



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA



Arquitectura de detección de actividades criminales basada  
en análisis de vídeo en tiempo real

Escuela Técnica Superior de Ingeniería de Telecomunicación  
Departamento de Comunicaciones  
Universitat Politècnica de València

Tesis presentada para la obtención del grado de  
*Doctor en Telecomunicación por la Universitat Politècnica de València*  
Valencia, Septiembre 2020

Autor:  
Julio Ernesto Suárez Páez

Director:  
Dr. Manuel Esteve Domingo



*A mi familia y a mi director de tesis quienes siempre me apoyaron*

# RESUMEN

Esta tesis doctoral propone el desarrollo de una arquitectura para sistema de detección de actividades criminales en vídeo aplicado a sistemas de mando y control para seguridad ciudadana. Este sistema está basado en la técnica de Deep Learning Faster R-CNN y tiene el novedoso enfoque de tratar las acciones criminales como los hurtos callejeros, en donde pueden ser identificados objetos como evidencia en una escena de vídeo. Esta tesis muestra el desarrollo de dicha aplicación, que demuestra ser efectiva, identificando la manera de reducir el coste computacional del análisis de video cuadro a cuadro obteniendo rendimientos congruentes con las tasas de cuadros por segundo generados por cámaras de sistema de vídeo vigilancia ciudadana. También es objeto de estudio una posible implementación en el sistema de seguridad ciudadana de la Policía Nacional de Colombia.

# RESUM

Esta tesi doctoral proposa el desenvolupament d'una arquitectura per a sistema de detecció d'activitats criminals en vídeo aplicat a sistemes de comandament i control per a seguretat ciutadana. Este sistema està basat en la tècnica de Deep Learning Faster R-CNN i té el nou enfocament de tractar les accions criminals com les afanades guies de carrers com a objectes que poden ser identificats en una escena de vídeo. Esta tesi mostra el desenvolupament de la dita aplicació, que demostra ser efectiva, identificant la manera de reduir el cost computacional de l'anàlisi de vídeo quadro a quadro obtenint rendiments congruents amb les taxes de cuadros per segon generats per cambres de sistema de vídeo vigilància ciutadana. També s'estudia una possible implementació en el sistema de seguretat ciutadana de la Policia Nacional de Colòmbia.

# ABSTRACT

This doctoral thesis proposes the development of a system to detect criminal activities in video applied to command and control systems for citizen security. This system is based on the Deep Learning technique called Faster R-CNN and has the novel approach of treating criminal actions like street thefts as objects that can be identified in a video scene. This thesis shows the development of this application and the way to reduce the computational cost of the video analysis frame by frame, obtaining performances congruent with the frame rate generated by citizen video surveillance system cameras. There is also a possible implementation in the citizen security system of the National Police of Colombia is being studied.







# AGRADECIMIENTOS

Agradezco a la Oficina de Telemática de la Policía Nacional de Colombia, a mi director de Tesis el Dr. Manuel Esteve Domingo, Catedrático de la Escuela Técnica Superior de Ingeniería de Telecomunicación de la UPV y al Dr. Jon Ander Gómez catedrático de la Escuela Técnica Superior de Ingeniería Informática de la UPV quienes prestaron su apoyo incondicional al desarrollo de esta tesis doctoral.



# ÍNDICE

RESUMEN.....	4
RESUM.....	5
ABSTRACT.....	6
AGRADECIMIENTOS.....	9
ÍNDICE.....	11
ÍNDICE DE FIGURAS.....	15
ACRÓNIMOS.....	21
1. MOTIVACIÓN Y OBJETIVOS .....	24
1.1    Introducción .....	24
1.1    Motivación .....	25
1.2    Objetivos de la tesis .....	27
1.3    Principales aportes de la tesis.....	28
1.3.1    Artículos científicos .....	28
1.3.2    Propiedad intelectual (COPYRIGHT).....	30
1.3.3    Organización de la memoria .....	31
2. ESTADO DEL ARTE TÉCNICO .....	34
2.1    Introducción .....	34
2.2    Sistemas de Información de mando y control en seguridad ciudadana .....	36
2.3    Sistemas de video vigilancia para seguridad ciudadana .....	41
2.4    Visión artificial.....	44
2.5    Deep Learning .....	45
2.6    Redes Neuronales Convolucionales .....	46

2.7	Arquitecturas de redes neuronales convolucionales .....	50
2.8	R-CNN (Region-Based Convolutional Neural Networks) .....	53
2.9	Fast R-CNN (Fast Region-Based Convolutional Neural Networks) .....	55
2.10	Faster R-CNN (Faster Region-Based Convolutional Neural Networks R-CNN).....	56
2.11	Aplicaciones de Deep Learning en seguridad ciudadana .....	57
<b>3.</b>	<b>ESCENARIO DE UTILIZACIÓN .....</b>	<b>63</b>
3.1	Modelo de desarrollo tecnológico de la Policía Nacional de Colombia .....	66
3.2	Centros de comando y control: Policía Nacional de Colombia .....	69
3.2.1	Planeación y ejecución de operaciones de Policía .....	74
3.3	Comando y control en el MNVCC.....	79
3.4	Sistemas de Video Vigilancia distribuidos en Colombia.....	85
<b>4</b>	<b>ARQUITECTURA DE DETECCIÓN DE ACTIVIDADES CRIMINALES</b>	<b>88</b>
4.1	Arquitectura genérica de detección y clasificación.....	88
4.2	Detector de armas de fuego cortas (revólveres y pistolas) .....	92
4.2.1	Fase 1: Red neuronal convolucional como clasificador de imágenes .....	93
4.2.2	Fase 2: reentrenamiento fino de AlexNet para la detección de armas de fuego cortas	95
4.2.3	Fase 3: Entrenamiento del predictor lineal binario para la detección de armas de fuego cortas.....	97
4.3	Detección de hurtos .....	98
4.3.1	Fase 1: Modelos pre-entrenados como clasificador de imágenes.....	100
4.3.2	Fase 2: Rentrenamiento fino de AlexNet, VGG16 y VGG19 para la detección de hurtos callejeros.....	101
4.3.3	Fase 3: Entrenamiento del clasificador lineal binario para la detección de hurto callejero	103
4.4	Sistema completo de detección de actividad criminales .....	106
4.4.1	Arquitectura detallada del sistema de detector de actividades criminales.....	108
4.4.2	Proceso de entrenamiento del detector de actividades criminales .....	111
4.4.3	Fase 1: Entrenamiento de la RPN usando en Dataset específico.....	112
4.4.4	Fase 2: Entrenamiento de una Fast R-CNN como detector de objetos. ....	113

4.4.5	Fase 3: Reentrenamiento fino de la RPN .....	113
4.4.6	Fase 4: Entrenamiento fino de laFast R-CNN .....	114
4.5	Resultados de la detección de actividades criminales analizar video de tiempo real 115	
4.6	Arquitectura final de detección de actividades criminales basada en análisis de vídeo en tiempo real .....	116
<b>5</b>	<b>RENDIMIETO DEL SISTEMA GENÉRICO DE DETECCIÓN Y CLASIFICACIÓN DE ACCIONES CRIMINALES .....</b>	<b>120</b>
<b>6</b>	<b>PROPUESTA DE APLICACIÓN DE LA ARQUITECTURA.....</b>	<b>126</b>
6.1	Sistemas embebidos.....	126
6.1.1	Particularización de la solución genérica en sistemas embebidos .....	127
6.1.2	Arquitectura de implementación en sistemas embebidos .....	128
6.2	Procesamiento de video centralizado .....	130
6.2.1	Implementación centralizada un centro de procesamiento de datos CPD.....	132
<b>7</b>	<b>INTEGRACIÓN EN EL SISTEMA DE MANDO Y CONTROL DE SEGURIDAD CIUDADANA .....</b>	<b>136</b>
7.1	Estructura del servicio de Policía .....	136
7.2	Estrategias operativas Policía Nacional de Colombia.....	137
7.3	Operaciones de Policía en Colombia.....	139
7.4	Infraestructura de Centros de comando y control Policía Nacional de Colombia ....	141
7.5	Arquitectura genérica del centro de comando y control Policía Nacional de Colombia 143	
7.5.1	SISTEMA DE INFORMACIÓN PARA EL SEGUIMIENTO Y CONTROL DE CASOS (SECAD) 145	
7.5.2	SUBSISTEMA DE ATENCIÓN A LLAMADAS DE EMERGENCIA 123.....	146
7.5.3	SUBSISTEMA DE VIDEO VIGILANCIA CIUDADANA.....	147
7.5.4	SUBSISTEMA DE DESPACHO .....	149
7.5.5	SALAS DE CRISIS / COMANDO .....	150
7.6	AUTOMATIZACIÓN DE ALARMAS EN POSIBLES CASOS POLICIALES DETECTADOS POR VIDEO 151	

7.7	INTEGRACIÓN CON EL SISTEMA DE COMANDO Y CONTROL DE SEGURIDAD CIUDADANA.....	152
8	CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN .....	159
8.1	CONCLUSIONES .....	159
8.2	FUTURAS LÍNEAS DE INVESTIGACIÓN .....	161
	REFERENCIAS .....	163

# ÍNDICE DE FIGURAS

Figura 1: Centro de Comando y Control Seguridad Ciudadana Policía Nacional de Colombia [5].....	37
Figura 2: Arquitectura del sistema de video vigilancia .....	43
Figura 3: Red de neuronas artificiales.....	47
Figura 4: Red Neuronal Convolutacional.....	48
Figura 5: Modelo AlexNet.....	51
Figura 6: Modelo VGG-16.....	52
Figura 7: Modelo VGG-19.....	52
Figura 8: Arquitectura R-CNN.....	53
Figura 9: Resultados de detección de objetos usando R-CNN .....	54
Figura 10: Arquitectura Fast R-CNN .....	55
Figura 11: Arquitectura Faster R-CNN .....	57
Figura 12: Modelo de desarrollo tecnológico (Policía Nacional de Colombia) .....	67
Figura 13: Prototipo de Centro de Comando y Control de Seguridad Ciudadana (Policía Nacional de Colombia).....	68

Figura 14: Centro de Comando y Control de Seguridad Ciudadana Ideal (Policía Nacional de Colombia) .....	70
Figura 15: Línea de mando Policía Nacional de Colombia [87] .....	71
Figura 16: Arquitectura funcional del Centro de Comando y Control de Seguridad Ciudadana (Policía Nacional de Colombia) .....	72
Figura 17: Georreferenciación de áreas de influencias de bandas criminales Localidad de Kennedy Bogotá D.C. [87] .....	76
Figura 18: Georreferenciación de actividades criminales Localidad de Kennedy Bogotá D.C. [87] .....	77
Figura 19: Informe Georreferenciado de actividades criminales Localidad de Kennedy Bogotá D.C. [87] .....	79
Figura 20: Mapa del Modelo Nacional de Vigilancia Comunitaria por Cuadrantes (MNVCC), [89] .....	81
Figura 21: Ejemplo de cuadrantes urbanos reales [87] .....	82
Figura 22: Ejemplo real de cuadrantes rurales, fluvial y viales [87] .....	83
Figura 23: Dimensiones del MNVCC [87] .....	84
Figura 24: Preparación de operaciones policiales rutinarias [87] .....	84
Figura 25: Arquitectura genérica de detección y clasificación de actividades criminales .....	88
Figura 26: Arquitectura genérica de detección de actividades criminales con varias señales de video .....	91
Figura 27: Modelo Pre-entrenado AlexNet .....	94



Figura 28: Entrenamiento fino de la Neuronal Convolutacional AlexNet ...	95
Figura 29: Muestra de imágenes del Dataset de reentrenamiento fino ..	96
Figura 30: Arquitectura del entrenamiento del predictor lineal de armas cortas .....	98
Figura 31: Entrenamiento AlexNet, VGG16 y VGG19.....	100
Figura 32: Muestra de imágenes del Dataset de hurtos para reentrenamiento fino.....	101
Figura 33: Reentrenamiento fino de AlexNet, VGG16 y VGG19 .....	102
Figura 34: Arquitectura del entrenamiento del predictor lineal de hurtos con modelos VGG16 y VGG19 .....	103
Figura 35: Arquitectura del entrenamiento del predictor lineal de hurtos con modelo AlexNet .....	104
Figura 36: Resultados positivos de la detección de hurtos callejeros..	106
Figura 37: Region Proposal Network (RPN).....	109
Figura 38: Arquitectura del detector de actividades criminales .....	110
Figura 39: Arquitectura detallada del detector de actividades criminales en video .....	111
Figura 40: Fase 1: entrenamiento de la RPN .....	112
Figura 41: Fase 1: entrenamiento del detector de objetos Fast R-CNN	113
Figura 42: Fase 3: entrenamiento fino de la RPN .....	114
Figura 43: Fase 4: Entrenamiento fino del detecto de objetos Fast R-CNN .....	115

Figura 44: Funcionamiento del sistema genérico de detección de actividades criminales.....	117
Figura 45: Comparación del coste computacional en función del número de cámaras desplegadas.....	122
Figura 46: Comparación del coste económico en función del número de cámaras desplegadas.....	123
Figura 47: Comparación del Consumo Energético en función del número de cámaras desplegadas.....	124
Figura 48: Arquitectura de detección de actividades criminales en un sistema embebido.....	127
Figura 49: Arquitectura propuesta para la Policía Nacional de Colombia basada en sistemas embebidos. ....	129
Figura 50: Arquitectura de Red Óptica internacional disponible en Colombia [116].....	131
Figura 51: Arquitectura propuesta para la Policía Nacional de Colombia basada en procesamiento centralizado.....	132
Figura 52: Estructura del servicio de Policía en Colombia [87] .....	137
Figura 53: Estrategias Operativas y Estrategia Institucional de Seguridad y Convivencia Ciudadana (EICOS) -Policía Nacional de Colombia [87]	139
Figura 54: Esquema de ejecución en operaciones Policiales - Policía Nacional de Colombia.....	140

Figura 55: Centro de comando y control (Planta Baja) - Policía Nacional de Colombia.....	141
Figura 56: Centro de comando y control (primera planta) - Policía Nacional de Colombia.....	143
Figura 57: Arquitectura genérica del centro de comando y control de seguridad ciudadana - Policía Nacional de Colombia .....	144
Figura 58: Arquitectura genérica del subsistema de video vigilancia....	149
Figura 59: Sala de crisis Dirección General de la Policía Nacional de Colombia - DIPON .....	151
Figura 60: Integración mediante sistemas embebidos .....	155
Figura 61: Integración mediante centro de procesamiento de datos CPD .....	156



# ACRÓNIMOS

<b>AWS</b>	Amazon Web Services
<b>C2IS</b>	Command and control information system
<b>CAD</b>	Centro Automático de Despacho
<b>CCCSC</b>	Centros de Comando y Control Seguridad Ciudadana
<b>CNN</b>	Convolutional Neural Network
<b>CPD</b>	Centro de Procesamiento de Datos
<b>CUDA</b>	Compute Unified Device Architecture
<b>Dataset</b>	Conjunto de datos
<b>DIPON</b>	Dirección General de la Policía Nacional
<b>EICOS</b>	Estrategia Institucional de Seguridad y Convivencia Ciudadana
<b>Fast R-CNN</b>	Fast Region-Based Convolutional Neural Network
<b>Faster R-CNN</b>	Faster Region-Based Convolutional Neural Networks
<b>FC</b>	Fully Connected
<b>GCP</b>	Google Cloud Platform
<b>GIS</b>	Geographic Information System
<b>GPU</b>	Graphics Processing Unit

<b>IDIGER</b>	Instituto Distrital de Gestión y Riesgos
<b>ILSVRC</b>	Large Scale Visual Recognition Challenge
<b>ML</b>	Machine Learni
<b>MNVCC</b>	Modelo nacional de vigilancia comunitaria por cuadrantes
<b>PIPSC-CV</b>	Plan Integral Policial para la Seguridad del Ciudadano “Corazón Verde”
<b>PONAL</b>	Policía Nacional de Colombia
<b>PSC</b>	Política de Seguridad y Convivencia
<b>R-CNN</b>	Region-Based Convolutional Neural Network
<b>ReLU</b>	Rectified Linear Unit
<b>RPN</b>	Region Proposal Network
<b>SECAD</b>	Sistema de Información para el Seguimiento y Control de Casos
<b>SVM</b>	Support Vector Machines
<b>VGG</b>	Visual Geometry Group



# 1.MOTIVACIÓN Y OBJETIVOS

## 1.1 Introducción

La seguridad ciudadana es un tema relevante para gobiernos en todo el mundo, pues la criminalidad en las ciudades es una cuestión que afecta directamente la calidad de vida de sus habitantes. Por esto las agencias de seguridad cuentan con diversos sistemas tecnológicos que los apoyan en sus tareas.

La información proveniente de los diversos sistemas tecnológicos de apoyo a la seguridad ciudadana es manejada por los sistemas de mando y control; C2IS [1] por sus siglas en inglés, el cual entrega la información en tiempo real y está enfocado en mejorar la conciencia situacional (*Situational Awareness*) para realizar una mejor toma de decisiones estratégicas en situaciones críticas que se pueden presentar en las ciudades como hurtos, escenarios de violencia, alteraciones de orden público y demás situaciones similares.

Uno de los sistemas tecnológicos más importantes que apoyan a las agencias de seguridad es el sistema de Video Vigilancia, el cual entrega video se monitoreo y seguridad ciudadana en tiempo real, sin embargo, en ocasiones la cantidad de cámaras instaladas en una ciudad es mayor al



que los operadores de sistema pueden manejar, haciendo que algunas veces algunos eventos captados por las cámaras de vigilancia no sean detectados por los operadores, impidiendo la toma de decisiones estratégicas en tiempo real, como la movilidad de unidades o prestar ayuda al ciudadano.

