de ocupación y generar reportes con información que facilite la gestión del flujo de usuarios y la prevención de situaciones de congestión.

1.1. Propósito

En este trabajo, se presenta el desarrollo de un sistema de monitoreo capaz de detectar aglomeraciones en espacios públicos utilizando cámaras y modelos de inteligencia artificial. Dicho sistema está pensado para procesar videos en tiempo real, pero su arquitectura no está limitada a la fuente de datos. Por lo tanto, es posible aplicarlo sobre videos grabados en distintos contextos, lo que permite realizar el conteo de personas en un espacio determinado y generar reportes con información que facilite el flujo de los usuarios.

1.2. Objetivos

El sistema desarrollado en este trabajo se orienta a los siguientes objetivos, que responden a la problemática de las aglomeraciones en espacios públicos y contribuyen a su gestión.

1.2.1. Objetivo General

Desarrollar e implementar un sistema de monitoreo inteligente de afluencia de personas en espacios públicos, que permita a entidades como operadores de transporte, gobiernos locales y organismos encargados de la gestión urbana, obtener información para apoyar la toma de decisiones operativas.

1.2.2. Objetivos Específicos

- Construir un conjunto de datos a partir de cámaras ubicadas en espacios públicos, con énfasis en Metro de Santiago, para entrenar un modelo de detección de personas.
- Implementar un modelo de detección de objetos, basado en visión por computador e inteligencia artificial, capaz de identificar y contar personas en entornos de alta densidad, mediante el uso de cámaras.
- Desarrollar una interfaz de monitoreo que permita visualizar en tiempo real el video procesado, el conteo de personas y la evolución del flujo a lo largo del tiempo.
- Generar reportes e indicadores de ocupación, variación temporal y aglomeraciones para apoyar la gestión de flujos y prevención de congestiones.

1.2.3. Contribuciones

Esta memoria contribuye al desarrollo de una plataforma inteligente de monitoreo de afluencia de personas en espacios publicos, integrando técnicas de inteligencia artificial y visión por computador para la detección, conteo y análisis de personas en tiempo real. Como parte del desarrollo, se construye un dataset con el objetivo de entrenar y validar el sistema, considerando imágenes representativas de escenarios con presencia de personas y posibles situaciones de aglomeración. A partir de este conjunto de datos, se entrena un modelo de detección de personas capaz de procesar información visual. Además, se desarrolla una aplicación web que permite la integración del modelo con una interfaz de monitoreo, facilitando la visualización de cámaras, el conteo de las personas detectadas, registros históricos, alertas de ocupación y generación de reportes. De esta manera, la solución no solo permite observar el estado actual de un espacio, sino que también permite disponer de información útil para apoyar la toma de decisiones operativas de seguridad.

La plataforma desarrollada aporta a las ciudades inteligentes y a la gestión eficiente de espacios públicos, pues permite transformar la información capturada por cámaras en datos que son procesables para la correcta supervisión de entornos con alta afluencia de personas. Esta solución puede ser aplicada en distintos escenarios, tales como, estaciones de metro, centros comerciales, estaciones de transporte, eventos masivos y otros lugares donde la concentración de personas requiere un monitoreo constante. En particular, en sistemas como el Metro de Santiago, esta herramienta puede contribuir a la detección temprana de aglomeraciones, apoyando la gestión del flujo de usuarios y la prevención de situaciones críticas.

Por último, la arquitectura propuesta constituye una contribución tecnológica escalable, pues permite la incorporación de nuevas cámaras, modelos de detección, métricas de análisis y funcionalidades asociadas a la gestión de afluencia, lo que posibilita que la plataforma pueda adaptarse a contextos distintos de monitoreo y evolucionar hacia sistemas más robustos de supervisión inteligente.

1.3. Estructura de la Memoria

La estructura de este documento está compuesta de la siguiente forma:

- El Capítulo 2 presenta la formalización del problema, profundizando en la motivación que da origen a la solución propuesta, describiendo el contexto y las necesidades específicas del entorno de espacios públicos.
- El Capítulo 3 se enfoca en el marco teórico y el estado del arte, abordando los conceptos fundamentales, técnicas y antecedentes relacionados que respaldan el desarrollo de este trabajo.

- El Capítulo 4 detalla la implementación de la solución, incluyendo la arquitectura del sistema, las etapas de desarrollo y las decisiones técnicas.
- El Capítulo 5 presenta los resultados obtenidos a partir de la implementación de la solución, incluyendo su análisis y evaluación mediante métricas relevantes.
- El Capítulo 6 expone las conclusiones del trabajo y resultados obtenidos a lo largo del desarrollo. Además, se identifican posibles mejoras y nuevos horizontes de investigación.

Capítulo 2

Formalización del Problema

En este capítulo se presenta la formalización del problema abordado en esta memoria, considerando el contexto actual del monitoreo de personas en espacios públicos y los desafíos asociados a la gestión de entornos con alta afluencia.

2.1. Contexto del Monitoreo en Espacios Públicos

En las últimas décadas, se ha visto un aumento significativo del fenómeno global de urbanización, según proyecciones de las Naciones Unidas (ONU) para el año 2050, la población que vive en áreas urbanas tendrá un aumento del 66 % [12]. Este crecimiento continuo implica que un volumen creciente de personas, haga su vida cotidiana en espacios compartidos, tales como estaciones de transporte público, centros comerciales, eventos deportivos, espacios públicos abiertos y zonas destinadas a eventos masivos. Como consecuencia de aquello, gestionar espacios públicos se vuelve un desafío complejo que exige nuevas herramientas de monitoreo, pues los flujos de personas varían según el horario, las condiciones del entorno y la actividad. Cuando la gestión no es eficiente o no existe preparación previa para escenarios de alta concurrencia, se pueden producir situaciones caóticas con consecuencias graves, incluyendo lesiones, fatalidades y daños a la propiedad pública [12], evidenciando la necesidad de comprender y gestionar de manera adecuada la concentración de personas en espacios públicos.

Monitorear espacios públicos abarca mucho más que solo observar personas. En entornos abiertos y cerrados, el análisis de multitudes es un área activa de investigación con aplicaciones en vigilancia. Dependiendo del contexto, las variables de interés pueden variar. En el transporte público, es relevante el conteo de personas y la estimación de densidad y flujo de pasajeros; en centros comerciales, la ocupación y distribución de las

personas; en eventos deportivos y masivos interesa identificar la concentración de los asistentes en entradas, salidas y zonas críticas. De la misma manera, en parques públicos y actividades recreativas el monitoreo permite conocer la afluencia y desplazamiento de las personas. Aunque estos escenarios son diferentes, existen variables en común: el conteo, la densidad, la distribución espacial y el flujo de personas.

A partir de estas variables, es necesario distinguir dos conceptos que a su vez son complementarios: el monitoreo de multitudes, orientado a la detección de personas, y el control de multitudes, enfocado en la predicción, toma de decisiones y aplicación de estrategias operativas para gestionar los eventos detectados [13]. Esta diferencia es relevante pues establece que el conteo de personas es solo el primer paso, ya que la información recopilada debe transformarse en datos que permitan a operadores anticipar situaciones críticas y tener una capacidad de respuesta eficiente.

2.2. Problemática de las Aglomeraciones en Entornos Urbanos

Las aglomeraciones en entornos urbanos pueden manifestarse en diversos contextos, dependiendo de la actividad desarrollada, el tipo de espacio y las condiciones del momento. Es por esto que resulta fundamental analizar los escenarios de alta afluencia como: recintos deportivos, conciertos, festivales, manifestaciones, aeropuertos, espacios públicos abiertos y sistemas de transporte público; ya que cada uno presenta riesgos y dinámicas particulares. A pesar de esto, todos estos escenarios comparten una misma problemática: la concentración elevada de personas genera congestión, pérdida de control de flujo, dificultades de evacuación y aumento de riesgo para los usuarios [12] [13].

A continuación, se revisan distintos casos que tienen por objetivo evidenciar las diferencias en cada una de estas situaciones, en las que puede ser presentado el mismo problema y la necesidad de contar con información para apoyar su gestión.

2.2.1. Dinámica de Multitudes en Eventos Deportivos

Los estadios y recintos deportivos forman parte de uno de los entornos más exigentes en términos de gestión de multitudes, pues concentran miles de personas en espacios reducidos y en intervalos de tiempo acotados. La evidencia histórica muestra que la ausencia de protocolos de control de acceso, puede terminar en tragedias de alta magnitud. En el año 2022, se registraron al menos 125 muertes en Indonesia, en el año 2023 se registraron 12 muertes en El Salvador y en el año 2024 se registraron 56 en Guinea [14]. Aunque existen casos que no presentan víctimas fatales, sí es necesario tenerlos en cuenta. En la final de la UEFA Champions League en el año 2022 los fallos en la gestión del acceso al Stade de France, produjeron congestiones que

pusieron en riesgo la integridad de los asistentes al evento, lo que evidencia que el problema no se encuentra limitado a escenarios extremos ni a recintos con infraestructura deficiente. Entender el por qué ocurren estas aglomeraciones implica entender los mecanismos físicos implícitos. En entornos con alta concurrencia, el flujo de personas se comporta de manera análoga a un fluido: cuando la densidad supera ciertos umbrales, la velocidad de desplazamiento cae drásticamente y el flujo se vuelve inestable. Este fenómeno se manifiesta en entradas, torniquetes, pasillos angostos, y es comúnmente llamado cuello de botella y ocurre cuando la geometría de un recinto fuerza la congestión de múltiples flujos en un espacio reducido. En estas zonas, los flujos generan fricción entre grupos que se desplazan en sentidos opuestos, y en condiciones de alta densidad pueden formarse arcos de bloqueo que interrumpen completamente el movimiento [14].



Figura 2.1: Representación de un cuello de botella [15].

Adicionalmente, los patrones temporales de llegada de los asistentes deben ser considerados, pues registros indican que los espectadores llegan entre 10 y 30 minutos previos al inicio del evento, lo que genera peaks de demanda de acceso que superan la capacidad de absorción de los puntos de entrada si es que no existe preparación previa. Este comportamiento produce una acumulación crítica de personas en los accesos, sobre todo cuando la presión por ingresar se vuelve mayor. Otro efecto a considerar al momento de la planificación de gestión de aglomeraciones, es el área corporal ocupada por cada individuo, su velocidad natural de desplazamiento y su tendencia a moverse en grupos que afectan directamente las tasas de flujo. Estudios realizados en un estadio en Reino Unido (no identificado en la publicación por razones de confidencialidad) indican que existen variaciones considerables según el tipo de evento que se realice. Los flujos máximos observados en condiciones de flujo libre alcanzaron 52,6 personas/m·min en partidos de fútbol, 42,7 personas/m·min en rugby y hasta 57,7 personas/m·min en un concierto de pop femenino. Estos valores se encuentran todos por debajo del máximo de 82 personas/m·min establecido en la sexta edición Guide to

Safety at Sports Grounds (Green Guide) publicada en 2018 y establecida como guía de referencia de seguridad en recintos deportivos del Reino Unido [16]. Esta diferencia sugiere que los parámetros normativos vigentes pueden sobrestimar las capacidades reales de flujo y que los modelos de planificación deben incorporar el perfil específico de cada evento para ser verdaderamente representativos.

En la Tabla 2.1 se muestran los valores de flujo y densidad, los cuales dependen del tipo evento y de la composición de multitud, reforzando la necesidad de considerar variables demográficas y conductuales en la planificación de accesos, salidas y evacuaciones en recintos deportivos.

Frente a estos desafíos, la simulación computacional y los protocolos operativos emergen como herramienta clave para anticipar y mitigar situaciones de riesgo antes de que ocurran [14].

Tipo de evento	**Composición predominante**	**Densidad máxima (pers/m^2)**	**Flujo máximo (pers/m·min)**
Partido de fútbol	Principalmente hombres adultos	1,20	52,6
Partido de rugby	Público más diverso y familiar	0,55	42,7
Concierto de cantante masculino	Parejas y grupos de adultos jóvenes	1,17	55,5
Concierto de artista pop femenina	Principalmente adolescentes	0,77	57,7

Tabla 2.1: Comparación de flujos y densidades según tipo de evento en un estadio [16].

2.2.2. Dinámica de Multitudes en Conciertos y Festivales

En conciertos y festivales, las aglomeraciones no se explican solo por la cantidad de asistentes, sino también por la forma en la que las personas se relacionan con el espacio, con el evento y con el resto del público. A diferencia de otros entornos cotidianos, estos eventos se construyen sobre una experiencia colectiva, en donde los asistentes comparten interés, emoción e identidad según el artista, espectáculo o festival. Esto puede hacer que la cercanía sea parte de la experiencia, lo que reduce la necesidad de mantener espacio personal y aumenta la confianza hacia otros asistentes del evento [17]. Por esta razón, una gran cantidad de personas puede no ser interpretada inmediatamente como riesgosa por el público. Esto vuelve más complejo el manejo de la multitud, ya que el riesgo no depende sólo del aforo permitido, sino también de cómo los asistentes perciben y aceptan la concentración de personas a su alrededor.

Además, la dinámica de las multitudes se encuentra condicionada por las características físicas del espacio en el que se desarrolla el evento. En recintos cerrados, la capacidad y distribución de público asistente

se encuentran determinadas por una infraestructura previamente definida, donde los accesos, pasillos, galerías y zonas destinadas al público condicionan la posibilidad de desplazamiento. Un ejemplo es el Movistar Arena de Santiago de Chile, recinto que recibe mas de 1,5 millones de personas al año y tiene una capacidad máxima aproximada de 15.500 personas, distribuida en butacas numeradas y sectores de cancha. Cuenta con diferentes configuraciones dependiendo del espectáculo y tiene un diseño que permite la evacuación del recinto en menos de 16 minutos [18]. En este tipo de espacios, una correcta gestión de los accesos es fundamental, ya que puede modificar la densidad y el comportamiento de los asistentes. Los corredores estrechos y bien estructurados favorecen la formación de filas y permiten que los asistentes identifiquen y respeten con mayor facilidad el orden de llegada. Por otro lado, cuando son corredores amplios frente a accesos reducidos, el orden de ingreso se ve afectado y pueden aparecer estrategias competitivas para avanzar. Este tipo de comportamiento puede intensificarse cuando existe una alta motivación por entrar al recinto, o cuando los asistentes quieren alcanzar una ubicación determinada, generando un aumento en la concentración de personas en los puntos de acceso. Por lo tanto, en los conciertos realizados en espacios cerrados, gestionar la aglomeración no depende únicamente del aforo permitido, sino también de cómo se organizan y distribuyen los flujos de asistentes durante el ingreso, permanencia y salida del evento [19].

En espacios abiertos, la dinámica de la multitud presenta características diferentes, ya que las personas tienen mayor superficie de desplazamiento y los flujos no se encuentran restringidos de la misma manera que en un recinto cerrado. Sin embargo, la formación de aglomeraciones puede darse de igual manera en un espacio más amplio. A medida que aumenta la densidad de personas, el comportamiento de la multitud puede cambiar y presentar fenómenos como la formación de carriles, oscilaciones en zonas de cuello de botella y obstrucciones en salidas. Cuando la congestión alcanza niveles críticos, el movimiento puede seguir patrones de parada y avance (del inglés *stop and go*), desplazamientos grupales involuntarios y en diferentes direcciones en zonas de alta densidad. De esta manera, incluso en eventos desarrollados en espacios abiertos, la concentración de asistentes en algunos sectores puede reducir la capacidad de desplazamiento y modificar el comportamiento general de la multitud [20].

Por otro lado, los conciertos gratuitos desarrollados en espacios públicos abiertos, forman parte de una dinámica diferente, pues la asistencia no se encuentra limitada por la capacidad asociada a la venta de entradas. En este tipo de eventos, pueden concurrir grandes volúmenes de personas, atraídas por el espectáculo y por formar parte de la experiencia colectiva, lo que hace más difícil estimar la cantidad de asistentes y su distribución en el espacio [20]. Un ejemplo corresponde al concierto gratuito realizado por Shakira en Copacabana - Brasil, el cual reunió aproximadamente 2 millones de personas y fue categorizado como el concierto más masivo de la historia protagonizado por una artista latina [21]. Por el contrario a un recinto cerrado, la multitud ocupa todo el espacio público, extendiéndose hacia sectores cercanos, generando grandes

flujos de llegada y salida que interactúan con la dinámica cotidiana de la ciudad, trasladando concentraciones de personas hacia las calles, accesos y sistemas de transporte. Por lo tanto, en eventos abiertos y gratuitos, la dinámica de la multitud depende de lo que ocurra en frente del escenario, de la forma en la que los asistentes llegan, se distribuyen y posteriormente abandonan el espacio utilizado para el evento.



Figura 2.2: Concierto de Shakira en Copacabana, 2026 [22].

Para gestionar esta problemática es necesario considerar factores físicos, operacionales y conductuales. Los organizadores deben planificar previamente la capacidad del recinto, la distribución de accesos y salidas, las rutas de evacuación, las zonas de atención médica, puntos de información y áreas que ayuden a descongestionar sectores con alta demanda, evitando la formación de cuellos de botella o zonas donde el público pierda capacidad de movimiento. Sin embargo, la planificación previa no es suficiente, ya que durante el desarrollo del evento los flujos de personas van cambiando por diversos factores. Es por esto, que la gestión en tiempo real se vuelve fundamental considerenado aspectos como identificación en los cambios de densidad, el movimiento de la multitud, la comunicación de instrucciones al público y la coordinación entre seguridad, producción y equipos de emergencia.

Desde esta perspectiva, la seguridad en conciertos y festivales se debe entender como un sistema con múltiples capas, donde se integra la planificación, evaluación de riesgos, control de accesos, monitoreo tecnológico, la educación del público, y la respuesta ante emergencias [23]. Con este enfoque, es posible comprender que un desastre mayor no es originado por una única falla, sino por la acumulación de debilidades en distintas barreras de prevención. Por lo tanto, los conciertos y festivales refuerzan la idea de contar con herramientas de apoyo, que anticipen aumentos en la densidad poblacional y que ayude con la toma de

decisiones en espacios públicos altamente concurridos.

2.2.3. Dinámica de Multitudes en Manifestaciones

A diferencia de los escenarios anteriores, las manifestaciones se caracterizan por la ausencia de un único organizador y de una infraestructura diseñada para recibir grandes cantidades de personas. En este tipo de eventos, la multitud ocupa un espacio público como una forma de expresión política y social. La toma de calles y plazas forma parte del medio principal para visibilizar las demandas de la ciudadanía y favorecer el encuentro entre quienes participan de la movilización.

Durante el estallido social ocurrido el 2019 en Chile, la apropiación del espacio permitió resignificar lugares emblemáticos de su capital, siendo el caso más representativo la Plaza Baquedano, rebautizada como Plaza Dignidad, además de transformar diversos elementos urbanos en símbolos del descontento social. En este contexto, el espacio público deja de ser solamente para la circulación de personas y pasa a convertirse en el elemento principal del desarrollo de la protesta [24].



(a) Estallido Social, Plaza Baquedano, 2019 [25]. (b) Marcha 8M, Plaza Baquedano, 2020 [26].

Figura 2.3: Manifestaciones realizadas en Plaza Baquedano, Santiago de Chile.

El comportamiento de la multitud en las protestas no ocurre al azar, sino que responde a la dinámica colectiva y a las condiciones del entorno en el que se desarrolla la manifestación. Las emociones influyen de manera directa en la conducta de los manifestantes, generando respuestas de confrontación, retirada o permanencia. Estas emociones pueden propagarse mediante contagio emocional, lo que conlleva a cambios rápidos en el comportamiento colectivo. Además, la tensión dentro de una protesta aumenta de manera progresiva antes de que aparezcan episodios de violencia y la configuración del espacio favorece o dificulta esa

transición [27].

Bajo esta premisa, la percepción de la justicia en el actuar policial, la comunicación, el respeto hacia los asistentes y la forma en la que se realizan las intervenciones influyen de manera directa en la respuesta de la multitud. Por otro lado, acciones percibidas como represivas incrementan la tensión social y favorecen el conflicto. Por lo tanto, la forma en la que se gestiona una protesta puede contribuir a disminuir o aumentar el riesgo de violencia [28].

De lo anterior, es posible notar que el escenario de las protestas corresponde a un entorno dinámico, pues el comportamiento colectivo depende de factores emocionales, sociales y operacionales que cambian de manera constante conforme se va desarrollando el evento. En consecuencia, anticipar la evolución del comportamiento de una manifestación se vuelve un trabajo difícil si hablamos de observación humana, lo que refuerza la idea de contar con herramientas de monitoreo que permitan gestionar la multitud y apoyen a una gestión más segura del espacio público.

La Tabla 2.2 resume las emociones principales que surgen en función del estado del manifestante y de diferentes eventos ocurridos durante una protesta. Es posible notar que el miedo se encuentra presente en la mayoría de las situaciones analizadas, mientras que la ira aparece principalmente cuando el manifestante nota acciones como intento de detención, evidenciando que las respuestas emocionales pueden modificar rápidamente la dinámica de la multitud.

Evento	**Pacífico**	**Violento**	**En retirada**
Incursión policial	Miedo	Miedo	Miedo intenso
Superado en número	Miedo	Miedo	Miedo intenso
Intento de arresto	Miedo intenso, ira	Miedo	Miedo intenso
Ser arrestado	Miedo intenso, ira	Miedo intenso	Miedo intenso
Estar rodeado	Miedo	Miedo	Miedo intenso

Tabla 2.2: Emociones en distintos eventos según estado del manifestante

2.2.4. Dinámica de Multitudes en Aeropuertos

Las aglomeraciones en aeropuertos se desarrollan en un entorno estructurado, donde el desplazamiento de los pasajeros sigue un recorrido definido. En este escenario, el comportamiento de los pasajeros se entiende como un flujo continuo de personas que avanza a través de distintos puntos de control, como se representa de manera general en la Figura 2.4.

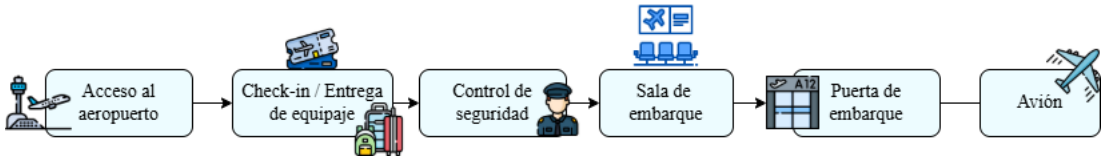


Figura 2.4: Diagrama del flujo de pasajeros en un aeropuerto.

El flujo inicia desde el acceso al terminal y continúa en etapas desde el check-in, control de seguridad finalizando en las puertas de embarque. Estas etapas son dependientes entre sí y su correcto funcionamiento permite mantener un tránsito continuo de pasajeros.

A medida que aumenta la cantidad de personas, la capacidad de atención de cada etapa cumple un rol fundamental, ya que cualquier situación que retrase la velocidad de atención puede generar acumulaciones y colas de espera. Esta situación ocurre principalmente en zonas de check-in y controles de seguridad, donde concurre la mayor parte de los pasajeros y en donde se desarrollan actividades que requieren de más tiempo de procesamiento (veáse Figura 2.5). Como consecuencia, una mala gestión de estos espacios genera mayor congestión y periodos de mayor concentración, influyendo en un aumento en los tiempos de espera y afectando la seguridad y experiencia de los usuarios [29].

Las variaciones en el flujo de pasajeros en el Aeropuerto Arturo Merino Benítez dependen de la época del año. Por ejemplo, en el año 2026, durante enero y febrero se registraron más de 5 millones de pasajeros, de los cuales aproximadamente 2,6 millones corresponden al primer mes del año, con distribución similar entre vuelos nacionales e internacionales [30]. Además, para el periodo de Navidad y Año Nuevo del año 2025 se concentran los flujos más altos de personas durante los días 23, 29 y 30 de diciembre proyectando más de 1,1 millones de pasajeros [31]. Lo anterior evidencia que en períodos de vacaciones y fechas de alta demanda se generan variaciones importantes en el volumen y distribución de los pasajeros.

En el caso particular de vuelos internacionales, el flujo de pasajeros se encuentra asociado a procesos de control migratorio. En Chile, esta labor está a cargo de la Policía de Investigaciones (PDI), institución que controla a las personas que ingresan o salen del país. Este proceso consta de la validación de documentos de identidad y detección de perfiles de riesgo [32]. Como estos controles forman parte del recorrido de los pasajeros cuando llegan a territorio nacional, la capacidad de atención de la policía adquiere relevancia en periodos de alta demanda, donde un aumento de flujo de personas puede provocar una congestión de pasajeros en este punto.

Por lo tanto el comportamiento de las multitudes en aeropuertos depende de la demanda de pasajeros y de la capacidad operativa del terminal, pues si se gestionan de manera correcta los recursos disponibles, la velocidad en la que los pasajeros avanzan a través del sistema no se ve afectada. La aparición de cuellos

de botella, tiempos largos de espera o grandes volúmenes de personas, puede comprometer la eficiencia y la seguridad de la operación. Es por esto que el monitoreo continuo de los flujos de pasajeros en este escenario se vuelve una herramienta fundamental para detectar de manera oportuna situaciones de congestión, optimizando la asignación de recursos, manteniendo un desplazamiento seguro y ordenado dentro del aeropuerto [9].



(a) Check-in de pasajeros en aeropuerto Arturo Merino Benítez [33].



(b) Embarque de pasajeros en aeropuerto Arturo Merino Benítez [34].

Figura 2.5: Etapas de pasajeros en el aeropuerto Arturo Merino Benítez.

2.2.5. Dinámica de Multitudes en Infraestructura de Transporte

La infraestructura de transporte es uno de los principales espacios en los que se pueden producir grandes concentraciones de personas debido al desplazamiento simultáneo de usuarios. Estaciones, terminales, paraderos y vehículos forman parte de una infraestructura en la que las personas circulan, esperan, ingresan o abandonan un medio de transporte. La cantidad de personas presentes en estos espacios varía según la demanda y el periodo en el que se realizan los desplazamientos.

Cuando aumenta la cantidad de pasajeros, los espacios que se destinan a la circulación y espera pueden concentrar mayor número de personas, lo que genera interacciones entre flujos que se desplazan en diferentes direcciones. Esto se puede dar en periodos habituales de alta demanda y en circunstancias particulares que aumentan temporalmente la cantidad de usuarios del sistema. La manera en la que se producen estas concentraciones depende del tipo de infraestructura y de las características del medio de transporte.

En el caso de transporte mediante buses, esta dinámica se manifiesta principalmente en los paraderos, durante el ascenso y descenso de pasajeros, en especial cuando la demanda se concentra en horarios determinados y la capacidad disponible se vuelve insuficiente. A medida que aumenta la cantidad de usuarios, aumenta la

interacción entre quienes esperan abordar, quienes permanecen en el bus y quienes descienden, dificultando el movimiento y aumentando los tiempos de parada [35].

Un ejemplo de esta situación se ha documentado en Concepción, Chile, donde estudiantes han reportado aglomeraciones entre las 18:00 y 20:00 horas, lo que coincide con la salida de trabajadores y estudiantes del sector. En este contexto, la cantidad de personas que usa el paradero genera dificultades asociadas a la capacidad del espacio para acoger a los usuarios que esperan el transporte público [36].

En los sistemas de transporte como trenes subterráneos, las aglomeraciones pueden producirse en distintos puntos de la infraestructura, incluyendo accesos, pasillos y andenes. En particular, las puertas de los trenes concentran flujos en direcciones opuestas, pues al igual que en los buses, existe interacción entre los pasajeros que quieren abandonar el tren y los que esperan ingresar. En condiciones de alta afluencia, esta interacción dificulta el abordaje, incrementa los tiempos de permanencia en los trenes y necesita de medidas de regulación para controlar la cantidad de personas que accede a los andenes [37].

Por otro lado, la realización de eventos masivos también genera aumentos significativos de pasajeros en sectores específicos de la ciudad. La llegada y salida de altas cantidades de personas en espacios acotados puede incrementar de manera simultánea la demanda en estaciones de metro, recorridos de buses y paraderos que se encuentran próximos a los recintos en los que se desarrollan este tipo de actividades. En consecuencia, la operación habitual del transporte puede necesitar modificaciones temporales orientadas a apoyar el desplazamiento de los asistentes y responder al aumento de usuarios.

Un caso particular y representativo corresponde a la estación de Metro en Santiago de Chile, Estadio Nacional, la cual se encuentra a los alrededores del recinto deportivo de mismo nombre, estación que fue diseñada considerando los flujos de pasajeros que se producen antes y después de un evento masivo. Esta estación cuenta con plataformas exteriores que permiten redirigir el desplazamiento de los asistentes y disminuir la concentración de personas en su interior [38]. Además, ante eventos de alta convocatoria se implementan refuerzos en recorridos de buses y paraderos del sector en el que se realiza el evento. Es necesario destacar que la concentración masiva de personas no ocurre solamente en el entorno del Estadio Nacional, pues otros sectores de Santiago presentan condiciones similares debido a la cercanía de recintos usados para actividades de alta convocatoria. En estos escenarios, Metro de Santiago puede extender su horario de funcionamiento y definir estaciones específicas para el ingreso y salida de pasajeros (véase Figura 2.6), lo que permite planificar de manera anticipada el desplazamiento de los asistentes y distribuir flujos entre distintos puntos de la red, contribuyendo a evitar concentraciones excesivas de pasajeros y posibles situaciones de congestión.

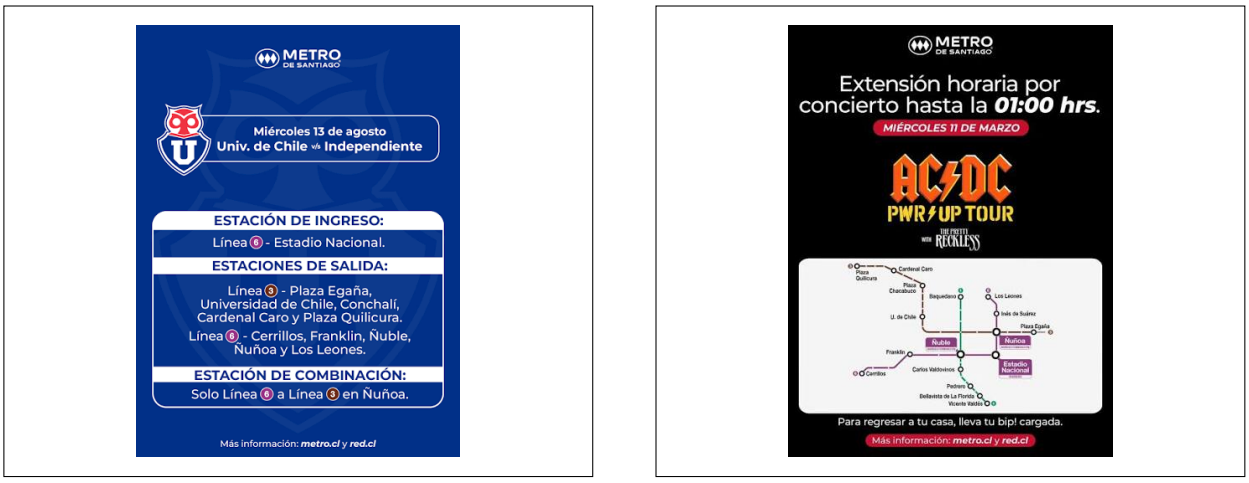


Figura 2.6: Ejemplos de medidas operacionales implementadas por Metro de Santiago en eventos de alta convocatoria [39], [40]

Por lo tanto, las aglomeraciones en la infraestructura de transporte pueden ocurrir tanto por variaciones en la demanda habitual como en situaciones que puedan modificar el funcionamiento del sistema. Su aparición en estaciones de metro, paraderos y buses vuelve necesaria la gestión de flujos de pasajeros para evitar niveles elevados de concentración.

2.2.6. Dinámica de Multitudes en Espacios Públicos Abiertos

Los espacios públicos son de libre acceso y forman parte de la vida cotidiana de las ciudades. Permiten a las personas permanecer libremente dentro de ellos y desarrollar diferentes actividades al aire libre, funcionando como lugares de encuentro, recreación y desplazamiento. Estos espacios comprenden parques, áreas verdes y sectores destinados a instalaciones recreativas.

La forma en que las personas utilizan estos espacios depende de la configuración física del entorno. Los senderos, las áreas abiertas y la vegetación influyen en la frecuencia de visita y la preferencia de actividades desarrolladas por los usuarios. La presencia de zonas de ejercicio, juegos, caminos y áreas de descanso atrae diversos grupos y favorece determinados usos del espacio [41].

En espacios verdes urbanos públicos, existen actividades asociadas al movimiento: caminar, correr, hacer ejercicio y además, existen actividades asociadas a la permanencia: leer, sentarse, descansar, conversar. Generando diversos comportamientos y en consecuencia, la ocupación de lugares puede variar temporalmente, pues las personas no necesariamente realizan una misma actividad durante toda la jornada [42] [43].