Teniendo en cuenta dicha problemática, en esta tesis doctoral se ha desarrollado una arquitectura de un sistema detección de eventos criminales para video en tiempo real el cual podría ser aplicado en sistemas de mando y control aplicados a seguridad ciudadana. Esta arquitectura se desarrolló usando técnicas de Deep Learning y fue desarrollado con el apoyo de la Policía Nacional de Colombia.

### **1.1 Motivación**

La arquitectura desarrollada en esta tesis se enfocó en la investigación aplicada, dirigida a solucionar casos de sistemas de mando y control aplicados a la seguridad ciudadana, cómo el presentado por la Policía Nacional de Colombia, la cual necesita un sistema de detección automática en video de eventos criminales. Teniendo en cuenta esto la motivación de esta tesis se centra en los siguientes aspectos:

- **Automatización en la detección de eventos criminales.**

Según la Policía Nacional de Colombia, en las ciudades colombianas hay desplegadas miles de cámaras de seguridad, sin embargo, en los centros de mando y control no hay una cantidad suficiente de Policías para hacer el monitoreo de todas las cámaras a la vez, lo que lleva a que algunas actividades criminales no sean detectadas, aunque si son capturadas por las cámaras de seguridad. Por tanto, es una necesidad de la República de Colombia contar con un sistema de detección que, en parte automatice la detección de eventos criminales y que sea aplicable al sistema de video vigilancia de la Policía Nacional.

- **Seguridad Ciudadana**

Los comandantes de las agencias de seguridad reaccionan según la información que tengan disponible, por tanto, si hay eventos los cuales no son detectados, las agencias de seguridad no podrán reaccionar a ellos, por ende, mejorar la conciencia situacional en los centros de mando y control tendrán una relación directa con la seguridad ciudadana.

- **Especificación de arquitectura aplicable en la Policía Nacional de Colombia**

La República de Colombia cuenta con una Policía Nacional bien preparada para responder a las necesidades de la población, sin embargo, es de gran importancia que el sistema sea aplicable en un caso de uso como el de la Policía Nacional de Colombia para optimizar tiempo y recursos.

## **1.2 Objetivos de la tesis**

En relación con la investigación realizada en esta tesis, se definieron los siguientes objetivos:

- Desarrollar una arquitectura de detección de actividades criminales basada en análisis de vídeo en tiempo real para en un futuro ser implementada en un sistema de mando y control para seguridad ciudadana.
- Identificar técnicas de Deep Learning adecuadas para la detección de eventos criminales en videos de seguridad ciudadana.
- Diseñar un sistema de detección de eventos criminales basado en Deep Learning en video de seguridad ciudadana.

- Diseñar una arquitectura que permita implementar el prototipo de detección de actividades criminales en video para la Policía Nacional de Colombia.
- Analizar una posible implementación del sistema en la Policía Nacional de Colombia.

### **1.3 Principales aportes de la tesis**

#### **1.3.1 Artículos científicos**

**Reduced computational cost prototype for street theft detection based on depth decrement in Convolutional Neural Network. Application to Command and Control Information Systems (C2IS) in the National Police of Colombia**

Revista: International Journal of Computational Intelligence Systems

ISSN: 1875-6883, Vol 12, Pag 123-130

País: Reino Unido

Editorial: Atlatis Press,

DOI: <https://doi.org/10.2991/ijcis.2018.25905186>

Indexación y Clasificación:

Q1 Scimago Journal & Country Rank (SJR)

Q2 Journal Citation Report (JCR)

**A Novel Low Processing Time System for Criminal Activities Detection Applied to Command and Control Citizen Security Centers**

Revista: Information

ISSN: 2078-2489, Vol 10, No. 10 Pag 365-384

País: Suiza

Editorial: MDPI, DOI: <https://doi.org/10.3390/info10120365>

Indexación y Clasificación:

Q3 Scimago Journal & Country Rank (SJR)

Emerging Sources Citation Index - Web of Science (Clarivate Analytics)

**A Novel Method of Spatiotemporal Dynamic Geo-Visualization of Criminal Data, Applied to Command and Control Centers for Public Safety**

Revista: ISPRS International Journal of Geo-Information

ISSN 2220-9964, Vol 10, No. 3 Pag 160-177

País: Suiza

Editorial: MDPI, <https://doi.org/10.3390/ijgi9030160>

Indexación y Clasificación:

Q1 Scimago Journal & Country Rank (SJR)

Q3 Journal Citation Report (JCR)

### 1.3.2 Propiedad intelectual (COPYRIGHT)

#### **DETECTOR DE ARMAS CORTAS DE FUEGO BASADO EN DEEP LEARNING.**

COPYRIGHT: REGISTRO DE SOPORTE LÓGICO - SOFTWARE

República de Colombia

Ministerio de Interior

Dirección Nacional de Derechos de Autor

Unidad Administrativa Especial

Oficina de Registro

Libro 13 Tomo 69 Partida 33

13 de septiembre de 2018

#### **DEEPPOLNET: DEEP LEARNING PROTOTYPE FOR VIDEO CRIME EVENTS DETECTION**

COPYRIGHT: REGISTRO DE SOPORTE LÓGICO - SOFTWARE

República de Colombia

Ministerio de Interior

Dirección Nacional de Derechos de Autor

Unidad Administrativa Especial

Oficina de Registro

Libro 13 Tomo 69 Partida 32

13 de septiembre de 2018

### **1.3.3 Organización de la memoria**

La memoria de la tesis está estructurada de la siguiente manera

#### **Primer capítulo:**

En el primer capítulo se realiza la introducción a la problemática que esta tesis pretende resolver, se muestran las motivaciones para realizar la tesis doctoral, así como los objetivos que se propone cumplir.

#### **Segundo capítulo:**

En este capítulo se realiza un estudio minucioso del estado del arte pertinente para el desarrollo de la tesis y se determinan cuáles son las técnicas más pertinentes para resolver los problemas planteados por esta tesis.

#### **Tercer capítulo:**

Se realiza una descripción del escenario de uso en donde tendrá una futura implementación, haciendo una descripción de la seguridad ciudadana en Colombia, se explica su complejidad y se describen los actores involucrados en el panorama de seguridad de Colombia.

Se explica el modelo de desarrollo tecnológico de la Policía Nacional de Colombia como responsable de la seguridad ciudadana en el territorio colombiano y se describe la implementación de los centros de mando y Control de la Policía Nacional de Colombia, llamados centros de comando y control, así como su modelo operativo y la línea de mando.

**Cuarto capítulo:**

En esta parte del documento se plantea la arquitectura genérica del sistema de detección de actividades criminales en video aplicado en sistemas de mando y control para seguridad ciudadana, se realiza el entrenamiento del sistema y se muestran sus resultados.

**Quinto capítulo:**

En el quinto capítulo se realiza un análisis de rendimiento del sistema planteado, desde el punto de vista del coste computacional, consumo energético y consumo de potencia eléctrica.

**Sexto capítulo:**

En este capítulo se mostrará el caso de estudio de una posible aplicación del sistema de detección de actividades criminales en video en la Policía Nacional de Colombia.

**Séptimo capítulo:**

En el sexto capítulo se realiza el análisis de integración del sistema planteado en esta tesis con los sistemas tácticos y tecnológicos de la Policía Nacional de Colombia.

**Octavo capítulo**

Presentan las conclusiones de la tesis y las futuras líneas de investigación, por último, se encuentran las referencias citadas.





## 2. ESTADO DEL ARTE TÉCNICO

### 2.1 Introducción

La seguridad ciudadana es definida por Naciones Unidas como *«el proceso de establecer, fortalecer y proteger el orden civil democrático, eliminando las amenazas de violencia en la población y permitiendo una coexistencia segura y pacífica [2].* Naciones Unidas también considera la seguridad ciudadana como un bien público por lo que es responsabilidad de los estados garantizarla en sus respectivos territorios.

Para garantizar la seguridad ciudadana, los estados han desarrollado políticas públicas de seguridad ciudadana, por ejemplo, en la República de Colombia el Ministerio de Defensa Nacional es el encargado de llevar a cabo la *«Política de Seguridad y Convivencia (PSC)»* la que ha sido desarrollada acorde con las necesidades de seguridad ciudadana de esta nación [3]. Con respecto al gobierno de España, en diciembre de 2017 se publicó la Estrategia de Seguridad Nacional, la cual entre varios objetivos propone *«promover una cultura de seguridad Nacional»* y *«Defender la legalidad y preservar de la seguridad ciudadana»* [4]. Estos dos ejemplos entre muchos otros muestran preocupación de los estados por garantizar la seguridad de sus ciudadanos.

Para dar cumplimiento a estas políticas los estados del mundo cuentan con diversas agencias de seguridad, y gran parte de la responsabilidad de garantizar la seguridad ciudadana recae en instituciones como la Policía Nacional y/o Policía Local, las cuales velan por el cumplimiento de las leyes y luchan contra elementos que afectan la seguridad ciudadana como la delincuencia común, las bandas criminales, narcotráfico, violencia doméstica, terrorismo extremista y demás amenazas.

Por los tamaños de las ciudades y las altas poblaciones de los países, las agencias de policía cuentan con diversas herramientas tecnológicas que facilitan el cumplimiento del deber, para preservar la seguridad ciudadana.

Buscando realizar una tesis doctoral que tenga impacto el país de origen del autor, esta tesis desarrolla una arquitectura de sistema de detección de actividades criminales en video aplicado en Sistemas de Mando y Control para seguridad ciudadana, el cual podrá ser implementado en un futuro en la Policía Nacional de Colombia.

En este capítulo se revisará el estado del arte de las tecnologías necesarias para el desarrollo de sistema de detección de actividades criminales en video aplicado en sistemas de mando y control para seguridad ciudadana propuesto en esta tesis doctoral.

## **2.2 Sistemas de Información de mando y control en seguridad ciudadana**

Las agencias estatales que deben preservar la seguridad ciudadana tienen que tomar decisiones estratégicas en tiempo real, sobre todo cuando las acciones de prevención y disuasión del delito no tiene los resultados esperados. Si se presentan situaciones inesperadas que amenacen la seguridad ciudadana los comandantes de las agencias de seguridad estatal deben tener información en tiempo real que les permita tomar las decisiones estratégicas adecuadas.

La herramienta tecnología que centraliza toda la información estratégica obtenida en tiempo real del territorio, es el sistema de información de mando y control C2IS [1] por su sigla en inglés, el cual mejora la conciencia situacional (*Situational Awareness*) [1] de los comandantes, permitiendo una toma efectiva de decisiones estratégicas en situaciones que comprometan negativamente la seguridad ciudadana como violencia, criminalidad, terrorismo y demás alteraciones del orden público.

Un sistema de información de mando y control muestra información estratégica la cual generalmente es desplegada en un lugar físico conocido como centro de mando y control [1], por ejemplo, la Policía Nacional de Colombia cuenta con *Centros de Comando y Control Seguridad Ciudadana* [5], los cuales centralizan información proveniente de múltiples fuentes como, llamadas de los ciudadanos a las líneas de emergencia, reportes de los agentes en el territorio, reportes de las redes de informantes, video en tiempo real proveniente del sistema de video vigilancia ciudadana entre otros.



Figura 1: Centro de Comando y Control Seguridad Ciudadana Policía Nacional de Colombia [5].

En estos centros de mando y control, los comandantes de las agencias de estatales encargadas de preservar la seguridad ciudadana, deben estar

conscientes de la situación presente en el territorio en tiempo real, y según este grado de conciencia situacional deberán tomar decisiones tácticas para cumplir su cometido. Si los comandantes no cuentan con la información suficiente, está errada o no está disponible en tiempo real, puede llevar a la toma incorrecta de decisiones o a la omisión de acciones críticas que en algunos casos pueden llevar a la pérdida de vidas humanas.

Como se ha mostrado la información disponible en C2IS es de gran relevancia para preservar la seguridad ciudadana a continuación, se enumeran las fuentes principales de información en un C2IS enfocado a seguridad ciudadana:

- **Llamadas de emergencia**

Esta información es recopilada telefónicamente por las líneas de emergencia (112, 123, 092, 911) en donde los ciudadanos reportan emergencias de todo tipo, esta información es almacenada en el sistema de información, la cual debe ser georreferenciada, clasificada y priorizada, según el tipo de caso, pues se da prioridad a emergencias que puedan implicar la pérdida de vidas.

- **Reportes de agentes en campo**

Las agencias estatales como la policía, cuenta con miles de agentes desplegados en el territorio nacional, estos tienen comunicación en

tiempo real con los centros de mando y control y al identificar cualquier situación anómala, esta es reportada y dicho informe se convierte en un caso de investigación, este es almacenado en el sistema de información la cual es georreferenciada, clasificada y priorizada.

- **Sistema de video vigilancia**

En la mayoría de grandes ciudades existe un gran despliegue de cámaras de seguridad las cuales son monitoreadas desde los centros de mando y control, cuando algún agente encargado del monitoreo detecta una perturbación a la seguridad ciudadana, procede a reportar el caso de inmediato en el sistema de información.

- **Redes de Informantes**

Es común que las agencias de policía cuenten con redes de colaboradores ciudadanos como taxistas, comerciantes o juntas vecinales. Cuando estas redes de apoyo detectan anomalías son reportadas al centro de mando y control, en donde esta información es georreferenciada, clasificada y priorizada en el sistema de información.

- **Sistemas de apoyo**

En ocasiones las agencias de seguridad cuentan con sistemas de apoyo para el reporte de situaciones críticas como botones de

pánico instalados en establecimiento comerciales o sitios estratégicos, sistemas de alarmas vecinales o residenciales instalados por privados y distintas aplicaciones móviles, toda esta información es llevada al sistema de información después de ser georreferenciada, clasificada y priorizada.

- **Análisis de datos previos**

Las agencias de seguridad realizan diversos análisis de datos, por ejemplo, datos criminalísticos, como resultado de estos análisis, es posible identificar zonas del territorio más propensas a alteraciones en fechas específicas, relacionadas con consumo de alcohol, días feriados etc.

- **Interconexión con otras agencias**

Los sistemas de mando y control dedicados a la seguridad ciudadana pueden recibir información de otras agencias estatales, por ejemplo, es posible recibir información de agencias militares, de inteligencia, federales, de seguridad nacional, forestales etc.



### **2.3 Sistemas de video vigilancia para seguridad ciudadana**

Como se enunció anteriormente la disponibilidad de información en tiempo real para los comandantes es de suma importancia para la efectividad de las acciones que se toman para preservar la seguridad ciudadana y una de las fuentes principales de información, es el sistema de video vigilancia desplegado en el territorio.

Sin embargo, en agencias de seguridad como la Policía Nacional de Colombia el monitoreo de miles de cámaras es realizado por agentes humanos sin ayudas tecnológicas para la detección de eventos, lo que dificulta el análisis de información proveniente de este sistema, ya que en relación existen muchas más cámaras que agentes disponibles para el monitoreo, lo que en ocasiones genera que existan eventos que no son detectados y por ende, no son reportados al sistema de información lo que reduce la efectividad de la Policía Nacional, en su labor de preservar la seguridad ciudadana.

Los sistemas de sistemas de video vigilancia están compuestos por los siguientes componentes:

- **Cámaras**

Son las encargadas de capturar las escenas de video y están diseñadas para trabajar en exteriores bajo condiciones climáticas

extremas, se dividen en dos grandes tipos cámaras de escena fija y cámaras de escena móvil.

- **Red de transporte de video**

Dada la gran dispersión geográfica de las cámaras y que todo este video debe ser llevado al centro de mando y control, debe existir una red que transporte el video desde las cámaras hasta el centro de mando y control, en el caso de Colombia esta red es una red IPv4, la cual no pertenece directamente a las Policía Nacional, sino que es propiedad de un operador de Telecomunicaciones con la cobertura adecuada.

- **Sistema de monitoreo**

Este sistema de monitoreo está compuesto por monitores en los cuales se despliega el video de las cámaras de vigilancia, este sistema es manejado por oficiales de policía, los cuales tienen la capacidad de controlar algunos aspectos de las cámaras como su orientación (solo en el caso del cámara con escena móvil) y el zoom de la imagen enfocada.

- **Sistema de grabación y almacenamiento de video**

Almacenar el video obtenido del sistema de video vigilancia es de gran importancia para agencias estatales como la Policía Nacional, pues este video puede ser usado para hacer seguimiento a bandas

criminales, terroristas, además de ser utilizado como material probatorio en procesos judiciales.

Dadas el gran espacio de almacenamiento requerido para almacenar esta gran cantidad de video cada agencia de seguridad determina con que características de calidad y por cuanto tiempo es almacenado este video, dependiendo de los requerimientos legales de cada país.