En la Tabla 2.3 se muestran variaciones en comportamientos observados en espacios verdes públicos

urbanos evidenciando que un mismo espacio puede presentar dinámicas diferentes a lo largo del día.

Comportamiento	**Mañana**	**Tarde**
Caminar	36 (64,7 %)	20 (35,3 %)
Atravesar el espacio	75 (94,9 %)	4 (5,1 %)
Ejercicio físico	26 (72,2 %)	10 (27,8 %)
Sentarse	60 (69,0 %)	27 (31,0 %)
Comer	15 (26,8 %)	41 (73,2 %)
Jugar	23 (56,1 %)	18 (43,9 %)
Leer	7 (29,2 %)	17 (70,8 %)
Estudiar	0 (0,0 %)	28 (100 %)

Tabla 2.3: Comportamientos observados en espacios verdes urbanos públicos según periodo del día [42].

En la ciudad de Santiago, el Parque Metropolitano y el Cerro San Cristóbal son un ejemplo de este tipo de espacios. En este entorno se combinan áreas verdes, senderos, y espacios para actividades al aire libre y cuenta con diferentes instalaciones y atractivos destinados a la recreación, permitiendo que dentro de un mismo parque coincidan personas que caminan, realizan actividad física, andan en bicicleta o se dirigen a puntos específicos del lugar: teleféricos, funiculares, jardín japonés, plazas de juegos, piscina Tupahue, etc. [44] (véase Figura 2.7).

Por otra parte, CicloRecreoVía, representa un uso temporalmente diferente del espacio. Es una iniciativa que tiene lugar en Santiago y Viña del Mar los días domingo, y que destina kilómetros de calles libres de vehículos motorizados (véase Figura 2.8). Esta iniciativa convoca aproximadamente a 35.000 personas cada fin de semana en Santiago y a 3.500 en Viña del Mar [45]. Este ejemplo evidencia que un espacio cotidianamente destinado al tránsito vehicular, puede concentrar grandes volúmenes de personas realizando diferentes actividades de manera simultánea al aire libre.

Por lo tanto, parques, áreas verdes y calles presentan usos que cambian según el sector, horario y las actividades realizadas. Comprender las variaciones del flujo de personas, resulta relevante para identificar la formación de aglomeraciones y su evolución en espacios urbanos de uso cotidiano.



Figura 2.7: Mapa del Parque Metropolitano de Santiago [46].



(a) CicloRecreoVía Santiago [45].



(b) CicloRecreoVía Viña del Mar [45].

Figura 2.8: Recorridos de CicloRecreoVía en Santiago y Viña del Mar.

2.3. Limitaciones de Métodos Tradicionales de Monitoreo

El monitoreo de multitudes permite obtener información de la cantidad de personas presentes en un espacio, su distribución y los patrones de movimiento que estas desarrollan. Lo anterior, adquiere especial relevancia en escenarios de alta concurrencia, donde los cambios en el comportamiento de las personas pueden modificar de manera rápida los las condiciones de seguridad. Para la obtención de estos antecedentes se han utilizado métodos basados en la observación, el conteo manual y sistemas de vigilancia mediante cámaras.

Dentro de los métodos tradicionales puede distinguirse principalmente el conteo manual de personas. El conteo manual depende directamente de operadores encargados de registrar o estimar la cantidad de personas

observadas en un espacio determinado. Una de las técnicas utilizadas es el método de Jacobs, útil para estimar la densidad de personas por unidad de superficie a partir de la cantidad de individuos que ocupan un área delimitada en una cuadrícula. Otra técnica, consiste en utilizar contadores manuales para registrar el número de personas que atraviesan puntos definidos, como accesos o salidas.

A pesar de que estos mecanismos continúan siendo útiles en espacios acotados, su uso presenta dificultades cuando aumenta la escala del espacio o de la multitud observada. Además, la información obtenida se ve condicionada según la zona que se esté observando. Por lo anterior, la capacidad de estas técnicas para presentar continuamente el estado general de una multitud, disminuye a medida que aumenta la complejidad del escenario.

Otra fuente relevante de información, constituye la observación directa mediante personal en terreno. Cuando se ejecuta una estrategia de gestión de multitudes, se debe observar el comportamiento de las personas para reconocer potenciales situaciones de emergencia y determinar el accionar correspondiente. Es por esto, que se necesitan guardias o personal encargado de supervisar el evento, pues la presencia de estos, permite observar de manera directa lo que ocurre en determinados sectores y comunicar situaciones que requieren una mayor atención [47].

La necesidad de mantener supervisión en áreas más grandes, ha impulsado el uso de herramientas capaces de ampliar la cobertura visual. Entre estas herramientas, el *Circuito Cerrado de Televisión (CCTV)* se convierte en uno de los mecanismos de vigilancia más utilizados en espacios públicos. Su uso permite observar de manera simultánea diferentes zonas mediante cámaras instaladas en puntos estratégicos, con el objetivo de entregar de manera continua información visual y complementando la vigilancia en terreno en los casos que sea necesario [48].

Aunque el CCTV permite superar de manera parcial las restricciones espaciales de observación, presenta una dificultad diferente, pues incorporar mayor número de cámaras, implica un aumento en la cantidad de información visual que debe ser observada. Por lo tanto, al generar grandes volúmenes de video, la evaluación manual de estos puede resultar compleja y ocupar demasiado tiempo cuando se observan escenarios de alta concentración de personas.

Cuando los sistemas de vigilancia dependen de la observación humana, los operadores deben distribuir su atención entre diferentes transmisiones y determinar cuáles necesitan una revisión más detallada (véase Figura 2.9). Las capacidades de atención sostenida y de memoria de trabajo son limitadas, por lo que alternar diferentes imágenes aumenta la posibilidad de pasar por alto situaciones que podrían ser relevantes durante la vigilancia [49].

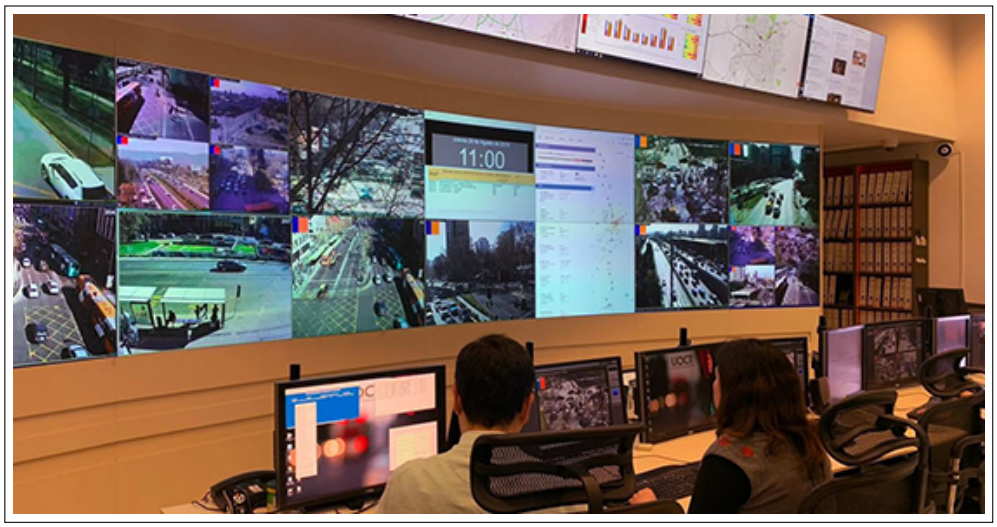


Figura 2.9: Operadores supervisando múltiples transmisiones de cámaras CCTV [50]

Por otro lado, el monitoreo mediante CCTV necesita un proceso de selección e interpretación de las imágenes y los operadores no mantienen el mismo nivel de observación sobre todos los grupos que aparecen de manera simultánea en las cámaras, sino que deciden en qué concentrar su atención y cuando un determinado comportamiento requiere de mayor seguimiento. Este proceso puede desarrollarse mediante vigilancia dirigida, manteniendo la atención sobre una persona o grupo determinado y así, seleccionar cuáles son las situaciones que requieren mayor atención y que deben ser reportadas. Según la experiencia de los operadores, pueden varias los niveles de eficiencia en la que distintas situaciones son identificadas. A lo anterior, se suman barreras como el tiempo transcurrido entre el reporte realizado en la sala de monitoreo y su despliegue en terreno [51].

La Tabla 2.4 sintetiza el funcionamiento, aporte y limitaciones del conteo manual, observación en terreno y CCTV con supervisión humana, con el objetivo de comparar las capacidades de cada método dentro del monitoreo de multitudes.

Método	**Funcionamiento**	**Principal aporte**	**Principales limitaciones**
Conteo manual [47]	Observadores estiman la cantidad de personas mediante técnicas como el método de Jacobs o contadores manuales en accesos y salidas.	Sencillo de implementar y puede utilizarse como referencia para validar otros sistemas de monitoreo.	Requiere intervención humana, presenta dificultades para escalar a grandes multitudes y su alcance se restringe a áreas o puntos específicos de control.
Observación en terreno [48]	Personal de seguridad supervisa directamente las condiciones de la multitud durante un evento.	Permite observar las condiciones del público e identificar situaciones que puedan requerir atención.	Su funcionamiento depende de la observación, interpretación y posterior comunicación de la información por parte del personal.
CCTV con supervisión humana [48], [51], [52]	Cámaras distribuidas en diferentes sectores transmiten imágenes que son observadas e interpretadas por operadores.	Permite ampliar la cobertura del monitoreo y observar diferentes sectores de un espacio.	El aumento de cámaras genera grandes volúmenes de información visual y mayores exigencias de supervisión. La atención, la selección de las imágenes observadas y las decisiones del operador pueden influir en el reporte de situaciones relevantes.

Tabla 2.4: Comparación de métodos tradicionales utilizados para el monitoreo de multitudes

En conjunto, los métodos tradicionales han permitido supervisar multitudes utilizando intervención humana y tecnológica. Por un lado, el conteo manual entrega mediciones útiles en zonas específicas, por otro lado, el personal en terreno aporta observación directa y por último, los sistemas CCTV permiten extender la cobertura del monitoreo. Sin embargo, al aumentar el tamaño del espacio o la cantidad de personas, estos métodos presentan dificultades para mantener una representación constante de las condiciones de la multitud. El problema no se encuentra sólo en obtener información, sino también en la forma de procesamiento, interpretación y de conversión a datos útiles para una correcta gestión.

2.4. Necesidades y Desafíos Asociados a Soluciones Inteligentes para la Gestión de Aglomeraciones

2.4.1. Necesidades Asociadas a Soluciones Inteligentes para la Gestión de Aglomeraciones

Con el desarrollo de la tecnología y con el objetivo de disminuir las limitaciones asociadas a los métodos convencionales de monitoreo, se ha impulsado el desarrollo de herramientas capaces de procesar de manera automática la información obtenida mediante cámaras. Bajo esta premisa, la visión por computador y el aprendizaje automático permiten disminuir la dependencia de los operadores y su observación continua, transformado las imágenes y videos capturados en información que puede ser útil para caracterizar una multitud. De esta manera, el monitoreo incorpora análisis, estimación de densidad, seguimiento y detección de comportamientos, permitiendo abordar distintas dimensiones de una aglomeración [53].

A continuación se enumeran las necesidades asociadas a estas soluciones:

1. **Análisis en tiempo real.** Incorporar estas herramientas resulta necesario ya que el comportamiento de las aglomeraciones es dinámico, pues la cantidad de personas puede variar temporalmente al igual que su distribución y forma de desplazamiento. Esto implica que se modifiquen de manera continua las condiciones del entorno, por lo que observar el espacio en un momento específico, entrega una representación puntual de una situación que puede evolucionar en minutos posteriores. El comportamiento dinámico genera la necesidad de contar con información actualizada mientras la aglomeración se desarrolla, por lo que es importante realizar un análisis continuo que entregue información en tiempo real a los responsables de la gestión, con el objetivo de contar con una gestión proactiva frente a situaciones que puedan requerir de atención [54]. De esta manera, las decisiones que deban tomar autoridades o personal de apoyo, se sustentan a partir de antecedentes obtenidos durante el desarrollo de una situación crítica y no después de que estas hayan sucedido.

2. **Detección temprana.** Si una situación es reconocida cuando la concentración de personas ya es evidente, las formas de intervención se ven limitadas. Por otro lado, conocer la manera en la que evoluciona la ocupación de un espacio permite concentrar la atención hacia sectores en los que se están produciendo cambios relevantes. Lo anterior, no implica necesariamente anticipar el comportamiento de una multitud, sino que es importante disponer de información suficiente para reconocer de manera temprana las condiciones que se están modificando.

3. **Distribución espacial.** Otro punto relevante, es que la cantidad total de personas no representa el

estado de una aglomeración, ya que dos espacios pueden contener una cantidad similar de individuos y presentar condiciones completamente diferentes dependiendo de la infraestructura del espacio. Una concentración en un acceso, pasillo o sectores reducidos, puede generar episodios distintos a otras en las que el número de personas se encuentra distribuido de manera uniforme.

4. **Conteo y densidad.** En este contexto, el conteo de personas y la estimación de densidad forman parte de dos tareas que resultan relevantes para el análisis automático de multitudes, determinando la cantidad de individuos presentes en una imagen o en un cuadro de video y representando la distribución espacial mediante un mapa asociado a la concentración de personas en distintas zonas de la imagen, respectivamente [55]. Esta información facilita la identificación de los sectores en los que existen mayores niveles de ocupación.

5. **Supervisión de espacios.** Contar con información sobre la cantidad y distribución de personas resulta útil cuando se usa en escenarios en los que la presencia de grandes cantidades de personas requiere de supervisión. En materia de seguridad pública, conocer el nivel de congestión de una multitud ayuda a identificar condiciones que requieran de mayor atención antes de que estas alcancen situaciones de riesgo. De la misma manera, en espacios de alta concurrencia, analizar el flujo de entrada y salida de personas permite estudiar el uso de los accesos y su relación con los tiempos de espera [55].

6. **Alertas automáticas.** Incorporar estas herramientas al monitoreo permite avanzar hacia sistemas capaces de generar información útil, como sistemas de monitoreo inteligente que alerten cuando una multitud alcanza niveles elevados de densidad. Desde este enfoque, las imágenes pasan a ser una fuente de datos desde la que pueden obtenerse indicadores útiles relacionados con la cantidad y concentración de personas [55].

7. **Registro histórico.** Esta integración permite mantener un registro de los datos acumulados, aportando antecedentes sobre horarios de mayor concurrencia, cambios en los niveles de ocupación o diferencias entre distintas jornadas. De esta manera, es posible comprender cómo se comportan los flujos de personas bajo diferentes condiciones, aportando información para la planificación y organización futura de un espacio.

8. **Ciudades inteligentes.** La evolución de estos sistemas se relaciona con el desarrollo de ciudades inteligentes donde la inteligencia artificial y el análisis de datos se pueden incorporar a diferentes procesos de gestión urbana y ayudar a comprender los cambios que ocurren dentro de una ciudad [56]. Desde esta perspectiva, la gestión de aglomeraciones es capaz de integrarse con otros procesos relacionados con el funcionamiento de los espacios urbanos.

2.4.2. Desafíos Técnicos Asociados a la Detección Automática de Personas

Incorporar nuevas tecnologías a la detección automática de personas presenta dificultades cuando se implementa en escenarios reales. La cantidad y distribución de personas, la posición respecto a la cámara, las características del entorno generan escenarios con diversos niveles de complejidad, logrando que los métodos de detección no dependan únicamente de reconocer personas.

A continuación se enumeran los desafíos técnicos asociados a esta detección:

1. **Oclusión:** Entre los principales desafíos se encuentra la oclusión, en particular, cuando existe una alta concentración de personas. A medida que aumenta la densidad de una multitud, los individuos empiezan a superponerse dentro del campo de visión de la cámara, provocando que algunas partes del cuerpo queden ocultas. Esta pérdida de información visual genera desafíos al identificar individualmente a cada persona, transformándolo en una tarea compleja para métodos que dependen de reconocer a cada individuo presente en la imagen [53] [57]. Además, el grado de oclusión puede variar dentro de una misma situación. Cuando se trata de sectores con menor concentración, las personas se diferencian de manera clara, por el contrario a zonas en las que se encuentran altamente congestionadas, sólo ciertas partes del cuerpo permanecen visibles. Aunque los métodos orientados a detectar el cuerpo completo tienen mejor funcionamiento en escenas dispersas, su desempeño disminuye cuando existe un aumento en la superposición de individuos [55]. Por lo tanto, aunque las técnicas utilizadas para analizar multitudes han evolucionado, una gran concentración de personas sigue siendo una dificultad relevante a la hora de obtener resultados precisos debido a la pérdida de información visual.

2. **Variaciones de iluminación:** Otro desafío a considerar son las variaciones de iluminación que se generan cuando las condiciones de captura cambian, incluso cuando la ubicación de la cámara y el espacio monitoreado no lo haga. Diferencias entre la intensidad y características de la luz pueden modificar la imagen, generando variaciones de color, brillo y contraste que se deben considerar durante el análisis y que pueden cambiar entre distintos momentos del día [57]. Por lo tanto, mantener un desempeño adecuado frente a diferentes condiciones de iluminación presenta un desafío para aplicar métodos automáticos en escenarios reales [57].

3. **Condiciones ambientales:** Además, las condiciones ambientales también alteran la calidad de las imágenes, en especial cuando el monitoreo se realiza en espacios abiertos. Condiciones como la lluvia, la niebla pueden producir desenfoque y disminuir la nitidez de la escenas, mientras que otras condiciones pueden incorporar interferencias visuales en el fondo, aumentando la dificultad para distinguir de manera correcta los elementos presentes en la imagen [57].

4. **Posición respecto a la cámara:** Por otro lado, la profundidad de campo, genera cambios en la forma en la que las personas se ven representadas. Si un individuo se ubica en un sector lejano a la cámara, ocupa una proporción menor que otro que está situado más cerca, generando que existan representaciones en escalas diferentes y aumentando la dificultad para lograr una identificación uniforme. Este problema adquiere mayor relevancia cuando una cámara cubre áreas extensas, pues las diferencias pueden provocar que algunos individuos aparezcan con más nivel de detalle que otros. Es por esto que el ángulo de instalación y la visión de la cámara influyen en las condiciones bajos las que debe realizarse la detección y los métodos de análisis [57].

5. **Fondo y entorno:** Otra dificultad se refleja en las características propias del entorno. Cuando el fondo presenta colores, texturas o patrones se combinan con las características de las personas identificadas en la imagen, generando falsos positivos y afectando los resultados obtenidos [57]. Lo anterior demuestra que los errores de detección también se asocian a la capacidad de diferenciar objetos y estructuras presentes en el espacio monitoreado. Una zona despejada presenta condiciones diferentes a otras que pueden tener estructuras, señaléticas o elementos distribuidos alrededor de la escena (véase Figura 2.10).

6. **Combinación de dificultades:** Las dificultades mencionadas pueden combinarse entre sí, aumentando la complejidad de una escena. Por ejemplo, una persona que está parcialmente oculta y se ubica en un fondo complejo, presenta condiciones diferentes a las de una persona que se encuentra cercana a la cámara y es completamente visible.

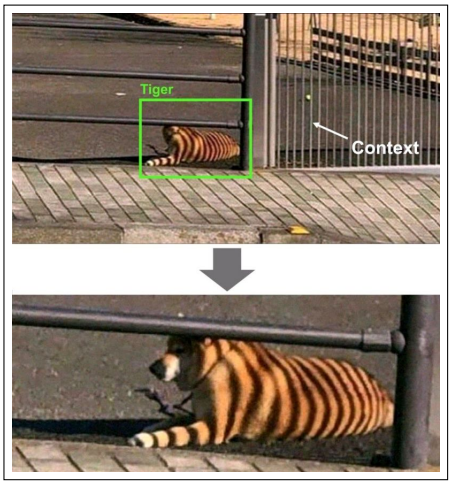


Figura 2.10: Ejemplo visual de dificultades en el fondo/entorno [58].

En la Tabla 2.5 se resumen las necesidades y desafíos asociados a la gestión de aglomeraciones.

Tipo	**Ítem**	**Descripción**
Necesidad	Análisis en tiempo real	Monitorear de forma continua para actuar de manera proactiva ante situaciones críticas.
	Detección temprana	Identificar cambios relevantes antes de que la concentración de personas sea evidente.
	Distribución espacial	Considerar cómo se distribuyen las personas en el espacio, no solo su cantidad total.
	Conteo y densidad	Determinar la cantidad de personas y su concentración en distintas zonas de la imagen.
	Supervisión de espacios	Apoyar la seguridad pública y el análisis de flujos en espacios de alta concurrencia.
	Alertas automáticas	Generar avisos cuando una multitud alcanza niveles elevados de densidad.
	Registro histórico	Acumular datos que permitan planificar y organizar el uso futuro de un espacio.
	Ciudades inteligentes	Integrar la gestión de aglomeraciones con otros procesos de gestión urbana.
Desafío	Oclusión	Superposición de personas que oculta partes del cuerpo y dificulta identificarlas individualmente.
	Variaciones de iluminación	Cambios de luz que alteran color, brillo y contraste de la imagen a lo largo del tiempo.
	Condiciones ambientales	Lluvia, niebla u otros factores que reducen la nitidez y agregan interferencias visuales.
	Posición respecto a la cámara	Diferencias de escala y nivel de detalle según la profundidad de campo y el ángulo de la cámara.
	Fondo y entorno	Colores, texturas o elementos del entorno que generan falsos positivos en la detección.
	Combinación de dificultades	Superposición de varios factores (oclusión, fondo complejo, distancia) en una misma escena.

Tabla 2.5: Resumen de necesidades y desafíos asociados a la gestión de aglomeraciones

En conclusión, la detección automática de aglomeraciones en escenarios reales, debe considerar las características de las multitudes y las condiciones bajo las que se obtiene la información. La oclusión limita la

visibilidad en escenarios complejos, la iluminación y las condiciones ambientales pueden modificar la apariencia de las imágenes y las características del fondo y la perspectiva genera diferencias de escala y pueden producir detecciones erradas. Estos factores forman parte de un desafío relevante para la implementación de sistemas de visión por computador dirigidos al monitoreo de aglomeraciones y se deben considerar al momento de evaluar su aplicación en condiciones reales.

2.5. Síntesis del Problema

Las aglomeraciones en espacios públicos se desarrollan bajo distintas condiciones y pueden cambiar según las características del entorno, la actividad realizada y la forma en la que las personas se distribuyen. Aunque cada escenario presenta condiciones especiales, tienen una problemática en común asociada a las grandes concentraciones de personas y a los cambios que pueden producirse durante el desarrollo de un evento. Como las aglomeraciones son de carácter dinámico, surge la necesidad de contar con información actualizada para reconocer variaciones en la ocupación de un lugar y gestionar espacios de alta afluencia.

A partir de lo señalado, la problemática abordada se centra en la necesidad de contar con herramientas que sean capaces de complementar los mecanismos existentes de monitoreo. Bajo este contexto, es necesario retomar la pregunta que guía el desarrollo de este trabajo: ***¿cómo es posible implementar un sistema de monitoreo que supere las limitaciones de enfoques tradicionales y permita una gestión proactiva del flujo de personas en espacios públicos de alta demanda?***

Capítulo 3

Estado del Arte

En este capítulo se revisa la evidencia científica de metodologías utilizadas para el desarrollo de sistemas inteligentes de monitoreo y sus aplicaciones. Es necesario organizarlas en base a distintas etapas de obtención, procesamiento y gestión de la información. Esta revisión permite conocer los métodos utilizados en la actualidad y las capacidades que ofrecen para el análisis y monitoreo de aglomeraciones. De esta manera, se fundamentan las decisiones asociadas al desarrollo de la solución propuesta.

3.1. Infraestructura para el Análisis de Información mediante Video

Los sistemas de videovigilancia utilizan cámaras instaladas en distintos puntos de un entorno con el objetivo de capturar contenido visual de las áreas que se encuentran bajo supervisión. Las secuencias de video obtenidas pueden ser utilizadas como fuente de información para tareas de detección, conteo de personas y estimación de multitudes, incorporando técnicas de visión por computador [59].

Sin embargo, la captura constituye sólo el punto de entrada de los datos, pues para obtener y presentar información a partir de secuencias de video, un sistema de monitoreo puede incorporar etapas de captura, transmisión, procesamiento, resultados del procesamiento y visualización de información [60]. La Figura 3.1 presenta el flujo general de estos módulos y la relación entre ellos.

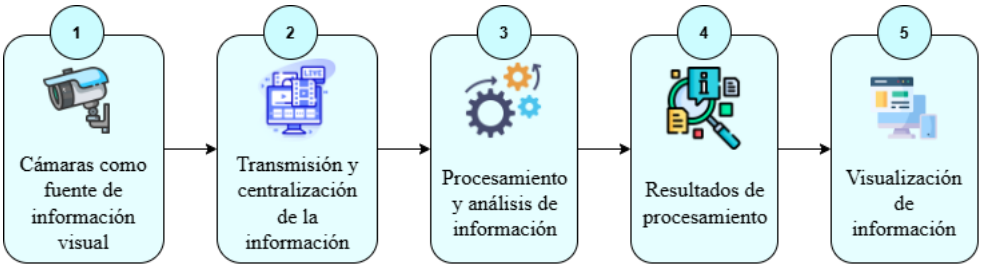


Figura 3.1: Flujo general de captura y procesamiento de información

A continuación, se detallan todos los módulos de la Figura 3.1 y su aplicación en videovigilancia.

3.1.1. Cámaras como Fuente de Información Visual

La obtención de información visual forma parte del punto de entrada de los sistemas de análisis de información de video, este puede realizarse mediante distintos tipos de cámaras y su configuración depende de las condiciones del entorno y de la aplicación en la que se utilicen. Las alternativas que se encuentran disponibles se diferencian por la resolución, frecuencia de imagen, campo de visión y por la forma en la se integran en las etapas posteriores de los sistemas de procesamiento. Algunas cámaras se utilizan para obtener y transmitir secuencias de video que se procesan con recursos computacionales externos, otras tienen la capacidad de operar en conjunto con dispositivos ubicados en el borde de la red y otras pueden incorporar capacidades de procesamiento directamente en el dispositivo. De esta manera, la captura y el análisis puede ser distribuido en distintos componentes de la infraestructura según los requerimientos del sistema [61].

Una de estas configuraciones corresponde al uso de cámaras tradicionales como fuente de video. En particular, una cámara estándar CCTV puede capturar secuencias de video para procesarlas posteriormente a través de equipos con GPU (del inglés, *Graphics Processing Unit*), en los que se ejecutan modelos de detección y algoritmos de seguimiento para la localización de personas, conteo y análisis de su desplazamiento. Con esta configuración, es posible incorporar capacidades de análisis en cámaras tradicionales sin necesidad de utilizar un dispositivo que cuente con los recursos computacionales necesarios para ejecutar este tipo de algoritmos [62].

Otro tipo de configuración mantiene las cámaras de videovigilancia como fuente de información, pero el procesamiento se hace mediante dispositivos *Edge* ubicados cerca del lugar de captura. Con este método es posible realizar de manera local las tareas de detección, seguimiento y conteo de objetos presentes en el entorno tales como, vehículos y peatones. Cuando las imágenes son procesadas, la información resultante es la transmitida, lo que evita el envío de fotogramas hacia servidores externos. Así, es posible usar menos

requerimientos de transmisión, mantener la información localizada y aprovechar las cámaras que ya se encuentran instaladas en el entorno que se desea monitorear [63].

Agregar nuevas capacidades computacionales en dispositivos cercanos a los dispositivos de captura, permite extender estas configuraciones hacia sistemas que puedan ejecutar de manera local los modelos de inteligencia artificial. Para lograrlo, es posible utilizar procesamiento Edge en conjunto con otras plataformas de procesamiento orientadas a inteligencia artificial, que logran acelerar tareas de visión por computador y pueden realizar tareas de detección, clasificación, seguimiento y reconocimiento de eventos sin la dependencia de recursos computacionales centralizados. Por otro lado, la captura puede realizarse mediante tecnologías diferentes a las de las cámaras tradicionales, lo que incluye cámaras térmicas y de visión nocturna, permitiendo obtener información visual en distintas condiciones del entorno [64].

Estas configuraciones se pueden encontrar en alternativas comerciales. Existen dispositivos encargados que permiten capturar y transmitir video en recursos computacionales externos que se encargan de su procesamiento. Como ejemplo comercial de esta categoría se encuentran Logitech C922, cámara con propósito general que captura video en resolución Full HD a 30 fps (del inglés *frames per second*, fotogramas por segundo) utilizando conexión USB [65].

Otra alternativa son las cámaras de acción, que se caracterizan por su portabilidad y su capacidad de realizar capturas en escenarios más complejos. Este tipo de cámaras se puede incorporar en sistemas de visión por computador para el análisis de espacios públicos, usando una cámara GoPro como fuente de video y un dispositivo Edge para ejecutar de manera local tareas de procesamiento. Al existir conexión inalámbrica entre ambos dispositivos, es posible trasmitir el video hacia el encargado de ejecutar los modelos, manteniendo separadas las funciones de captura y procesamiento de la información [66]. Como ejemplo comercial de este tipo de dispositivos se encuentra la GoPro HERO 13, capaz de capturar video en resolución 5.3K a 60 fps, con conexión Wi-Fi 6 para transferir información [67].

Las cámaras 360 representan otra alternativa para obtener información de los alrededores del dispositivo de captura. Su uso en sistemas de análisis de video permite extender la cobertura espacial y puede utilizarse en detección de peatones, intersecciones viales y otros elementos presentes en el entorno, con el objetivo de observar distintas direcciones de circulación desde una misma ubicación, y analizar los objetos presentes en la escena [68]. Como ejemplo comercial de esta categoría se encuentra la Insta360 X5, que usa dos lentes para obtener videos en 360° y permite realizar capturas con resolución máxima de 8K a 30 fps [69].

Por último, existen cámaras que añaden recursos especializados para realizar tareas de análisis de manera directa en el dispositivo de captura. Con esta configuración es posible ejecutar tareas sin la necesidad de depender de recursos computacionales [64]. Un ejemplo comercial es la cámara AXIS Q1805-LE, que cuenta

con resolución de 1080p e incorpora análisis basado en aprendizaje profundo, además cuenta con estabilización electrónica de la imágen y funciones de ciberseguridad integradas [70].

Los ejemplos comerciales presentados permiten observar las características de distintos dispositivos utilizados para la obtención de información visual. Mientras que algunas cámaras están orientadas principalmente a la captura y transmisión de video, otras permiten añadir capacidades de análisis en el mismo dispositivo. Con estas diferencias es posible comparar las alternativas consideradas según sus características de captura, conectividad, capacidad de procesamiento y precio referencial, las cuales se sintetizan en la Tabla 3.1.

Dispositivo	**Tipo**	**Resolución / FPS**	**Conectividad**	**Procesamiento**	**Precio ref. (CLP)**
Logitech C922 [65]	Propósito general	Full HD / 30 fps	USB	Procesamiento externo	$112.990 [71]
GoPro HERO13 [67]	Cámara de acción	5.3K / 60 fps	Wi-Fi 6	Procesamiento externo / Edge	$441.990 [72]
Insta360 X5 [69]	Cámara 360°	8K / 30 fps	Wi-Fi / Bluetooth	Procesamiento externo	$539.990 [73]
AXIS Q1805-LE [70]	Cámara IP con analítica	1080p	Ethernet / PoE	Analítica integrada	$2.536.785 [74]

Tabla 3.1: Comparación de dispositivos comerciales para la captura de información visual, ordenados de menor a mayor precio referencial.

3.2. Transmisión y Centralización de la Información

Cuando se obtiene la información visual, las secuencias de video deben transferirse hacia los recursos que se encargan de su procesamiento, su forma de transferencia puede variar según la ubicación de los componentes y el medio que se emplea para establecer la comunicación. En sistemas de video se utilizan conexiones cableadas e inalámbricas y los recursos computacionales se pueden distribuir entre la cámara, dispositivos Edge, servidores locales, centros de datos y servicios en la nube, permitiendo que las funciones del sistema se ejecuten en diferentes niveles de la infraestructura [61].

Para el procesamiento de video local, una de las configuraciones utilizadas consiste en conectar directamente el dispositivo de captura al recurso computacional que se encarga de recibir la información. Este tipo de configuración se implementa mediante cámaras USB conectadas a computadores de placa única, lo que permite obtener el flujo de video sin necesidad de añadir una red entre ambos componentes. Un ejemplo de este tipo de sistema corresponde al uso de una cámara Logitech C922 conectada mediante USB una Raspberry Pi 4, desde donde se obtiene el video en tiempo real y luego se procesan las imágenes capturadas [75].

Por otro lado, en sistemas donde las cámaras y los recursos de procesamiento se conectan mediante una red, la información puede ser recibida a través de transmisiones de video que vienen de cámaras IP. Para lograr la comunicación se utilizan protocolos RTSP (del inglés *Real Time Streaming Protocol*), a través del cual un dispositivo puede acceder al flujo de una cámara y obtener de manera consecutiva los fotogramas para que estos sean procesados. Esta configuración se implementa utilizando cámaras IP conectadas a una red local y dispositivos Edge que se encargan de recibir y procesar las transmisiones RTSP. Además, permite integrar diferentes cámaras en una misma infraestructura y así, mantener múltiples flujos de video de manera simultánea [76].

Cuando aumenta la cantidad de cámaras y de información generada, es posible incorporar en la infraestructura dispositivos que se encarguen de la centralización de las transmisiones recibidas. Un ejemplo corresponde a los NVR (del inglés *Network Video Recorder*) que conforman una alternativa que se utiliza para recibir información desde cámaras IP e integra funciones de almacenamiento y visualización de un video en tiempo real. Es posible observar su aplicación en sistemas de video vigilancia, compuesto por múltiples cámaras conectadas mediante una red, donde las grabaciones se pueden concentrar inicialmente en estos dispositivos y posteriormente transferirse hacia otros sistemas almacenamiento cuando es necesario contar con mayores capacidades para conservar y recuperar la información [77].

En infraestructuras de mayor escala, la centralización se extiende a sistemas de gestión de video o VMS (del inglés *Video Management System*). Estos sistemas permiten la administración conjunta de múltiples cámaras, servidores, usuarios y configuraciones, distribuyendo funciones entre componentes especializados. Implementaciones de este tipo agregan servidores de gestión que se encargan de la configuración general y de servidores de grabación que reciben y almacenan las transmisiones de las cámaras, mientras que distintos clientes pueden acceder a los recursos que son administrados por el sistema. De esta forma, VMS permite ampliar la centralización desde el almacenamiento de las secuencias hacia la gestión de los diferentes componentes que forman parte de la infraestructura de video [78].

Con estas configuraciones, los sistemas actuales son capaces de distribuir sus recursos de procesamiento en distintos niveles de la infraestructura. El procesamiento mediante dispositivos Edge permite ejecutar operaciones antes de que se transfiera la información hacia los recursos externos, disminuyendo el volumen de datos que pasa por la red y la dependencia de conexiones de mayor alcance. Los servidores locales concentran el procesamiento de múltiples cámaras en una misma red, mientras que los servicios en la nube aportan una mayor capacidad de almacenamiento a cambio de mayor dependencia de conectividad. Es por esto, que también se utilizan arquitecturas híbridas *Edge-Cloud*, en donde las tareas se distribuyen entre recursos cercanos a la captura y recursos remotos según los requerimientos [61].

Las implementaciones revisadas en conjunto, muestran distintos niveles de transmisión y centralización

de la información, desde conexiones directas entre una cámara y un recurso local hasta infraestructuras capaces de integrar múltiples cámaras, servidores y recursos remotos. La elección entre estas modifica la forma en que el video es transferido, administrado y puesto a disposición en las etapas posteriores del procesamiento.

En la Tabla 3.2 se sintetizan las principales características de las configuraciones revisadas para la transmisión y centralización de la información, considerando el medio utilizado, el nivel de centralización y la ubicación de los recursos.

Configuración	**Medio de transferencia**	**Centralización**	**Ubicación del procesamiento**	**Aplicación observada**
Conexión directa [75]	USB	No requerida	Equipo local	Transferencia directa entre la cámara y el recurso computacional.
Cámara IP / RTSP [76]	Red IP	No requerida	Edge o servidor	Transmisión de uno o múltiples flujos de video hacia recursos de procesamiento.
NVR [77]	Red IP	Centralizada	Local	Recepción, almacenamiento y visualización de información proveniente de múltiples cámaras IP.
VMS [78]	Red IP	Centralizada	Distribuida	Gestión conjunta de cámaras, servidores, usuarios y transmisiones de video.
Arquitectura Edge-Cloud [61]	Red local / Internet	Distribuida	Edge, servidor local y/o Cloud	Distribución del procesamiento y almacenamiento entre distintos niveles de la infraestructura.

Tabla 3.2: Comparación de configuraciones para la transmisión y centralización de información visual.

3.3. Procesamiento y Análisis de Información

Procesar información visual permite transformar imágenes o secuencias de video en datos que pueden ser utilizados posteriormente para análisis. Para esto, se pueden incorporar distintas operaciones en un flujo de procesamiento que incluye: preprocesamiento, extracción de características, detección y postprocesamiento. Estas pueden ser implementadas utilizando herramientas orientadas al procesamiento de información visual.

Ejecutar estas operaciones se encuentra condicionada por los recursos computacionales disponibles. Los modelos de aprendizaje profundo utilizados para el análisis de video necesitan de una capacidad de cómputo elevada, especialmente cuando el procesamiento de secuencias es de manera continua. Es por ello que se necesita de aceleradores como GPU y TPU (del inglés *Tensor Processing Unit*), que se encuentran disponibles en diferentes niveles de la infraestructura, desde dispositivos Edge hasta estaciones de trabajo y servidores. La complejidad de los modelos que pueden ejecutarse y el tiempo requerido para procesar la información, se ven condicionados por la capacidad de estos recursos [61].

A partir de estas herramientas y recursos computacionales, es posible implementar técnicas diferentes que se encuentran orientadas a interpretar el contenido presente en imágenes y secuencias de video. En este contexto, la visión por computador entrega métodos para la identificación y análisis de elementos presentes en una escena, constituyendo la base para la clasificación, localización, detección y segmentación de objetos.

3.3.1. Visión por Computador Aplicada al Análisis de Multitudes

La visión por computador corresponde a un campo orientado a la interpretación de contenido digital como imágenes y videos. Su aplicación permite el desarrollo de diferentes tareas entre las que se encuentran: clasificación de imágenes, localización, detección de objetos y segmentación de instancias. La clasificación permite detectar una categoría asociada a una imagen, mientras que la localización permite determinar la posición de un objeto dentro de ella. Por otro lado, la detección combina la identificación y localización de objetos que se encuentran presentes en una escena, mientras que la segmentación permite su distinción a nivel de píxeles.

Para el análisis y monitoreo de multitudes, la detección de objetos resulta de especial interés, pues permite considerar a las personas como elementos individuales en una escena. Mediante esta tarea, es posible identificar su presencia y determinar su ubicación en imágenes o secuencias de video, lo que permite representar individualmente a las personas presentes [79].

Cuando se consideran secuencias de imágenes consecutivas para el análisis, es posible complementar la información presente en cada cuadro con una dimensión temporal. De esta manera, se amplían las posibilidades de análisis gracias al procesamiento de video, incorporando información asociada a la evolución de los elementos a lo largo del tiempo [80].

La implementación de estas tareas se puede realizar mediante distintas herramientas de software orientadas al procesamiento y análisis de la información visual. Algunas bibliotecas integran capacidades de visión por computador con herramientas de aprendizaje profundo y su clasificación puede variar según el alcance de cada tecnología. En esta sección se consideran en su aplicación al procesamiento de imágenes y

visión por computador, independiente de que algunas incorporen funcionalidades adicionales.

Una de las herramientas utilizadas es `OpenCV` (del inglés *Open Source Computer Vision*), biblioteca de código abierto de visión por computador, que entrega funcionalidades para el procesamiento de imágenes y video. Está desarrollada en C/C++ con el propósito de entregar implementaciones de alta velocidad, lo que favorece su uso en aplicaciones que requieren de procesamiento en tiempo real. Además, la biblioteca permite implementar distintas técnicas de visión por computador y trabajar tanto con métodos tradicionales como con modelos de aprendizaje profundo para tareas de detección de objetos [81].

Otra herramienta utilizada corresponde a `scikit-image`, biblioteca de código abierto que se encuentra desarrollada principalmente en `Python` y orientada al procesamiento de imágenes. `Scikit-image` se encuentra orientada al análisis de imágenes en un ecosistema de computación científica de `Python`. Utiliza `NumPy` para la representación de información visual, lo que facilita su compatibilidad con otras herramientas del mismo entorno, incluyendo `OpenCV`. Entre sus funcionalidades se encuentran operaciones de segmentación, detección de bordes, análisis de textura, detección de características y transformaciones aplicadas a imágenes.

Existen diferencias en la implementación de ambas tecnologías. Mientras que `OpenCV` dispone de una interfaz para `Python`, `scikit-image` utiliza `Python` con `Cython` para operaciones que necesitan de mayor rendimiento. De la misma manera, con `scikit-image` es posible trabajar con arreglos multidimensionales, a diferencia del módulo de procesamiento de imágenes para `OpenCV`, que se encuentra orientado principalmente para imágenes bidimensionales. Estas características muestran que con las herramientas disponibles es posible presentar diferentes orientaciones incluso cuando comparten funcionalidades relacionadas al procesamiento y análisis de información visual [82].

Una tercera herramienta corresponde a `Dlib`, biblioteca de código abierto y multiplataforma, usada en herramientas de aprendizaje automático y procesamiento de imágenes. Desde su desarrollo, la biblioteca fue creada para proporcionar componentes reutilizables que puedan integrarse en aplicaciones y trabajar con distintos tipos de datos, entre ellos imágenes [83]. Además, sus funcionalidades se pueden utilizar para la extracción y análisis de características presentes en una imagen, integrándose con otras herramientas de procesamiento dentro de aplicaciones de visión por computador [84].

En conjunto, estas tecnologías presentan distintos alcances dentro del procesamiento de información visual. Por un lado, `OpenCV` se encuentra principalmente orientado a aplicaciones de visión por computador y procesamiento en tiempo real. `Scikit-image` por su parte, entrega herramientas de procesamiento de análisis de imágenes integradas al ecosistema científico de `Python`. Por otro lado, `Dlib` incorpora herramientas de procesamiento de imágenes junto con funcionalidades de aprendizaje automático. A partir de lo anterior, es posible utilizar las bibliotecas de acuerdo con los requerimientos sin necesidad de que su uso sea excluyente

dentro de un mismo flujo de procesamiento. En la Tabla 3.3 se sintetizan las herramientas descritas y su aplicación.

Tecnología	**Orientación**	**Funcionalidades**	**Características principales**
`OpenCV` [81], [82]	Visión por computador y procesamiento de imágenes y video.	Preprocesamiento, extracción de características y detección de objetos mediante diferentes técnicas de visión por computador.	Biblioteca de código abierto orientada a implementaciones de alta velocidad y aplicaciones en tiempo real.
`scikit-image` [82]	Procesamiento y análisis de imágenes dentro del ecosistema científico de Python.	Segmentación, detección de bordes, análisis de textura, detección de características y transformaciones de imágenes.	Biblioteca de código abierto integrada con `NumPy` y otras herramientas científicas de `Python`, con soporte para datos multidimensionales.
`Dlib` [83], [84]	Procesamiento de imágenes y aprendizaje automático aplicado al análisis visual.	Extracción y análisis de características presentes en imágenes mediante herramientas de procesamiento y aprendizaje automático.	Biblioteca de código abierto y multiplataforma que integra herramientas de procesamiento de imágenes con funcionalidades de aprendizaje automático.

Tabla 3.3: Comparación de herramientas utilizadas para el procesamiento de imágenes y visión por computador.

3.3.2. Análisis de Objetos en Secuencias de Video

El análisis de objetos en secuencias de video incluye una dimensión temporal que permite relacionar observaciones obtenidas en distintos fotogramas. A diferencia del análisis de una imagen donde se puede identificar y localizar los elementos presentes, el procesamiento de video permite mantener información sobre estos elementos a medida que la secuencia transcurre. Este tipo de análisis puede ser aplicado a objetos de distinta naturaleza, como personas, vehículos o animales, dependiendo del entorno y del propósito del sistema.

Para el seguimiento de múltiples objetos se distinguen dos enfoques según la forma en la que son inicializados. En el seguimiento basado en detección, se localizan los objetos de interés presentes en cada fotograma y posteriormente se asocian las detecciones obtenidas a través de la secuencia para reconstruir sus trayectorias. Por otro lado, en el seguimiento libre de detección, los objetos se inicializan de manera manual en el primer fotograma y luego son localizados en los fotogramas que siguen. Entre ambos enfoques, el seguimiento basado en detección es el más utilizado, pues permite relacionar las tareas de detección y seguimiento dentro del análisis de secuencias de video [85]. La Figura 3.2 representa de manera general este flujo de análisis, incluyendo oclusiones como una dificultad que puede afectar su desarrollo.

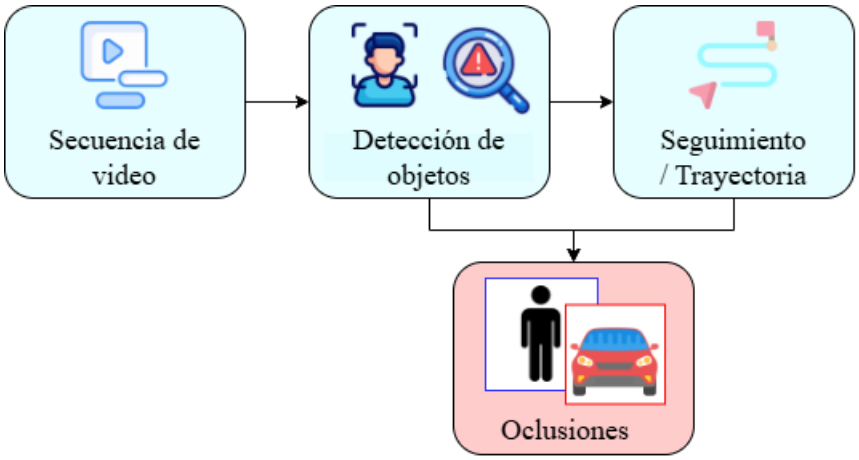
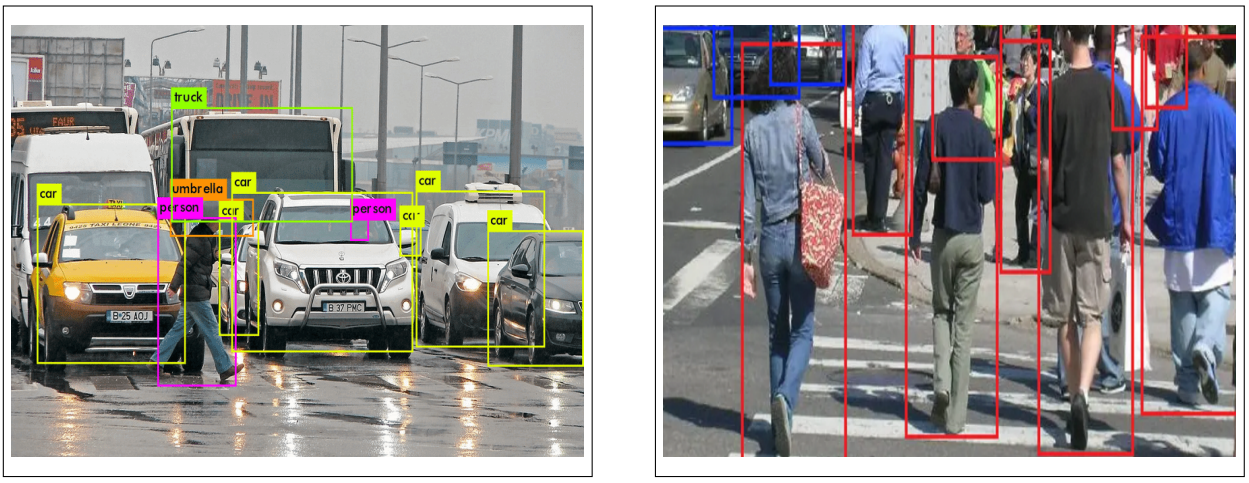


Figura 3.2: Flujo general para el análisis de objetos en secuencias de video.

La detección es el punto de entrada para la identificación y localización de objetos que luego serán analizados. Su aplicación permite determinar la presencia y región que ocupa un objeto en un fotograma y así, entregar una representación espacial de los elementos de interés que se encuentran presentes en una escena. En sistemas de videovigilancia, este proceso se puede realizar de manera consecutiva sobre los fotogramas que forman parte del video, entregando observaciones que luego se pueden relacionar durante el seguimiento.

Por otro lado, el seguimiento agrega continuidad temporal a las detecciones obtenidas y asocia las observaciones que pertenecen a un mismo objeto entre fotogramas consecutivos. Para lograrlo, se mantiene un identificador asociado a cada elemento mientras se encuentra en la escena, lo que permite diferenciarlo de los otros objetos presentes. Como resultado, se obtiene la ubicación en un instante determinado y además, es posible contar con un registro de su historial de posiciones y representar su trayectoria a lo largo del tiempo [85], [86].

La continuidad de este proceso se puede ver afectada por las oclusiones que ocurren cuando un objeto queda parcial o totalmente oculto dentro de una escena. Estas se pueden producir por la presencia de elementos del entorno que bloquean su visualización, lo que corresponde a una oclusión inter-clase, o por la superposición entre objetos de una misma clase, que corresponde a una oclusión intra-clase (veáse Figura 3.3). Este último caso, es de especial relevancia en escenas con grupos o multitudes donde la cercanía entre los elementos presentes aumenta la posibilidad de que unos oculten a otros. Como resultado, se puede perder parte de las características visuales que se utilizan para realizar la detección y se dificulta la asociación del mismo objeto entre distintos fotogramas [87].



(a) Oclusión inter-clase [88]. (b) Oclusión intra-clase [89].

Figura 3.3: Ejemplos de oclusión inter-clase e intra-clase.

Para manejar las oclusiones, se han desarrollado enfoques orientados a mantener la capacidad de identificar los objetos cuando su información visual se encuentra incompleta. Entre ellos se considera el uso de las partes visibles del objeto, la estimación de las regiones que permanecen disponibles y el aprovechamiento de la información que proviene de fotogramas consecutivos y del contexto de la escena. Estas estrategias buscan disminuir la pérdida de detecciones y las dificultades para mantener la continuidad de los objetos durante el análisis.

En conjunto, las tareas descritas permiten pasar desde la identificación de los objetos que se encuentran en una escena hasta su análisis a lo largo de una secuencia de video. Por un lado, la detección entrega información asociada a su presencia y ubicación, por otro el seguimiento permite mantener su identidad y representar su trayectoria, considerando las dificultades que se pueden producir por las oclusiones durante este proceso [85], [86], [87]. De esta manera, el análisis visual permite generar información sobre los elementos que se encuentran presentes en un espacio y su evolución a lo largo del tiempo.

3.4. Resultados del Procesamiento para el Análisis de Aglomeraciones

A partir del procesamiento de las imágenes y secuencias de video es posible obtener información que permite definir la presencia, distribución y movimiento de los elementos identificados en el espacio observado. En el contexto de análisis de multitudes, esta información se puede expresar mediante características medibles

que se obtienen a nivel individual o colectivo y que permite representar la concentración de personas y su dinámica a lo largo del tiempo. Entre las características usadas principalmente se encuentran el conteo, la densidad, la velocidad, la dirección y el flujo de movimiento.

Con el conteo es posible representar a través de un valor la cantidad total de personas presentes en una escena, pero no entrega información sobre cómo las personas se distribuyen dentro del espacio observado, pues una misma cantidad se puede concentrar en una zona determinada o puede distribuirse a lo largo de todo el espacio. Para agregar la componente espacial se utiliza la estimación de densidad, que permite representar la concentración de personas en diferentes regiones de una imagen y complementar la información que entrega el conteo global [90].

Los mapas de densidad son una forma de visualizar esta distribución, en esta representación, se asocian valores a distintas regiones los cuales permiten reflejar la concentración local de personas dentro de una imagen. La suma de estos valores sobre una región determinada permite obtener el número de personas presentes en la zona, mientras que su aplicación en el mapa completo permite obtener el conteo de la escena. De esta forma, el mapa permite conservar la ubicación espacial de la multitud y permite identificar sectores con distintos niveles de concentración [91].

Además, el análisis de secuencia de video permite incorporar información que se relaciona con el movimiento. Mientras que la velocidad representa la rapidez con la que se mueven los individuos o la multitud, la dirección permite identificar la orientación seguida durante el desplazamiento. Ambas características pueden ser analizadas a nivel individual, considerando el movimiento de cada persona, o a nivel colectivo, y así, identificar las tendencias que predominan en la multitud.

El flujo, en cambio, permite representar patrones de desplazamiento que se desarrollan a través del espacio y del tiempo. Su análisis permite observar la forma en que la multitud circula por un entorno e identificar los cambios respecto a los patrones que son predominantes. Estas características se relacionan con la información temporal que se obtiene mediante el seguimiento, pues las variaciones de posición a través de los fotogramas consecutivos permiten definir aspectos como la velocidad, dirección y trayectoria de los elementos que se observan.

Las características que se obtienen no deben ser necesariamente interpretadas de forma aislada. Su análisis en conjunto permite representar distintos estados de una multitud considerando concentración y movimiento. Por ejemplo, una condición de alta densidad que se acompaña de baja velocidad se puede caracterizar como una situación de congestión, mientras que los cambios en la velocidad o desviaciones de las direcciones predominantes permiten identificar alteraciones en la dinámica observada [92]. De esta manera, caracterizar una aglomeración puede considerar distintas dimensiones de la información obtenida y

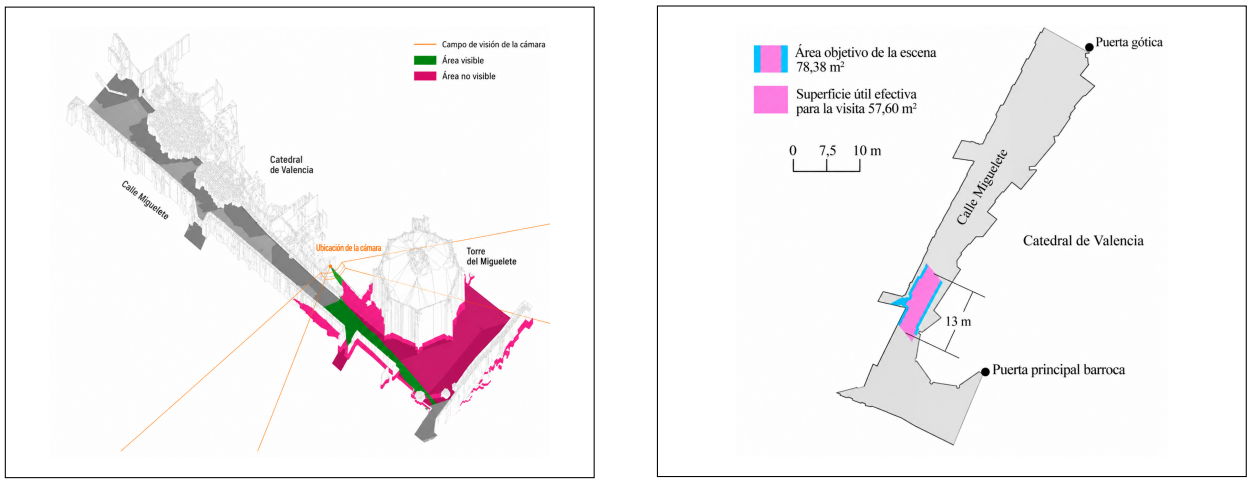
no únicamente la cantidad total de personas presentes en la escena.

Para interpretar las condiciones que presenta una multitud, los valores que se obtienen mediante estas métricas se pueden asociar a niveles o umbrales de referencia. Uno de los criterios utilizados corresponde al Nivel de Servicio (LOS, del inglés *Level of Service*), que relaciona variables como la densidad y el flujo con las condiciones de desplazamiento presentes en el entorno analizado. Esta clasificación permite distinguir distintos estados, desde condiciones de circulación libre hasta niveles en los que aumenta la interacción entre las personas y el movimiento se encuentra restringido. Así, las métricas obtenidas permiten cuantificar la presencia y el movimiento de las personas e interpretar las condiciones de ocupación observadas [93].

Para establecer estos niveles, los umbrales utilizados pueden variar según las características del espacio donde se realiza el análisis. En aplicaciones de LOS, se establecen rangos diferentes según el tipo de infraestructura, pues las condiciones de desplazamiento y ocupación no son equivalentes entre diferentes espacios. Por esta razón, identificar condiciones de alta ocupación o congestión se asocia con criterios definidos según las características del entorno [94].

Definir el espacio sobre el cual se obtienen las métricas es un aspecto relevante para interpretarlas. En el estudio realizado por Viñals et al. se implementa un sistema de monitoreo con una cámara de video 2D con reconocimiento de personas. La cámara se instala a 12 [m] de altura en la Catedral de Valencia y su campo visual abarca un área aproximada de 155 [m^2]. Aunque el análisis no se realiza directamente en toda la superficie capturada, sino que se delimita una zona específica que se encuentra destinada al tránsito peatonal [95].

Para establecer esta zona, se define inicialmente un área de 78,38 [m^2] considerando un tramo de 13 [m] dentro de la calle. Luego, se descuentan sectores asociados a obstáculos y márgenes que no se encuentran disponibles para el desplazamiento de peatones, obteniendo una superficie útil de 57,60 [m^2]. Con esta delimitación es posible establecer una relación entre las personas que se identifican en la escena y la superficie disponible para su circulación, tal como se observa en la Figura 3.4



(a) Campo visual de la cámara.

(b) Área delimitada para el análisis.

Figura 3.4: Campo visual y delimitación del área de análisis en la calle del Miguelete, Valencia [95].

A partir de esta representación espacial, la densidad se expresa mediante el número de personas presentes por metro cuadrado. Además, se consideran distancias interpersonales para establecer los diferentes niveles de ocupación del espacio y relacionar la cantidad de personas detectadas con la superficie disponible.

En este caso, los niveles de ocupación se establecen considerando la cantidad de personas presentes de manera simultánea y la distancia disponible dentro de la superficie útil. Cuando aumenta la cantidad de personas, disminuye la distancia disponible entre ellas provocando un cambio en la condición de ocupación del espacio. Se distinguen niveles que varían desde condiciones que se consideran adecuadas, asociadas a mayores distancias interpersonales, hasta condiciones que se consideran críticas cuando esta distancia se reduce. Para el área analizada, las distancias entre 0,45 y 0,80 [m] se asocian a un nivel crítico de ocupación, que puede alcanzar hasta 216 personas presentes simultáneamente. De esta manera, la cantidad de personas y la densidad obtenida se pueden interpretar según niveles que permiten caracterizar el estado de ocupación del espacio y su relación con las distancias interpersonales se observa en la Figura 3.5

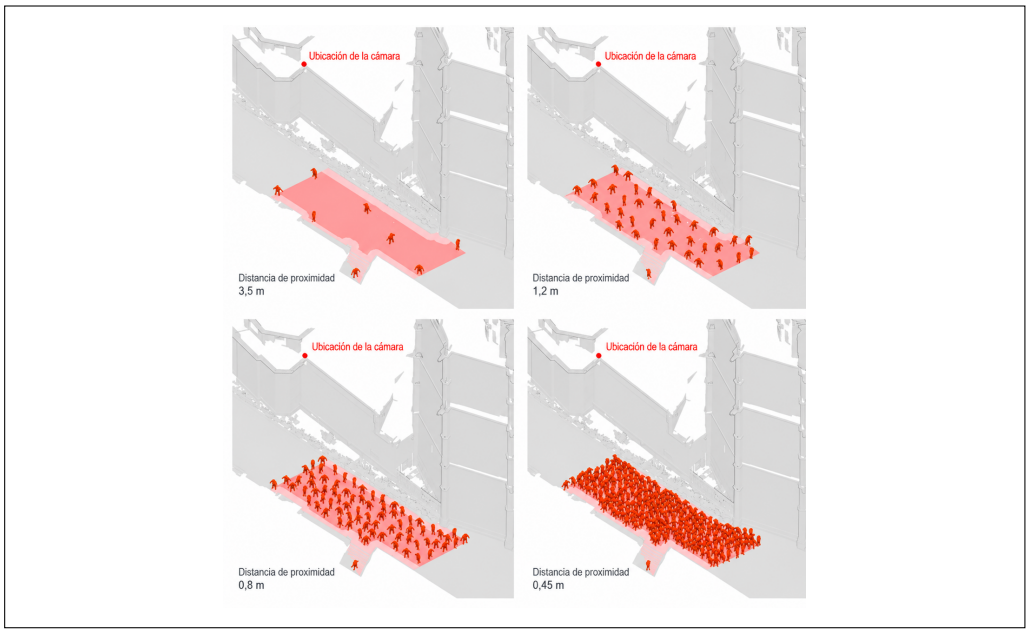


Figura 3.5: Representación de las distancias interpersonales utilizadas para establecer los niveles de ocupación [95].

El monitoreo es realizado de manera continua en intervalos de 10 segundos, lo que permite observar la evolución temporal de la afluencia y reconocer incrementos que se asocian a actividades desarrolladas en el entorno. En las jornadas estudiadas se identifican aumentos de personas asociados a eventos y festividades locales, y se alcanzan diferentes niveles de ocupación de la superficie disponible. A partir de estas variaciones, el estudio plantea el uso de la información obtenida para reconocer episodios de saturación y apoyar a las medidas preventivas relacionadas con la gestión de flujos peatonales [95].

Otro caso de aplicación de estas métricas corresponde a un análisis realizado durante un evento de alta concurrencia en Puri, India. En este estudio se analiza el comportamiento de la multitud del sector de Medical Chhak a través de una cámara instalada a 14 [m] de altura. Además, se selecciona una región fija de 20x20 [m] dentro de la vía, lo que corresponde a una superficie de 400 $[m^2]$ sobre la que se estudian las condiciones de desplazamiento de los asistentes [96].

En este caso, la imagen presenta variaciones de escala producidas por la perspectiva debido a la inclinación de la cámara, donde los elementos ubicados a mayor distancia ocupan menor proporción de la imagen. Para relacionar las observaciones con las dimensiones físicas del terreno, la imagen se rectifica hasta obtener una representación ortogonal del área que se está analizando. Luego, la superficie se divide en una cuadrícula de 20 x 20 celdas, donde cada celda representa 1 $[m^2]$. Con esta transformación es posible utilizar una referencia espacial homogénea para calcular la densidad y analizar el movimiento de la multitud, como se

observa en la Figura 3.6

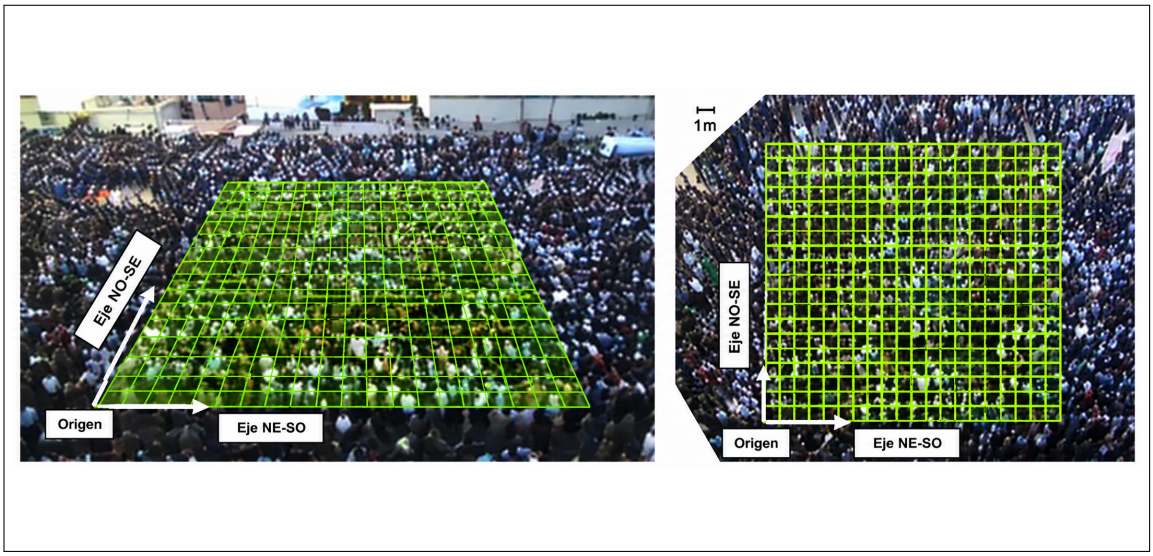


Figura 3.6: Representación en perspectiva y vista ortogonal del área de análisis en Puri, India [96].