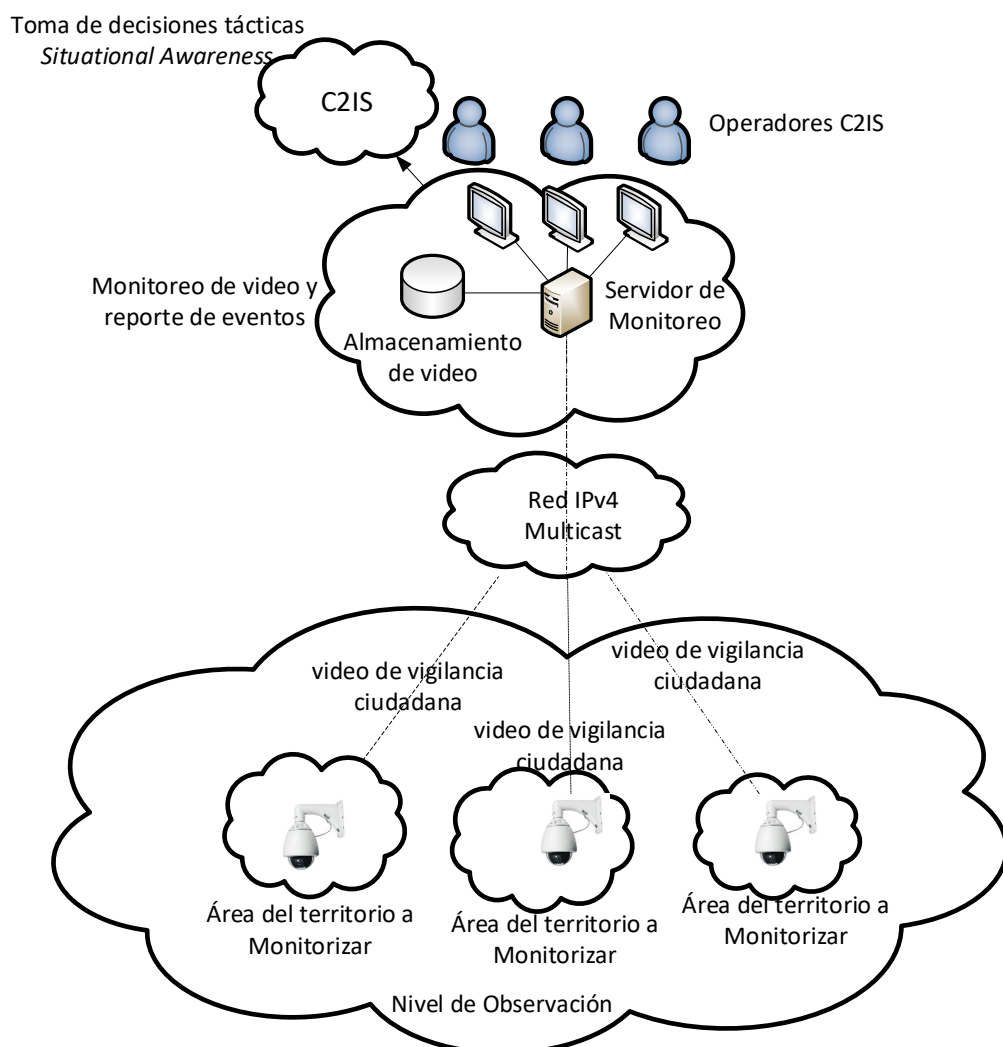


Figura 2: Arquitectura del sistema de video vigilancia

## 2.4 Visión artificial

Dada la problemática que esta tesis doctoral plantea resolver, la técnica adecuada para lograr los objetivos se enmarca en la visión artificial, campo de estudio que pretende que los sistemas sean capaces de darle sentido a las imágenes y que no se reduzcan a matrices de bits [6].

Este campo ha tenido grandes avances durante los últimos años con la inclusión de técnicas innovadoras, que aprovechan en reciente desarrollo del hardware incluido en las tarjetas gráficas para procesamiento paralelo [7] [8].

Estos avances están liderados por las técnicas basadas en Deep Learning [7], las cuales han tenido grandes logros. Uno de los éxitos más notables es el logrado en Large Scale Visual Recognition Challenge (ILSVRC), en el cual usando *Deep Learnig* se ha logrado reducir el margen de error de clasificación de imágenes a valores cercanos al 3.0% lo cual es comparable con el error humano; valor 8.5 veces más bajo al obtenido con otras metodologías, que es cercano al 26% [9], teniendo en cuenta estos resultados y la gran investigación en el campo esta tesis doctoral se centrará en Deep Learning para solucionar la problemática planteada.

## 2.5 Deep Learning

Para definir el aprendizaje profundo (Deep Learning), primero se debe definir el concepto de aprendizaje automático pues, el Deep Learning puede ser definido con un subconjunto de técnicas contenidas dentro del aprendizaje automático (Machine Learning) [7].

Por esto, se empezará definiendo el Machine Learning como la rama de la inteligencia artificial, que tiene por objeto hacer que los computadores sean capaces de generalizar comportamientos que matemáticamente no se pueden realizar por otros medios, es decir, que a partir de datos iniciales se puedan reconocer patrones, para luego ser reconocidos en datos no estudiados previamente [10].

Con esta tecnología es posible realizar múltiples aplicaciones imposibles de realizar con otras técnicas como predictores [11] [12], clasificadores [13] [14], sistemas de ayuda para pilotos [15], y muchas otras, siendo el soporte para el procesamiento masivo de datos Big Data [16] [17].

Ciertos algoritmos de Learning Machine tienen un alto grado de complejidad, los cuales intentan modelar problemas de alta complejidad, como por ejemplo la visión por computador, en donde cada imagen a color es una matriz tridimensional y una señal video en promedio, genera 30 imágenes por segundo, lo que implica grandes cantidades de información que deben ser procesadas en lapsos muy cortos de tiempo.

Por la complejidad de estos algoritmos, son llamados de aprendizaje profundo o Deep Learning [7], y el ejemplo más representativo son las redes neuronales convolucionales [8], las cuales fueron mencionadas por primera vez en Japón por Kunihiko Fukushima en la década de 1980 [18] [19].

## **2.6 Redes Neuronales Convolucionales**

Como se mencionó antes, las redes neuronales convolucionales son técnicas de aprendizaje automático de alta complejidad por lo que son clasificadas como técnicas de aprendizaje profundo o Deep Learning [7].

Las redes neuronales convolucionales o CNN (*convolutional neural network*) por su sigla en inglés, fueron diseñadas desde sus inicios para ser aplicadas en visión por computador [18] [19] y actualmente gracias a su implementación en hardware del alto poder para procesamiento en paralelo como las GPU (*graphics processing unit*), se han logrado implementar modelos complejos con resultados excelentes en visión artificial [7] [8] [9].

Como su nombre lo indica las CNN, se basan en redes neuronales artificiales y son clasificadas como profundas dado que las redes neuronales artificiales en muchas aplicaciones usan un promedio de 2 o 3

capas, mientras que las redes neuronales convolucionales pueden tener cientos de capas [7] [8].

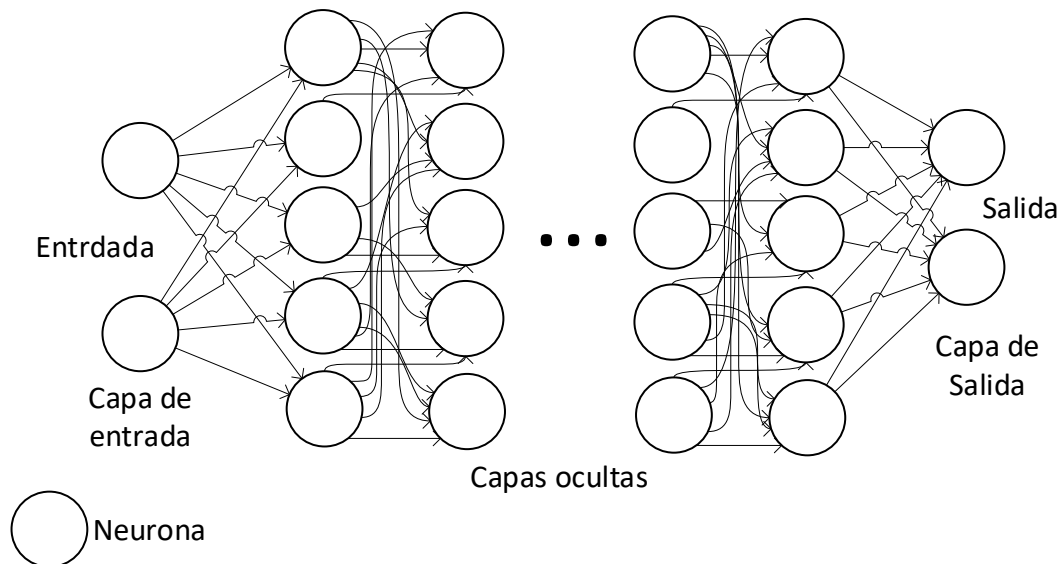


Figura 3: Red de neuronas artificiales.

Estas redes profundas combinan múltiples capas de procesamiento no lineal, usando elementos de operación en paralelo inspirados en el sistema nervioso biológico, conocido como «neuronas», estas neuronas se interconectan entre sí formando redes, las cuales a partir de un proceso repetitivo de prueba y error llamado entrenamiento, pueden aprender las características de un determinado conjunto de datos (Dataset) conocido, una vez aprendidas estas características, dichas redes de neuronas pueden identificar características similares a las aprendidas en el proceso de entrenamiento en casos que nunca han sido analizados por dicha red de neuronas [10].

Las redes neuronales convolucionales están formadas por tres partes fundamentales, la capa de entrada, la sección de detección de características y la sección de clasificación, que enseguida serán descritas [8].

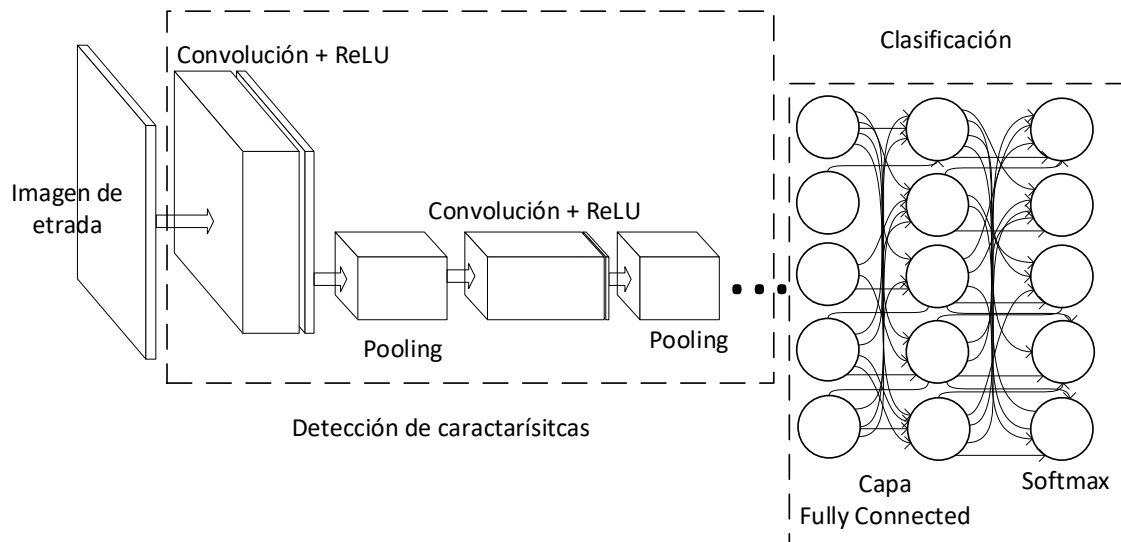


Figura 4: Red Neuronal Convolucional

- **Capa de entrada**

En esta capa se ingresan los datos (imágenes) a ser procesados por la red neuronal convolucional, aquí se definen ítems como el tamaño y profundidad de color de las imágenes [7] [8].

- **Detección de características.**

El objetivo de esta sección es identificar características de los datos que están siendo procesados, por ejemplo, si la red está aprendiendo a identificar seres humanos, la red aprenderá las



características principales de los seres humanos; como tener un torso, cuatro extremidades y cabeza.

La detección de características cuenta con tres operaciones fundamentales las cuales son:

**Convolución:** realiza un filtrado convolucional a las imágenes para la identificación de características de estas.

**Pooling:** simplifica la salida realizando operaciones no lineales de downsampling, reduciendo la cantidad de parámetros que la red tiene que aprender.

**Rectified Linear Unit (ReLU):** permite un entrenamiento más rápido y efectivo al mapear valores negativos a cero y mantener valores positivos.

Según la topología de red, se varía cantidad, posición y características de estas tres operaciones fundamentales, las cuales tienen efectos directos sobre la detección de características [7] [8].

- **Clasificación**

En la fase final la red neuronal convolucional procede a realizar la clasificación de objetos según las características aprendidas anteriormente y para eso cuenta con una capa **Fully Connected**, la cual tiene como salida un vector de dimensión  $k$ , en donde  $k$  es el número de clases que la red tiene para clasificar, este vector

contiene las probabilidades de cada clase para cada objeto clasificado.

La capa final de una arquitectura CNN usa una función **softmax** (**normalized exponential function**), para entregar la salida de la clasificación [7] [8].

## 2.7 Arquitecturas de redes neuronales convolucionales

Como se ha mencionado anteriormente las CNN han logrado grandes avances en el campo de la visión artificial, especialmente en la clasificación de imágenes logrado en el *Large Scale Visual Recognition Challenge (ILSVRC)*, reducir el margen de error de clasificación de a valores cercanos al 3.0 % lo cual es comparable con el error humano [9]. Para lograr estos resultados desde 2012 se han desarrollado gran variedad de topologías de redes neuronales convolucionales que son variaciones de la arquitectura básica descrita anteriormente, las cuales se entrenan usando un conjunto de datos de un millón doscientos mil imágenes divididas en mil clases diferentes.

El proceso de entrenamiento de estos modelos demanda un gran poder computacional de procesamiento en paralelo y toma un tiempo considerable, pero como se ha dicho antes con excelentes resultados.

Durante los últimos años se han desarrollado arquitecturas de redes neuronales convolucionales, las cuales son cada vez más profundas llegando a los cientos de capas.

Entre estos modelos pre-entrenados se destacan AlexNet, el cual fue desarrollado en 2012 por Alex Krizhevsky [20], este modelo marcó un hito en la investigación del Deep Learning, pues en el ILSVRC de 2012 mostró una gran mejoría en comparación con otras técnicas de aprendizaje automático. AlexNet es relativamente sencillo y solo está formado por 5 capas de convolución, tres capas MaxPool, dos fases de normalización, 3 capas fully connected y una capa Softmax, la tasa de error obtenida en ILSVRC fue de 15.3% [9].

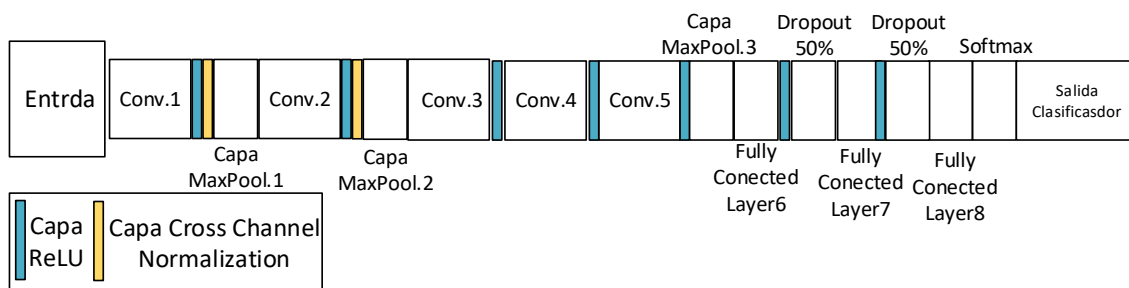


Figura 5: Modelo AlexNet.

AlexNet, ha sido muy popular siendo usado en varios estudios científicos [21] [22] [23], presentando mejoras como ZFNet la cual mejora los hiperparámetros de AlexNet obteniendo mejores resultados en la detección [24].

En 2014 Karen Simonyan y Andrew Zisserman de la Universidad de Oxford presentaron los modelos denominados “muy profundos” VGG-16 y VGG-19 [25], los cuales son más complejos como se ve en la figura 6, sin embargo en el ILSVRC logran reducir el margen de error al 7.3% [9], este mismo año Google presentó el modelo GoogleNet [26], la cual tiene una complejidad un poco mayor los modelos VGG pero obtiene una tasa de error de 6.67% [9]; sin embargo esta mejora en el margen de error tiene un mayor costo computacional debido a la gran complejidad de la red como se mostrará más adelante.

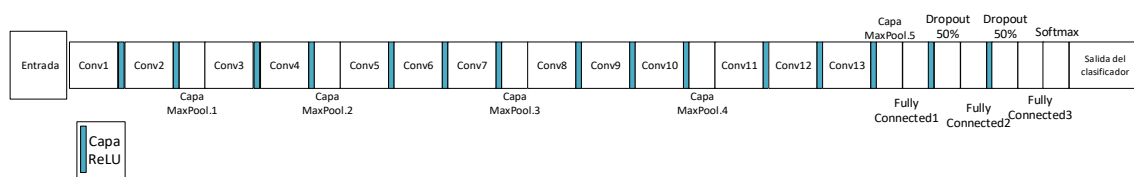


Figura 6: Modelo VGG-16.

Los últimos modelos desarrollados han llevado la complejidad de las redes mucho más allá, pues por ejemplo el modelo ResNet [27] presentado por Kaiming He, Xiangyu Zhang, Shaoqing Ren y Jian Sun tiene 152 capas y en el ILSVRC logra un margen de error del 3.57% [9], lo que es comparable con el margen de error de un ser humano promedio.

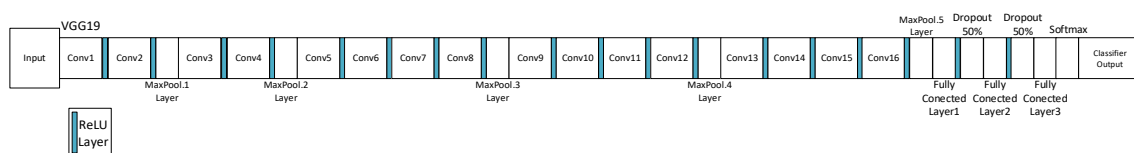


Figura 7: Modelo VGG-19.