A partir del área delimitada, la densidad se expresa mediante el número de personas presentes por metro cuadrado, mientras que el movimiento se caracteriza utilizando la velocidad de desplazamiento en metros por segundo. Además, el estudio incorpora el flujo peatonal que se expresa según la cantidad de personas que atraviesan una sección de la vía por minuto y por metro de ancho. Así, las tres métricas permiten representar dimensiones diferentes de la aglomeración: la densidad describe su concentración espacial, la velocidad caracteriza el desplazamiento de los individuos y el flujo representa el movimiento de personas a través de una sección determinada del espacio.

A partir de estos casos, es posible observar que las métricas obtenidas se pueden interpretar mediante niveles o umbrales que permiten caracterizar los distintos estados de una aglomeración. Estos valores dependen de las características del espacio analizado y de los criterios utilizados para establecerlos, por lo que no necesariamente corresponden a límites generales aplicables a cualquier escenario. En Valencia, los niveles de ocupación se establecen considerando la cantidad de personas presentes, la superficie disponible y las distancias interpersonales, mientras que en Puri la relación entre densidad, velocidad y flujo permite caracterizar las distintas condiciones de desplazamiento de la multitud [96]. De esta manera, las métricas permiten cuantificar las características de una aglomeración e interpretar su estado según las condiciones particulares en las que se desarrolla.

La Tabla 3.4 sintetiza las métricas principales y criterios utilizados que caracterizan una aglomeración.

Métrica / criterio	**Qué representa**	**Aplicación en el análisis**
Conteo [90]	Cantidad de personas presentes en la escena o en una región determinada.	Permite cuantificar la presencia de personas y observar variaciones en la afluencia a lo largo del tiempo.
Densidad [90], [91]	Concentración y distribución espacial de las personas dentro del área observada.	Permite identificar sectores con diferentes niveles de concentración y relacionar la cantidad de personas con el espacio disponible.
Velocidad [92]	Rapidez con la que se desplazan los individuos o la multitud.	Permite caracterizar el movimiento e identificar cambios en las condiciones de desplazamiento.
Dirección [92]	Orientación seguida por los individuos o por el movimiento predominante de la multitud.	Permite identificar tendencias de desplazamiento y desviaciones respecto del movimiento predominante.
Flujo [92], [96]	Movimiento de personas a través de un espacio durante un período determinado.	Permite analizar la circulación de la multitud y reconocer cambios en los patrones de desplazamiento.
Nivel de ocupación / LOS [93], [94], [95]	Criterio de interpretación que relaciona las métricas obtenidas con diferentes condiciones de ocupación o desplazamiento.	Permite establecer niveles o umbrales para interpretar condiciones de circulación, alta ocupación, congestión o saturación según las características del espacio analizado.

Tabla 3.4: Características utilizadas para el análisis de multitudes.

3.5. Visualización de Información en Sistemas de Monitoreo

Los resultados obtenidos durante el procesamiento de imágenes y secuencias de video se pueden presentar de distintas formas para facilitar su interpretación dentro de un sistema de monitoreo. La información se puede incorporar de manera directa sobre la imagen capturada a través de elementos gráficos o se puede representar mediante paneles que permiten resumir el estado y la evolución de las variables analizadas. Estas formas de representación se pueden utilizar de manera conjunta, lo que permite visualizar información individual asociada a los elementos detectados, y al mismo tiempo, entregar una representación global de las condiciones presentes en el entorno monitoreado.

Una de las formas utilizadas corresponde a la superposición de información sobre los fotogramas de video. En este tipo de interfaz se mantiene la imagen capturada como referencia y sobre ella se añaden elementos como cajas delimitadoras, identificadores, líneas, regiones de interés, trayectorias y valores numéricos. Con

esto es posible mantener una relación espacial entre los resultados de procesamiento y los elementos a los que se encuentran asociados, con el objetivo que el usuario pueda reconocer dónde ocurre una condición determinada dentro de la escena y observar su evolución durante el monitoreo.

Zhao, Jiang y Wang presentan un sistema desarrollado para el monitoreo de una pasarela peatonal, cuya localización específica no es mencionada debido a restricciones de privacidad establecidas por el organismo responsable de la infraestructura. La interfaz integra los resultados de detección, seguimiento y análisis sobre la transmisión de video. Los peatones se representan en cajas delimitadoras y con identificadores, además es posible registrar el conteo según la dirección de desplazamiento a través de una línea trazada sobre la pasarela. Para el análisis de densidad se agrega una región de interés que delimita la superficie que se utiliza para el cálculo. Por otro lado, para representar el movimiento, se representan las trayectorias y la velocidad asociada a los peatones. De esta manera, distintos resultados de procesamiento se presentan sobre una misma escena y se mantienen asociados al espacio donde fueron obtenidos.

La interfaz también utiliza elementos visuales para mostrar el estado de las variables monitoreadas. Los niveles de densidad definidos en la región de interés permiten generar advertencias cuando la ocupación alcanza condiciones determinadas, mientras que la velocidad se utiliza para identificar estados de movimiento. Estas alertas aparecen sobre la misma transmisión junto con los valores obtenidos, lo que permite observar de manera simultánea la escena, la variable analizada y su estado [97]. La Figura 3.7 muestra estas formas de representación para el conteo, densidad y análisis de velocidad.



Figura 3.7: Visualización de información asociada al conteo, densidad y velocidad en un sistema de monitoreo peatonal [97].

Además las interfaces pueden incorporar información que representa el comportamiento de los elementos a lo largo del tiempo. En este contexto, las trayectorias se pueden presentar de manera conjunta para observar patrones de desplazamiento y facilitar la identificación de los sectores o direcciones que predominan.

Un ejemplo de su aplicación corresponde al análisis de flujo peatonal que se realiza en un cruce urbano ubicado en el sector de Hongdae, en Seúl, Corea del Sur, donde las trayectorias que se obtienen son representadas sobre la imagen y se diferencian según la dirección que siguen los peatones. Además, el espacio se organiza mediante zonas asociadas al cruce y a los sectores de espera, lo que permite relacionar recorridos con áreas determinadas de la escena. Cuando se acumulan las trayectorias registradas, su visualización permite reconocer los sectores con mayor presencia de recorridos y observar las direcciones que predominan, como se observa en la Figura 3.8.

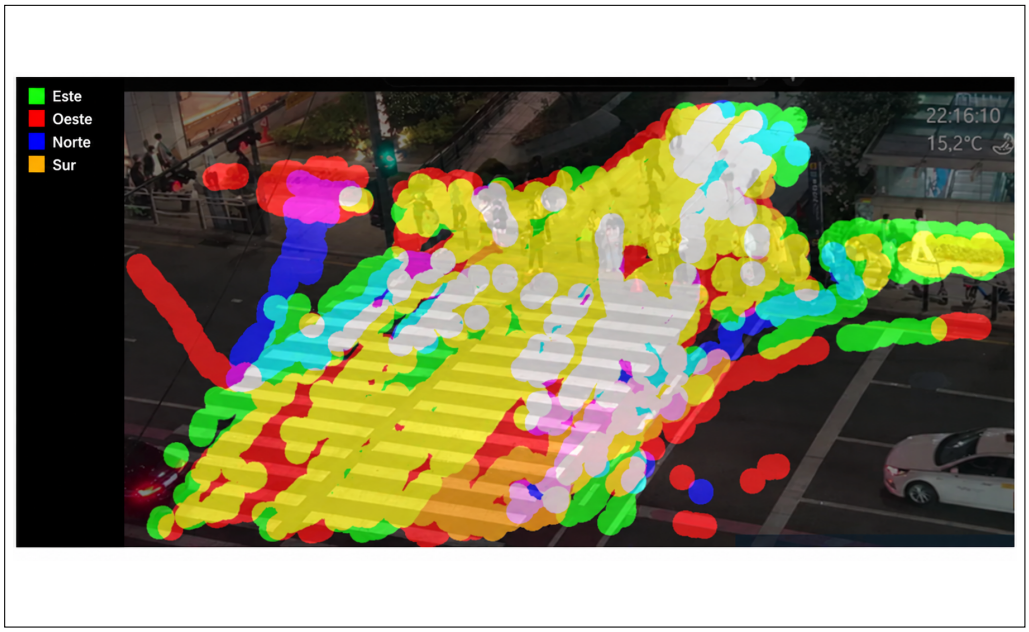


Figura 3.8: Visualización de trayectorias y direcciones de desplazamiento peatonal en un cruce urbano[98].

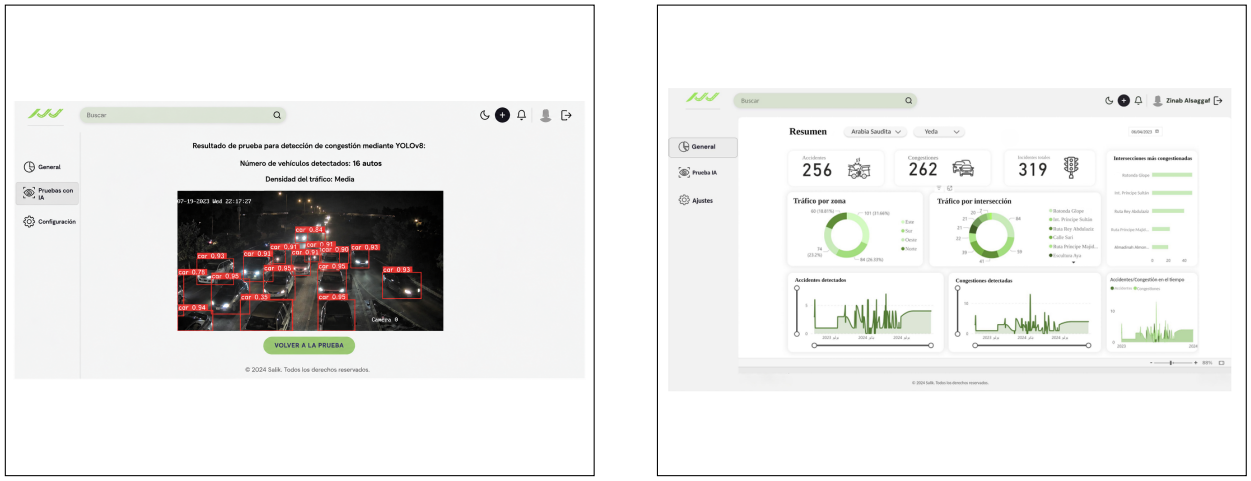
En este mismo caso de estudio, la interfaz combina información como la cantidad de personas activas, con información acumulada asociada al total de peatones registrados, las direcciones de desplazamiento y la ocupación de distintas zonas. Además, a partir de estos datos se generan representaciones temporales y estadísticas que permiten observar cambios en el flujo, tiempos de espera y variaciones en el uso del espacio, logrando que la información se pueda organizar en distintos niveles [98].

A partir de estas aplicaciones, es posible observar que una interfaz de monitoreo es capaz de combinar representación directa de los resultados sobe una escena, junto con información resumida fuera de ella. La representación sobre la escena permite relacionar los resultados con los elementos o zonas observadas, mientras

que los indicadores y gráficos permiten sintetizar las observaciones obtenidas durante periodos de tiempo más extensos. De esta manera, las distintas formas de visualización se pueden complementar para entregar una representación más completa de las condiciones identificadas y de sus variaciones a lo largo del tiempo.

Estas formas de visualización también se pueden adaptar a los tipos de elementos que se quieran monitorear y a la información que se busca representar. En ese sentido, resulta relevante observar como se presenta la información en sistemas orientados al tráfico vehicular, donde las variables que son monitoreadas y su representación responden a las condiciones propias de circulación. Alzamzami et al. presentan un sistema orientado al monitoreo vehicular que muestra como puede ser presentado este tipo de información en una interfaz mediante el análisis automático del tráfico. Este sistema utiliza detecciones sobre imágenes para identificar vehículos y accidentes, mientras que los resultados se organizan en una plataforma que incluye indicadores, gráficos y opciones para consultar la información registrada. Así, es posible visualizar información directamente en la escena e incorporar componentes que permiten resumir el estado general del tráfico.

Para monitorear la congestión, la interfaz muestra la imagen procesada en conjunto con la cantidad de vehículos detectados y los distintos niveles de densidad vehicular que se asocian a la escena, lo que permite interpretar estados según el contexto observado. La plataforma también agrega información sobre el tráfico e incidentes y permite consultar los resultados según variables como la ubicación y periodo seleccionado [99]. En Figura 3.9 se muestra distintas formas de presentación de la información utilizadas en la interfaz, considerando resultados asociados a la escena y su representación en un panel de información.



(a) Visualización de resultados sobre la escena .

(b) Panel para la visualización de información del tráfico.

Figura 3.9: Formas de visualización de información en un sistema de monitoreo vehicular [99].

En conjunto, los sistemas revisados muestran una combinación de recursos para presentar la información generada durante un monitoreo y evidencian que la visualización puede adaptarse según las características del

entorno, los elementos monitoreados y el propósito del sistema. Por lo tanto, su presentación puede integrar distintos recursos que ayudan a la interpretación de resultados y su uso dentro del monitoreo [97], [98], [99].

3.6. Sistemas Inteligentes de Monitoreo de Afluencia

Los sistemas inteligentes de monitoreo pueden integrar los módulos que forman el flujo general de captura y procesamiento de información presentado en la Sección 3.1, considerando captura, transmisión y centralización, procesamiento, resultados del procesamiento y visualización de la información. La manera en la que estos se integran depende del propósito del sistema, del entorno que se monitorea y de las tecnologías que se utilizan.

Integrar estos módulos permite que la información obtenida durante el monitoreo se utilice para identificar y anticipar determinadas condiciones. Wong y Law desarrollan un estudio orientado al monitoreo de congestión que utiliza la captura de video combinada con planos del espacio observado para la detección y seguimiento de peatones, para que posteriormente se relacionen sus posiciones con las coordenadas del espacio físico. A partir de este procesamiento, es posible obtener información individual asociada al desplazamiento de personas e información colectiva relacionada con el conteo, densidad y flujo peatonal en diferentes regiones.

En este mismo caso, se organizan los registros espaciales y temporales del flujo de personas a través de una estructura que representa distintas regiones y su conectividad. Con la información registrada, es posible analizar la evolución del flujo y estimar su comportamiento futuro. Por otro lado, con los resultados, es posible representar los sectores con mayor concentración e identificar las condiciones actuales de congestión. En conjunto, el sistema integra funciones asociadas a la captura, procesamiento, resultados del procesamiento y visualización de la información para analizar condiciones de congestión [100]. La Figura 3.10 muestra el flujo general que se utiliza para la integración de estas funciones.

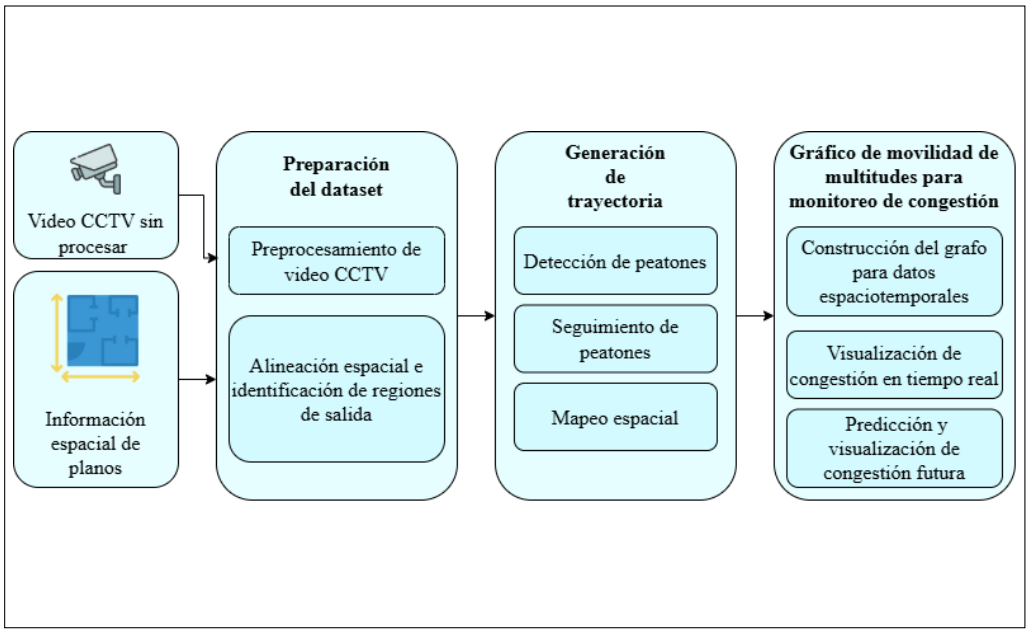


Figura 3.10: Flujo de monitoreo de congestión [100].

La información que se utiliza para monitorear la presencia y concentración de personas también se puede obtener mediante tecnologías diferentes a las cámaras. En el estudio realizado por Wiangwiset et al. en Tailandia, se utilizan puntos de acceso Wi-Fi que se distribuyen en distintas ubicaciones para obtener registros que se asocian a los dispositivos presentes en las zonas monitoreadas. En esta arquitectura, los puntos de acceso cumplen la función de sensado y la información recopilada se transmite hacia recursos de cómputo para realizar su procesamiento.

El sistema considera dominios de sensado, comunicación y cómputo, lo que permite centralizar los datos obtenidos desde las distintas ubicaciones y estimar la cantidad de personas presentes en estas. Los resultados se almacenan y luego se presentan en una aplicación web que permite consultar las condiciones registradas en diferentes sectores y periodos de tiempo. Así, se pueden identificar funciones asociadas a la infraestructura completa, pero con una configuración basada en infraestructura Wi-Fi como fuente de información [101]. La Figura 3.11 muestra la organización de esta arquitectura.

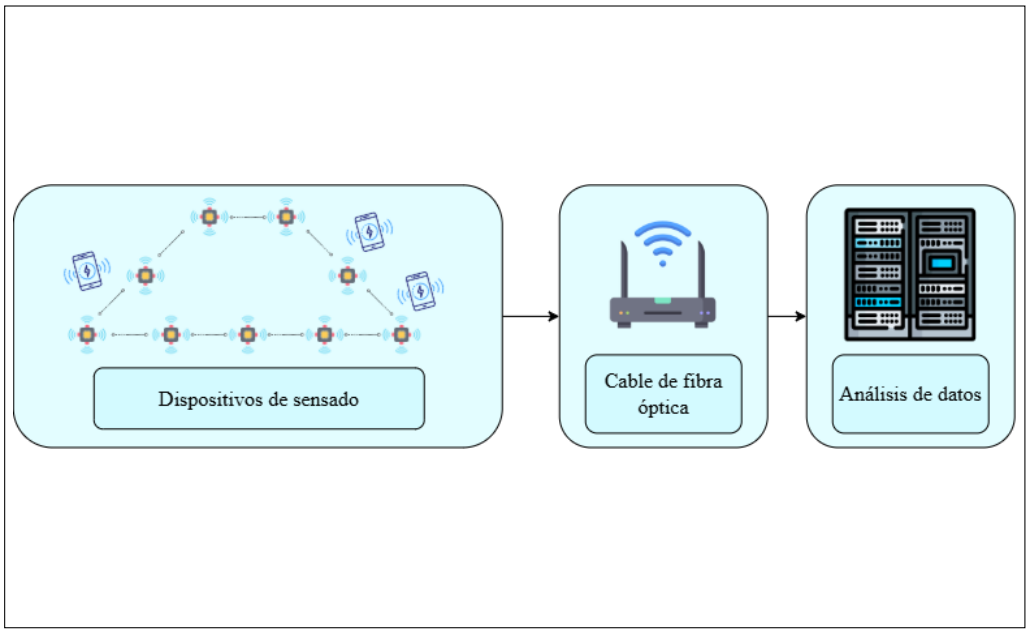


Figura 3.11: Flujo de información para el monitoreo de afluencia mediante Wi-Fi [101].

Es posible observar las distintas funciones de monitoreo cuando la información obtenida se utiliza para generar respuesta frente a las condiciones que son identificadas. El caso de estudio revisado con anterioridad y orientado al tráfico vehicular, utiliza cámaras CCTV instaladas en instersecciones para obtener imágenes que se procesan con el objetivo de determinar la cantidad de vehículos, clasificar niveles de congestión e identificar accidentes.

Con los resultados obtenidos durante el procesamiento, el sistema es capaz de generar respuestas según las condiciones identificadas. Cuando se detecta congestión, la cantidad de vehículos y el número de carriles se utilizan para ajustar la duración de los tiempos del semáforo ubicado en la intersección, lo que permite adaptar su funcionamiento según las condiciones observadas. Además, el sistema integra una comunicación inalámbrica entre el semáforo y los vehículos cercanos a través de microcontroladores y conexión Wi-Fi, lo que permite enviar advertencias ante situaciones de congestión o accidentes y presentarlas en una pantalla instalada en el vehículo. De manera simultánea, los eventos detectados se envían a una plataforma central y se almacenan en una base de datos, donde los administradores pueden consultar registros e indicadores asociados al tráfico [99]. La Figura 3.12 muestra el flujo utilizado para la detección de las condiciones del tráfico desde el procesamiento hasta el registro y visualización de la información.

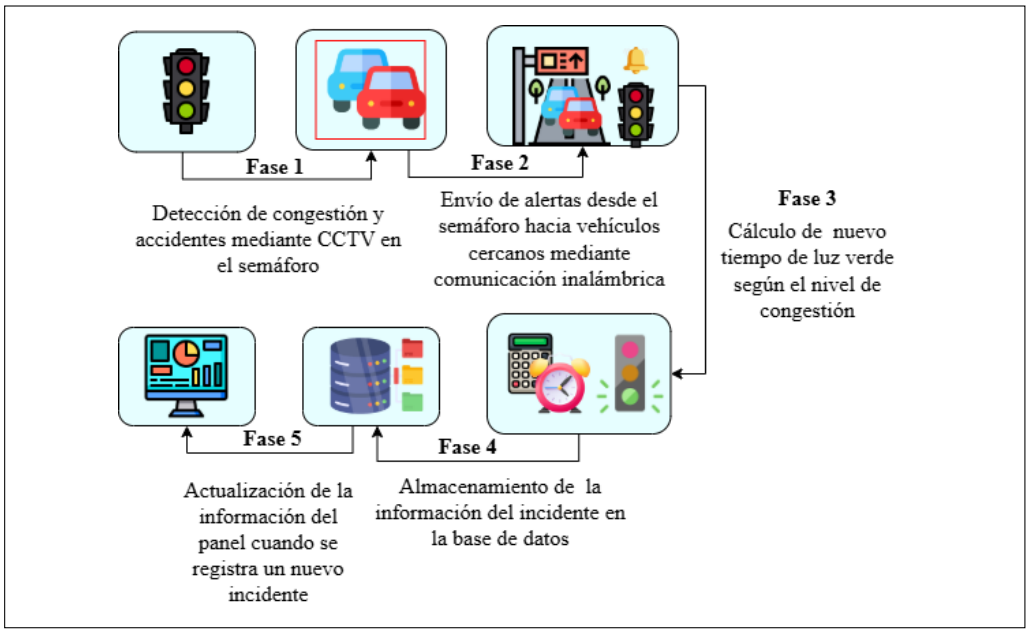


Figura 3.12: Flujo utilizado para la detección y gestión de condiciones del tráfico [99].

Estas funciones también pueden integrarse en soluciones comerciales. Un ejemplo corresponde a StreetFlow, una plataforma de videoanalítica que utiliza cámaras IP como fuente de información y modelos de inteligencia artificial para reconocer distintos elementos presentes en una escena. A partir de este análisis, la plataforma permite obtener información asociada al conteo, clasificación, dirección y velocidad de loe elementos detectados, además de identificar eventos o condiciones relevantes para el monitoreo, entre ellas la presencia de aglomeraciones.

El procesamiento de la información se puede realizar de manera local o en un servidor, y complementarse con reglas que determinan las acciones asociadas a los eventos identificados. Cuando ocurre un evento que se define como relevante, la plataforma es capaz de generar una alerta y conservar el registro con la evidencia correspondiente [102]. La Figura muestra un ejemplo del panel principal de la plataforma donde se presentan los resultados del análisis de video junto con información gráfica y alertas asociadas al monitoreo.

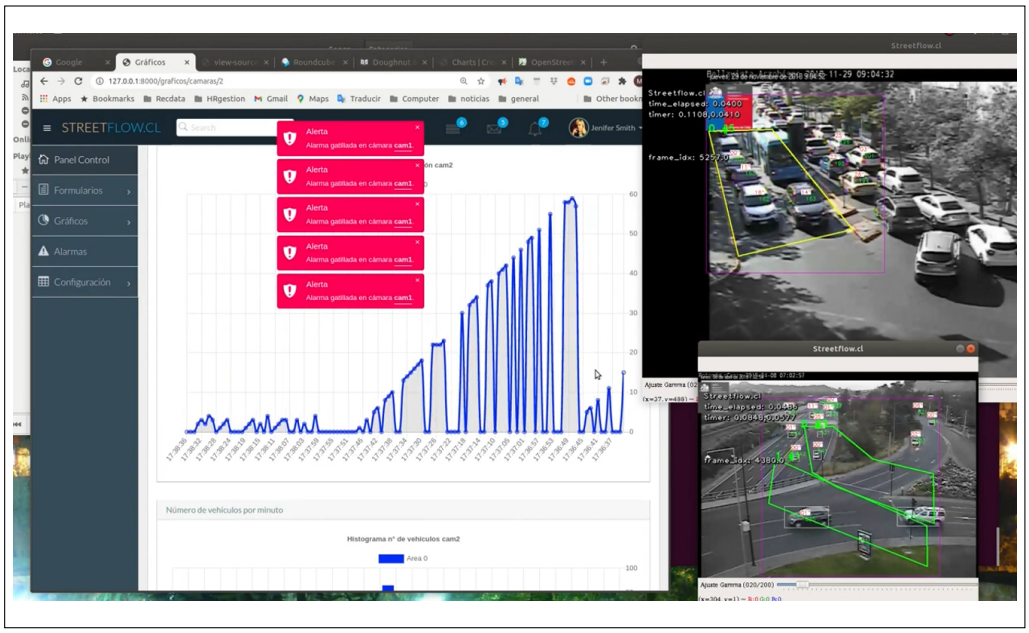


Figura 3.13: Panel principal de StreetFlow [102].

Los sistemas revisados muestran las distintas formas de integrar las funciones que participan en el monitoreo. A partir de sus características, se pueden relacionar las funciones identificadas con los módulos de captura, transmisión y centralización, procesamiento, resultados del procesamiento y visualización de la información descritos, como se muestra en la Tabla 3.5.

Caso	**Captura**	**Transmisión y centralización**	**Procesamiento**	**Resultados del procesamiento**	**Visualización**
Monitoreo de congestión peatonal [100]	Video CCTV e información espacial del entorno.	No se detalla en el estudio.	Detección, seguimiento y análisis espacial de peatones.	Conteo, densidad, flujo peatonal y estimación de condiciones futuras.	Seguimiento individual, mapas de calor y representación de condiciones de congestión.
Monitoreo de afluencia mediante Wi-Fi [101]	Puntos de acceso Wi-Fi distribuidos en las zonas monitoreadas.	Infraestructura de comunicación para centralizar los registros obtenidos.	Procesamiento de registros asociados a los dispositivos detectados.	Cantidad de personas presentes en las distintas ubicaciones.	Aplicación web con información por ubicación y períodos de tiempo.
Monitoreo vehicular [99]	Cámaras CCTV instaladas en intersecciones.	Plataforma central y comunicación inalámbrica.	Detección de vehículos, congestión y accidentes.	Conteo de vehículos, niveles de congestión e identificación de accidentes.	Dashboard con indicadores, gráficos e imágenes procesadas.
StreetFlow [102]	Cámaras IP.	Transmisión hacia recursos de procesamiento local o central.	Modelos de inteligencia artificial y reglas asociadas a los eventos identificados.	Conteo, clasificación, dirección, velocidad y detección de eventos.	Plataforma de monitoreo con video procesado, gráficos y alertas.

Tabla 3.5: Relación entre módulos del flujo general y los sistemas inteligentes de monitoreo revisados.

En conjunto, los casos revisados permiten observar la aplicación de estos módulos en sistemas desarrollados para diferentes entornos, donde la información se utiliza según las condiciones que se busca monitorear. El análisis de estos casos permite fundamentar las bases del diseño de la solución que se propone en el siguiente capítulo.

Capítulo 4

Diseño de la Solución

Este capítulo describe la propuesta y diseño de la solución para abordar el monitoreo de aglomeraciones en espacios públicos. A partir del análisis presentado en los capítulos anteriores, se establece la estructura y componentes necesarios para el desarrollo de la solución propuesta.

4.1. Requisitos del Sistema

En esta sección se definen los requisitos que orientan el desarrollo de la solución propuesta y que establecen las condiciones necesarias para su funcionamiento. Estos se clasifican en requisitos funcionales, no funcionales, de ambiente y de hardware. Además, se define el perfil de usuario considerado para la interacción con el sistema.

4.1.1. Requisitos Funcionales

Los requisitos funcionales definen las acciones que el sistema debe ser capaz de realizar. En este trabajo, la solución propuesta debe permitir la captura y procesamiento de la información para detectar y monitorear la presencia de personas en espacios públicos. A continuación se describen los requisitos funcionales considerados.

- **RF1 - Cobertura del área de monitoreo:** La cámara debe ser capaz de detectar completamente el área donde se realizará el monitoreo de forma clara.
- **RF2 - Detección y conteo de personas:** El sistema debe tener un modelo predictivo capaz de mantener un conteo preciso incluso en zonas con alto flujo de personas.

- **RF3 - Almacenamiento de resultados:** El sistema debe ser capaz de almacenar los resultados de la predicción.
- **RF4 - Visualización de información:** El sistema deberá mostrar en la interfaz web información clara y procesada sobre la afluencia detectada. Esto incluye:
  - Conteo de personas en tiempo real por cámara.
  - Señalización de niveles de ocupación o umbrales definidos para detectar posibles aglomeraciones.
  - Generación de reportes históricos descargables en formato CSV con fecha, hora y número de personas detectadas.
- **RF5 - Generación de alertas:** El sistema debe notificar mediante alertas visuales los eventos críticos.
- **RF6 - Generación de reportes:** El sistema debe ser capaz de generar reportes por parte del usuario donde se detalle fecha, hora y conteo de personas de un día específico.
- **RF7 - Control de acceso:** El sistema debe tener un sistema de acceso por credenciales.
- **RF8 - Monitoreo multicámara:** El sistema debe ser capaz de soportar múltiples cámaras en tiempo real.

4.1.2. Requisitos No Funcionales

Los requisitos no funcionales definen las condiciones que el sistema debe cumplir para asegurar su correcto funcionamiento. Estos complementan las funciones definidas con anterioridad y establecen criterios asociados a la calidad de solución. A continuación se presentan los requisitos no funcionales considerados.

- **RNF1 - Calidad de captura:** El sistema debe grabar de forma clara.
- **RNF2 - Usabilidad de la interfaz:** La interfaz debe ser simple e intuitiva para usuarios no técnicos.
- **RNF3 - Adaptabilidad del entorno:** El sistema debe funcionar en ambientes abiertos y cerrados sin ajustes manuales.
- **RNF4 - Tiempo de respuesta:** El sistema debe operar en tiempo real, con una latencia máxima de 1 segundo entre la captura de la imagen y la visualización del resultado procesado en la interfaz.
- **RNF5 - Disponibilidad del sistema:** Se espera una disponibilidad de al menos un 95 % del tiempo de operación, dado que el monitoreo debe mantenerse activo de forma continua mientras las cámaras estén transmitiendo.

- **RNF6 - Seguridad de acceso:** Las credenciales de acceso deberán gestionarse adecuadamente para evitar accesos no autorizados.

4.1.3. Requisitos de Interfaces

Los requisitos de interfaces definen la forma en la que el sistema interactúa con los elementos externos que se involucran en su funcionamiento. Para ello, se consideran los eventos que se pueden producir durante el monitoreo y la respuesta asociada a estos.

4.1.3.1. Eventos Externos

Los eventos externos se pueden originar a partir de cámaras de videovigilancia, usuarios o acciones programadas. A continuación, se describen los principales eventos considerados.

1. **Inicio de captura:** Se inicia la transmisión de video desde una cámara para comenzar el monitoreo. Este evento puede ser iniciado por el usuario o de forma programada por el sistema y considera la cámara y la hora de inicio.
2. **Solicitud de visualización:** El usuario solicita visualizar el video y la información obtenida durante el monitoreo de una cámara o ubicación determinada.
3. **Detección de personas:** El modelo procesa las imágenes capturadas e identifica la presencia de personas considerando el umbral de confianza definido.
4. **Generación de alerta:** Se genera cuando el conteo de personas alcanza o supera uno de los umbrales definidos para identificar los niveles de ocupación.
5. **Solicitud de reporte:** El usuario solicita información histórica de afluencia correspondiente a una cámara y un período determinado.
6. **Acceso de usuario:** El usuario ingresa sus credenciales para solicitar el acceso al sistema de monitoreo.

4.1.3.2. Respuestas del Sistema

A partir de los eventos definidos, el sistema ejecuta diferentes acciones asociadas al procesamiento, visualización y registro de la información. A continuación, se describen las principales respuestas consideradas.

1. **Activación del procesamiento:** El sistema inicia el procesamiento de las imágenes provenientes de la cámara para realizar la detección y conteo de personas.

2. **Visualización de información:** El sistema presenta en la interfaz el video de la cámara junto con la información obtenida durante el monitoreo.
3. **Actualización y registro del conteo:** El sistema actualiza el número de personas detectadas y almacena el registro correspondiente junto con la fecha, hora y cámara asociada.
4. **Generación de alertas:** El sistema presenta una alerta visual cuando el conteo de personas alcanza o supera alguno de los umbrales establecidos.
5. **Generación de reporte:** El sistema genera un reporte a partir del período y la cámara seleccionados por el usuario, incorporando la información histórica registrada.
6. **Ingreso al sistema:** El sistema valida las credenciales proporcionadas y, cuando estas son correctas, permite al usuario acceder a la plataforma.

4.1.4. Requisitos de Ambiente

Los requisitos de ambiente establecen las condiciones necesarias que garantizan el correcto funcionamiento y despliegue de la solución propuesta, considerando recursos de hardware y software requeridos.

4.1.4.1. Requisitos de Hardware

Los requisitos de hardware establecen las características mínimas necesarias para la captura de video y la ejecución de los componentes asociados al funcionamiento del sistema.

1. **Equipo de procesamiento:**
   - Procesador con capacidad para ejecutar tareas de procesamiento de video.
   - Memoria RAM suficiente para soportar la ejecución simultánea de los componentes del sistema.
   - Unidad de procesamiento gráfico (GPU) compatible con la ejecución del modelo de detección.
   - Conectividad para recibir los flujos de video y permitir la comunicación con los demás componentes del sistema.
2. **Dispositivo de captura de video:**
   - Cámara digital compatible con el sistema.
   - Resolución mínima Full HD (1920x1080 píxeles).
   - Frecuencia de captura de al menos 30 fps.

- Capacidad para proporcionar el flujo de video al sistema mediante una conexión compatible.

4.1.4.2. Requisitos de Software

Los requisitos de software consideran las herramientas necesarias para el desarrollo de los distintos componentes del sistema.

Lenguajes de programación

- **Python:** Lenguaje de programación interpretado, de alto nivel y propósito general, utilizado en diferentes áreas de desarrollo y caracterizado por favorecer la legibilidad y el desarrollo rápido de aplicaciones [103].
- **C++:** Lenguaje de programación de propósito general que permite desarrollar aplicaciones que requieren un manejo eficiente de los recursos computacionales [104].
- **JavaScript:** Lenguaje de programación utilizado en el desarrollo de aplicaciones web y asociado a la construcción de componentes interactivos [105].

Bibliotecas para Procesamiento y Análisis de Datos

Para apoyar las tareas de procesamiento y análisis existen diferentes bibliotecas que proporcionan herramientas para el manejo de datos, operaciones numéricas y aplicaciones asociadas al aprendizaje automático y visión por computador. Entre las opciones disponibles se encuentran:

- **NumPy:** Biblioteca esencial para la computación científica. Entrega estructuras de datos eficientes que facilitan el procesamiento numérico y el análisis de datos [106].
- **Pandas:** Biblioteca usada para la manipulación y análisis de datos en Python [107].
- **scikit-image:** Biblioteca orientada al procesamiento de imágenes que proporciona algoritmos para realizar diferentes operaciones de análisis y transformación sobre imágenes digitales [82].
- **Dlib:** Biblioteca que proporciona herramientas orientadas al aprendizaje automático y visión por computador [83].

Herramientas para Captura y Procesamiento de Video

Para la adquisición y gestión de flujos de video existen diferentes herramientas que permiten trabajar con dispositivos de captura y distintas fuentes de video. Entre las alternativas disponibles se encuentran:

- **OpenCV:** Biblioteca de visión por computador que incorpora funcionalidades para acceder a dispositivos de captura y obtener cuadros desde diferentes fuentes de video [81].

- **FFmpeg:** Framework multimedia que permite trabajar con contenido de video mediante funciones de codificación, decodificación y gestión de diferentes formatos y flujos multimedia [108].
- **GStreamer:** Framework multimedia basado en una arquitectura de elementos conectados mediante tuberías, que permite construir flujos para la captura y procesamiento de video [109].

Modelos para Detección de Objetos

Para realizar la detección de objetos existen diferentes modelos que se pueden utilizar según los requerimientos de procesamiento de una aplicación. Entre ellos se encuentran:

- **Faster R-CNN:** Modelo de detección de dos etapas que incorpora una red de propuesta de regiones (RPN) para identificar regiones candidatas, las cuales posteriormente son utilizadas para realizar la localización y clasificación de los objetos.
- **SSD:** Modelo de detección de una etapa que realiza la identificación de objetos utilizando mapas de características de diferentes escalas, sin incorporar una etapa independiente para la generación de regiones candidatas.
- **YOLO:** Familia de modelos de detección de una etapa que realiza directamente sobre la imagen las tareas de localización y clasificación de los objetos mediante una única red.

Las características descritas corresponden a los principales enfoques utilizados para la detección de objetos mediante modelos de una y dos etapas [110].

Bases de Datos

Para el almacenamiento de la información existen diferentes sistemas de gestión de bases de datos que permiten organizar y conservar los registros generados. Entre las alternativas disponibles se encuentran:

- **MySQL:** Sistema de gestión de bases de datos relacional que organiza la información mediante tablas compuestas por filas y columnas y permite establecer relaciones entre los datos [111].
- **PostgreSQL:** Sistema de gestión de bases de datos relacional que permite almacenar y gestionar información estructurada mediante tablas y relaciones entre los datos **figueroa2017comparison**.
- **MongoDB:** Sistema de gestión de bases de datos NoSQL orientado a documentos, que organiza la información mediante colecciones y documentos en formatos JSON/BSON [111].

Backend

Para el desarrollo de servicios del lado del servidor existen diferentes tecnologías y frameworks. Entre las alternativas disponibles se encuentran:

- **FastAPI:** Framework desarrollado en Python orientado a la creación de aplicaciones y APIs, con soporte para programación asíncrona en la gestión de solicitudes [112].
- **Flask:** Framework desarrollado en Python que permite construir aplicaciones y servicios web e incorporar funcionalidades adicionales mediante extensiones [112].
- **Django REST Framework:** Framework basado en Django orientado a la construcción de APIs, que incorpora funcionalidades relacionadas con serialización, autenticación, filtrado e interacción con bases de datos [112].
- **Node.js:** Entorno de ejecución de JavaScript que utiliza un modelo basado en eventos y operaciones de entrada y salida asíncronas, permitiendo desarrollar aplicaciones y servicios del lado del servidor [113].

Frontend

Para el desarrollo de interfaces web existen diferentes tecnologías orientadas a la construcción y actualización de los componentes presentados al usuario. Entre las alternativas disponibles se encuentran:

- **React:** Biblioteca de JavaScript y TypeScript orientada al desarrollo de interfaces de usuario mediante componentes reutilizables.
- **Angular:** Framework basado en TypeScript que utiliza una arquitectura basada en componentes e incorpora herramientas para funcionalidades como enrutamiento, solicitudes HTTP y gestión de formularios.
- **Vue.js:** Framework progresivo basado en JavaScript y TypeScript que permite desarrollar interfaces mediante componentes e incorporar sus funcionalidades de manera gradual.

Las características mencionadas corresponden a los principales enfoques utilizados para el desarrollo de interfaces web basadas en componentes [114].

4.1.5. Perfil de Usuario

El sistema está orientado a un único perfil de usuario: el encargado del monitoreo de personas en espacios públicos. Este usuario requiere una interfaz clara, que permita el acceso a la información del monitoreo en tiempo real, la visualización de alertas y la consulta de registros históricos que apoye la supervisión de la afluencia de personas.

- **Perfil:** Encargado de monitoreo.

- **Características socioculturales:** Usuario con conocimientos básicos en el uso de herramientas tecnológicas y sistemas de monitoreo.
- **Características ocupacionales:** Personal encargado de labores de seguridad, vigilancia o supervisión de espacios públicos.
- **Rango etario:** Orientado a usuarios adultos, independientemente de su rango etario.
- **Características asociadas al uso:** Capacidad para interpretar la información presentada durante el monitoreo y actuar frente a las alertas generadas por el sistema.
- **Otras características:** Disponibilidad para desempeñar funciones en distintos turnos y acceder a la información de monitoreo desde diferentes ubicaciones.

4.2. Arquitectura del Sistema

El diseño de la arquitectura del sistema establece la organización de los elementos que conforman la solución y las interacciones necesarias para su funcionamiento. Para ello, se utiliza un diagrama de contexto que permite identificar la relación del sistema con elementos externos y un diagrama de arquitectura que presenta la organización interna de los módulos que forman parte de la solución.

4.2.1. Diagrama de Contexto

El diagrama de contexto presenta una visión general de la solución propuesta y su interacción con los principales elementos involucrados en el monitoreo. En la Figura 4.1 se muestra la relación del sistema con el encargado de monitoreo, las cámaras en tiempo real, el sistema de alertas y la base de datos. De la misma manera, se muestran las interacciones principales asociadas al acceso mediante credenciales, la exportación de métricas y el almacenamiento de la información generada durante el funcionamiento del sistema.

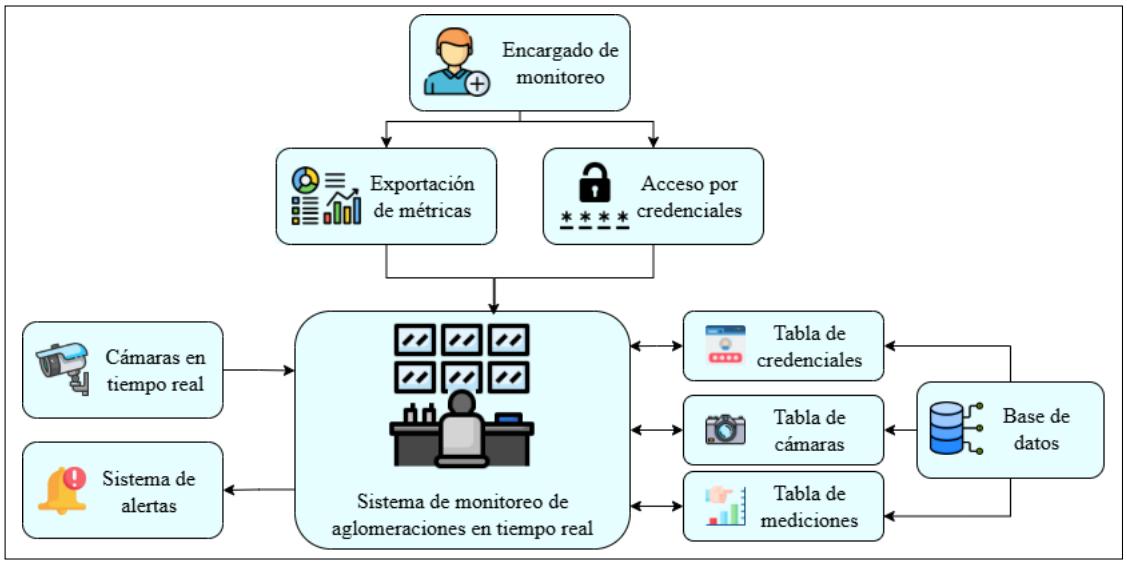


Figura 4.1: Diagrama de contexto del sistema.

4.2.2. Diagrama de Arquitectura

El diagrama de arquitectura presenta la organización interna de la solución propuesta y la relación entre los módulos que la conforman, como se observa en la Figura 4.2. Las conexiones entre los módulos permiten identificar el recorrido de la información durante las distintas etapas de monitoreo.

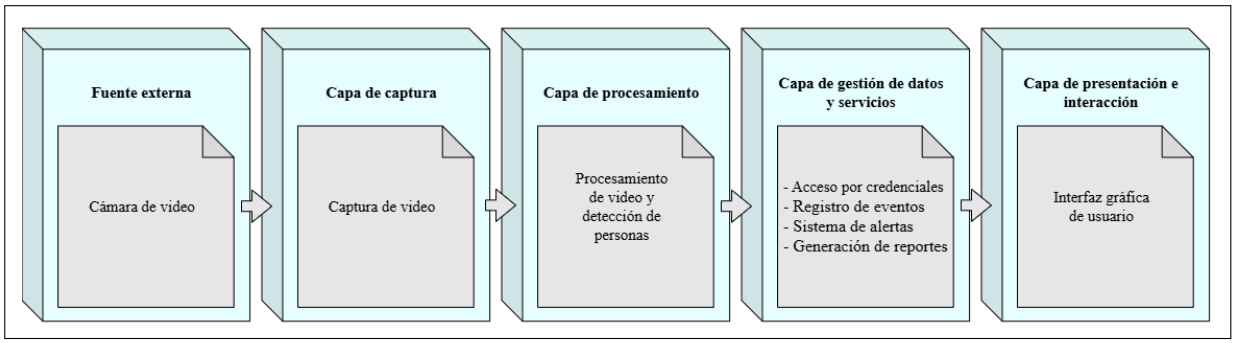


Figura 4.2: Arquitectura general del sistema.

A partir de la estructura presentada, la arquitectura del sistema se organiza mediante un enfoque modular, en el que cada componente cumple una función específica dentro de la solución y mantiene la interacción con los otros elementos necesarios para el monitoreo. Con esta organización, se facilita el mantenimiento, escalabilidad y adaptación del sistema a diferentes condiciones de operación, considerando su aplicación en espacios como recintos deportivos, aeropuertos, estaciones de metro, terminales de buses o centros comerciales.

4.3. Módulos del Sistema

La arquitectura presentada anteriormente se compone de módulos conectados entre sí, cada uno encargado de cumplir una función específica dentro del proceso de monitoreo. En la Figura 4.3 se presentan los módulos que conforman el sistema y las principales relaciones establecidas entre ellos.

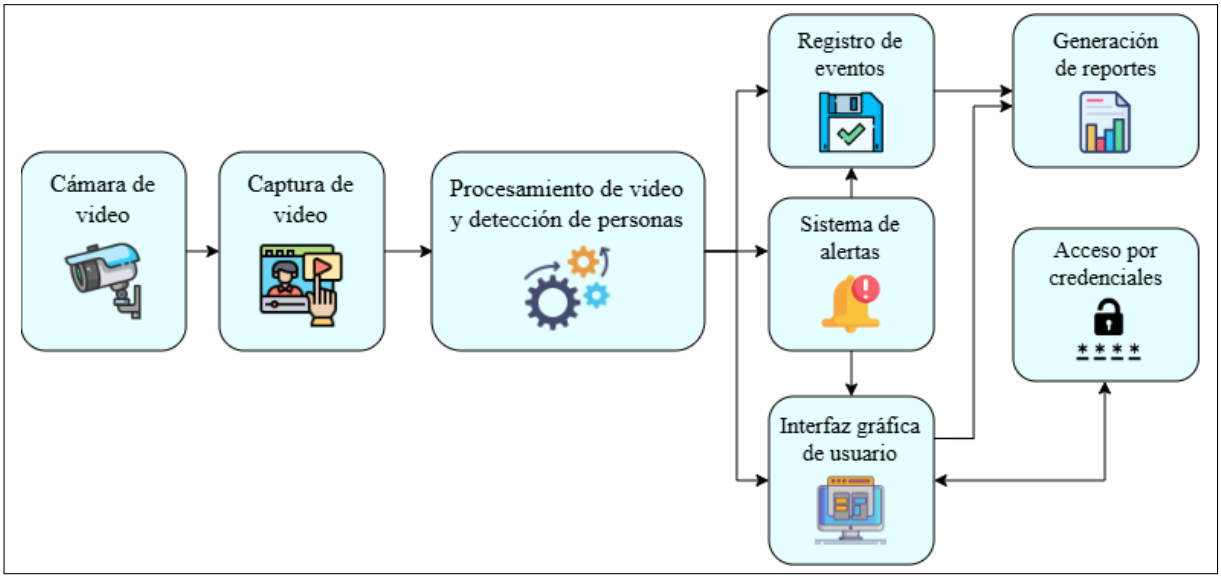


Figura 4.3: Diagrama de módulos del sistema.

A continuación, se describe el propósito de cada módulo y su participación dentro de la solución propuesta.

4.3.1. Capa de Captura

Corresponde al punto de entrada de la información utilizada durante el monitoreo y comprende los elementos encargados de obtener el flujo de video proveniente de las cámaras para su posterior procesamiento.

4.3.1.1. Captura de Video

El módulo de captura de video se encarga de obtener el flujo de video proveniente de la cámara instalada en el espacio donde se realiza el monitoreo. La información capturada constituye la entrada utilizada posteriormente por el módulo de procesamiento para realizar la detección de personas.

- **Propósito:** Recibir el flujo de video en tiempo real proveniente de la cámara conectada al sistema y entregar cada cuadro al módulo de procesamiento.
- **Alcance:** Corresponde al módulo de entrada del sistema y se encuentra al inicio del flujo de información. Su función se limita a la obtención y entrega de las imágenes al módulo de procesamiento, sin realizar tareas de detección o análisis sobre estas.
- **Dependencias:**
  - Requiere conexión con el dispositivo de captura de video.
  - Debe estar vinculado al módulo de procesamiento de video y detección de personas para proporcionar los cuadros capturados durante el monitoreo.
- **Supuestos:**
  - La cámara se encuentra encendida y correctamente conectada.
  - El sistema operativo permite el acceso al dispositivo de captura de video.
- **Restricciones:**
  - Resolución mínima requerida: 1920x1080 píxeles.
  - Frecuencia mínima de captura: 30 fps.
- **Estructura general:**
  - **Entrada:** Flujo de video proveniente de la cámara mediante conexión USB o RTSP.
  - **Proceso:**
    - Inicialización de la cámara.
    - Configuración de la resolución y frecuencia de captura.
    - Captura de cuadros.
    - Envío de los cuadros al módulo de procesamiento de video y detección de personas.
  - **Salida:** Cuadro individual (*frame*) destinado al módulo de procesamiento.

4.3.2. Capa de Procesamiento

La capa de procesamiento se encarga de analizar la información obtenida durante la captura con el propósito de identificar y contar las personas presentes en las imágenes.

4.3.2.1. Procesamiento de Video y Detección de Personas

El módulo de procesamiento del modelo se encarga de procesar el flujo de video recibido desde el módulo de captura. Mediante el uso de herramientas de inteligencia artificial, se realiza la detección y conteo de personas presentes en las imágenes, generando los resultados utilizados por los demás módulos del sistema.

- **Propósito:** Procesar los cuadros recibidos mediante un modelo de detección de objetos para identificar y contar personas durante el monitoreo.
- **Alcance:** Recibe los cuadros provenientes del módulo de captura de video y realiza su procesamiento mediante técnicas de visión por computador. A partir de las detecciones obtenidas, determina el número de personas presentes y proporciona los resultados requeridos por los demás módulos del sistema.
- **Dependencias:**
  - Módulo de captura de video para recibir los cuadros capturados.
  - Módulo de interfaz gráfica de usuario para proporcionar los resultados obtenidos durante el procesamiento.
  - Módulo de registro de eventos para almacenar los resultados de las detecciones realizadas.
  - Módulo de alertas del sistema para proporcionar el conteo utilizado en la evaluación de los umbrales definidos.
- **Supuestos:**
  - El modelo permite detectar personas presentes en las imágenes capturadas.
  - El hardware dispone de los recursos computacionales necesarios para ejecutar el procesamiento durante el monitoreo.
- **Restricciones:**
  - Precisión mínima requerida del modelo: 90 %.
  - Intervalo máximo de procesamiento entre cuadros: 1 segundo.
- **Estructura general:**
  - **Entrada:** Cuadro individual (*frame*) proveniente del módulo de captura de video.
  - **Proceso:**
    - Recibir el cuadro capturado.

- Ejecutar el modelo de detección.
- Identificar las personas detectadas.
- Determinar el conteo de personas.
- Proporcionar los resultados a los módulos correspondientes.

- **Salida:** Detecciones realizadas, conteo de personas y fecha y hora asociadas al procesamiento.

4.3.3. Capa de Gestión de Datos y Servicios

Esta capa reúne los módulos encargados de gestionar la información generada durante el monitoreo, mantener los registros necesarios para su posterior consulta y atender las funciones asociadas a alertas, reportes y control de acceso.

4.3.3.1. Registro de Eventos

El módulo del registro de eventos se encarga de almacenar los resultados obtenidos durante el procesamiento. Cada registro mantiene la información asociada al conteo realizado, junto con la fecha y hora correspondiente, permitiendo su posterior consulta.

- **Propósito:** Registrar y almacenar la información obtenida durante el monitoreo para permitir su posterior consulta, incluyendo el conteo de personas, alertas generadas, fecha, hora y zona asociada.
- **Alcance:** El módulo de registro de eventos se encarga del almacenamiento de la información generada durante el monitoreo. Los registros almacenados permiten mantener un historial de los resultados obtenidos y proporcionan la información requerida posteriormente por el módulo de generación de reportes.
- **Dependencias:**
  - Módulo de procesamiento de video y detección de personas para recibir los resultados obtenidos durante el procesamiento.
  - Módulo de alertas del sistema para registrar las alertas generadas durante el monitoreo.
  - Módulo de generación de reportes para proporcionar la información histórica requerida para la generación de reportes.
- **Supuestos:**

- Existe capacidad de almacenamiento suficiente para mantener los registros generados.
- El sistema dispone de acceso a la base de datos utilizada para almacenar la información.

- **Restricciones:**
  - La información debe almacenarse de manera estructurada en la base de datos.
  - El registro de la información se realiza en intervalos de 10 segundos.
- **Estructura general:**
  - **Entrada:** Conteo de personas, alertas generadas, fecha, hora y zona asociada.
  - **Proceso:**
    - Recibir la información generada durante el monitoreo.
    - Estructurar los datos correspondientes al registro.
    - Almacenar la información en la base de datos.
  - **Salida:** Registros almacenados y disponibles para su posterior consulta.

4.3.3.2. Sistema de Alertas

El módulo de alertas del sistema se encarga de generar alertas cuando el conteo de personas alcanza o supera los umbrales previamente definidos. Estas alertas se almacenan en la base de datos junto con la información asociada al evento, permitiendo consultar su fecha y gestionar su estado. Además, son presentadas mediante la interfaz de usuario para informar al encargado de monitoreo sobre posibles situaciones de aglomeración.

- **Propósito:** Generar alertas a partir de los resultados obtenidos durante el procesamiento y de los umbrales definidos, permitiendo informar al usuario sobre situaciones que requieren atención.
- **Alcance:** Recibe los resultados generados durante el monitoreo y evalúa el cumplimiento de los umbrales establecidos. Cuando se genera una alerta, esta se registra en la base de datos junto con la información asociada al evento, permitiendo consultar su fecha y gestionar su estado como pendiente o cerrada.
- **Dependencias:**
  - Módulo de procesamiento de video para recibir los resultados del monitoreo.
  - Módulo de registro de eventos para almacenar la información asociada a las alertas generadas.
  - Módulo de interfaz gráfica de usuario para presentar y gestionar las alertas.

- **Supuestos:**
  - Los umbrales utilizados para la generación de alertas se encuentran previamente definidos en el sistema.
  - Los resultados necesarios para evaluar los umbrales se encuentran disponibles durante el monitoreo.
- **Restricciones:**
  - El tiempo máximo considerado para generar una alerta una vez detectada la condición correspondiente es de 1 segundo.
- **Estructura general:**
  - **Entrada:** Resultados del procesamiento y umbrales definidos.
  - **Proceso:** Evaluación de los resultados, generación de la alerta y registro de la información asociada.
  - **Salida:** Alerta registrada y disponible para su visualización y gestión.

4.3.3.3. Generación de Reportes

El módulo de reportes permite generar reportes a partir de la información almacenada durante el monitoreo. El usuario puede consultar los registros correspondientes a un período determinado y obtener la información asociada al conteo de personas.

- **Propósito:** Generar reportes de afluencia a partir de la información histórica almacenada durante el monitoreo.
- **Alcance:** El módulo de generación de reportes opera a partir de una solicitud realizada por el usuario mediante la interfaz. El módulo consulta los registros almacenados en el módulo de registro de eventos, selecciona la información correspondiente al período solicitado y genera un reporte en formato CSV.
- **Dependencias:**
  - Módulo de interfaz gráfica de usuario para recibir la solicitud de generación del reporte.
  - Módulo de registro de eventos para acceder a los registros históricos del monitoreo.
- **Supuestos:**
  - El módulo de registro de eventos contiene los registros correspondientes al período solicitado.
  - La información almacenada se encuentra disponible para su consulta.
- **Restricciones:**

- Los reportes se generan a partir de los registros correspondientes al día seleccionado por el usuario.
- El formato de exportación utilizado es CSV.

- **Estructura general:**
  - **Entrada:** Solicitud de reporte correspondiente a una fecha seleccionada por el usuario.
  - **Proceso:**
    - Recibir la solicitud de generación del reporte.
    - Consultar los registros almacenados en el módulo de registro de eventos.
    - Seleccionar la información correspondiente a la fecha solicitada.
    - Generar el archivo en formato CSV.
  - **Salida:** Reporte de afluencia en formato CSV.

4.3.3.4. Acceso por Credenciales

El módulo de acceso por credenciales se encarga de controlar el ingreso al sistema mediante la autenticación del usuario. De esta manera, el acceso a la información y a las funciones de monitoreo queda restringido a usuarios autorizados.

- **Propósito:** Controlar el acceso al sistema mediante la autenticación de usuarios, permitiendo que únicamente usuarios autorizados puedan acceder a las funcionalidades disponibles.
- **Alcance:** El módulo de acceso por credenciales se encarga de controlar el acceso inicial al sistema mediante la validación de las credenciales ingresadas. Una vez autenticado, el usuario puede acceder a la interfaz y utilizar las funcionalidades asociadas al monitoreo.
- **Dependencias:**
  - Módulo de interfaz gráfica de usuario para presentar la interfaz de inicio de sesión y permitir el ingreso de las credenciales.
  - Base de datos para acceder a la información requerida durante la validación del usuario.
- **Supuestos:**
  - Las credenciales de acceso se encuentran previamente registradas.
  - La información necesaria para realizar la autenticación se encuentra disponible en la base de datos.

- **Restricciones:**
  - No se contempla el registro público de nuevos usuarios.
  - Las credenciales de acceso deben mantenerse protegidas para evitar accesos no autorizados.
- **Estructura general:**
  - **Entrada:** Nombre de usuario y contraseña.
  - **Proceso:**
    - Recibir las credenciales ingresadas por el usuario.
    - Validar las credenciales con la información almacenada.
    - Autorizar o rechazar el acceso según el resultado de la validación.
  - **Salida:** Acceso autorizado o denegado al sistema.

4.3.4. Capa de Presentación e Interacción

La capa de presentación e interacción constituye el punto de comunicación entre el usuario y las funcionalidades disponibles en el sistema. Permite presentar la información obtenida durante el monitoreo y entrega acceso a las funciones disponibles mediante una interfaz web.

4.3.4.1. Interfaz Gráfica de Usuario

El módulo de interfaz de usuario permite al encargado de monitoreo visualizar la información obtenida durante el funcionamiento del sistema. A través de esta interfaz se presenta el video procesado junto con los resultados asociados al monitoreo.

- **Propósito:** Presentar al usuario el video procesado junto con el conteo de personas, las alertas generadas y las opciones disponibles para la consulta y generación de reportes.
- **Alcance:** El módulo de interfaz gráfica de usuario corresponde al módulo de interacción entre el usuario y el sistema. Su función es presentar la información obtenida durante el monitoreo y permitir el acceso a las funcionalidades asociadas a la visualización de alertas y generación de reportes.
- **Dependencias:**
  - Módulo de procesamiento de video y detección de personas para obtener el video procesado y el conteo de personas.

- Módulo de alertas del sistema para presentar las alertas generadas durante el monitoreo.
- Módulo de generación de reportes para solicitar y presentar los reportes generados por el usuario.

- **Supuestos:**
  - El usuario dispone de un dispositivo que permita acceder y visualizar la interfaz.
  - Existe conectividad entre la interfaz y los módulos asociados a la información presentada.
- **Restricciones:**
  - La información presentada se actualiza mediante la recarga de la interfaz, a excepción de la visualización del video.
  - La interfaz debe permitir su utilización por usuarios sin conocimientos técnicos especializados.
- **Estructura general:**
  - **Entrada:** Video procesado, conteo de personas, alertas e información asociada a los reportes.
  - **Proceso:**
    - Presentar el video correspondiente al monitoreo.
    - Mostrar el conteo de personas detectadas.
    - Visualizar las alertas generadas.
    - Representar gráficamente la evolución del conteo.
    - Permitir la solicitud y descarga de reportes.
  - **Salida:** Información del monitoreo presentada al usuario mediante la interfaz.

La organización de los módulos permite establecer de manera estructurada el flujo de información y las relaciones entre los componentes que conforman la solución propuesta.

4.4. Matriz de Requisitos Funcionales y Módulos

La matriz de requisitos funcionales y módulos permite identificar los módulos que participan en el cumplimiento de cada requisito, estableciendo su relación con la arquitectura propuesta.

Para facilitar su representación, se utilizan abreviaturas CAP para Captura de Video, PROC para Procesamiento de Video y Detección de Personas, GUI para Interfaz Gráfica de Usuario, LOG para Registro

de Eventos, ALR para Sistema de Alertas, REP para Generación de Reportes y USR para Acceso por Credenciales. En la Tabla 4.1 se presenta la relación correspondiente.

	CAP	**PROC**	**GUI**	**LOG**	**ALR**	**REP**	**USR**
RF1	X						
RF2	X	X					
RF3				X			
RF4		X	X	X	X	X	
RF5		X	X		X		
RF6			X	X		X	
RF7			X				X
RF8	X	X	X				
RNF1	X						
RNF2			X				
RNF3	X	X					
RNF4	X	X	X		X		
RNF5	X	X	X	X	X	X	X
RNF6			X				X

Tabla 4.1: Matriz de requisitos funcionales y no funcionales y módulos del sistema.

4.5. Diseño de la Interfaz de Usuario

El diseño de la interfaz de usuario busca facilitar la visualización de la información y el acceso a las funcionalidades principales asociadas al monitoreo. Para mantener una organización clara entre las distintas vistas, se considera una estructura visual común, utilizando una barra lateral de navegación y una distribución de los elementos orientada a destacar información relevante, como el conteo de personas, niveles de ocupación y alertas generadas.

4.5.1. Modelo de Navegación

El modelo de navegación del sistema muestra de manera general cómo el usuario interactúa con la plataforma una vez ha accedido correctamente.

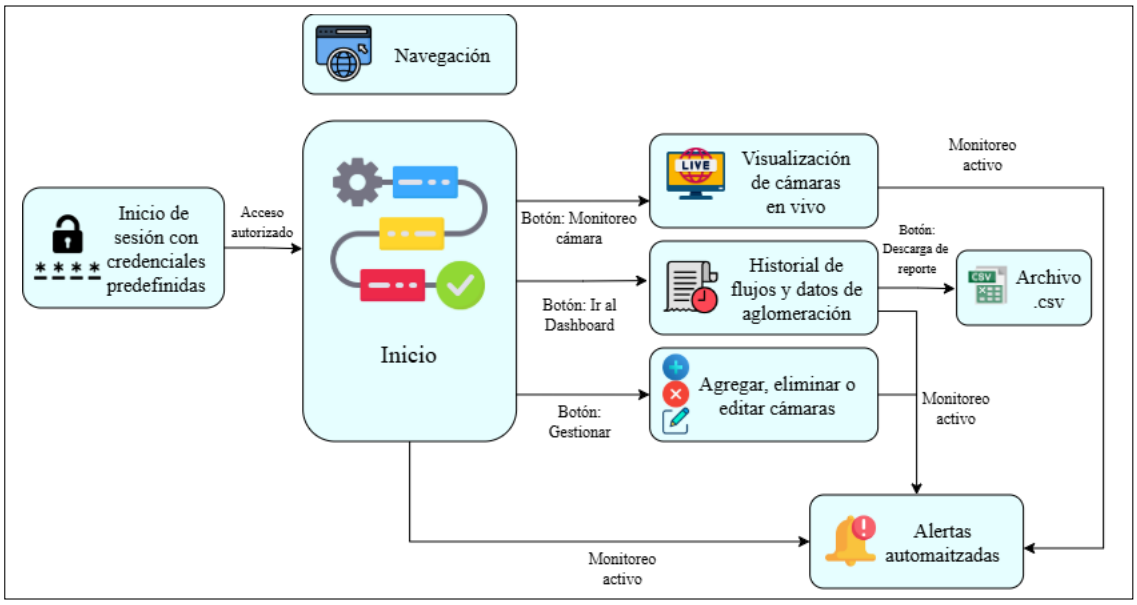


Figura 4.4: Modelo de navegación del sistema.