## 2.8 R-CNN (Region-Based Convolutional Neural Networks)

Como se ha venido mencionando, con respecto a la visión artificial as técnicas de aprendizaje automático cuentan con varias aplicaciones por ejemplo clasificadores de imágenes [28] [25] [29] el reconocimiento de emociones [30], entre otras, lo que hace que el desarrollo de Deep Learning no se haya detenido, pues en los últimos dos años han existido desarrollos de gran relevancia que constituyen el estado del arte en esta temática.

Parte importante del estado del arte en este tema, es la técnica de Deep Learning conocida como R-CNN (*Region-Based Convolutional Neural Networks*) publicada en el *IEEE Transactions on Pattern Analysis and Machine Intelligence* en 2016 por Ross Girshick [31], la cual propone analizar imágenes por regiones y no imágenes completas como las CNN descritas anteriormente, gracias a este estudio por regiones es posible detectar varios objetos o situaciones de interés en la misma imagen.

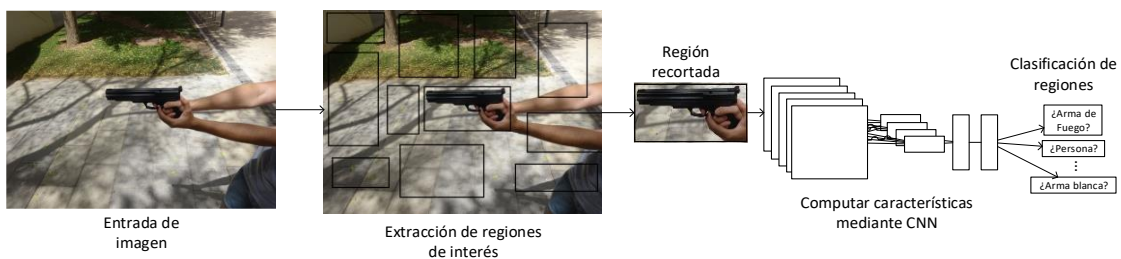


Figura 8: Arquitectura R-CNN

Para lograr esto R-CNN (*Region-Based Convolutional Neural Networks*) define regiones de interés o *Region Proposals* las cuales son analizadas por una Red Neuronal Convolutiva que actúa como clasificador de imágenes, estas CNN usan modelos como los descritos en el numeral 2.7, los cuales han sido reentrenados para cada caso específico. Gracias a este método se han podido solucionar problemas de visión artificial que con otros métodos eran casi imposibles de abordar [6] [31], sin embargo el tiempo de procesamiento de cada imagen está por el orden de las decenas de segundos lo que impide su aplicación en video de tiempo real.



Figura 9: Resultados de detección de objetos usando R-CNN

## 2.9 Fast R-CNN (Fast Region-Based Convolutional Neural Networks)

Para solucionar el problema del alto tiempo de procesamiento Ross Girshick propuso una mejora sobre R-CNN que fue llamado Fast R-CNN [32], el cual reduce el tiempo de procesamiento de cada imagen en alrededor de 20 veces con respecto al R-CNN, con lo que el tiempo de procesamiento por imagen está en el orden de los segundos.

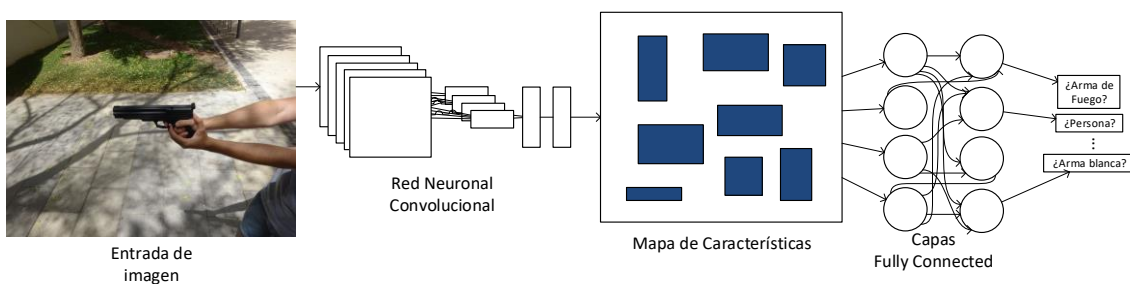


Figura 10: Arquitectura Fast R-CNN

Fast R-CNN ha sido usando en variedad de estudios interesantes en el campo de la robótica [33], conducción autónoma [34] [35], localización y clasificación de objetos en video [36], detección facial [37] y muchos más mostrando su efectividad, la cual se logra agregar a un mapa de características extraído por una Red Neuronal Convolutiva, lo que reduce en gran medida el tiempo de procesamiento de imagen.

Con esta técnica sería posible hacer procesamiento de video en tiempo real, sin embargo, se debería usar hardware de muy alto rendimiento para procesar un solo flujo de video.

## **2.10 Faster R-CNN (Faster Region-Based Convolutional Neural Networks R-CNN)**

Una última mejora a la técnica fue publicada en 2017 en el *IEEE Transactions on Pattern Analysis and Machine Intelligence* por Shaoqing Ren; Kaiming He; Ross Girshick y Jian Sun, que fue llamada Faster R-CNN [38], en donde se potenciamn las mejoras obtenidas en Fast R-CNN [32] implementado una RPN (*region proposal network*) y eliminando los métodos para calcular las propuestas de regiones de interés.

Al aplicar estas mejoras al algoritmo el grupo de Shaoqing Ren y Ross Girshick, logró que Faster R-CNN obtuviera tiempos de procesamiento de imágenes de fracciones de segundo, mejorando 10 veces los tiempos obtenidos por Fast R-CNN, convirtiendo a Faster R-CNN en estado del arte en Deep Learning aplicado a Visión por computador.



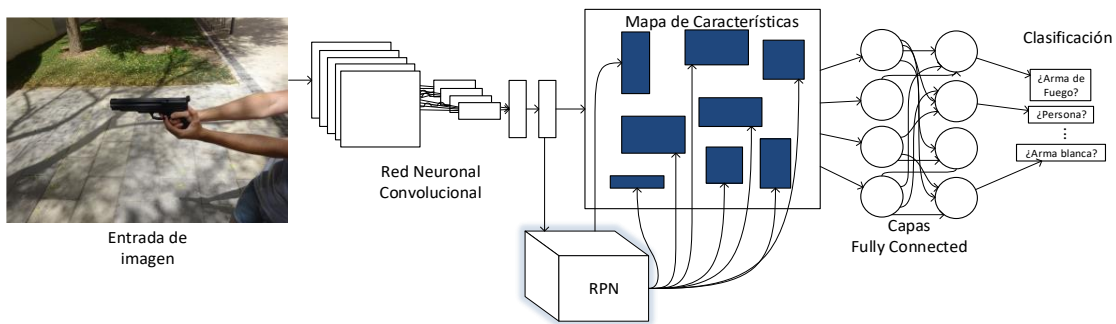


Figura 11: Arquitectura Faster R-CNN

Como se ve en la figura 11, Faster R-CNN es una mejora de Fast R-CNN la cual agrega un novedosa RPN [38], dicha red tiene por objeto predecir la ubicación de las regiones de interés apodando al mapa de características. Aa red neuronal convolucional usada como base para la operación de Faster R-CNN en [38] fue VGG-16 [25], con lo que se obtuvieron los resultados mostrados anteriormente.

## 2.11 Aplicaciones de Deep Learning en seguridad ciudadana

Como se ha venido mostrando el estado del arte de aplicaciones de visión artificial está regido por Deep Learning, y específicamente por las técnicas basadas en R-CNN, Fast R-CNN y Faster R-CNN, a continuación, se mostrarán algunas de las aplicaciones desarrolladas por otros autores, más recientes y relevantes con respecto a aplicaciones de estas

tecnologías en sistemas de mando y control o sistemas de seguridad ciudadana.

- **Detección de peatones**

La detección de peatones en los sistemas de seguridad ciudadana puede ser de gran relevancia, por ejemplo, para detectar la presencia de peatones en zonas restringidas y es el primer paso para aplicaciones más avanzadas como seguimiento de peatones. Al revisar el las bases de indexación se encuentran cientos de trabajos relacionados con el tema, entre los cuales se destacan: [39] publicado en 2018 el cual propone un detector de peatones basado en Fast R-CNN en escenas urbanas, también es interesante el trabajo presentado en [40], en el cual los investigadores desarrollan una técnica basada en CNN con el objeto de detectar trayectorias de peatones para ser aplicado en sistemas de detección de riesgos viales. En este mismo tema resalta el trabajo presentado en [41], en el cual se trabaja un problema típico en la detección de peatones, que es la confusión entre peatón y ciclista. Este trabajo propone un modelo basado en Fast R-CNN el cual puede identificar tanto a ciclistas, motociclistas y peatones, lo que podría ser muy interesante en situaciones de seguridad en donde se cometan crímenes en motocicleta o bicicleta [42]. Finalmente es de destacar el trabajo presentado en Consumer Electronics (ICCE), 2018 [43] en el cual se

muestra un sistema basado en Faster R-CNN para la detección de peatones a larga distancia desde imágenes capturadas desde una aeronave no tripulada.

- **Detección de individuos**

En los sistemas de video vigilancia existe la necesidad de identificar individuos sospechosos, sin embargo por las condiciones del sistema es imposible usar sistemas biométricos de reconocimiento facial, pues los individuos sospechosos evitan las cámaras de seguridad y además cubren sus rostros con sombreros y capotas, por tanto hay trabajos que usan CNN enfocados en la identificación de individuos según su caminar [44] [45] [46]. También existen trabajos interesantes centrados en la detección de peatones basados en sus atributos [47] [48], lo que es un buen comienzo para sistemas de identificación basado en descripción de sospechosos.

- **Detección vehículos e interpretación de matrículas en automóviles**

En las alteraciones a la seguridad ciudadana a menudo hay vehículos involucrados, su identificación es de gran relevancia para las agencias de seguridad pues una vez identificados los vehículos sospechosos se pueden tomar las decisiones tácticas

correspondientes, si estos vehículos son captados por el sistema de video vigilancia Deep Learning puede ser de gran ayuda y existen varios trabajos enfocados en la identificación de vehículos [49] [50] [51] y de sus números de matrícula [52] [53] [54] [55].

- **Detección de emociones**

Otra aplicación de Deep Learning que puede ser de utilidad para las agencias de seguridad es la detección de emociones expresadas facialmente, pues en sitios estratégicos como aeropuertos u oficinas estatales se podrían instalar cámaras de seguridad con sistemas de Deep Learning entrenados para la detección de emociones [56] [57] [58] como ansiedad o miedo, los cuales pueden ser precursores de alteraciones a la seguridad ciudadana.

- **Detección de escenas violentas**

Sin lugar a duda la violencia genera alteraciones a la seguridad ciudadana, por tanto, la detección de actividades violentas como peleas callejeas, peleas en escenarios deportivos, etc., son de gran interés para las agencias de seguridad. En este sentido es importante resaltar trabajos como [59] [60] en donde se detecta la violencia en secuencias de video o trabajos en donde se detectan gritos de auxilio [61] [62].

Como se mostró en este capítulo Deep Learning agrupa varias técnicas que conforman el estado del arte en temáticas de visión artificial. Por esto esta tesis doctoral desarrolla la arquitectura de detección de actividades criminales en video aplicado en sistemas de mando y control para seguridad ciudadana con base en Deep Learning.



### 3. ESCENARIO DE UTILIZACIÓN

Colombia es una nación con aproximadamente 49 millones de habitantes de los cuales el 77% viven en ciudades [63]. En el país se ha venido disminuyendo en índice de pobreza de su población pasando del 49.7% en 2002 al 27.8% en 2015 [64]. Históricamente, Colombia ha sido golpeada por la violencia y otros fenómenos típicos de países en desarrollo como la inseguridad en las ciudades, la cual se ha venido disminuyendo durante la última década, pero todavía sigue siendo considerable.

Diferentes problemas sociales como el desempleo (11.7% enero 2018) [65], la inmigración de países vecinos [66] [67], violencia en áreas rurales [68], ha llevado a las ciudades Colombianas a tener altas tasas de criminalidad que aunque se han venido disminuyendo durante los últimos años, se siguen manteniendo altas [69] [70], y en 2017 la Policía Nacional de Colombia oficialmente registró 90.281 casos de hurto a personas y 11.983 casos de homicidios en todo el territorio Nacional [71].

Para afrontar esa situación, el gobierno de Colombia cuenta con diferentes programas sociales y además cuenta con apoyo internacional para mejorar el índice de desarrollo humano de la población, y dar una solución de fondo a las problemáticas que tiene el país. Como ejemplo de estos

programas se destacan: Programa de las Naciones Unidas para el Desarrollo [72], Delegación de la Unión Europea en Colombia [73], USAID Colombia [74], programas de desarrollo del departamento de Planeación Nacional [75], entre otros.

Estos programas nacionales e internacionales han tenido excelentes resultados pues en 2018 el índice de desarrollo humano de Colombia es de 0.747 clasificando al país en la categoría «alta» de este importante indicador de naciones unidas [76].

A pesar de este positivo panorama, es innegable que en las ciudades colombianas existe inseguridad y es responsabilidad del estado Colombiano mantener la seguridad ciudadana de la población, por lo que el país cuenta con distintas fuerzas de seguridad que están formadas por: Ejército Nacional [77], Armada Nacional [78] y Fuerza Aérea Colombiana [79] quienes son responsables de mantener la soberanía y defender las fronteras del país, mientras que la Policía Nacional de Colombia [80] tiene la responsabilidad de mantener la seguridad ciudadana en las ciudades y hacer cumplir las leyes del país.

Por otro lado, la situación social en Colombia es dinámica pues durante los últimos años el país ha acogido a más de un millón de inmigrantes que se refugian de la situación económica y social del vecino país Venezuela [81], lo que hace que la fuerza pública tenga que aumentar los esfuerzos para brindar condiciones para todos los residentes del país.



El estado colombiano ha implementado políticas para acoger a los refugiados venezolanos, por ejemplo, se han entregado permisos de residencia y trabajo a los refugiados quienes hayan entrado al país de forma legal [82].

Muchos de los migrantes residentes en Colombia han denunciado conductas de xenofobia, discriminación y explotación laboral [83] [84], por parte de ciudadanos colombianos, pues así como el flujo constante de refugiados al país ha logrado despertar compasión y solidaridad en la mayoría de colombianos, otros ven en esta población migrante oportunidades para obtener mano de obra de bajo costo y otros culpan a los migrantes del desempleo y la inseguridad. Estos fenómenos de xenofobia, discriminación y explotación laboral son comunes en los procesos migratorios resultados de conflictos armados, inestabilidad política o crisis económica.

También se debe tener en cuenta que muchas familias venezolanas migrantes tienen antepasados colombianos que salieron del país el siglo pasado en búsqueda de mejores oportunidades o huyendo de la violencia fruto del conflicto armado colombiano o del Narcotráfico y ahora regresan al país.

Como en cualquier proceso migratorio la seguridad ciudadana se ve afectada tanto como para la población migrante por fenómenos de xenofobia, discriminación y explotación laboral como para la población originaria por posibles alteraciones en el orden público y mantener la seguridad ciudadana en el territorio es responsabilidad de la Policía Nacional, quien se debe valer de todas las herramientas tecnológicas para el cumplimiento de sus objetivos misionales.

Como respuesta a la cambiante situación social de Colombia, con sus problemáticas de terrorismo, delincuencia, etc., la Policía Nacional implementa diversas soluciones humanas y tecnológicas para poder mantener la seguridad ciudadana en Colombia como sistemas de mando y control y sistemas de video vigilancia ciudadana; herramientas que permiten mejorar los tiempos respuestas en diferentes operaciones policíacas en el territorio colombiano.

### **3.1 Modelo de desarrollo tecnológico de la Policía Nacional de Colombia**

Como se ha dicho con anterioridad, la Policía Nacional de Colombia, tiene la responsabilidad de garantizar la seguridad ciudadana en todo el territorio colombiano pues en Colombia no existen las figuras de policía Local o departamental.

Para cumplir con sus objetivos institucionales la Policía Nacional de Colombia tiene un modelo de desarrollo tecnológico centrado en la implementación final de un modelo de ciudad inteligente pasando por el modelo de ciudad segura y empezando por la implementación de centros de comando y control.