El modelo de navegación del sistema se estructura para facilitar el acceso del operador a las distintas funcionalidades disponibles, manteniendo organización entre las vistas que conforman la interfaz.

4.5.2. Vistas del Prototipo

A partir del modelo de navegación, se presentan los prototipos correspondientes a las principales vistas consideradas para la interfaz de usuario. Estas representan la distribución de los elementos y las funcionalidades que el operador tiene disponibles durante la interacción con el sistema y su diseño responde a los requisitos de interfaz definidos en la Subsección 4.1.3. Los objetivos específicos de cada una se detallan a continuación.

- **Vista de Inicio de Sesión:** Permite al operador ingresar sus credenciales y acceder a las distintas funciones del sistema (véase Figura 4.5).



Figura 4.5: Vista inicio de sesión.

- **Vista de Panel Pricipal:** Corresponde a la vista inicial luego de acceder al sistema y permite al operador dirigirse a las otras secciones disponibles para el monitoreo (véase Figura 4.6).

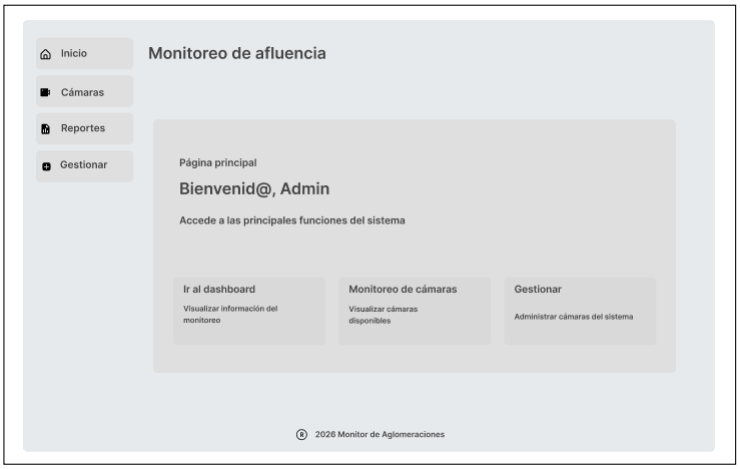


Figura 4.6: Vista Panel Principal.

- **Vista de Monitoreo de Cámaras:** Permite visualizar las cámaras disponibles para el monitoreo, junto con la información asociada al conteo de personas y al nivel de ocupación de cada una (véase Figura 4.7).

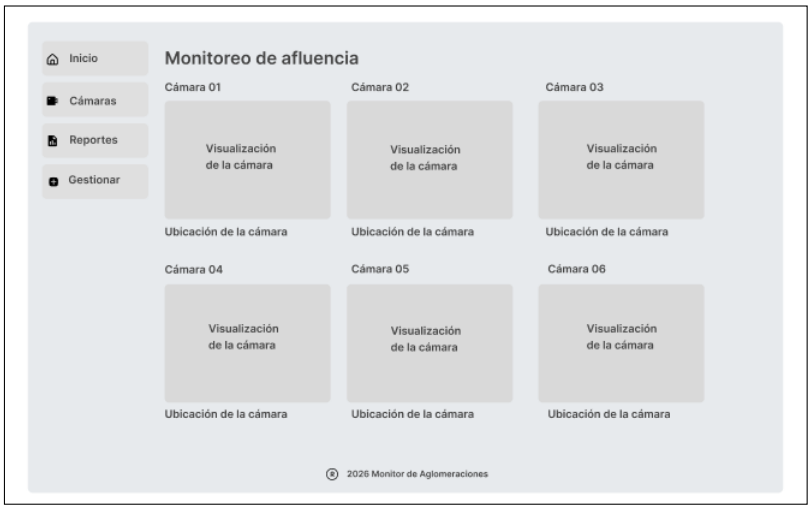


Figura 4.7: Vista monitoreo de cámaras.

- **Vista de Reportes:** Permite consultar la información histórica, gráficos de afluencia y generación de reportes (véase Figura 4.8).



Figura 4.8: Vista de métricas.

- **Vista de Gestión de Cámaras:** Permite administrar las cámaras, configurar la información y los umbrales asociados a cada una (véase Figura 4.9).

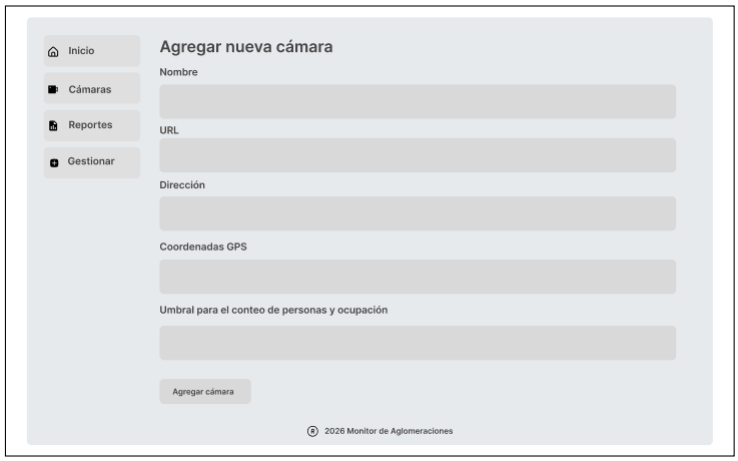


Figura 4.9: Vista gestión de cámaras.

- **Visualización de Alertas:** Permite informar al operador cuando se detectan condiciones asociadas a los umbrales definidos y presenta la información correspondiente al evento generado (véase Figura 4.10).

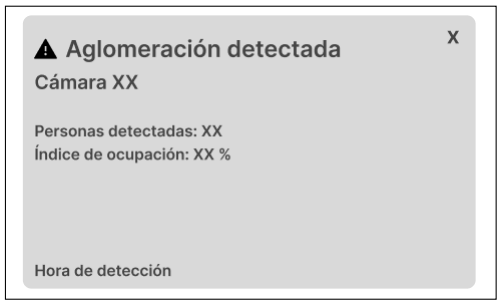


Figura 4.10: Visualización de alertas.

4.6. Transición hacia la implementación

El diseño presentado en este capítulo establece la estructura general de la solución propuesta, considerando los requisitos del sistema, su arquitectura, los módulos que la componen y el diseño de la interfaz de usuario. A partir de estos elementos se define la organización y funcionamiento esperado del sistema de monitoreo, proporcionando una base estructurada para el desarrollo de la solución.

Capítulo 5

Implementación de la Solución

El presente capítulo muestra el proceso de implementación de la solución propuesta, que consiste en el desarrollo de un sistema de monitoreo inteligente de afluencia de personas en espacios públicos, con énfasis en Metro de Santiago como caso de estudio. Para su desarrollo, se llevan a cabo distintas etapas asociadas a la construcción e integración de los módulos definidos en el diseño de la solución, considerando desde la captura de la información, hasta su visualización en una interfaz web. Asimismo, se describe la implementación de las funcionalidades destinadas al monitoreo de la afluencia, la generación de alertas y elaboración de reportes a partir de la información obtenida durante el funcionamiento del sistema.

5.1. Caso de Estudio

El caso de estudio de esta investigación se centra en el monitoreo de afluencia de personas en espacios públicos de alta demanda, considerando el Metro de Santiago como caso de referencia debido a las concentraciones de pasajeros que se pueden producir en sus estaciones durante distintos periodos de la operación. En este contexto, es posible abordar el análisis de espacios en los que la cantidad de personas puede variar continuamente, requiriendo de mecanismos que permitan obtener información sobre los distintos niveles de afluencia que existen.

El enfoque de la implementación consiste en desarrollar un sistema de monitoreo inteligente capaz de procesar información proveniente de distintas fuentes de video, para la detección y conteo de personas presentes en un área establecida. Con esta información, es posible obtener métricas asociadas a la afluencia y el índice de aglomeración, además de generar alertas cuando los umbrales establecidos por cada cámara

son sobrepasados. Para ello, se implementan los componentes definidos en el Capítulo 4, integrando distintas herramientas y tecnologías de acuerdo con las funciones requeridas en cada componente del sistema.

5.1.1. Herramientas Utilizadas

La implementación del prototipo integra la captura y transmisión de video, su procesamiento mediante técnicas de inteligencia artificial y visión por computador, el almacenamiento de la información obtenida y su posterior visualización en una interfaz web. Además, la implementación considera la generación de reportes en formato `CSV` y el uso de grabaciones correspondientes a espacios del Metro de Santiago para evaluar el funcionamiento del prototipo en escenarios con alta presencia de personas. En la Tabla 5.1 se presentan las principales herramientas utilizadas durante la implementación del prototipo. Por su parte, la Figura 5.1 muestra la integración de los componentes desarrollados, las herramientas asociadas y su comunicación dentro del sistema.

Componente	**Herramientas y tecnologías utilizadas**
Captura de video	Cámaras y fuentes de video
Procesamiento de video	Python / OpenCV / Ultralytics / YOLOv8s / YOLOv8m-seg
Backend y almacenamiento	FastAPI / MySQL
Interfaz web	React

Tabla 5.1: Herramientas y tecnologías utilizadas en la implementación de la solución.

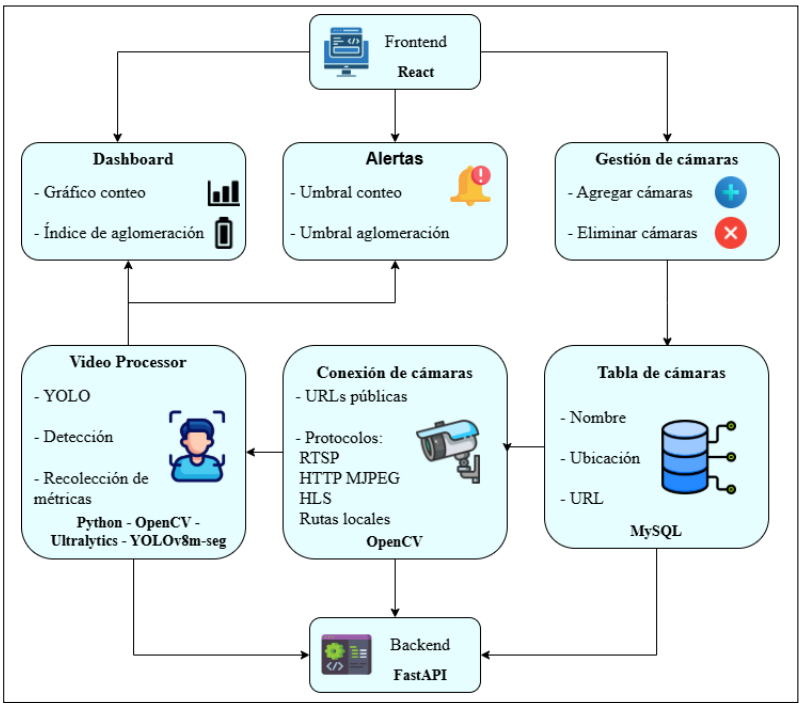


Figura 5.1: Estructuración de componentes del sistema.

5.1.1.1. Captura de Video

Es el módulo encargado de recibir las transmisiones que provienen de las cámaras y permite la conexión simultánea de múltiples fuentes de video. Para que estas sean incorporadas al prototipo se consideran direcciones URL (del inglés *Uniform Resource Locator*) públicas y rutas locales. Las transmisiones URL utilizan los protocolos `RTSP`, `HTTP-MJPEG` y `HLS`. Los flujos recibidos se utilizan posteriormente en el módulo de procesamiento para el procesamiento de información.

5.1.1.2. Procesamiento de Video

El módulo de procesamiento implementa el modelo de detección de YOLO (del inglés *You Only Look Once*) y es utilizado para identificar y contar personas en tiempo real. A partir de las detecciones obtenidas, el sistema permite calcular el índice de aglomeración e identificar condiciones críticas durante el monitoreo.

YOLO corresponde a la familia de modelos orientados a la detección de objetos que permite identificar y localizar elementos presentes en una imagen. Su funcionamiento presenta un equilibrio entre velocidad y precisión, favoreciendo su aplicación en tareas que requieren procesamiento en tiempo real [115].

Para la implementación del prototipo se considera `YOLOv8`, disponible en distintas configuraciones

que modifican la complejidad del modelo de acuerdo con los recursos y requerimientos de una aplicación: `YOLOv8n`, `YOLOv8s`, `YOLOv8m`, `YOLOv8l` y `YOLOv8x`. A medida que aumenta el tamaño de estas configuraciones, también aumenta la cantidad de parámetros y operaciones requeridas, permitiendo alcanzar mayores niveles de precisión a cambio de una mayor demanda computacional [116]. Para comparar las configuraciones de `YOLOv8` se considera la precisión media promedio (del inglés *mean Average Precision* y su abreviación mAP), junto con la cantidad de parámetros, las operaciones de punto flotante (del inglés *Floating Point Operations* y su abreviación FLOPs) y la velocidad de procesamiento. En la Tabla 5.2 se presentan las características de las configuraciones disponibles.

Modelo	**Tamaño (píxeles)**	**mAP (0.5–0.95)**	**Parámetros (M)**	**FLOPs (G)**	**Velocidad (ms)**
YOLOv8n	640 × 640	37,3	3,2	8,7	0,99
YOLOv8s	640 × 640	44,9	11,2	11,2	1,20
YOLOv8m	640 × 640	50,2	25,9	25,9	1,83
YOLOv8l	640 × 640	52,9	43,7	43,7	2,39
YOLOv8x	640 × 640	53,9	68,2	68,2	3,53

Tabla 5.2: Comparación de rendimiento entre configuraciones de YOLOv8 [116].

Considerando las diferencias presentadas, las configuraciones de `YOLOv8` evidencian distintos niveles de detección, velocidad y complejidad computacional. A medida que aumenta el tamaño del modelo, se obtienen mayores valores de mAP, pero también aumenta la cantidad de parámetros, operaciones y tiempo de procesamiento [116]. En función de estas características, se selecciona `YOLOv8s` para la implementación del prototipo, pues permite mantener equilibrio entre el rendimiento de detección y los requerimientos computacionales necesarios para el procesamiento de video en tiempo real.

5.1.1.3. Backend y Almacenamiento

El backend del prototipo se desarrolla en `FastAPI`, un framework basado en Python orientado a la creación de aplicaciones y API, que permite gestionar las solicitudes realizadas al sistema [112]. Este componente coordina la comunicación entre los distintos módulos del sistema, interactúa con la base de datos y entrega los servicios necesarios para autenticar usuarios, gestionar el procesamiento de video y coordinar la generación de reportes en formato CSV.

Para almacenar la información se utiliza `MySQL`, un sistema de gestión de bases de datos relacional que organiza los datos mediante tablas y permite realizar operaciones a través del lenguaje SQL [111]. Su implementación permite almacenar y gestionar la información utilizada por los distintos componentes del

prototipo, manteniendo la comunicación con el backend desarrollado en FastAPI.

5.1.1.4. Interfaz Web

Para la implementación de la interfaz web se utiliza `React`, una biblioteca de `JavaScript` orientada al desarrollo de interfaces de usuario mediante una arquitectura basada en componentes reutilizables [114]. Esta herramienta permite estructurar la información presentada al usuario y actualizar los elementos visuales de acuerdo con los datos obtenidos durante el funcionamiento del sistema.

En el desarrollo del prototipo, el módulo de interfaz web permite la interacción directa con el usuario e incluye un inicio de sesión, un panel principal para el monitoreo de las múltiples cámaras almacenadas en la base de datos, la visualización de métricas en tiempo real, alertas automáticas y la descarga de reportes. Su diseño prioriza la claridad visual y la adaptabilidad a distintos dispositivos.

5.2. Análisis del Flujo de Video

El proceso de monitoreo comienza con la recepción de las fuentes de video utilizadas por el sistema. El módulo de captura se encarga de recibir estas transmisiones y proporcionar los cuadros necesarios para su análisis posterior, lo que permite trabajar con múltiples fuentes de manera simultánea. Esta etapa es parte de la entrada del flujo de procesamiento y mantiene separada la obtención del video de las tareas asociadas a la detección de personas.

Para cada cámara registrada en el sistema se mantiene una instancia de procesamiento asociada a su identificador, permitiendo administrar de manera independiente los flujos incorporados al monitoreo. Cuando se agrega una cámara, el sistema crea el procesador utilizando la fuente de video configurada y los parámetros asociados a dicha cámara. Así, incorporar nuevas fuentes no modifica el funcionamiento de las cámaras existentes y permite mantener el procesamiento de múltiples transmisiones en una misma plataforma.

Una vez obtenidos los cuadros, estos se procesan para identificar a las personas presentes en la escena. En función de estas características, se selecciona inicialmente `YOLOv8s` como configuración base para el prototipo, pues permite mantener un equilibrio entre el rendimiento de detección y los requerimientos computacionales necesarios para el procesamiento de video en tiempo real. En la versión implementada se utiliza `YOLOv8m-seg`, lo que permite complementar la detección de personas con información espacial de las áreas identificadas en cada cuadro.

A partir de los resultados obtenidos durante el procesamiento, se determina la cantidad de personas

presentes en la escena y el índice de aglomeración asociado al flujo analizado, que representa el nivel de ocupación observado en la escena. Estas métricas se actualizan durante el monitoreo y se mantienen vinculadas a la cámara que corresponde, permitiendo que cada fuente tenga sus propios resultados. Además, cada cámara mantiene umbrales asociados al conteo de personas y al índice de aglomeración. Estos valores se obtienen desde la configuración almacenada para la cámara y se asignan a su procesador al momento de iniciar el monitoreo. Durante el procesamiento, estas métricas se comparan con el umbral, lo que permite identificar cuando existen eventos con sus condiciones establecidas sobrepasadas. Cuando esto ocurre, se genera una alerta asociada a la cámara en la que se origina el evento, incluyendo la información correspondiente al conteo e índice de aglomeración detectados.

Los resultados generados, se utilizan de manera simultánea por distintos componentes del sistema. El flujo procesado permite presentar el video durante el monitoreo, mientras que las métricas obtenidas son utilizadas para actualizar los gráficos e indicadores disponibles en la interfaz. Además, esta información se registra con el objetivo de mantener un historial de las mediciones realizadas, lo que permite consultar posteriormente su evolución y generar los reportes correspondientes. En la Figura 5.2 se presenta un ejemplo del procesamiento realizado sobre una de las grabaciones utilizadas durante las pruebas del sistema.

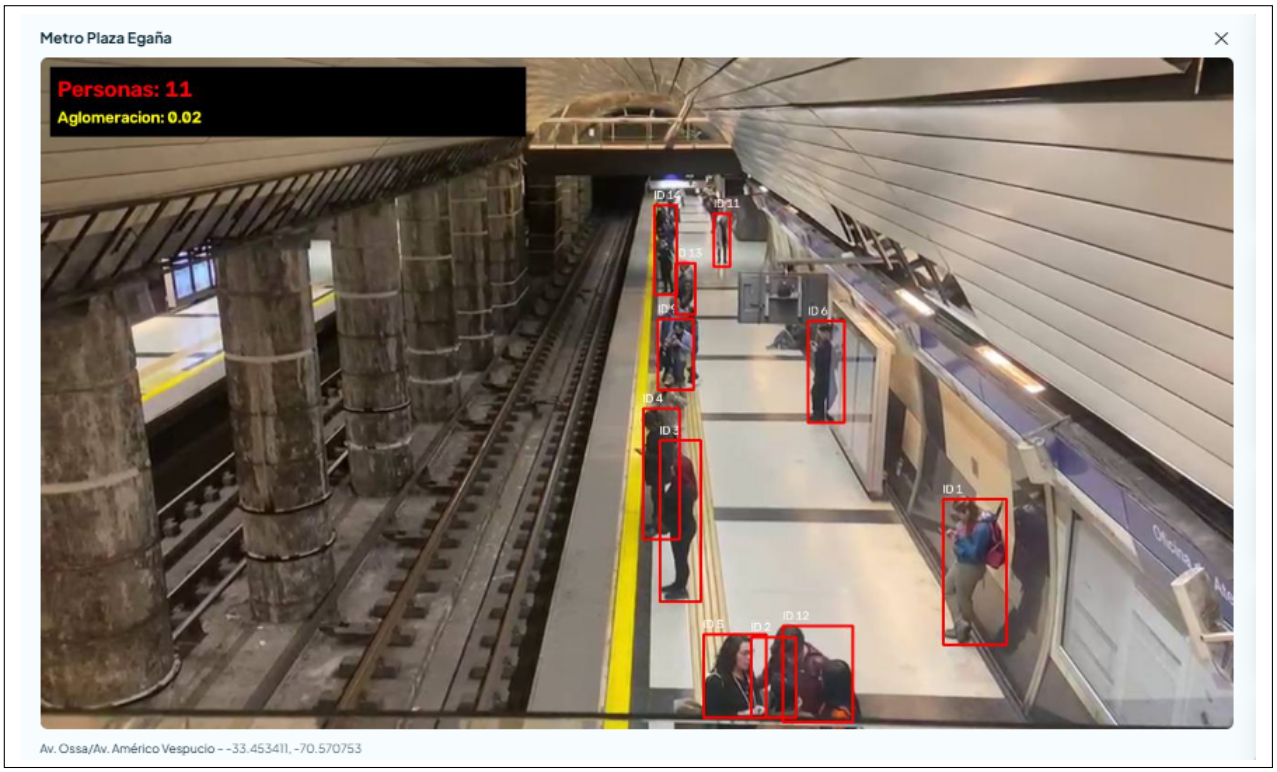


Figura 5.2: Procesamiento del flujo de video para la detección de personas.

5.2.1. Resultados del Procesamiento

Con el objetivo de analizar los resultados obtenidos durante el monitoreo, se consideran escenas procesadas bajo diferentes condiciones de visualización. Para ello, se utilizan los valores generados por el sistema correspondientes al conteo de personas y al índice de aglomeración.

El índice de aglomeración se obtiene a partir de las máscaras de segmentación generadas por `YOLOv8m-seg` para las personas identificadas en cada cuadro. Para su cálculo, se considera el área ocupada por las máscaras correspondientes a la clase persona respecto del área total analizada, obteniendo un valor normalizado entre 0 y 1. Se calcula como:

$$I_{\text{aglomeración}} = \frac{A_{\text{personas}}}{A_{\text{total}}},$$

donde un valor cercano a 0 representa una menor ocupación de la escena y su aumento indica una mayor proporción del área ocupada por personas.

Para reducir variaciones entre cuadros consecutivos, el valor obtenido se estabiliza mediante una media móvil exponencial. Por lo tanto, el índice de aglomeración no depende únicamente de la cantidad de personas detectadas, sino que también del área que estas ocupan dentro de la imagen.

Para observar cómo se comporta el procesamiento se consideran dos escenas con características diferentes. La primera corresponde al escenario de Metro Plaza Egaña, utilizado como parte del caso de estudio. En esta escena, el sistema obtiene un conteo de 9 personas y un índice de aglomeración de 0,01. En una revisión manual del cuadro se identifican 12 personas visibles, donde se observan diferencias en sectores donde existe cercanía entre personas, oclusión parcial y elementos de la infraestructura que dificultan la visualización completa de algunos individuos. En este caso, la diferencia entre el conteo manual y el obtenido por el sistema corresponde a 3 personas, equivalente a un error de conteo del 25 %.

La segunda escena corresponde al video utilizado durante las pruebas locales del prototipo. En este caso, el sistema muestra un conteo de 16 personas y un índice de aglomeración de 0,12. En contraste con la escena anterior, las personas ocupan una mayor proporción del cuadro debido a su cercanía a la cámara, lo que se refleja en un índice de aglomeración superior. Esta diferencia permite observar que el indicador considera la ocupación espacial obtenida mediante la segmentación y no solamente la cantidad de personas presentes. En la Tabla 5.3 se sintetizan los resultados del procesamiento obtenidos.

Escenario	**Referencia manual**	**Conteo sistema**	**Error conteo**	**Índice**
Metro Plaza Egaña	12	9	25 %	0,01
Prueba Local	18	16	11,1 %	0,12

Tabla 5.3: Resultados obtenidos durante el procesamiento de los flujos de video.

Las diferencias observadas entre ambos escenarios se encuentran relacionadas con las condiciones de cada escena, donde factores como la distancia respecto de la cámara, la agrupación de personas y las oclusiones parciales pueden dificultar su identificación.

5.3. Gestión y Almacenamiento de Datos

La gestión de la información utilizada durante el funcionamiento del sistema, se realiza mediante la base de datos MySQL, utilizada para centralizar los datos asociados a las cámaras y resultados generados durante el monitoreo. MySQL corresponde a un sistema de gestión de base de datos relacional que organiza la información mediante tablas con una estructura previamente definida y permite establecer relaciones entre los datos almacenados [111]. Su elección responde a la estructura de la información utilizada por el sistema, ya que los registros generados durante el monitoreo tienen relación entre sí. De esta manera, la información que corresponde a las cámaras, se puede asociar con los registros históricos, manteniendo una organización estructurada de los datos y permitiendo su posterior consulta desde los diferentes componentes del sistema. A partir de esta organización, en la Figura 5.3 se presenta el esquema relacional de la base de datos implementada, identificando las tablas que la conforman y las relaciones establecidas entre ellas.

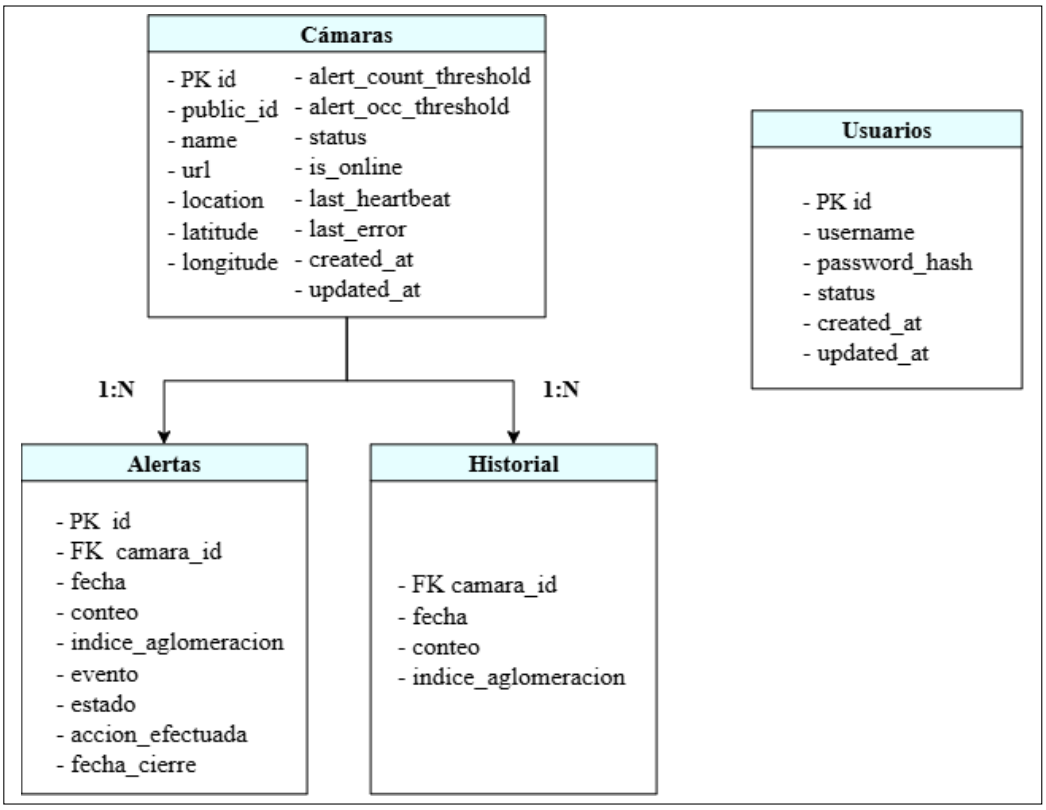


Figura 5.3: Esquema relacional de la base de datos del sistema.

Para la gestión de cámaras se utiliza una tabla de cámaras, que almacena la información necesaria para identificar, localizar y establecer la conexión con cada una de las fuentes incorporadas al sistema. Cada registro considera un identificador interno y uno público, el nombre de la cámara, la dirección URL, su ubicación y sus coordenadas geográficas. Además, almacena los umbrales definidos para la generación de alertas, información asociada a su estado y disponibilidad, junto con los registros relacionados con su creación y actualización. En la Tabla 5.4 se presentan los datos almacenados para la gestión de las cámaras.

Campo	Descripción
`id`	Identificador interno de la cámara.
`public_id`	Identificador utilizado por la interfaz web.
`name`	Nombre asignado a la cámara.
`url`	Dirección utilizada para acceder a la fuente de video.
`location`	Ubicación asociada a la cámara.
`latitude` / `longitude`	Coordenadas geográficas asociadas a la ubicación de la cámara.
`alert_count_threshold`	Umbral de cantidad de personas definido para la generación de alertas.
`alert_occ_threshold`	Umbral del índice de aglomeración definido para la generación de alertas.
`status`	Estado de la cámara dentro del sistema.
`is_online`	Indica la disponibilidad de la cámara.
`last_heartbeat`	Último registro de disponibilidad de la cámara.
`last_error`	Último error registrado para la cámara.
`created_at` / `updated_at`	Fecha de creación y última actualización del registro.

Tabla 5.4: Estructura de datos utilizada para el almacenamiento de cámaras.

Por otro lado, la información generada durante el análisis de flujo de video se almacena en una tabla de historial. Cada registro se asocia a una cámara y contiene la fecha y hora de medición, el conteo de personas y el índice de aglomeración obtenido durante el procesamiento. Con esta estructura, es posible conservar los resultados generados a lo largo del tiempo y luego recuperarlos para su consulta y uso en la generación de reportes. En la Tabla 5.5 se presenta la estructura utilizada para almacenar estos registros.

Campo	Descripción
`fecha`	Fecha y hora asociada al registro.
`camara_id`	Identificador de la cámara asociada al registro.
`conteo`	Cantidad de personas obtenida durante el procesamiento.
`indice_aglomeracion`	Índice de aglomeración calculado para el registro.

Tabla 5.5: Estructura de datos utilizada para el almacenamiento del historial de monitoreo.

Además, se implementa una tabla destinada al almacenamiento de alertas generadas durante el monitoreo. Cuando las condiciones previamente establecidas para una cámara levantan una alerta, el evento se asocia a la fuente correspondiente y se almacenan las métricas existentes al momento de su detección. Inicialmente estos registros se mantienen en estado pendiente hasta que el usuario realiza su gestión desde la interfaz.

La gestión de una alerta permite registrar la acción que realiza el usuario y modificar su estado,

conservando fecha y hora en la que se realiza el cierre. Así, el sistema es capaz de mantener un registro de los eventos detectados y de las acciones realizadas posteriormente. La Tabla 5.6 muestra los campos utilizados en la base de datos.

Campo	**Descripción**
`id`	Identificador interno de la alerta.
`camara_id`	Identificador de la cámara que generó la alerta.
`fecha`	Fecha y hora en que se generó la alerta.
`conteo`	Cantidad de personas detectadas al momento de generar la alerta.
`indice_aglomeracion`	Índice de aglomeración registrado al momento de generar la alerta.
`evento`	Tipo de evento asociado a la alerta generada.
`estado`	Estado de gestión de la alerta, correspondiente a pendiente o cerrada.
`accion_efectuada`	Descripción de la acción registrada por el usuario al gestionar la alerta.
`fecha_cierre`	Fecha y hora en que la alerta fue cerrada.

Tabla 5.6: Estructura de datos utilizada para el almacenamiento y gestión de alertas.

Por último, para controlar el acceso al sistema se utiliza una tabla de usuarios, encargada de almacenar la información necesaria para la autenticación. Cada usuario dispone de un identificador interno y un nombre de usuario único, junto con la contraseña almacenada. Además, se mantiene el estado de la cuenta y las fechas asociadas a la creación y actualización del registro. En la Tabla 5.7 se presentan los campos utilizados para el almacenamiento de esta información.

Campo	**Descripción**
`id`	Identificador interno del usuario.
`username`	Nombre de usuario utilizado para el acceso al sistema.
`password_hash`	Contraseña almacenada mediante su valor hash.
`status`	Estado del usuario dentro del sistema, correspondiente a activo o inactivo.
`created_at` / `updated_at`	Fecha de creación y última actualización del registro.

Tabla 5.7: Estructura de datos utilizada para el almacenamiento de usuarios.