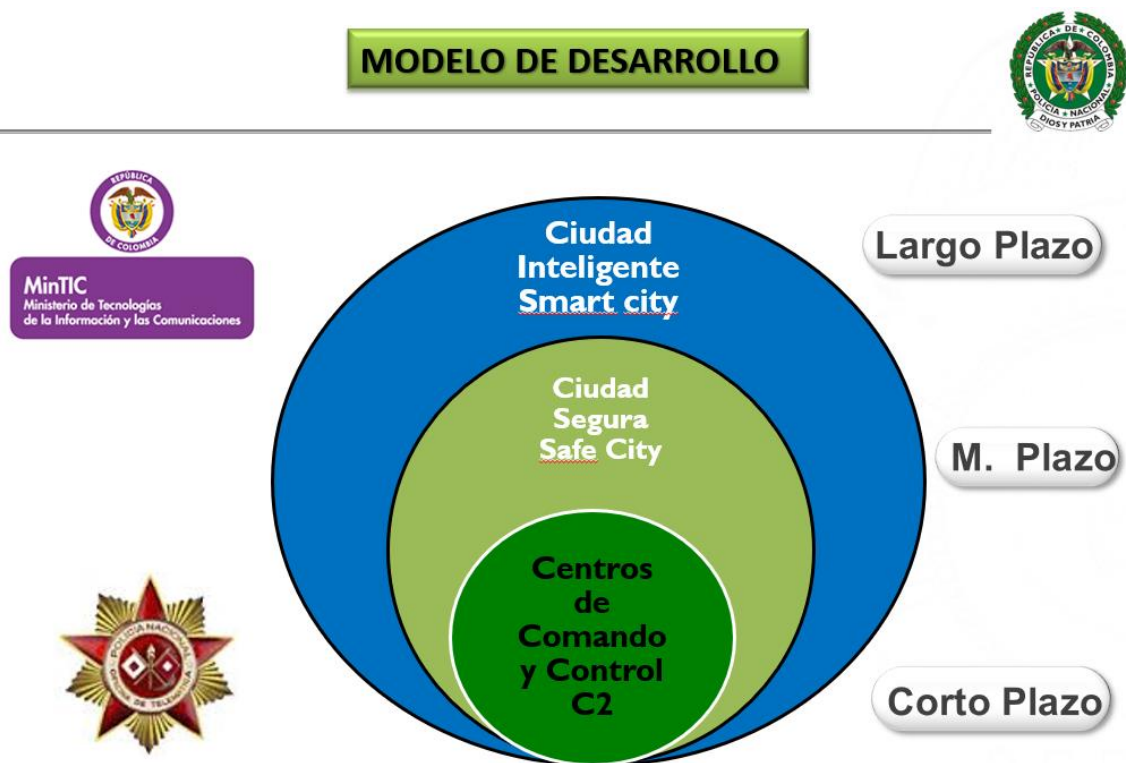


Figura 12: Modelo de desarrollo tecnológico (Policía Nacional de Colombia)

Para el desarrollo de este plan, el estado colombiano a través de sus diferentes organismos ha venido implementando diferentes centros de mando y control, por ejemplo en la capital del país Bogotá D.C. en 2017

se inauguró el Centro de Comando, Control, Comunicaciones y Cómputo C-4, en el cual convergen la Línea de Seguridad y Emergencias 123, al Cuerpo Oficial de Bomberos del Distrito Capital, al Instituto Distrital de Gestión y Riesgos IDIGER, y el Centro Automático de Despacho CAD de la Policía Nacional el cual recibe alrededor de 41.000 llamadas diarias [85].

Por otra parte, en la ciudad de Villavicencio también se puso en operación un centro de Comando y control para la seguridad ciudadana, el cual es operado por la policía nacional y recibió el apoyo tecnológico del Ministerio de Tecnologías de la Información y las Comunicaciones [86].



Figura 13: Prototipo de Centro de Comando y Control de Seguridad Ciudadana (Policía Nacional de Colombia)

Actualmente para complementar los centros de comando y control, la policía Nacional de Colombia cuenta con varios *Centros Automáticos de Despacho* (CAD), los cuales están distribuidos por el territorio colombiano y están encargados de la gestión y despacho de las unidades policiales frente a los casos de policía reportados por diferentes medios siendo el más relevante la línea única de atención a emergencias 123.

El panorama de desarrollo tecnológico con respecto a la seguridad ciudadana está centrado en la implementación a corto plazo, de igual modo la consolidación de centros de comando y control en las principales ciudades de Colombia, a mediano plazo se tiene previsto la implementación de esquemas de ciudades seguras y a largo plazo con el apoyo del Ministerio de Tecnologías de la Información y las Comunicaciones, la implementación de ciudades inteligentes.

### **3.2 Centros de comando y control: Policía Nacional de Colombia**

Para garantizar la seguridad ciudadana la Policía Nacional de Colombia [80] cuenta con 178.000 oficiales de policía, los cuales están desplegados por todo el territorio nacional. La Policía Nacional cuenta con gran variedad de herramientas tecnológicas como los sistemas de información de mando y control (C2IS) [1], los cuales centralizan toda la información estratégica en tiempo real mejorando la conciencia situacional de los comandantes de

policía, mejorando la toma de decisiones estratégicas como la localización de oficiales de Policía en las calles, movilidad de las unidades motorizadas, ubicación de unidades antidisturbios etc.



Figura 14: Centro de Comando y Control de Seguridad Ciudadana Ideal  
(Policía Nacional de Colombia)

Los sistemas de información de mando y control centralizan la información estratégica de seguridad ciudadana en un lugar físico llamado Centro de Comando y Control de Seguridad Ciudadana, en donde bajo una estricta línea de mando la información se trasmite de los operadores del C2IS a los comandantes de Policía para tomar las más importantes decisiones estratégicas en el mejor tiempo posible.

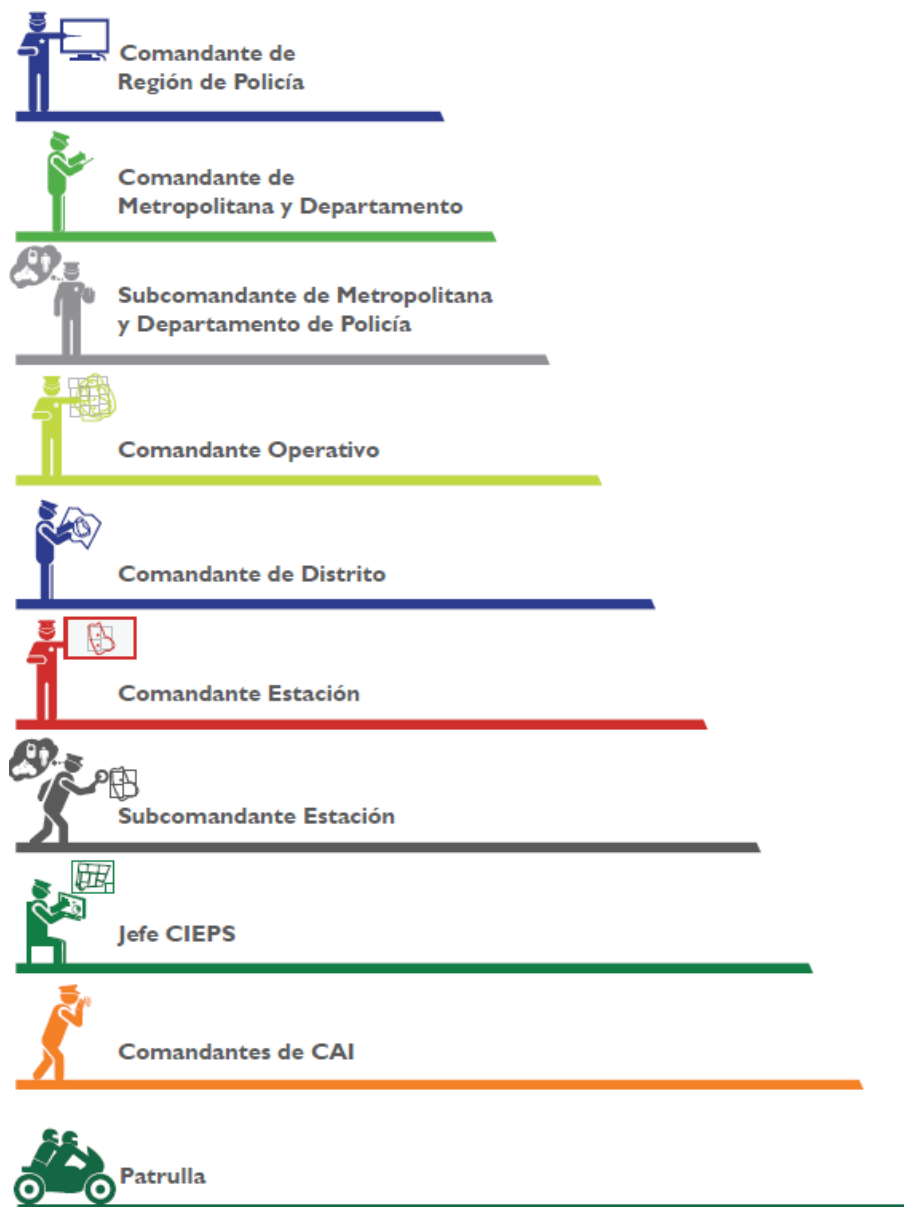


Figura 15: Línea de mando Policía Nacional de Colombia [87]

En la República de Colombia la Policía Nacional ha implementado un sistema de mando y control en el cual se define:

«Comando y control: Es el ejercicio de la autoridad, conducción y seguimiento del servicio de Policía a las tareas y acciones previamente planeadas para garantizar la efectividad del servicio en cada cuadrante, con el debido uso y articulación de los medios y herramientas tecnológicas disponibles. [88]

Con base en este concepto de comando y control; la Policía Nacional adoptó la siguiente arquitectura operativa para los sistemas de mando y control a ser implementados en la república de Colombia.

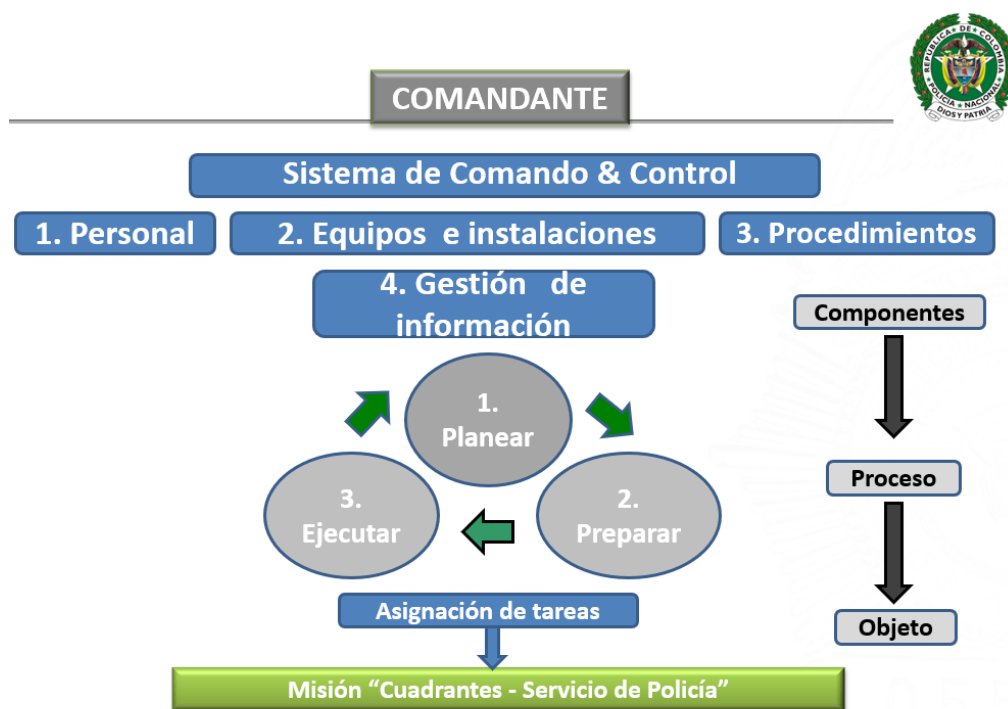


Figura 16: Arquitectura funcional del Centro de Comando y Control de Seguridad Ciudadana (Policía Nacional de Colombia)



Dentro de esta arquitectura existen diferentes roles y procesos los cuales son:

- **Comandante**

El comandante del centro de comando y control (comandante operativo) es el oficial de policía de más alto grado y autoridad, es la máxima autoridad y el cual es el responsable de la toma de decisiones, plantear estrategias, planear operaciones e impartir ordenes con el objeto de cumplir las funciones misionales de la Policía Nacional y de seguridad ciudadana en su área de operaciones.

- **Personal**

Este aspecto se refiere al personal policiaco disponible en el centro de comando y control el cual cuenta con todo el entrenamiento policial y se encuentra bajo estricta línea de mando.

- **Procedimientos**

Para la correcta ejecución de las operaciones policiales, el alto mando de la institución ha definido una serie de procedimientos diseñados para obtener la máxima efectividad en las operaciones policiales, bajo el estricto control y aplicación de las leyes de la república de Colombia, estos procedimientos están en constante revisión y actualización debido al proceso de mejora continua de la

institución y posibles actualizaciones legales como nuevas leyes, normas, códigos civiles o de policía.

- **Gestión de la Información**

Otro proceso de suma importancia en el centro de comando y control es la gestión de la información, la cual siempre debe fluir y ser resguardada para posibles investigaciones posteriores, por ejemplo, las ordenes provenientes del alto mando tienen diversos niveles de confidencialidad y son entregadas por escrito o por medios digitales a los comandantes, estas órdenes son gestionadas mediante el sistema de gestión documental de la Policía Nacional.

### **3.2.1 Planeación y ejecución de operaciones de Policía**

La planeación y ejecución de operaciones de Policía en los centros de comando y control de la Policía Nacional de Colombia está basada en un proceso estructurado en tres fases bajo estricta línea de mando de la siguiente manera:

- **Planeación:** en esta fase se definen los objetivos estratégicos que se desean alcanzar con las operaciones, por ejemplo, si se planean operaciones contra bandas criminales se definen los objetivos de captura, planes de acción, armamento, posiciones estratégicas, línea de mando en campo, etc. y en

ocasiones se pueden planear operaciones conjuntas con otras fuerzas como el ejército nacional etc.

La planeación es responsabilidad del comandante, quien siempre sigue la línea de mando y reporta ante sus superiores.

En caso de situaciones de crisis, las operaciones de un centro de comando y control de una ciudad pueden ser dirigidas y/o monitoreadas directamente por el alto mando de la Policía Nacional desde la sala de crisis ubicada en la Dirección General de la Policía Nacional en la ciudad de Bogotá.

Si embargo para la gran mayoría de operaciones u operaciones rutinarias son planeadas por el comandante del centro de comando y control.

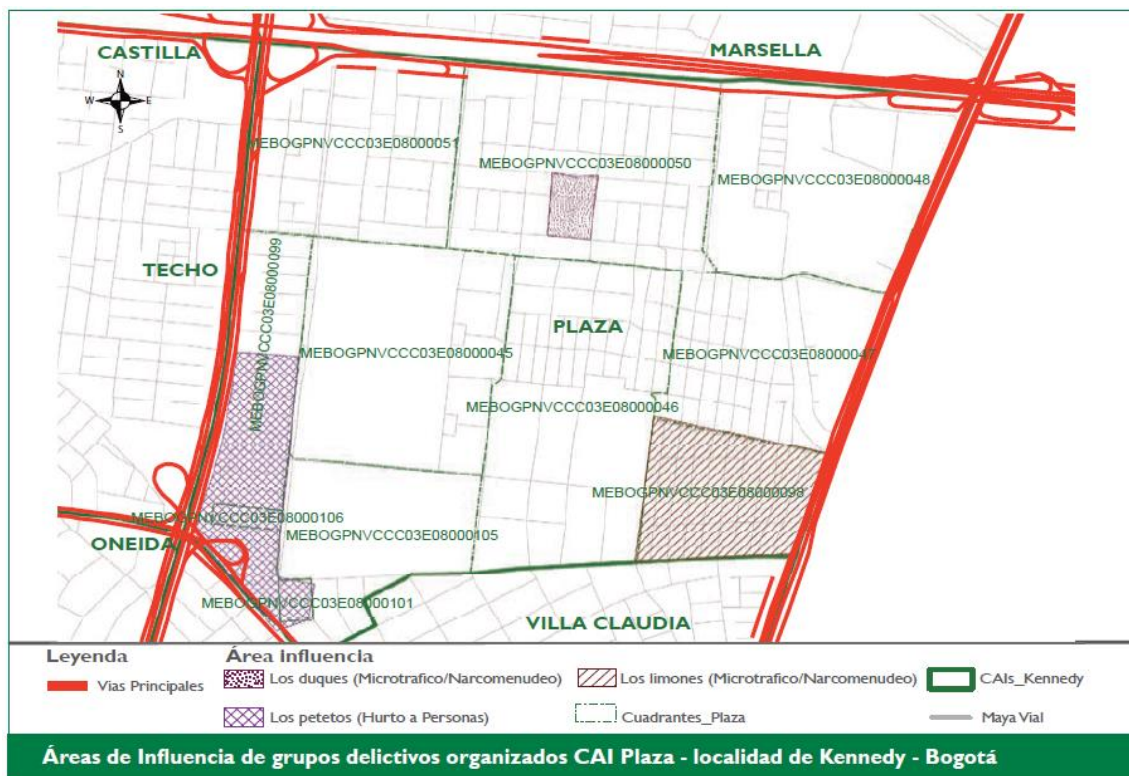


Figura 17: Georreferenciación de áreas de influencias de bandas criminales Localidad de Kennedy Bogotá D.C. [87]

- **Preparación:** una vez planeadas las operaciones policiales se deben hacer despliegues logísticos en las zonas geográficas a ser intervenidas. Por ejemplo, si se planean ejecutar órdenes de captura a determinada banda criminal, horas antes de la operación los agentes deben preparar las órdenes judiciales de captura. Aparte del respectivo soporte legal para la operación, los policías deben preparar el armamento, las protecciones corporales, dispositivos de comunicación y los vehículos a usar en la operación,

finalmente deben desplazarse y tomar las posiciones estratégicas previamente planeadas para evitar fugas y posibles enfrentamientos, siempre velando por la seguridad de la población.

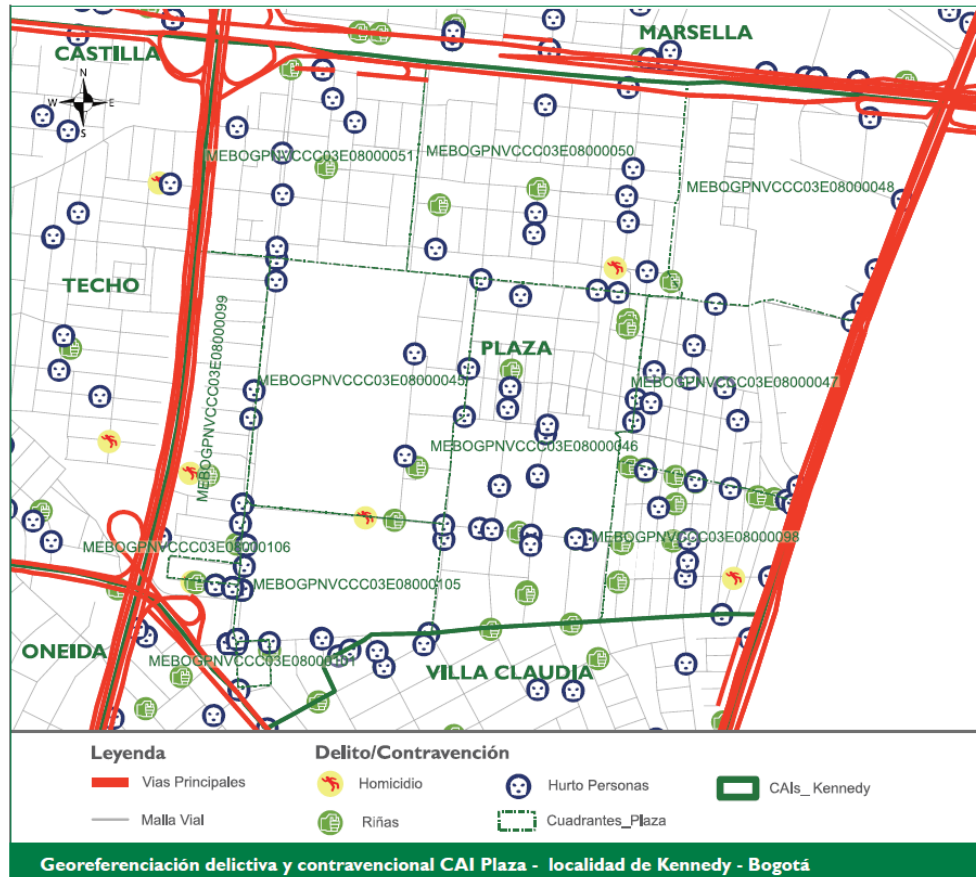


Figura 18: Georeferenciación de actividades criminales Localidad de Kennedy Bogotá D.C. [87]

- **Ejecución:** Una correcta y eficiente ejecución de operaciones de policía es uno de los objetivos fundamentales del centro de comando y control, y las fases anteriores del planeación y

preparación de las operaciones solo tienen por objeto cumplir los objetivos de las operaciones en la fase de ejecución de las operaciones de Policial.

Durante la fase de ejecución de operaciones es común que los policías en campo arriesguen sus vidas, pues, por ejemplo, si las operaciones tienen como objeto desarticular bandas criminales, es posible que existan enfrentamientos con armas de fuego, con consecuencias letales.

Por eso, en la ejecución de las operaciones policiales los sistemas de comunicación en tiempo real crítico entre el centro de comando y control con los policías en campo es crucial para asegurar el éxito de la operación.

Para esto la Policía Nacional de Colombia cuenta con diversos sistemas de comunicaciones y algunos de ellos utilizan asignaciones de espectro electromagnético exclusivas.

Para el comandante del centro de comando y control es fundamental contar con una conciencia situacional (*situational awareness*) excepcional para poder tomar las decisiones e impartir las ordenes adecuadas para cumplir los objetivos operacionales siempre salvaguardando la vida de la población y de los policías en campo.

Para esto el sistema de comando y control está alimentado con diversas fuentes de información en tiempo real, como el sistema de video vigilancia, y georreferenciación en tiempo real de las unidades policíacas en el terreno.

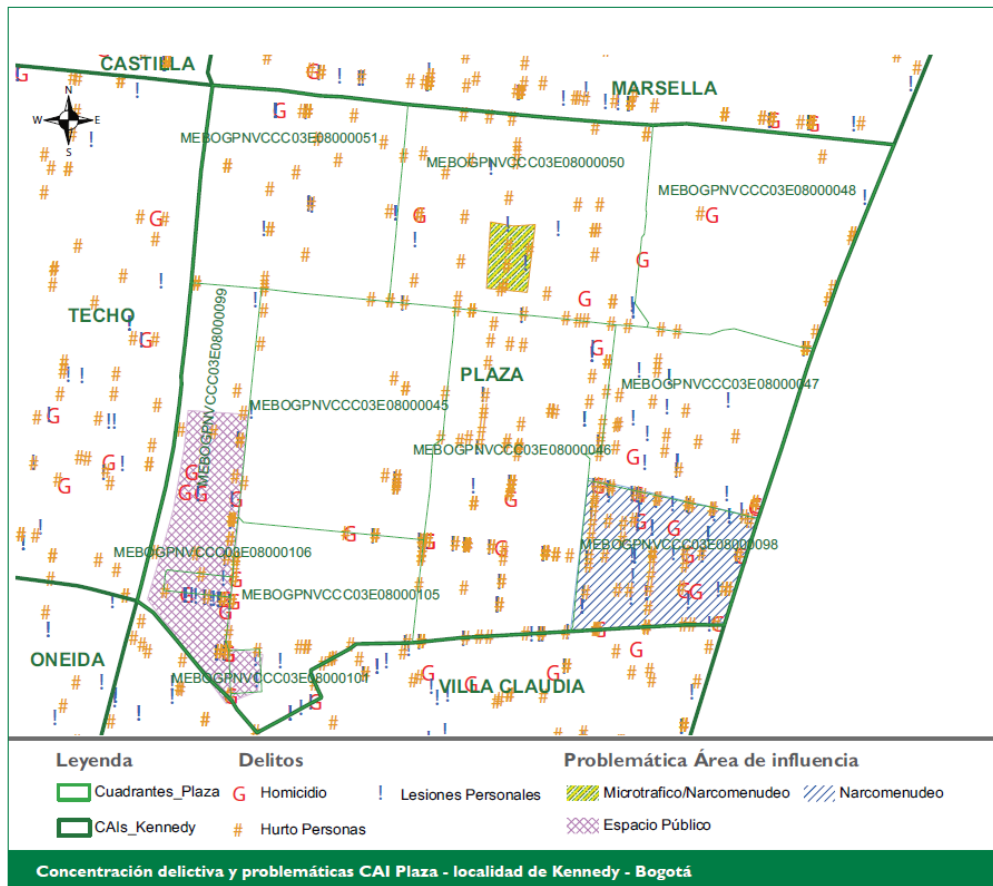


Figura 19: Informe Georreferenciado de actividades criminales Localidad de Kennedy Bogotá D.C. [87]

### 3.3 Comando y control en el MNVCC

El comando y control de las operaciones policíacas en Colombia es aplicado al Modelo Nacional de Vigilancia Comunitaria por Cuadrantes

(MNVCC), el cual es el modelo adoptado por la Policía Nacional de Colombia para asegurar la seguridad ciudadana en el territorio nacional y él es definido por la Policía como un modelo en el cual *«El personal uniformado de los cuadrantes, busca identificar las problemáticas y manifestaciones de violencia y criminalidad que atentan contra la convivencia y seguridad ciudadana en lo local y a partir de allí generar herramientas de corresponsabilidad que permitan mitigar estos fenómenos»* [89].





Figura 20: Mapa del Modelo Nacional de Vigilancia Comunitaria por Cuadrantes (MNVCC), [89].

Dentro del MNVCC, se definen zonas geográficas llamadas cuadrantes las cuales son definidas por la Policía Nacional como «*un sector geográfico*

*fijo que a partir de sus características delictivas, contravencionales, sociales, demográficas, geográficas y económicas recibe distintos tipos de atención de servicio policía» [89], es decir que el territorio de la república de Colombia es de vivido por la policía en zonas geográficas llamas cuadrantes dentro de las cuales se ejercerá el control policial para garantizar la convivencia y la seguridad ciudadana.*



Figura 21: Ejemplo de cuadrantes urbanos reales [87]

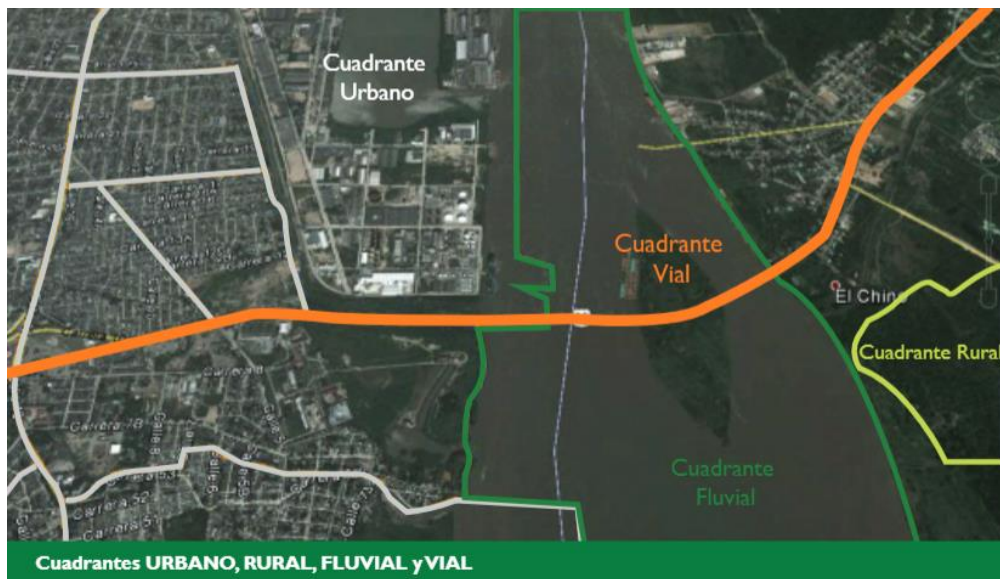


Figura 22: Ejemplo real de cuadrantes rurales, fluvial y viales [87]

El Modelo Nacional de Vigilancia Comunitaria por Cuadrantes (MNVCC), está formando por tres dimensiones fundamentales las cuales estratégica, práctica y operacional. Estas dimensiones están enfocadas en cumplir los objetivos misionales de la Policía Nacional centrados en preservar la seguridad ciudadana y asegurar la convivencia.



Figura 23: Dimensiones del MNVCC [87]

En este marco los centros de comando y control se convierten en la herramienta para planear, preparar y ejecutar las operaciones policiales que permitan cumplir los objetivos estratégicos del MNVCC y su vez la Policía Nacional de Colombia.



Figura 24: Preparación de operaciones policiales rutinarias [87]

### **3.4 Sistemas de Video Vigilancia distribuidos en Colombia**

Para los comandantes operativos en los centros de comando y control de la Policía Nacional de Colombia es fundamental contar con una excelente conciencia situacional (Situational Awareness) [1], para poder planear, preparar y ejecutar operaciones policiales acorde al Modelo Nacional de Vigilancia Comunitaria por Cuadrantes que permita cumplir con los objetivos estratégicos, tácticos y operacionales de la Policía Nacional de Colombia.

Una de las fuentes de información en tiempo real más importantes en los centros de comando y control de la Policía son los sistemas de video vigilancia, pues entregan una visión realista de la situación de seguridad ciudadana.

En Colombia estos sistemas están desplegados en las ciudades y principales carreteras del país, y solo en la ciudad de Bogotá D.C. se cuentan con más 3.300 cámaras instaladas a finales de 2018 [90], lo que mejora en gran medida la conciencia situacional en los centros de comando y control.

Si embargo estos sistemas tecnológicos tienen un punto débil, pues en ocasiones los operadores del centro de comando y control, encargados del sistema de video vigilancia no tienen la capacidad de revisar todas las cámaras simultáneamente dado que en una ciudad como Bogotá D.C. con más de 8 millones de habitantes [91] con más de 3.300 cámaras operativas todas ellas no pueden ser supervisadas simultáneamente lo que hace que

en ocasiones se registren actividades criminales que no son reportadas en tiempo real al C2IS, dado que este proceso se hace manualmente.

Esta situación motiva el desarrollo de la arquitectura de detección de actividades criminales en video aplicado en Sistemas de Mando y Control para seguridad ciudadana presentado en esta tesis doctoral.



## 4 ARQUITECTURA DE DETECCIÓN DE ACTIVIDADES CRIMINALES

La arquitectura propuesta para la detección de actividades criminales deberá tener la capacidad de desplegarse en un sistema de Video Vigilancia de seguridad ciudadana, para esto deberá tener con ciertas características de sencillez y coste computacional, para la cual se propone un sistema genérico de detección y clasificación de actividades criminales descrito a continuación.

### 4.1 Arquitectura genérica de detección y clasificación

La arquitectura genérica de detección y clasificación de actividades criminales propuesta se muestra a continuación:

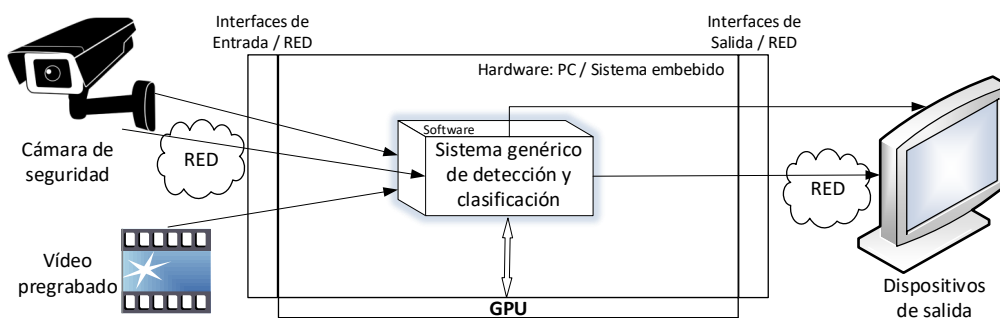


Figura 25: Arquitectura genérica de detección y clasificación de actividades criminales



Esta arquitectura genérica, está compuesta por los siguientes componentes:

- Sistema Genérico de Detección y Clasificación: Sistema de detección a actividades criminales basado en Deep Learning el cual usará procesamiento en paralelo para realizar procesamiento de video en tiempo real.
- Plataforma de Hardware: es necesaria para la ejecución del Sistema Genérico de Detección y Clasificación, esta plataforma cuenta con interfaces de entrada y salida de video, capacidad de procesamiento por CPU y GPU. Esta plataforma de hardware no tiene limitaciones en cuanto al tipo de infraestructura, es decir los recursos de hardware de la plataforma podrían estar en infraestructuras, Cloud, On-premise (infraestructura local) o en sistemas embebidos.
- Fuentes de Video: Dentro de la arquitectura genérica de detección y clasificación, se contempla poder analizar video proveniente de las cámaras del sistema de video vigilancia o de fuentes con video pregrabado.
- Dispositivos de salida: Como componente final del sistema genérico de detección y clasificación plantea mostrar las detecciones de acciones criminales a los operarios del sistema de video vigilancia

de tal forma que con esta información se puedan tomar acciones necesarias en el menor tiempo posible.

Es de tener en cuenta que se plantea que el sistema de detección de actividades y clasificación de actividades criminales es software basado en Deep Learning, éste puede ser ejecutado en distintos tipos de hardware ya sean, servidores, PC de escritorio o sistemas embebidos; sin embargo, el hardware debe contar con una GPU con la potencia suficiente, compatible con CUDA® de nVidia® en la que se puedan ejecutar algoritmos Deep Learning.

Las fuentes de video a utilizar, pueden ser cámaras de seguridad o videos pregrabados, las cuales pueden provenir de cualquier interfaz de entrada de hardware ya sea local como (por ejemplo, puertos USB) o provenir de una red de comunicaciones TCP/IP, una vez el vídeo sea adquirido este será analizado como se muestra en la figura 25, aprovechando el procesamiento en paralelo aportado por la GPU.

Una vez el vídeo sea procesado las detecciones del sistema pueden ser mostradas localmente mediante los dispositivos de salida como el monitor, sin embargo, según el diseño global del sistema las detecciones pueden ser enviadas por las interfaces de red a donde requieran ser visualizadas o procesadas.

Es de gran importancia indicar que cada señal de video a analizar por cada cámara o cada video pregrabado, deberá contar con un sistema individual para la detección de actividades criminales, con sus propios recursos de hardware y software para realizar el procesamiento del video, y para un posible despliegue se tendría una arquitectura como la mostrada a continuación.