En conjunto, las estructuras implementadas permiten relacionar la configuración de las cámaras con los registros obtenidos durante el monitoreo. Esta información es gestionada por el `backend` y utilizada por la interfaz para su consulta y visualización.

5.4. Generación de Alertas

La generación de alertas permite identificar situaciones en las que los resultados obtenidos durante el procesamiento alcanzan los límites configurados para cada cámara. Cuando alguno de estos es alcanzado o superado, se genera un evento asociado a la cámara que corresponde, el cual queda registrado junto con la información obtenida en ese instante y su estado de gestión.

Las escenas analizadas en Metro de Santiago no representan una condición de aglomeración. Sin embargo, con el propósito de verificar el funcionamiento del mecanismo de generación de alertas, se realiza una prueba controlada utilizando el escenario de Metro Plaza Egaña. Para ello, se establece temporalmente un umbral de 8 personas y un umbral de ocupación de 100 %. Estos valores son seleccionados exclusivamente para provocar una condición de alerta asociada al conteo de personas, por lo que no representan umbrales reales de aglomeración para la estación.

Durante la prueba, el conteo de personas supera el umbral establecido, generándose las alertas correspondientes mientras la condición permanece activa. Los eventos registran tanto el número de personas detectadas como el índice de aglomeración observado en cada instante. Asimismo, las alertas permanecen almacenadas una vez finalizado el procesamiento, permitiendo su consulta y posterior gestión.

Durante el monitoreo, las alertas generadas se presentan mediante notificaciones emergentes en la interfaz, permitiendo al operador identificar la cámara asociada, el número de personas detectadas y el porcentaje de ocupación. La Figura 5.4 muestra estas notificaciones durante el funcionamiento del sistema.

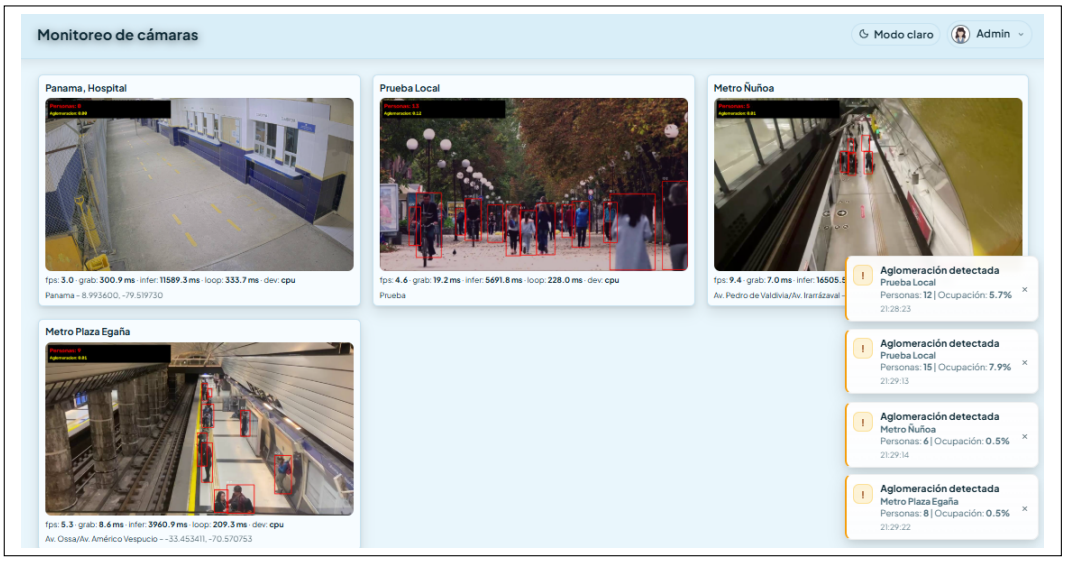


Figura 5.4: Notificaciones emergentes de alertas durante el monitoreo.

Como parte de la prueba, se verifica también el ciclo de gestión de una alerta. En el evento que se muestra en la Figura 5.5, se registran 16 personas y un índice de aglomeración de 0,032. Inicialmente, el evento se encuentra en estado pendiente, pero luego de registrar la acción efectuada, su estado cambia a cerrado, conservando tanto la acción realizada como la fecha y hora de cierre.



(a) Alerta en estado pendiente.

(b) Alerta gestionada y cerrada.

Figura 5.5: Generación y gestión de una alerta.

En conjunto, los resultados obtenidos permiten comprobar el comportamiento de las alertas durante su generación y posterior gestión.

5.5. Desarrollo de la Interfaz Web

La interfaz web es el componente mediante el cual el usuario interactúa con las funcionalidades implementadas en el sistema. Su desarrollo permite integrar en un mismo entorno la visualización de las fuentes de video, consulta de información, la gestión de alertas y la administración de las cámaras registradas.

La implementación de la interfaz considera la comunicación entre el `frontend` y el `backend` para acceder a la información procesada y almacenada por el sistema. Con esta comunicación es posible presentar al usuario los datos de monitoreo mediante distintas vistas, y acceder a las funciones disponibles mediante una estructura de navegación común.

5.5.1. Vistas Principales del Sistema

La interfaz desarrollada se organiza en diferentes vistas que permiten al usuario acceder a las funciones disponibles durante la operación del sistema. A continuación se presentan las principales vistas implementadas en el prototipo y los elementos que conforman cada una de ellas.

5.5.1.1. Vista Inicio de Sesión

La vista de inicio de sesión mantiene la estructura definida durante la etapa de diseño e incorpora la autenticación mediante credenciales registradas por el usuario administrador. Una vez validadas, el sistema permite el ingreso a la plataforma y dirige al usuario hacia la vista principal. La interfaz añade modos de visualización claro y oscuro, los cuales mantienen la distribución de los componentes y adaptan sus características visuales según la configuración seleccionada. En la Figura 5.6 se muestra la implementación de esta vista en modo claro.



Figura 5.6: Vista implementada de inicio de sesión.

5.5.1.2. Vista de Panel Principal

Corresponde a la primera interfaz presentada luego del acceso al sistema. En su implementación se incorpora una barra lateral que mantiene disponibles las secciones principales de la plataforma, correspondientes al monitoreo de cámaras, reportes, gestión de alertas y gestión de cámaras. Además, la vista principal cuenta con accesos directos a estas funcionalidades y presenta la cantidad de cámaras registradas en el sistema. En la Figura 5.7 se presenta la vista implementada utilizando el modo claro.



Figura 5.7: Vista implementada del panel principal.

5.5.1.3. Vista de Monitoreo de Cámaras

La vista de monitoreo de cámaras muestra las fuentes de video registradas en el sistema y permite acceder de manera individual a su visualización durante el procesamiento. Para cada cámara se muestra su nombre, ubicación y coordenadas geográficas, junto con los indicadores obtenidos durante el análisis correspondientes al conteo de personas y al índice de aglomeración.

Durante el monitoreo, las detecciones realizadas se presentan de manera directa sobre el flujo de video, lo que permite identificar visualmente a las personas detectadas en cada escena. A su vez, el conteo de personas y el índice de aglomeración se actualizan según los resultados que se obtienen durante el procesamiento. De esta manera, el usuario puede observar en una misma vista la fuente de video procesada, las detecciones realizadas y los indicadores asociados al monitoreo.

Al seleccionar una cámara, el flujo procesado se presenta en una vista ampliada, manteniendo disponible las detecciones realizadas y los indicadores asociados al monitoreo. En la Figura 5.8 se presenta la vista implementada utilizando el modo claro.

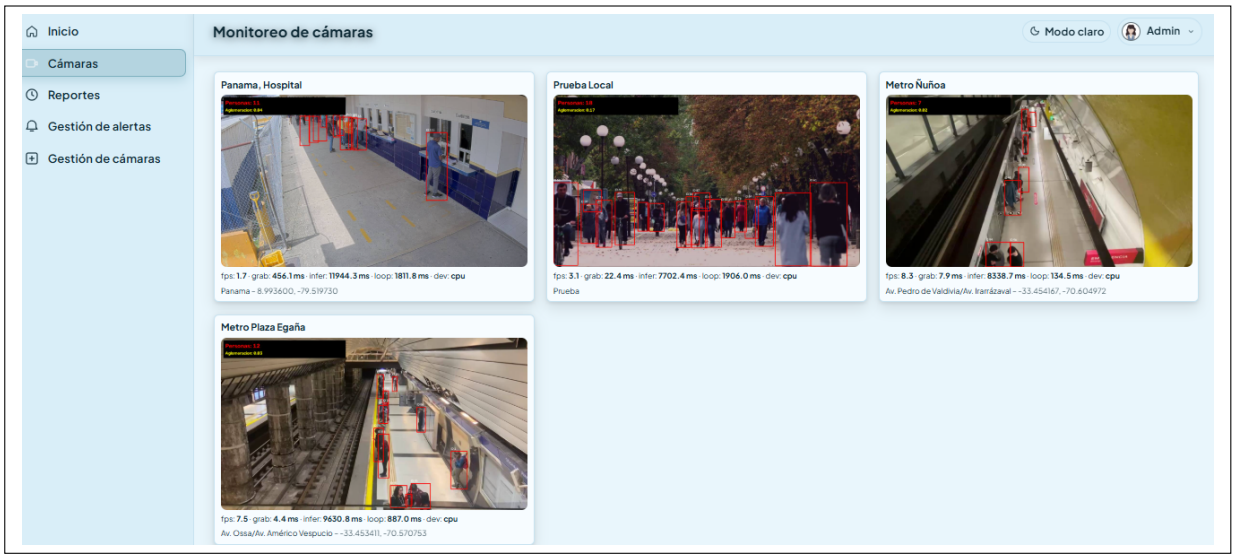


Figura 5.8: Vista implementada para el monitoreo de las cámaras registradas.

5.5.1.4. Vista de Reportes

La vista de reportes presenta los indicadores principales obtenidos durante el monitoreo para cada una de las cámaras registradas. Para cada fuente se presenta la evolución temporal del conteo de personas y el índice de aglomeración, permitiendo al usuario observar los valores obtenidos y sus variaciones durante el periodo analizado.

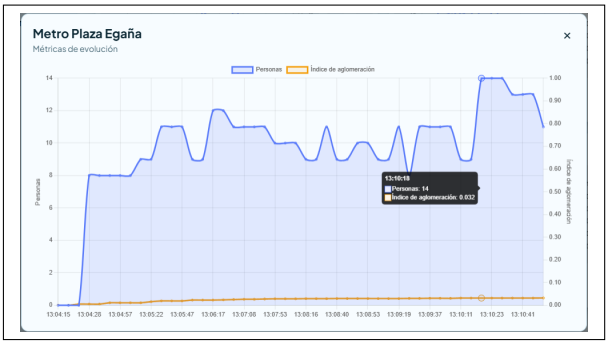
El conteo de personas se representa mediante una línea de evolución que permite identificar los cambios en la cantidad detectada a través del tiempo. Por su parte, el índice de aglomeración se presenta en una segunda serie dentro del mismo gráfico y se complementa con un indicador vertical que muestra su valor actual. Este último incorpora una representación visual mediante colores, lo que permite reconocer el nivel asociado al indicador sin depender únicamente de su valor numérico. Así, ambos indicadores clave de rendimiento (KPI del inglés *Key Performance Indicator*) son presentados de manera complementaria: el conteo muestra la cantidad de personas mientras que el índice presenta la proporción del área de la imagen ocupada por estas.

La visualización de los KPI es interactiva, pues el usuario puede ampliar el gráfico que corresponde a cada cámara y consultar los valores registrados en diferentes instantes del monitoreo. Al posicionarse sobre un punto, la interfaz muestra la hora del registro junto con el conteo de personas y el índice de aglomeración correspondiente. Además, cada cámara cuenta con la opción de descargar los registros en formato `CSV`, lo que permite conservar y consultar posteriormente la información presentada en el dashboard. En la Figura 5.9, correspondiente al escenario observado en Metro Plaza Egaña, se presenta la vista general de los indicadores y su visualización ampliada, en donde se observa un registro con 14 personas y un índice de aglomeración

de 0,032. Esta representación permite relacionar ambos indicadores en un mismo instante y analizar su comportamiento a lo largo del tiempo.



(a) Vista general de los indicadores.



(b) Visualización ampliada de los indicadores.

Figura 5.9: Visualización de los indicadores de monitoreo para Metro Plaza Egaña.

5.5.1.5. Vista de Gestión de Cámaras

La vista de gestión de cámaras permite administrar las fuentes de video incorporadas al sistema. En su implementación se presenta el listado de cámaras configuradas junto con sus principales datos de identificación y ubicación, permitiendo acceder a las operaciones disponibles sobre cada registro.

Para añadir o editar cámaras se realiza un formulario que permite ingresar la información requerida para conectar la fuente de video, incluyendo su nombre, dirección URL, ubicación y coordenadas geográficas. Además, es posible establecer los umbrales asociados al conteo de personas y al índice de aglomeración, los cuales son utilizados por el sistema para identificar las condiciones que generan una alerta. Estos parámetros pueden ser modificados mediante la opción de edición.

Las operaciones realizadas desde esta vista se mantienen asociadas al registro correspondiente en la base de datos, permitiendo incorporar, modificar o eliminar las fuentes utilizadas por el sistema. En la Figura 5.10 se presenta la implementación de la vista de gestión de cámaras utilizando el modo claro.

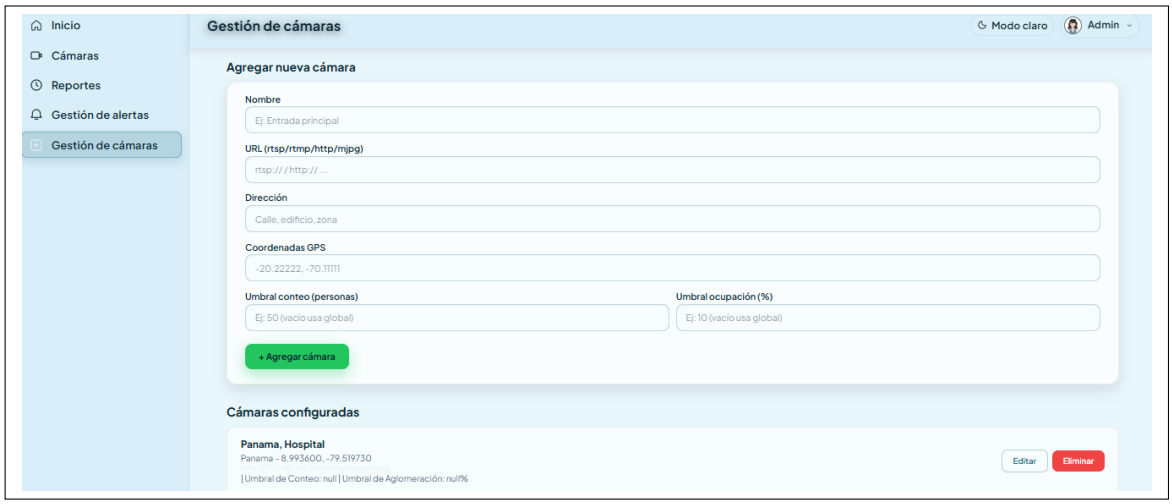


Figura 5.10: Vista implementada para la gestión de cámaras.

5.5.1.6. Vista de Gestión de Alertas

Esta vista concentra los eventos registrados por el sistema y permite al usuario consultar su información desde un mismo lugar. Las alertas se presentan mediante un listado que identifica la cámara asociada, los valores registrados, la fecha y hora del evento y su estado de gestión. Además, la interfaz incorpora filtros por estado y cámara, lo que facilita la consulta de los eventos de acuerdo con las necesidades del usuario.

Los estados de las alertas se presentan mediante indicadores visuales que permiten distinguir rápidamente los eventos pendientes de aquellos que ya han sido gestionados. Al seleccionar una alerta, la interfaz permite acceder a su información detallada y a las acciones disponibles según su estado. En la Figura 5.11 se muestra la vista general de gestión de alertas utilizando el modo claro.

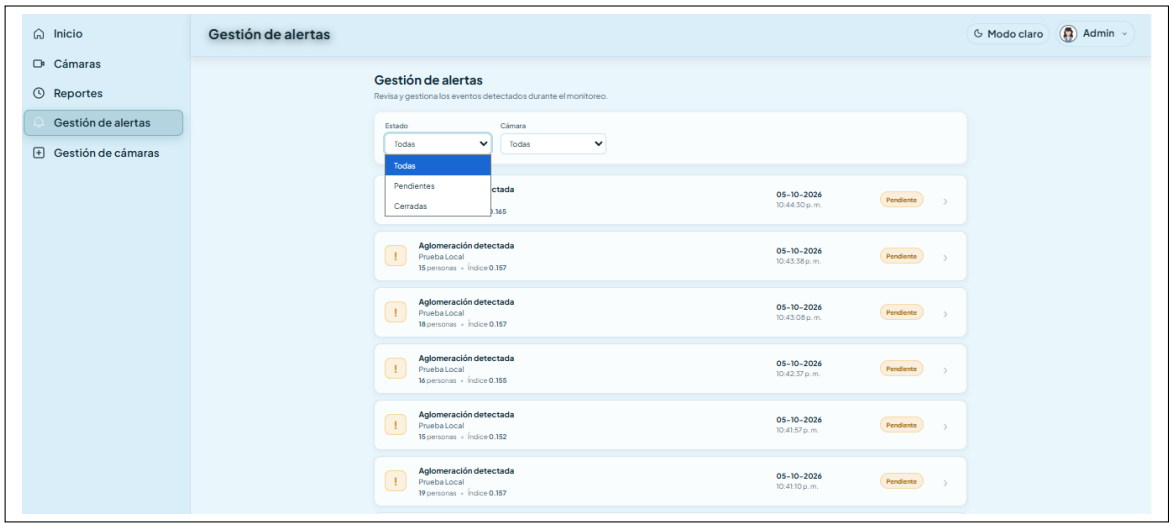


Figura 5.11: Vista implementada para la consulta y gestión de alertas.

5.6. Verificación y Validación

Luego de implementar los componentes del sistema, se realizan pruebas orientadas a verificar las principales funcionalidades definidas para el sistema y validar su comportamiento durante los escenarios utilizados en su ejecución. Para ello, se consideran las funciones disponibles para el usuario y los resultados obtenidos durante el procesamiento de las fuentes de video.

5.6.1. Verificación

La verificación se encuentra orientada a comprobar el funcionamiento de los módulos implementados de acuerdo con los requisitos definidos para el sistema. Las pruebas realizadas consideran los siguientes aspectos:

- Comprobar el acceso al sistema mediante credenciales preestablecidas, verificando que el ingreso se realice mediante un usuario autorizado.
- Verificar la visualización de las fuentes de video incorporadas al sistema desde la interfaz web.
- Comprobar que el módulo de procesamiento realice la detección y conteo de personas y genere el índice de aglomeración correspondiente.
- Verificar la generación y descarga de reportes en formato `CSV`, manteniendo los registros obtenidos durante el monitoreo.

- Evaluar la actualización de la información generada por el sistema bajo las condiciones utilizadas durante las pruebas del prototipo.
- Verificar la generación de alertas a partir de los umbrales configurados y su posterior gestión desde la interfaz web.

Estas pruebas permiten comprobar la integración de los componentes implementados y el funcionamiento del flujo establecido desde la recepción de las fuentes de video hasta la presentación de los resultados en la interfaz web.

5.6.2. Validación

La validación se orientó a comprobar el comportamiento del prototipo respecto de las funciones consideradas para el monitoreo de afluencia. Para ello, se utilizaron las fuentes de video incorporadas durante la implementación y se observaron los resultados generados durante su procesamiento.

- El conteo de personas permite representar la cantidad de detecciones obtenidas durante el procesamiento de las fuentes de video.
- El índice de aglomeración presenta variaciones de acuerdo con los resultados obtenidos durante el monitoreo.
- La interfaz permite acceder a las funciones implementadas y consultar la información generada por el sistema.
- La generación de reportes en formato `CSV` permite recuperar los registros obtenidos para su posterior consulta.
- La configuración individual de umbrales permite establecer condiciones de alerta diferenciadas para las cámaras incorporadas al sistema.

A partir de estas pruebas es posible validar el funcionamiento integrado del sistema y su capacidad para presentar al usuario la información obtenida durante el monitoreo.

5.7. Análisis Crítico y Evaluación de Riesgos

El funcionamiento observado durante la implementación permitió comprobar la integración de los módulos desarrollados y analizar aquellos aspectos que pueden influir en el desempeño del sistema. Si bien

el prototipo permite ejecutar las funciones consideradas para el monitoreo, existen factores asociados a las condiciones de operación que deben ser considerados ante una eventual utilización en escenarios reales.

- **Precisión del conteo:** la detección de personas puede verse afectada en escenarios de mayor densidad debido a la presencia de oclusiones entre personas, así como por las condiciones de iluminación y el ángulo desde el cual se obtiene la imagen.
- **Rendimiento en un contexto real:** el funcionamiento del sistema fue comprobado bajo las condiciones utilizadas durante el desarrollo del prototipo. Una implementación con una mayor cantidad de fuentes de video o bajo condiciones diferentes de conectividad puede modificar los tiempos requeridos para el procesamiento y presentación de la información.
- **Actualización del modelo:** el desempeño de la detección depende de las capacidades del modelo utilizado y de su comportamiento frente a las características de las imágenes procesadas. Por esta razón, una futura implementación puede requerir la evaluación del modelo frente a nuevos escenarios y condiciones de operación.
- **Gestión de datos:** el almacenamiento de los registros permite conservar la información generada durante el monitoreo. Sin embargo, ante una implementación en un entorno real resulta necesario considerar mecanismos destinados a resguardar la disponibilidad e integridad de la información frente a posibles fallas.

Estos aspectos corresponden principalmente a observaciones asociadas al uso del sistema fuera de las condiciones consideradas durante su desarrollo. Su identificación permite establecer los principales elementos que deben ser evaluados antes de ampliar la operación del sistema hacia escenarios con una mayor cantidad de cámaras, usuarios o condiciones de monitoreo.

5.8. Síntesis de la Implementación

El proceso de implementación presentado en este capítulo permite materializar el diseño propuesto e integrar la funcionalidades del sistema. Las pruebas realizadas permiten verificar su funcionamiento e identificar las limitaciones y aspectos que se deben considerar en su aplicación en escenarios reales. De esta manera, se establece el alcance de la solución desarrollada y las posibilidades de continuidad y mejora del sistema.

Capítulo 6

Conclusiones y Trabajo Futuro

El desarrollo de esta investigación permite abordar el problema asociado al monitoreo de la afluencia de personas en espacios públicos de alta demanda mediante el diseño e implementación de una solución basada en técnicas de visión por computador. A partir del trabajo realizado, es posible construir un sistema funcional orientado a la detección y conteo de personas.

En este capítulo se presentan las principales conclusiones obtenidas durante el desarrollo del sistema, considerando el grado de cumplimiento de los objetivos planteados y las limitaciones identificadas durante el proceso. De la misma manera, se establecen las líneas de trabajo futuro que permiten continuar y fortalecer la solución desarrollada, finalizando con una reflexión respecto del aporte de la investigación.

6.1. Cumplimiento de Objetivos

Los objetivos establecidos para el desarrollo de esta investigación son alcanzados en términos generales. El trabajo realizado permite desarrollar e implementar un sistema de monitoreo inteligente de afluencia de personas en espacios públicos, capaz de obtener información a partir del procesamiento de fuentes de video y presentarla mediante una interfaz orientada a la visualización y análisis de los resultados. Por lo tanto, se cumple el objetivo general de contar con una herramienta que genere información destinada a apoyar la toma de decisiones operativas en espacios con presencia de aglomeraciones.

A continuación, se presenta el cumplimiento de cada uno de los objetivos específicos:

- **Construcción de un conjunto de datos:** Este objetivo se cumple mediante la construcción y etiquetado de un conjunto de datos compuesto por imágenes representativas de escenarios con presencia

de personas. La información obtenida es utilizada para realizar el entrenamiento del modelo de detección aplicado durante el desarrollo del sistema.

- **Implementación del modelo de detección:** Este objetivo se cumple mediante la implementación de un modelo basado en visión por computador e inteligencia artificial para identificar y contar personas en las fuentes de video. En las pruebas realizadas, para Metro Plaza Egaña se obtuvo un conteo de 9 personas frente a una referencia manual de 12, correspondiente a un error de conteo de 25 %, mientras que en Prueba Local se registraron 16 personas frente a una referencia manual de 18, con un error de 11,1 %. Estos resultados permiten cuantificar el comportamiento del sistema frente a las condiciones presentes en los escenarios evaluados.
- **Desarrollo de la interfaz de monitoreo:** Este objetivo se cumple mediante el desarrollo de una interfaz web que permite visualizar las fuentes de video procesadas y los resultados obtenidos. La implementación presenta los dos KPI principales del sistema, correspondientes al conteo de personas y al índice de aglomeración, permitiendo consultar sus valores actuales y su evolución temporal mediante gráficos interactivos. Además, integra la visualización de las detecciones, la gestión de cámaras y la consulta de alertas desde un mismo entorno.
- **Generación de reportes e indicadores:** Este objetivo se cumple mediante la generación y almacenamiento del conteo de personas y del índice de aglomeración, lo que permite conservar su evolución temporal y consultar los registros mediante gráficos y reportes en formato `CSV`. Además, a partir de estos indicadores se implementan alertas asociadas a umbrales configurables para cada cámara. Durante la prueba funcional realizada en Metro Plaza Egaña se utilizó temporalmente un umbral de 8 personas y 100 % de ocupación, con el objetivo de comprobar la generación, almacenamiento y posterior gestión de los eventos.

6.2. Limitaciones y Dificultades

Durante el desarrollo e implementación del sistema se presentan algunas limitaciones y dificultades que condicionan el alcance de las pruebas realizadas, entre las que se destacan las siguientes:

- **Acceso a las transmisiones del Metro de Santiago:** Inicialmente se considera la posibilidad de utilizar transmisiones provenientes de las cámaras del Metro de Santiago. Sin embargo, luego de una reunión con representantes de la organización, se informa que no es posible entregar acceso a estos registros debido a restricciones de confidencialidad y a la naturaleza de las situaciones que pueden ser captadas por sus sistemas de videovigilancia. Frente a esta limitación, se realizan grabaciones propias

en sus instalaciones, las cuales permiten evaluar el comportamiento del prototipo frente a diferentes condiciones de presencia y distribución de personas.

- **Evaluación mediante grabaciones del caso de estudio:** El uso de grabaciones propias permite incorporar al sistema imágenes obtenidas directamente en el entorno considerado como caso de estudio. Sin embargo, esta alternativa no permite evaluar la integración del sistema con la infraestructura ni observar su funcionamiento continuo bajo las condiciones propias de una operación real. Por lo tanto, las pruebas realizadas permiten evaluar el comportamiento del sistema, pero no corresponden a un despliegue operacional dentro de la red de Metro.
- **Recursos computacionales durante el procesamiento:** El procesamiento simultáneo de distintas fuentes de video y la actualización de la información asociada requieren recursos computacionales que pueden influir en los tiempos de carga y procesamiento observados durante la ejecución. Las pruebas finales del sistema se realizan utilizando CPU, por lo que los resultados obtenidos bajo este entorno no permiten caracterizar el rendimiento que podría alcanzarse mediante una infraestructura con GPU ni evaluar el comportamiento del sistema frente a una cantidad mayor de cámaras funcionando de manera simultánea.

6.3. Trabajo Futuro

A partir de las funcionalidades implementadas y de las limitaciones identificadas durante el desarrollo, se establecen distintas líneas de trabajo futuro orientadas a ampliar el alcance del sistema y evaluar su comportamiento en condiciones de operación de mayor escala, las cuales se presentan a continuación:

- **Integración con infraestructura real:** Evaluar el sistema mediante su conexión directa con cámaras pertenecientes a un entorno operacional corresponde a una de las principales líneas de continuidad. Esto permite analizar su funcionamiento a partir de transmisiones provenientes directamente de la infraestructura utilizada y complementar las pruebas realizadas mediante las grabaciones del caso de estudio.
- **Evaluación con una mayor cantidad de cámaras:** Durante las pruebas realizadas se ejecutan simultáneamente cuatro fuentes de video. Como trabajo futuro, se propone ampliar esta cantidad con el propósito de evaluar el comportamiento del sistema ante una mayor carga de procesamiento y determinar los recursos computacionales necesarios para escalar progresivamente su funcionamiento.
- **Fortalecimiento del modelo de detección:** Los resultados obtenidos muestran diferencias entre el conteo generado por el sistema y la referencia manual en los escenarios evaluados. Como trabajo futuro,

se propone ampliar el conjunto de datos con imágenes que representen distintas condiciones de densidad, iluminación, movimiento, perspectiva y oclusión, con el propósito de fortalecer el entrenamiento del modelo y posteriormente evaluar su comportamiento mediante una mayor cantidad de escenarios de prueba.

- **Ampliación del análisis de la información:** Los registros históricos generados por el sistema pueden ser utilizados para incorporar nuevas formas de análisis sobre la evolución de la afluencia de personas. Como continuidad del trabajo desarrollado, estos datos podrían emplearse para identificar patrones temporales y estudiar la incorporación de mecanismos predictivos orientados a anticipar condiciones de alta ocupación.

6.4. Reflexión Final

En conclusión, el desarrollo de esta investigación permite evidenciar las posibilidades que ofrecen la inteligencia artificial y la visión por computador para abordar problemáticas asociadas a la gestión de espacios públicos con alta afluencia de personas. La transformación de las imágenes capturadas por cámaras en información disponible para consultas y análisis permite avanzar desde una supervisión basada principalmente en la observación hacia herramientas que entregan apoyo adicional para comprender las condiciones presentes en estos espacios.

El sistema desarrollado forma parte de una aproximación a esta problemática mediante una solución que integra análisis automatizado de las fuentes de video con herramientas orientadas al análisis de afluencia. Sin embargo, existen aspectos que requieren ser evaluados en condiciones operacionales reales, aunque el trabajo realizado permite establecer una base sobre la cual continuar estudiando la aplicación de estas tecnologías en contextos donde la concentración y desplazamiento de personas requieren de supervisión constante.

Finalmente, esta investigación permite reconocer que la incorporación de tecnologías inteligentes en la gestión de espacios públicos no depende solo de la capacidad de detección de personas, sino también de la posibilidad de transformar detecciones en información útil para la toma de decisiones. En ese sentido, el trabajo desarrollado representa un punto de partida para continuar avanzando hacia herramientas que contribuyen a una gestión más informada de la afluencia de personas y que, además, se adapten a las necesidades de distintos entornos de alta demanda.

Bibliografía

[1] Soobeen Park y Shaoqing Zhang. "A pilot study of circulation layout based on perceived retail crowding". En: *Journal of Retailing and Consumer Services* 49 (2019), págs. 305-315. DOI: 10.1016/j.jretconser.2019.04.008.

[2] Xingya Yan y Yujiao Ding. "Stadium crowd counting method based on peer to peer network". En: *Xi'an University of Posts and Telecommunications (XUPT)* (2023). Supported by Shaanxi Science and Technology Program (2023-YBGY-156).

[3] Andrea Schneider, Eva Krueger, Beat Vollenwyder, Jasmin Thurau y Achim Elfering. "Understanding the relations between crowd density, safety perception and risk-taking behavior on train station platforms: A case study from Switzerland". En: *Transportation Research Interdisciplinary Perspectives* 10 (2021), pág. 100390. DOI: 10.1016/j.trip.2021.100390.

[4] Tanvi Daware y Tripti Dhote. "Enhancing Public Safety Through Real-Time Crowd Density Analysis and Management". En: *2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA)*. 2023, págs. 1040-1046. DOI: 10.1109/ICIRCA57980.2023.10220731.

[5] Ricardo Giesen, Darío Hidalgo y Raúl Ramos. "Public transport: Status and research priorities for Latin America and the Caribbean". En: *Latin American Transport Studies* 3 (2025), pág. 100025. DOI: 10.1016/j.latran.2025.100025.

[6] Paul Basnak, Ricardo Giesen y Juan Carlos Muñoz. "Do Chilean cities have the right set of public transport modes?" En: *Research in Transportation Business & Management* 57 (2024), pág. 101244. DOI: 10.1016/j.rtbm.2024.101244.

[7] Vicente Valdivieso, Felipe González, Louis de Grange y Rodrigo Troncoso. "Infraestructura dedicada para buses y calidad del servicio: el caso de Santiago, Chile". En: *Documento de investigación* (2016).

[8] Verónica Vistoso Martínez. "Efectos urbanos de los terminales de transporte interurbano sobre el espacio público barrial. El caso de la comuna de Estación Central". Tesis para optar al grado de Magíster en Urbanismo. Santiago, Chile: Facultad de Arquitectura y Urbanismo, Universidad de Chile, mar. de 2020.