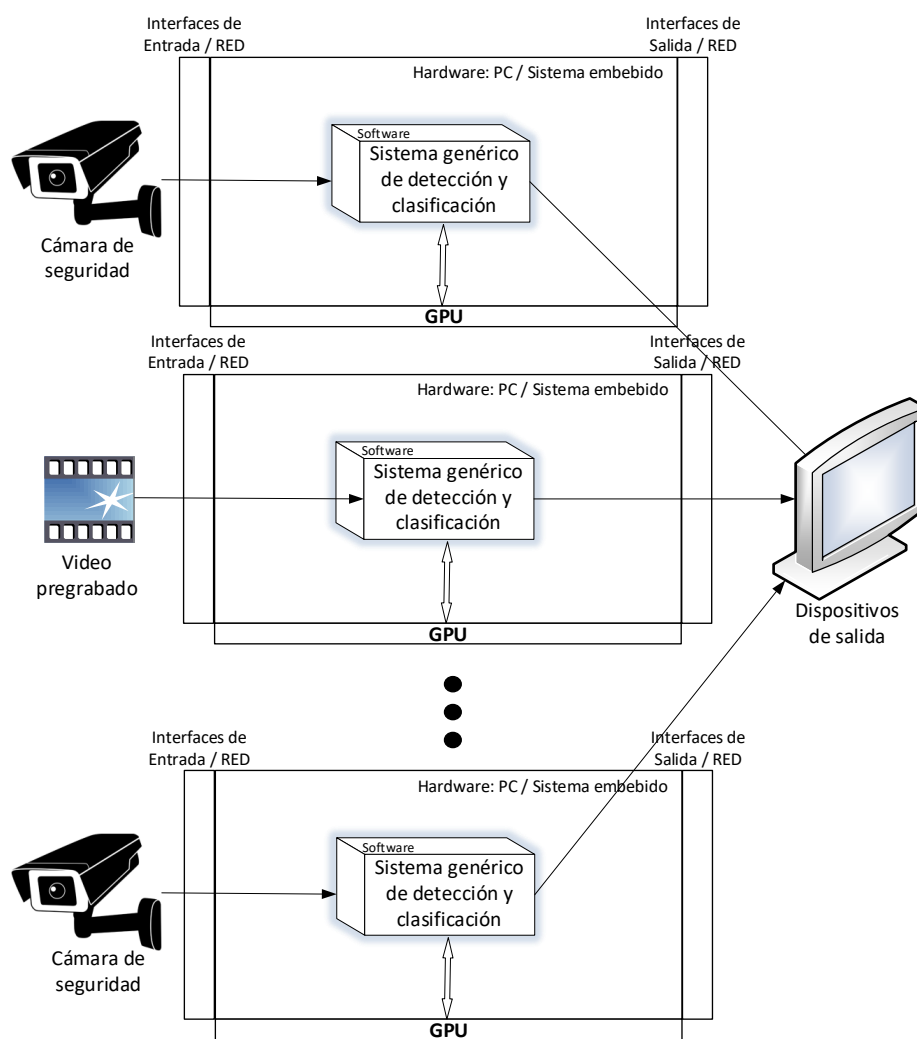


Figura 26: Arquitectura genérica de detección de actividades criminales con varias señales de video

Una vez mostrada la arquitectura genérica de detección y clasificación de actividades criminales, se procederá a plantear el sistema de detección y clasificación de actividades criminales, el cual utilizarán técnicas de Deep Learning según lo revisado en el capítulo 2.

El diseño y entrenamiento del sistema de detección y clasificación basado en Deep Learning se explicará de la siguiente manera en los numerales **4.2 (Detección de armas de fuego)**, **4.3 (Detección de hurtos)** y **4.4 (Sistema completo de detección de actividades criminales)**.

#### **4.2 Detector de armas de fuego cortas (revólveres y pistolas)**

Para adaptar las técnicas de Deep Learning al problema de la detección de actividades criminales, se tomará el enfoque de la detección de objetos usando técnicas por regiones analizadas en capítulos anteriores.

Dado que la presencia de armas de fuego en eventos delictivos es un acto que debe tener atención inmediata de las autoridades, la primera fase del proceso de esta tesis doctoral es realizar un detector de armas de fuego cortas utilizando la técnica R-CNN.

Para este detector se utilizará la arquitectura propuesta por Ross Girshick en [31] la cual será implementada en Matlab 2019b, usando un computador portátil Intel Core I7 7700HQ, 16 GB de RAM DDR4, con una GPU NVIDIA Geforce 1070 GTX en formato MXM la cual consta de 2048 núcleos CUDA a 1442 MHz de frecuencia base y 8 GB de memoria VRAM DDR5 [92].

El entrenamiento de detector de armas consta de tres fases las cuales serán descritas a continuación:

#### **4.2.1 Fase 1: Red neuronal convolucional como clasificador de imágenes**

Como se mostró en el capítulo anterior las redes neuronales convolucionales tienen un gran desempeño como clasificadores de imágenes [9] sin embargo, para tener una buena tasa de clasificación estos modelos como GoogLeNet [26] o ResNet [27] tienen una complejidad de cientos de capas lo que implica un consumo computacional elevado, por esta razón y al realizar pruebas preliminares con varios modelos la implementación se hizo con la CNN «AlexNet» [20], la cual tiene la capacidad de clasificar mil clases de imágenes diferentes. El proceso de entrenamiento de

AlexNet [20] fue realizado con cerca de un millón doscientas mil imágenes pertenecientes al famoso *Dataset* ImageNet [93].

Por estas razones se usó AlexNet dada su menor complejidad que modelos como VGG16, VGG19 [25], GoogleNet [26] o ResNet [27] lo que tiene implicaciones en el costo computacional y el uso de VRAM en la GPU que en el caso de la GTX 1070 MXM es de 8 GB.

La arquitectura de la red Neuronal Convolutiva “AlexNet” usada en se muestra páginas atrás en la figura 5, y el proceso de entrenamiento se ilustra en la figura 25.

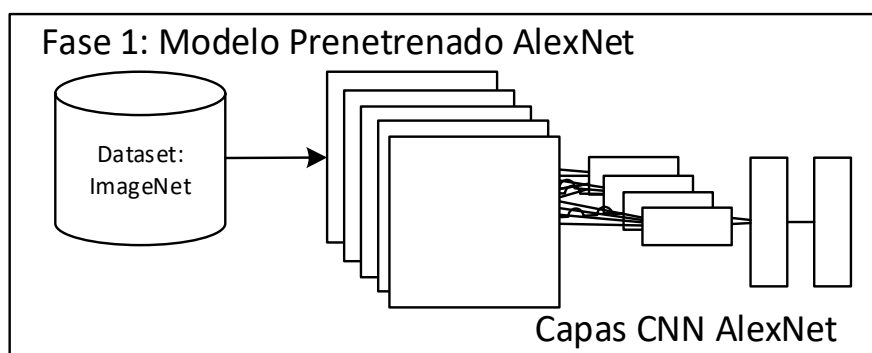


Figura 27: Modelo Pre-entrenado AlexNet.

#### 4.2.2 Fase 2: reentrenamiento fino de AlexNet para la detección de armas de fuego cortas

Una vez se cuenta con el modelo pre-entrenado AlexNet se procede a realizar un reentrenamiento fino del mismo, sin embargo, en esta fase no se usa el Dataset ImageNet [93] sino que se creó un Dataset único para el desarrollo de esta tesis doctoral

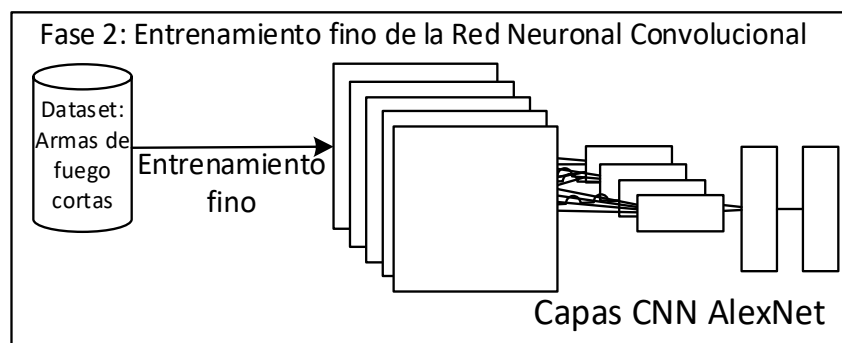


Figura 28: Entrenamiento fino de la Neuronal Convolutiva AlexNet

Este *Dataset* fue construido con diversas imágenes de armas cortas, empuñadas por diversos usuarios, en diferentes ángulos, armas sin ser empuñadas y en diversos escenarios. En estas imágenes las armas cortas fueron etiquetadas manualmente por el autor de la tesis y está compuesta por una pistola calibre 9 milímetros y un revolver calibre 38L *COLT Detective Special* [94], ya que son comúnmente usadas en acciones criminales.

Una vez el Dataset fue construido y etiquetado se procede a realizar el reentrenamiento fino de AlexNet, analizado el *Dataset* de armas cortas un total de 100 épocas, como se ve en la figura 26.

Después de realizar varios ajustes, los parámetros de entrenamiento se ajustaron para el uso de los 8 GB de VRAM del GPU usando una tasa inicial de aprendizaje de 0.001, con un factor de caída de tasa de aprendizaje 0.1 por cada época.



Figura 29: Muestra de imágenes del Dataset de reentrenamiento fino

Este reentrenamiento fino tomó alrededor de 16 horas de procesamiento. Este tiempo puede ser reducido usando esquemas Multi-GPU, o un clúster de varios equipos que cuenten GPU.



### 4.2.3 Fase 3: Entrenamiento del predictor lineal binario para la detección de armas de fuego cortas

En la última fase de entrenamiento del detector de armas cortas se procede a entrenar un Clasificador Binario el cual usa una Máquina de Vector Soporte (SVM por su sigla en inglés) con *kernel* lineal [10] [95] [96] cuyo objeto es clasificar dos clases, “arma” y “no arma”.

El entrenamiento de una máquina de vector soporte usada como clasificador binario debe contar con dos clases de datos para el entrenamiento, en este caso las clases son “arma” y “no arma”, como datos de entrenamiento de la clase “arma” se usan las características de las imágenes del *Dataset* construido de armas cortas, obtenidas como resultado del reentrenamiento fino de AlexNet (fase 2) y de las propuestas de regiones de interés etiquetadas manualmente en *Dataset* de entrenamiento fino como se muestra en la figura 28. Para la clase “no arma” se usan trozos de imágenes extraídas de *Dataset* de armas cortas, diferentes a las etiquetadas manualmente y a las obtenidas en la fase 2 del proceso de entrenamiento de la R-CNN.

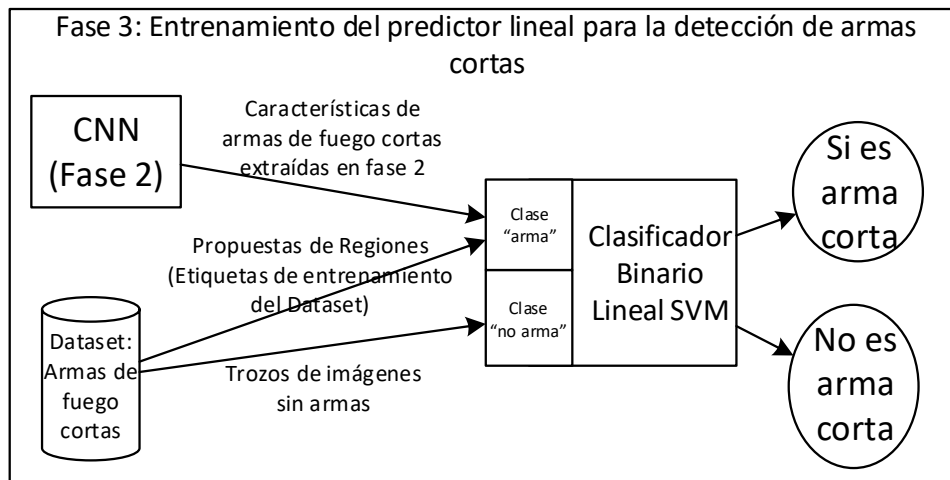


Figura 30: Arquitectura del entrenamiento del predictor lineal de armas cortas

De esta manera se entrenó un detector de armas de fuego cortas basado en Deep Learning usado R-CNN, en la figura 9 se pueden observar detecciones positivas realizadas por el detector.

### 4.3 Detección de hurtos

Una vez implementado un detector de armas cortas funcional, se procede a desarrollar un detector de hurtos callejeros y a medir el tiempo de procesamiento por imagen de AlexNet contra VGG16 con el fin de desarrollar un modelo que pueda ser usando en video de tiempo real y que pueda ser implementado en un sistema de mando y control de seguridad ciudadana.

Para desarrollar el detector de hurtos callejero esta tesis doctoral parte de una premisa novedosa en la cual a la fecha de escritura de la misma no se han encontrado publicaciones al respecto, la cual es tomar los hurtos como si fueran objetos a detectar, partiendo de la premisa que los hurtos callejeros tienen características únicas las cuales pueden ser identificadas por medio de Deep Learning.

De esta manera se logró hacer un detector de hurtos que funciona cuadro a cuadro, lo que hace este detector indiferente ante movimientos bruscos de cámara y no necesitaría hacer análisis de trayectorias ni identificación de individuos.

Teniendo en cuenta esta premisa novedosa de tomar los hurtos callejeros como objetos a detectar, se entrena una R-CNN para la detección de hurtos, sin embargo, en esta oportunidad se usarán dos modelos VGG16 y AlexNet con el objeto de medir que tanto impacta la complejidad del modelo usado al costo computacional, esto con el objeto de identificar de una mejor manera el modelo a usar en la aplicación final.

Como se mostró en el numeral anterior el entrenamiento de una R-CNN está compuesto de tres fases, las cuales se describirán a continuación.

### 4.3.1 Fase 1: Modelos pre-entrenados como clasificador de imágenes

En esta fase se analizarán los modelos AlexNet, VGG16 y VGG19 pues como se ve en las figuras 6 y 7 tienen diferencias con respecto a su complejidad. Estos dos modelos han sido entrenados con cerca de un millón doscientas mil imágenes divididas en mil clases pertenecientes al Dataset ImageNet [63].

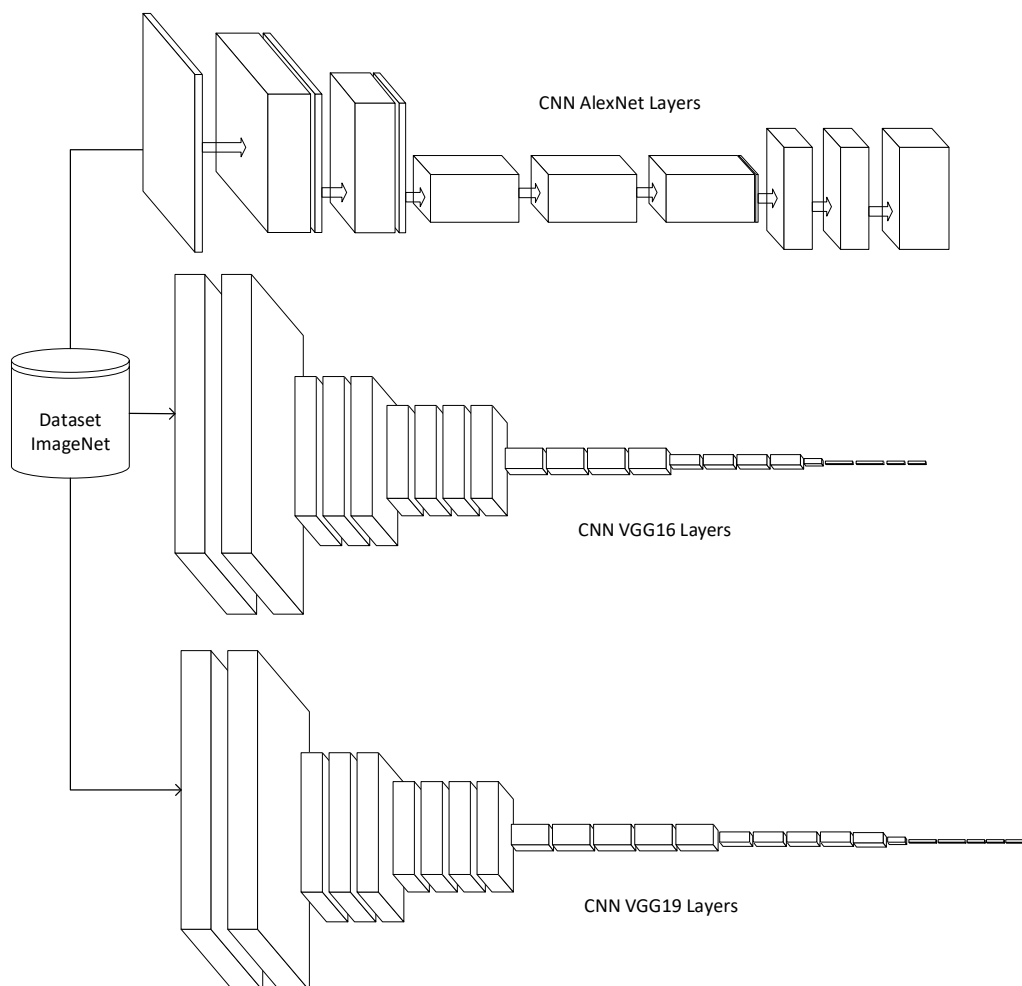


Figura 31: Entrenamiento AlexNet, VGG16 y VGG19

#### 4.3.2 Fase 2: Rentrenamiento fino de AlexNet, VGG16 y VGG19 para la detección de hurtos callejeros

Para esta fase también se debió construir un Dataset de hurtos callejeros, esa Dataset fue realizado con la ayuda de la Oficina de Telemática de la Policía Nacional de Colombia, quienes facilitaron bajo acuerdo de confidencialidad videos de seguridad ciudadana obtenidos en varias ciudades de Colombia. De algunos de estos se extrajeron una serie de imágenes en donde exista hurto callejero, las cuales fueron etiquetadas manualmente para construir el Dataset de hurtos callejeros.



Figura 32: Muestra de imágenes del Dataset de hurtos para reentrenamiento fino.

Una vez construido el Dataset se procede a realizar el reentrenamiento fino de los modelos AlexNet, VGG16 y VGG19, observando que el consumo de VRAM en el GPU para VGG16 y VGG19 usando los mismos parámetros de entrenamiento que AlexNet.