[9] Cristina Oprea, Mircea Rosca, Eugen Rosca, Ilona Costea, Anamaria Ilie, Oana Dinu y Aura Ruscă. "Analyzing Passenger Flows in an Airport Terminal: A Discrete Simulation Model". En: *Computation* 12.11 (2024). DOI: 10.3390/computation12110223. URL: https://doi.org/10.3390/computation12110223.

[10] Marcos Medina-Tapia, Francesc Robusté y Miquel Estrada. "Spatial analysis of public transportation infrastructure in Santiago, Chile, using the continuous approximation method". En: *Transportation Research Procedia* 58 (2021), págs. 309-316. DOI: 10.1016/j.trpro.2021.11.042.

[11] Jin Su Kim, Min-Gu Kim y Sung Bum Pan. "A study on implementation of real-time intelligent video surveillance system based on embedded module". En: *EURASIP Journal on Image and Video Processing* 2021.35 (2021), págs. 1-22. DOI: 10.1186/s13640-021-00576-0.

[12] Andrew J. Park, Ryan Ficocelli, Lee D. Patterson, Valerie Spicer, Frank Dodich y Herbert H. Tsang. "Modelling Crowd Dynamics and Crowd Management Strategies". En: *2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*. 2021, págs. 0627-0632. DOI: 10.1109/IEMCON53756.2021.9623214.

[13] Donatella Darsena, Giacinto Gelli, Ivan Iudice y Francesco Verde. "Sensing Technologies for Crowd Management, Adaptation, and Information Dissemination in Public Transportation Systems: A Review". En: *IEEE Sensors Journal* (2022). Manuscript received December 1, 2021; accepted November 14, 2022.

[14] Ander García, Bruno V. Guerrero, Dariel Hernández-Delfin, Jae-Jin Lee y Marco Ellero. "Modeling crowd flow during stadium ingress under alternative access protocols: San Mamés as a case study". En: *Simulation Modelling Practice and Theory* 149 (2026), pág. 103279. DOI: 10.1016/j.simpat.2026.103279.

[15] Cotransa Group. *Cuellos de botella: ¿qué son y cuál es su origen?* https://cotransagroup.com/cuellos-de-botella-que-son-y-cual-es-su-origen/. Consultado el 25 de agosto de 2026.

[16] Anna Larsson, Elin Ranudd, Enrico Ronchi, Aoife Hunt y Steve Gwynne. "The impact of crowd composition on egress performance". En: *Fire Safety Journal* 120 (2021), pág. 103040. DOI: 10.1016/j.firesaf.2020.103040.

[17] Thomas A. Morton y Séamus A. Power. "Coming together after standing apart: What predicts felt safety in the post-coronavirus crowd?" En: *Social Science & Medicine* 293 (2022), pág. 114649. DOI: 10.1016/j.socscimed.2021.114649.

[18] Movistar Arena. *Sobre Nosotros*. https://www.movistararena.cl/movistar/sobre-nosotros. Consultado el 11 de agosto de 2026. 2026.

[19] Juliane Adrian, Armin Seyfried y Anna Sieben. "Crowds in front of bottlenecks at entrances from the perspective of physics and social psychology". En: *arXiv preprint arXiv:2007.06396* (2020). URL: https://arxiv.org/abs/2007.06396.

[20] Keith Still, Marina Papalexi, Yiyi Fan y David Bamford. "Place crowd safety, crowd science? Case studies and application". En: *Journal of Place Management and Development* 13.4 (2020), págs. 385-407. ISSN: 1753-8335. DOI: 10.1108/JPMD-10-2019-0090.

[21] HSB Noticias. *¡Para la historia! Shakira superó con su show gratis en Copacabana a Madonna y Lady Gaga*. https://www.hsbnoticias.com/para-la-historia-shakira-supero-con-su-show-gratis-en-copacabana-a-madonna-y-lady-gaga/. Consultado el 11 de agosto de 2026. 2026.

[22] Central Web. *Shakira hizo historia junto a Corona ante millones de personas en Copacabana*. Mayo de 2026. URL: https://www.centralweb.cl/shakira-hizo-historia-junto-a-corona-ante-millones-de-personas-en-copacabana/ (visitado 26-08-2026).

[23] Milad Haghani, Matt Coughlan, Ben Crabb, Anton Dierickx, Claudio Feliciani, Roderick van Gelder, Paul Geoerg, Nazli Hocaoglu, Steve Laws, Ruggiero Lovreglio, Zoe Miles, Alexandre Nicolas, William J. O'Toole, Syan Schaap, Travis Semmens, Zahra Shahhoseini, Ramon Spaaij, Andrew Tatrai, John Webster y Alan Wilson. "A roadmap for the future of crowd safety research and practice: Introducing the Swiss Cheese Model of Crowd Safety and the imperative of a Vision Zero target". En: *Safety Science* 168 (2023), pág. 106292. DOI: 10.1016/j.ssci.2023.106292.

[24] Guillermo Rivera-Aguilera, Miguel Imas y Luis Jiménez-Díaz. "Jóvenes, multitud y estallido social en Chile". En: *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud* 19.2 (2021), págs. 1-23. DOI: 10.11600/rlcsnj.19.2.4543. URL: https://dx.doi.org/10.11600/rlcsnj.19.2.4543.

[25] Galería CIMA. *La Marcha Mas Grande De Chile — Timelapse Oficial*. Video de YouTube. URL: https://www.youtube.com/watch?v=Vl8-rtU_J-E (visitado 12-08-2026).

[26] El Mostrador. *Movimiento feminista sigue haciendo historia: dos millones de mujeres marcharon en Santiago y regiones en el 8M*. Mar. de 2020. URL: https://www.elmostrador.cl/destacado/2020/03/08/movimiento-feminista-sigue-haciendo-historia-dos-millones-de-mujeres-marcharon-en-santiago-y-regiones-en-el-8m/ (visitado 12-08-2026).

[27] Julien Rosenberger, Julien Saunier y Nicolas Sabouret. "Multi-agent Simulation of Violence Emergence in Protests". En: *3rd International Conference on Human and Artificial Rationalities (HAR 2024)*. HAL Id: hal-04842616. Paris, France, sep. de 2024, págs. 35-51. DOI: 10.1007/978-3-031-84595-6_3.

[28] Nancy Rodríguez y David White. "An analysis of protesting activity and trauma through mathematical and statistical models". En: *Crime Science* 12.17 (2023). DOI: 10.1186/s40163-023-00197-0. URL: https://doi.org/10.1186/s40163-023-00197-0.

[29] Sultan Alodhaibi, Robert L. Burdett y Prasad K. D. V. Yarlagadda. "Impact of Passenger-Arrival Patterns in Outbound Processes of Airports". En: *Procedia Manufacturing* 30 (2019), págs. 323-330. DOI: 10.1016/j.promfg.2019.02.046.

[30] Ministerio de Obras Públicas de Chile. *Subsecretario MOP informa que temporada de verano 2026 superó los 5 millones de pasajeros en el Aeropuerto Santiago*. https://www.mop.gob.cl/subsecretario-mop-informa-que-temporada-de-verano-2026-supero-los-5-millones-de-pasajeros-en-el-aeropuerto-santiago/. Consultado el 13 de agosto de 2026. 2026.

[31] Ministerio de Obras Públicas de Chile. *MOP, RREE y Aeropuerto proyectan flujo de más 1,1 millones de pasajeros para vuelos nacionales e internacionales entre Navidad y Año Nuevo*. https://www.mop.gob.cl/mop-rree-y-aeropuerto-proyectan-flujo-de-mas-11-millones-de-pasajeros-para-vuelos-nacionales-e-internacionales-entre-navidad-y-ano-nuevo/. Consultado el 13 de agosto de 2026. 2025.

[32] Policía de Investigaciones de Chile. *Migraciones y Policía Internacional*. Consultado el 13 de agosto de 2026. 2026. URL: https://pdichile.cl/institucion/unidades/migraciones.

[33] Paulina Ortega. *2,5 millones de pasajeros: tráfico aéreo del aeropuerto de Santiago parte el año con niveles superiores a los prepandemia*. La Tercera. Consultado el 16 de agosto de 2026. 2024. URL: https://www.latercera.com/pulso-pm/noticia/25-millones-de-pasajeros-trafico-aereo-del-aeropuerto-de-santiago-parte-el-ano-con-niveles-superiores-a-los-prepandemia/QFJMJE4MQZHFJKUXAPISTADI7Q/.

[34] El Aéreo. *Aeropuerto Santiago de Chile anuncia que 10 rutas tendrán más frecuencias de vuelos en 2026*. Consultado el 16 de agosto de 2026. 2026. URL: https://elaereo.com/2026/04/aeropuerto-santiago-de-chile-anuncia-que-10-rutas-tendran-mas-frecuencias-de-vuelos-en-2026/.

[35] Lijun Sun, Alejandro Tirachini, Kay W. Axhausen, Alexander Erath y Der-Horng Lee. "Models of bus boarding and alighting dynamics". En: *Transportation Research Part A: Policy and Practice* 69 (2014), págs. 447-460. ISSN: 0965-8564. DOI: 10.1016/j.tra.2014.09.007.

[36] Diario Concepción. *Soluciones para 2030: estudiantes aseguran vivir inseguridad por aglomeración en paradero de Paicaví*. 2 de jun. de 2026. URL: https://www.diarioconcepcion.cl/ciudad/2026/06/02/soluciones-para-2030-estudiantes-aseguran-vivir-inseguridad-por-aglomeracion-en-paradero-de-paicavi.html (visitado 01-09-2026).

[37] Yunchao Qu, Yao Xiao, Hao Liu, Haodong Yin, Jianjun Wu, Qiushi Qu, Daqing Li y Tao Tang. "Analyzing crowd dynamic characteristics of boarding and alighting process in urban metro stations". En: *Physica A: Statistical Mechanics and its Applications* 526 (2019), pág. 121075. ISSN: 0378-4371. DOI: 10.1016/j.physa.2019.121075.

[38] La Tercera. *El diseño para multitudes que tendrá la estación Estadio Nacional del Metro*. 2017. URL: https://www.latercera.com/diario-impreso/el-diseno-para-multitudes-que-tendra-la-estacion-estadio-nacional-del-metro/ (visitado 01-09-2026).

[39] Metro de Santiago. *Extensión horaria por partido Universidad de Chile vs. Independiente*. Instagram. 2025. URL: https://www.instagram.com/p/DNQ4D5jx4x8/ (visitado 01-09-2026).

[40] Metro de Santiago. *Extensión horaria por concierto de AC/DC*. Instagram. 2026. URL: https://www.instagram.com/p/DVwYjBYFB0f/ (visitado 01-09-2026).

[41] Shuyan Han, Yu Ye, Yifan Song, Shurui Yan, Feng Shi, Yuhao Zhang, Xiaodong Liu, Hu Du y Dexuan Song. "A Systematic Review of Objective Factors Influencing Behavior in Public Open Spaces". En: *Frontiers in Public Health* 10 (2022), pág. 898136. DOI: 10.3389/fpubh.2022.898136. URL: https://doi.org/10.3389/fpubh.2022.898136.

[42] Diogo Guedes Vidal, Catarina Patoilo Teixeira, Cláudia Oliveira Fernandes, Agnieszka Olszewska-Guizzo, Ricardo Cunha Dias, Helena Vilaça, Nelson Barros y Rui Leandro Maia. "Patterns of human behaviour in public urban green spaces: On the influence of users' profiles, surrounding environment, and space design". En: *Urban Forestry & Urban Greening* 74 (2022), pág. 127668. ISSN: 1618-8667. DOI: 10.1016/j.ufug.2022.127668.

[43] Daniel Parra-Ovalle, Carme Miralles-Guasch y Oriol Marquet. "Pedestrian street behavior mapping using unmanned aerial vehicles. A case study in Santiago de Chile". En: *PLOS ONE* 18.3 (2023), e0282024. DOI: 10.1371/journal.pone.0282024. URL: https://doi.org/10.1371/journal.pone.0282024.

[44] Parque Metropolitano de Santiago. *Parque Metropolitano de Santiago*. Sitio web oficial. Consultado el 17 de agosto de 2026. URL: https://parquemet.cl/.

[45] CicloRecreoVía. *CicloRecreoVía*. Sitio web oficial. Consultado el 16 de agosto de 2026. URL: https://ciclorecreovia.cl/.

[46] Parque Metropolitano de Santiago. *Recorre y descubre Parquemet*. https://mapas.parquemet.cl/doc/MapaParque.pdf. Mapa del Parque Metropolitano de Santiago. Consultado el 25 de agosto de 2026.

[47] Mehdi Moussaıd, Victor R. Schinazi, Mubbasir Kapadia y Tyler Thrash. "Virtual Sensing and Virtual Reality: How New Technologies Can Boost Research on Crowd Dynamics". En: *Frontiers in Robotics and AI* 5 (2018), pág. 82. DOI: 10.3389/frobt.2018.00082. URL: https://doi.org/10.3389/frobt.2018.00082.

[48] Ali M. Al-Shaery, Shroug S. Alshehri, Norah S. Farooqi y Mohamed O. Khozium. "In-Depth Survey to Detect, Monitor and Manage Crowd". En: *IEEE Access* 8 (2020), págs. 209008-209019. DOI: 10.1109/ACCESS.2020.3038334.

[49] Jonas De Bruyne, Jamil Joundi, Jessica Morton, Aleksandra Zheleva, Niels Van Kets, Glenn Van Wallendael, Durk Talsma, Jelle Saldien, Lieven De Marez, Wouter Durnez y Klaas Bombeke. "I spy with my AI: The effects of AI-based visual cueing on human operators' performance and cognitive load in CCTV control rooms". En: *International Journal of Industrial Ergonomics* 95 (2023), pág. 103444. ISSN: 0169-8141. DOI: 10.1016/j.ergon.2023.103444.

[50] Ministerio de Transportes y Telecomunicaciones. *Conmemoramos 30 años de la UOCT con agenda enfocada en innovación tecnológica*. 2025. URL: https://mtt.gob.cl/conmemoramos-30-anos-de-la-uoct-con-agenda-enfocada-en-innovacion-tecnologica/ (visitado 01-09-2026).

[51] Eric L. Piza y Lauren N. Moton. "Proactive monitoring and operator discretion: A systematic social observation of CCTV control room operations". En: *Journal of Criminal Justice* 86 (2023), pág. 102071. DOI: 10.1016/j.jcrimjus.2023.102071.

[52] Amnah Aldayri y Waleed Albattah. "Taxonomy of Anomaly Detection Techniques in Crowd Scenes". En: *Sensors* 22.16 (2022). ISSN: 1424-8220. DOI: 10.3390/s22166080. URL: https://www.mdpi.com/1424-8220/22/16/6080.

[53] Akbar Khan, Jawad Ali Shah, Kushsairy Kadir, Waleed Albattah y Faizullah Khan. "Crowd Monitoring and Localization Using Deep Convolutional Neural Network: A Review". En: *Applied Sciences* 10.14 (2020), pág. 4781. DOI: 10.3390/app10144781.

[54] Ali M. Al-Shaery, Hamad Aljassmi, Soha G. Ahmed, Norah S. Farooqi, Abdullah N. Al-Hawsawi, Mohammed Moussa, Abdessamad Tridane y Md. Didarul Alam. "Real-Time Pilgrims Management Using Wearable Physiological Sensors, Mobile Technology and Artificial Intelligence". En: *IEEE Access* (2022). DOI: 10.1109/ACCESS.2022.3221771.

[55] Zizhu Fan, Hong Zhang, Zheng Zhang, Guangming Lu, Yudong Zhang y Yaowei Wang. "A survey of crowd counting and density estimation based on convolutional neural network". En: *Neurocomputing* 472 (2022), págs. 224-251. ISSN: 0925-2312. DOI: 10.1016/j.neucom.2021.02.103.

[56] Aisha M. Alasmari, Norah S. Farooqi y Youseef A. Alotaibi. "Recent trends in crowd management using deep learning techniques: a systematic literature review". En: *Journal of Umm Al-Qura University for Engineering and Architecture* 15 (jun. de 2024), págs. 355-383. DOI: 10.1007/s43995-024-00071-3.

[57] Jianyong Wang, Mingliang Gao, Qilei Li, Hyunbum Kim y Gwanggil Jeon. "A Survey on Supervised, Unsupervised, and Semi-Supervised Approaches in Crowd Counting". En: *Computers, Materials & Continua* 81.3 (2024), págs. 3561-3576. DOI: 10.32604/cmc.2024.058637.

[58] Anxious_City_7864. *Data without context is noise! (With Zoom)*. Publicación en Reddit, r/datascience. 2022. URL: https://www.reddit.com/r/datascience/comments/tq93vt/data_without_context_is_noise_with_zoom/ (visitado 26-08-2026).

[59] Mostafa Ahmed Ezzat, Mohamed A. Abd El Ghany, Sultan Almotairi y Mohammed A.-M. Salem. "Horizontal Review on Video Surveillance for Smart Cities: Edge Devices, Applications, Datasets, and Future Trends". En: *Sensors* 21.9 (2021). DOI: 10.3390/s21093222. URL: https://doi.org/10.3390/s21093222.

[60] Omar Elharrouss, Noor Almaadeed y Somaya Al-Maadeed. "A review of video surveillance systems". En: *Journal of Visual Communication and Image Representation* 77 (2021), pág. 103116. ISSN: 1047-3203. DOI: 10.1016/j.jvcir.2021.103116.

[61] Arun A. Ravindran. "Internet-of-Things Edge Computing Systems for Streaming Video Analytics: Trails Behind and the Paths Ahead". En: *IoT* 4.4 (2023), págs. 486-513. DOI: 10.3390/iot4040021.

[62] Sergio A. Velastin, Rodrigo Fernández, Jorge E. Espinosa y Alessandro Bay. "Detecting, Tracking and Counting People Getting On/Off a Metropolitan Train Using a Standard Video Camera". En: *Sensors* 20.21 (2020), pág. 6251. DOI: 10.3390/s20216251.

[63] Sam Leroux, Bo Li y Pieter Simoens. "Automated training of location-specific edge models for traffic counting". En: *Computers and Electrical Engineering* 99 (2022), pág. 107763. DOI: 10.1016/j.compeleceng.2022.107763.

[64] Marek Kulbacki, Jakub Segen, Zenon Chaczko, Jerzy W. Rozenblit, Michał Kulbacki, Ryszard Klempous y Konrad Wojciechowski. "Intelligent Video Analytics for Human Action Recognition: The State of Knowledge". En: *Sensors* 23.9 (2023), pág. 4258. DOI: 10.3390/s23094258.

[65] Logitech. *C922 Pro HD Stream Webcam*. URL: https://www.logitech.com/es-es/shop/p/c922-pro-stream-webcam (visitado 07-09-2026).

[66] Sarah Williams y Minwook Kang. "Striking a Pose: DIY Computer Vision Sensor Kit to Measure Public Life Using Pose Estimation Enhanced Action Recognition Model". En: *Smart Cities* 8.6 (2025), pág. 183. DOI: 10.3390/smartcities8060183.

[67] GoPro. *HERO13 Black*. 2024. URL: https://gopro.com/en/us/shop/cameras/learn/hero13black/CHDHX-131-master.html (visitado 22-09-2026).

[68] Chinthaka Premachandra, Shohei Ueda y Yuya Suzuki. "Detection and Tracking of Moving Objects at Road Intersections Using a 360-Degree Camera for Driver Assistance and Automated Driving". En: *IEEE Access* 8 (2020), págs. 135652-135660. DOI: 10.1109/ACCESS.2020.3011430.

[69] Insta360. *Insta360 X5*. 2025. URL: https://www.insta360.com/es/product/insta360-x5 (visitado 22-09-2026).

[70] Axis Communications. *AXIS Q1805-LE Bullet Camera*. URL: https://www.axis.com/es-es/products/axis-q1805-le (visitado 07-09-2026).

[71] Notebook Store. *Logitech HD Pro Webcam C922*. 2026. URL: https://notebookstore.cl/logitech-hd-pro-webcam-c922-camara-web-color-720p-1080p-h-264 (visitado 22-09-2026).

[72] Notebook Store. *Cámara de acción GoPro HERO13 Black*. 2026. URL: https://notebookstore.cl/gopro-hero-13-black (visitado 22-09-2026).

[73] Falabella. *Cámara de acción Insta360 X5 Paquete Estándar 360° 8K*. 2026. URL: https://www.falabella.com/falabella-cl/product/143170997/Camara-de-accion-Insta360-X5-Paquete-Estandar-360%C2%B0-8K/143170998 (visitado 22-09-2026).

[74] WiAutomation. *AXIS Q1805-LE*. 2026. URL: https://cl.wiautomation.com/axis/productos-generales/otre/q1805le (visitado 22-09-2026).

[75] Turki M. Alanazi. "Embedded System Based Raspberry Pi 4 for Text Detection and Recognition". En: *Intelligent Automation & Soft Computing* 36.3 (2023), págs. 3343-3354. DOI: 10.32604/iasc.2023.036411.

[76] Daniele Berardini, Lucia Migliorelli, Alessandro Galdelli, Emanuele Frontoni, Adriano Mancini y Sara Moccia. "A deep-learning framework running on edge devices for handgun and knife detection from indoor video-surveillance cameras". En: *Multimedia Tools and Applications* 83 (2024), págs. 19109-19127. DOI: 10.1007/s11042-023-16231-x.

[77] Shengcheng Ma, Xin Chen, Zhuo Li y Yingjie Yang. "A Retrieval Optimized Surveillance Video Storage System for Campus Application Scenarios". En: *Journal of Electrical and Computer Engineering* 2018 (2018), pág. 3839104. DOI: 10.1155/2018/3839104.

[78] Muhammad H. El-Saba. "Combining Supervisory Control and Data Acquisition (SCADA) with Artificial Intelligence (AI) as a Video Management System". En: *Intelligent Video Surveillance - New Perspectives*. IntechOpen, 2022. DOI: 10.5772/intechopen.104766.

[79] B. Ganga, B. T. Lata y K. R. Venugopal. "Object detection and crowd analysis using deep learning techniques: Comprehensive review and future directions". En: *Neurocomputing* 597 (2024), pág. 127932. ISSN: 0925-2312. DOI: 10.1016/j.neucom.2024.127932.

[80] Rongyong Zhao, Feng Hua, Bingyu Wei, Cuiling Li, Yulong Ma, Eric S. W. Wong y Fengnian Liu. "A Review of Abnormal Crowd Behavior Recognition Technology Based on Computer Vision". En: *Applied Sciences* 14.21 (2024), pág. 9758. DOI: 10.3390/app14219758. URL: https://doi.org/10.3390/app14219758.

[81] Divya Rajawat, Arihant Srivastava, Bhanu Prakash Lohani, Prabhat Yadav, Ajay Rana y Shubhi Gupta. "Object Detection in Images and Videos Using OpenCV: A Comparative Study of Deep Learning and Traditional Computer Vision Techniques". En: *2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*. IEEE, 2023.

[82] Stéfan van der Walt, Johannes L. Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D. Warner, Neil Yager, Emmanuelle Gouillart, Tony Yu y scikit-image contributors. "scikit-image: image processing in Python". En: *PeerJ* 2 (2014), e453. DOI: 10.7717/peerj.453.

[83] Davis E. King. "Dlib-ml: A Machine Learning Toolkit". En: *Journal of Machine Learning Research* 10 (2009), págs. 1755-1758.

[84] Kuang Shen, Rusyaizila Ramli, Joong Huang Chuah y Grace Mei Ting Chai. "Driver Fatigue Detection Using OpenCV and Dlib Library". En: *2023 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*. 2023, págs. 1-6. DOI: 10.1109/CSDE59766.2023.10487775.

[85] Wenhan Luo, Junliang Xing, Anton Milan, Xiaoqin Zhang, Wei Liu y Tae-Kyun Kim. "Multiple object tracking: A literature review". En: *Artificial Intelligence* 293 (2021), pág. 103448. DOI: 10.1016/j.artint.2020.103448.

[86] Ujwalla Gawande, Kamal Hajari y Yogesh Golhar. "Pedestrian Detection and Tracking in Video Surveillance System: Issues, Comprehensive Review, and Challenges". En: *Video Surveillance Systems and Applications*. IntechOpen, 2020. DOI: 10.5772/intechopen.90810. URL: https://doi.org/10.5772/intechopen.90810.

[87] Chen Ning, Li Menglu, Yuan Hao, Su Xueping y Li Yunhong. "Survey of pedestrian detection with occlusion". En: *Complex & Intelligent Systems* 7 (2021), págs. 577-587. DOI: 10.1007/s40747-020-00206-8.

[88] WTT. *Overlay Object Detection with Insufficient Data*. 2023. URL: https://wttech.blog/blog/2023/overlay-object-detection-with-insufficient-data/ (visitado 24-09-2026).

[89] Picsellia. *Ensuring the Success of a Computer Vision Project*. URL: https://www.picsellia.com/post/ensuring-the-success-of-a-computer-vision-project (visitado 24-09-2026).

[90] Muhammad Asif Khan, Hamid Menouar y Ridha Hamila. "Visual crowd analysis: Open research problems". En: *AI Magazine* 44.3 (2023), págs. 296-311. DOI: 10.1002/aaai.12117. URL: https://doi.org/10.1002/aaai.12117.

[91] Vishwanath A. Sindagi y Vishal M. Patel. "A survey of recent advances in CNN-based single image crowd counting and density estimation". En: *Pattern Recognition Letters* 107 (2018), págs. 3-16. ISSN: 0167-8655. DOI: 10.1016/j.patrec.2017.07.007.

[92] Francisco Luque Sánchez, Isabelle Hupont, Siham Tabik y Francisco Herrera. "Revisiting crowd behaviour analysis through deep learning: Taxonomy, anomaly detection, crowd emotions, datasets, opportunities and prospects". En: *Information Fusion* 64 (2020), págs. 318-335. ISSN: 1566-2535. DOI: 10.1016/j.inffus.2020.07.008.

[93] S. Peckover, A. Raineri y A. T. Scanlan. "Implementation of Congestion-Related Controls Improves Runner Density, Flow Rate, Perceived Safety, and Satisfaction during an Australian Running Event". En: *Sports* 10.9 (2022), pág. 132. DOI: 10.3390/sports10090132.

[94] Z. Huang, Y. Xu, Y. Sun y X. Zeng. "Dynamic Assessment of Pedestrian Level of Service to Support Spatial Retrofit and Operational Interventions in an Underground Transport Hub: A Case Study of Chongqing Shapingba Station". En: *Buildings* 16.16 (2026), pág. 3140. DOI: 10.3390/buildings16163140.

[95] María José Viñals, Patricio R. Orozco Carpio, Penélope Teruel y José M. Gandía-Romero. "Real-Time Monitoring of Visitor Carrying Capacity in Crowded Historic Streets Through Digital Technologies". En: *Urban Science* 8.4 (2024), pág. 190. DOI: 10.3390/urbansci8040190. URL: https://doi.org/10.3390/urbansci8040190.

[96] Biswanath Basak y Sumana Gupta. "Developing an agent-based model for pilgrim evacuation using visual intelligence: A case study of Ratha Yatra at Puri". En: *Computers, Environment and Urban Systems* 64 (2017), págs. 118-131. ISSN: 0198-9715. DOI: 10.1016/j.compenvurbsys.2017.01.006.

[97] Can Zhao, Yiyang Jiang y Jinfeng Wang. "A Computer Vision-Based Pedestrian Flow Management System for Footbridges and Its Applications". En: *Infrastructures* 10.9 (2025), pág. 247. DOI: 10.3390/infrastructures10090247.

[98] Farkhod Akhmedov, Doston Khasanov, Sarvarbek Sodikovich Yusupov, Oybek Usmankulovich Mallaev, Halimjon Ergashevich Khujamatov, Toshtemir Abdikhafizovich Khujakulov y Young Im Cho. "AH-YOLOFlow: Vision-Based Pedestrian Flow Analytics for Smart Urban Crosswalk Performance Monitoring". En: *Sustainability* 18 (2026), pág. 9040.

[99] Ohoud Alzamzami, Zainab Alsaggaf, Reema AlMalki, Rawan Alghamdi, Amal Babour y Lama Al Khuzayem. "Passable: An Intelligent Traffic Light System with Integrated Incident Detection and Vehicle Alerting". En: *Sensors* 25.18 (2025), pág. 5760.

[100] Vivian W. H. Wong y Kincho H. Law. "Fusion of CCTV Video and Spatial Information for Automated Crowd Congestion Monitoring in Public Urban Spaces". En: *Algorithms* 16.3 (2023), pág. 154. DOI: 10.3390/a16030154. URL: https://doi.org/10.3390/a16030154.

[101] Thalerngsak Wiangwiset, Chayada Surawanitkun, Wullapa Wongsinlatam, Tawun Remsungnen, Apirat Siritaratiwat, Chavis Srichan, Prachya Thepparat, Weerasak Bunsuk, Aekkaphan Kaewchan y Ariya Namvong. "Design and Implementation of a Real-Time Crowd Monitoring System Based on Public Wi-Fi Infrastructure: A Case Study on the Sri Chiang Mai Smart City". En: *Smart Cities* 6.2 (2023), págs. 987-1008. DOI: 10.3390/smartcities6020048. URL: https://doi.org/10.3390/smartcities6020048.

[102] StreetFlow. *StreetFlow: Video Analítica con Inteligencia Artificial*. 2026. URL: https://streetflow.cl/ (visitado 28-09-2026).

[103] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant et al. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". En: *Nature Methods* 17 (2020), págs. 261-272. DOI: 10.1038/s41592-019-0686-2.

[104] Bjarne Stroustrup. "Thriving in a Crowded and Changing World: C++ 2006–2020". En: *Proceedings of the ACM on Programming Languages* 4.HOPL (2020). DOI: 10.1145/3386320.

[105] Allen Wirfs-Brock y Brendan Eich. "JavaScript: The First 20 Years". En: *Proceedings of the ACM on Programming Languages* 4.HOPL (2020), págs. 1-189. DOI: 10.1145/3386327.

[106] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt et al. "Array programming with NumPy". En: *Nature* 585 (2020), págs. 357-362. DOI: 10.1038/s41586-020-2649-2.

[107] Wes McKinney. "Data Structures for Statistical Computing in Python". En: *Proceedings of the 9th Python in Science Conference*. 2010, págs. 56-61. DOI: 10.25080/Majora-92bf1922-00a.

[108] Xintian Wu, Pengfei Qu, Shaofei Wang, Lin Xie y Jie Dong. "Extend the FFmpeg Framework to Analyze Media Content". En: *arXiv preprint arXiv:2103.03539* (2021). arXiv: 2103.03539.

[109] Edip Demirbilek y Jean-Charles Grégoire. "Multimedia Communication Quality Assessment Testbeds". En: *arXiv preprint arXiv:1609.06612* (2016). arXiv: 1609.06612 [cs.MM].

[110] Zonghui Li, Yongsheng Dong, Longchao Shen, Yafeng Liu, Yuanhua Pei, Haotian Yang, Lintao Zheng y Jinwen Ma. "Development and challenges of object detection: A survey". En: *Neurocomputing* 598 (2024), pág. 128102. ISSN: 0925-2312. DOI: 10.1016/j.neucom.2024.128102.

[111] Benymol Jose y Sajimon Abraham. "Performance analysis of NoSQL and relational databases with MongoDB and MySQL". En: *Materials Today: Proceedings* 24 (2020), págs. 2036-2043. ISSN: 2214-7853. DOI: 10.1016/j.matpr.2020.03.633.

[112] Bartłomiej Bednarz y Marek Miłosz. "Benchmarking the performance of Python web frameworks". En: *Journal of Computer Sciences Institute* 36 (2025), págs. 336-341.

[113] Bartosz Miłosierny y Mariusz Dzieńkowski. "The comparative analysis of web applications frameworks in the Node.js ecosystem". En: *Journal of Computer Sciences Institute* 18 (2021), págs. 42-48.

[114] Damian Bogusz, Paweł Ciszewski y Beata Pańczyk. "Performance analysis of web application client layer development tools using Angular, React and Vue as examples". En: *Journal of Computer Sciences Institute* 32 (2024), págs. 223-230.

[115] Juan Terven, Diana-Margarita Córdova-Esparza y Julio-Alejandro Romero-González. "A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS". En: *Machine Learning and Knowledge Extraction* 5 (2023), págs. 1680-1716. DOI: 10.3390/make5040083. URL: https://doi.org/10.3390/make5040083.

[116] Yao Huang, Wenzhu Tan, Liu Li y Lijuan Wu. "WFRE-YOLOv8s: A New Type of Defect Detector for Steel Surfaces". En: *Coatings* 13.12 (2023), pág. 2011. DOI: 10.3390/coatings13122011. URL: https://doi.org/10.3390/coatings13122011.