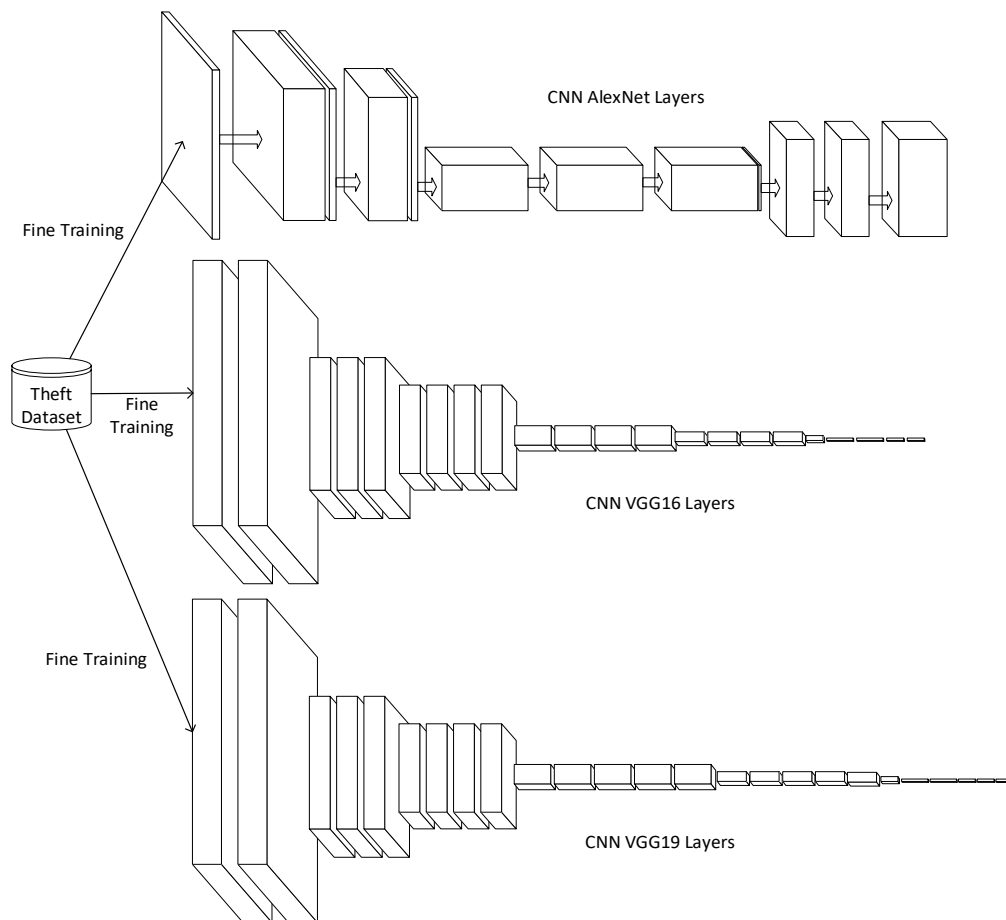


Figura 33: Reentrenamiento fino de AlexNet, VGG16 y VGG19

### 4.3.3 Fase 3: Entrenamiento del clasificador lineal binario para la detección de hurto callejero

Como fase final de entrenamiento del detector de hurto callejero cortas se procede a entrenar un Clasificador Binario el cual usa una Máquina de Vector Soporte (SVM por su sigla en inglés) con *kernel* lineal [10] [95] [96] cuyo objeto es clasificar dos clases, “Hurto” y “No Hurto”.

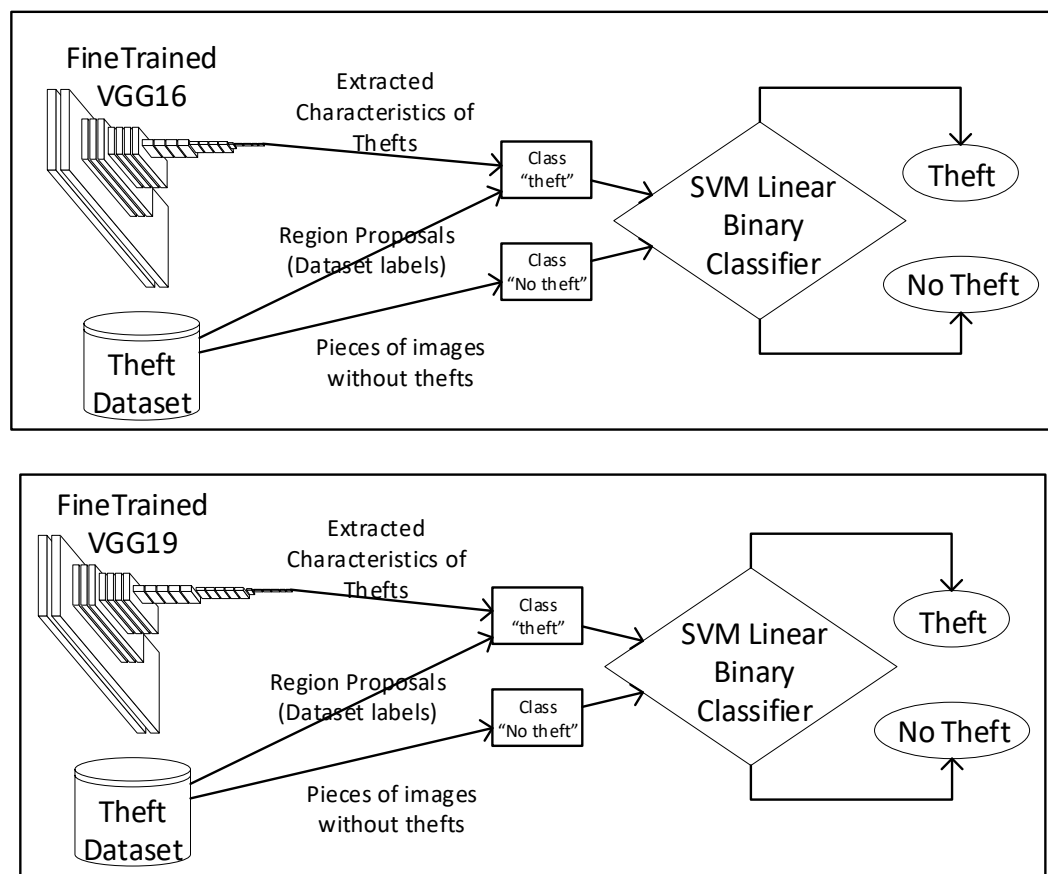


Figura 34: Arquitectura del entrenamiento del predictor lineal de hurtos con modelos VGG16 y VGG19

El entrenamiento de una máquina de vector soporte usada como clasificador binario debe contar con dos clases de datos para el entrenamiento, en este caso las clases son «Hurto» y «No Hurto», como datos de entrenamiento de la clase «Hurto» se usan las características de las imágenes del *Dataset* de Hurtos construido con el apoyo de la Policía Nacional de Colombia, obtenidas como resultado del reentrenamiento fino de AlexNet, VGG16 y VGG19 y de las propuestas de regiones de interés etiquetadas manualmente en *Dataset* de entrenamiento fino como se muestra en la figura 33 para VGG16, VGG19 y 20 para AlexNet. Para la clase «No Hurto» se usan trozos de imágenes extraídas de *Dataset* de Hurtos, diferentes a las etiquetadas manualmente y a las obtenidas en la fase 2 del proceso de entrenamiento de la R-CNN, tal como se ve en las figuras 34 y 35.

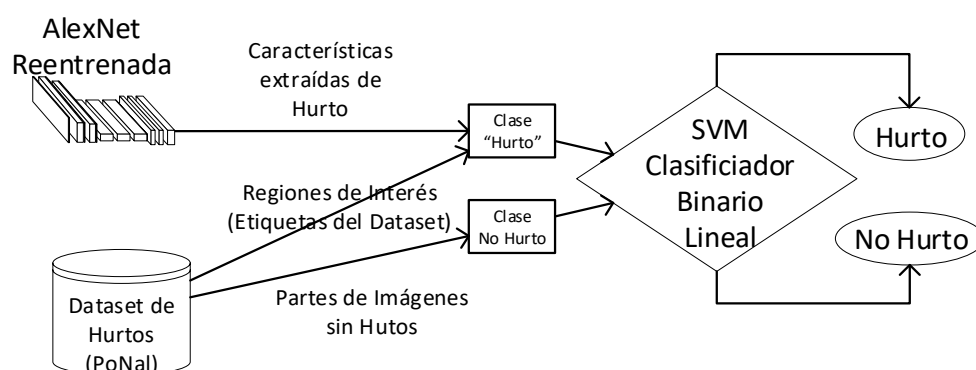


Figura 35: Arquitectura del entrenamiento del predictor lineal de hurtos con modelo AlexNet



Los resultados de detector de hurtos callejeros R-CNN son positivos, pues en las pruebas realizadas logan un margen de detección de 78% usando el modelo VGG19 y del 74% usando AlexNet, sin embargo, el tiempo de procesamiento de imágenes es muy diferente entre los dos modelos, pues el detector de hurtos basado en AlexNet tiene un tiempo de procesamiento de imagen promedio de 2.3 segundos en comparación con VGG16 que cuenta con un tiempo procesamiento promedio de imagen de 23.8 segundos.

Este resultado es concluyente pues el detector basado en AlexNet tiene un tiempo de procesamiento 10 veces menor al detector basado en VGG16, y dado que la única diferencia entre los dos detectores es la complejidad del modelo se concluye que la complejidad del modelo tiene un impacto directo en el costo computacional y en el tiempo de procesamiento de cada imagen.



Figura 36: Resultados positivos de la detección de hurtos callejeros.

#### 4.4 Sistema completo de detección de actividad criminales

En visión artificial existen técnicas para la detección de violencia y crimines realizando análisis de video [97] [98] [99] [100], sin embargo su aplicación en el caso específico de los sistemas de mando y control para seguridad ciudadana, no está generalizada pues la gran movilidad de las escenas de video, presenta un gran reto para el procesamiento de este tipo de video.

Después de confirmar que es posible detectar hurtos con técnicas R-CNN y de identificar cual es la relación entre la complejidad del modelo CNN utilizado y el costo computacional, se procede a diseñar un modelo de detección de actividades criminales que puedas procesar imágenes en un

tiempo lo suficientemente corto como para ser aplicado en secuencias de video.

Para la detección de actividades criminales se utilizará la técnica Faster R-CNN [38], que como se ha mostrado antes constituye el estado del arte en la detección de objetos, además es una mejora de R-CNN y de Fast R-CNN.

Por otro lado Faster R-CNN ha mostrado buen desempeño en la detección de objetos en secuencias de video [101] [102] [103] [104], siendo hasta 25 veces más rápido que Fast R-CNN y 250 veces más rápido que R-CNN [105] lo que es de gran relevancia para su aplicación en sistemas de mando y control de seguridad ciudadana.

La arquitectura de detección de actividades criminales propuesta en esta tesis doctoral estará centrada en un detector Faster R-CNN [38], la cual se entrenará para la detección de objetos usados común mente en actividades criminales, los cuales serán armas blancas y armas cortas de fuego. Como aspecto novedoso se entrenará para la detección de hurtos callejeros como si fueran objetos dentro del detector, y para beneficiar el rendimiento del sistema se usará el modelo de CNN AlexNet.

#### **4.4.1 Arquitectura detallada del sistema de detector de actividades criminales**

Como se mencionó anteriormente, la arquitectura usada para el detector de actividades criminales será Faster Region-Based Convolutional Network (Faster R-CNN), la cual está compuesta por dos partes principales: la region proposal network (RPN) y la Fast R-CNN [38], las cuales serán descritas a continuación:

##### **Region Proposal Network**

El propósito de la RPN [38] es identificar regiones de interés en cada cuadro de video, esta red usa AlexNet y el mapa de características extraído de dicha CNN para calcular las posibles regiones de interés de cada imagen de video, sin embargo por cómo se mostró en el numeral anterior se usa AlexNet en cambio de VGG19 con el objeto de reducir el costo computacional a diferencia de lo realizado en [38]. La RPN está enfocada en identificar regiones las cuales tengan armas blancas, armas de fuego cortas y hurtos callejeros los cuales serán tratados como objetos por identificar.

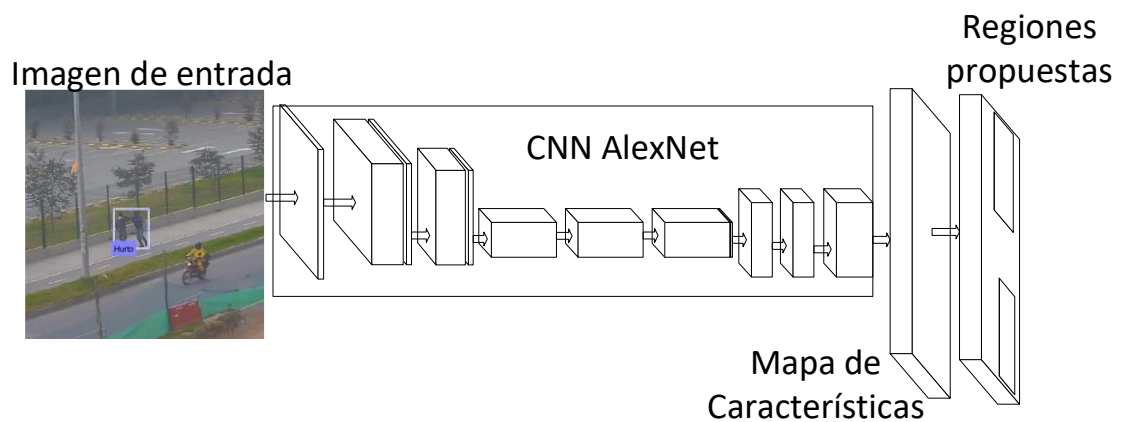


Figura 37: Region Proposal Network (RPN)

### Fast Region-Based Convolutional Network

La segunda parte del detector de actividades criminales está compuesta por una Fast R-CNN [32], que actúa como un detector de objetos, la cual usa las regiones propuestas por la RPN y también utiliza la CNN AlexNet [21] como modelo base.

Esta Fast R-CNN [32] será entrenada para la detección de armas blancas, armas cortas de fuego y hurtos callejeros.

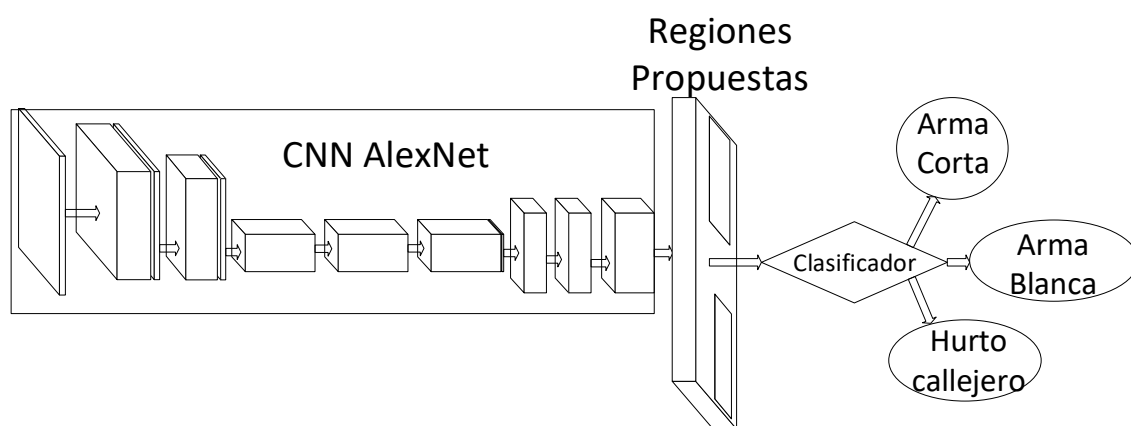


Figura 38: Arquitectura del detector de actividades criminales

### **Faster Region-Based Convolutional Network de detección de actividades criminales**

La arquitectura completa del detector de actividades criminales se basa en Faster R-CNN por lo que usa tanto la RPN y la Fast R-CNN descritas anteriormente de como se muestra en la figura 37, como se muestra en [38], y cambiando el modelo CNN VGG19 por AlexNet según las mediciones mostradas anteriormente, para esta aplicación específica Faster R-CNN tendría la capacidad de analizar video en tiempo real.

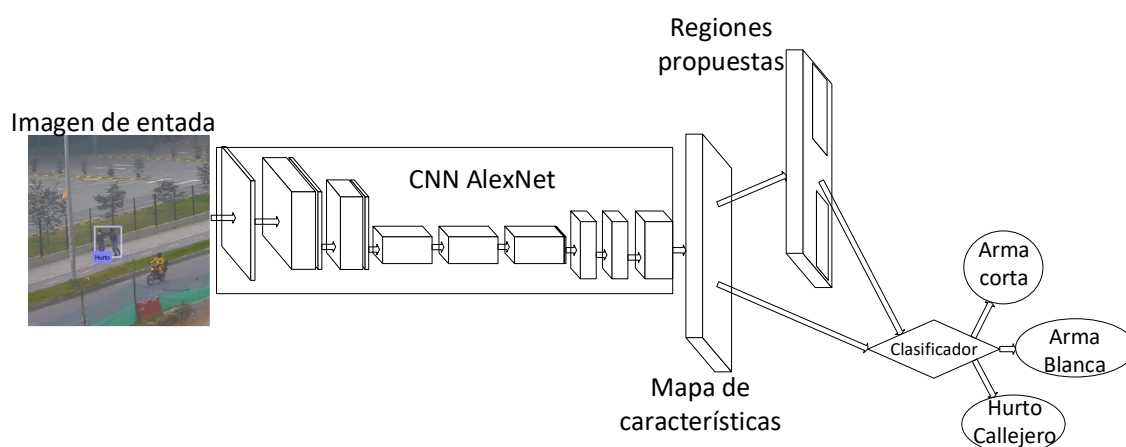


Figura 39: Arquitectura detallada del detector de actividades criminales en video

#### 4.4.2 Proceso de entrenamiento del detector de actividades criminales

Una vez descrita la arquitectura que se utilizará para el detector de actividades criminales en video, se procede a describir el proceso de entrenamiento, al cual se realizó usando Matlab 2019b en un computador con una GPU NVIDIA GTX 1070 [106] usando Windows 10. Este proceso de entrenamiento tiene cuatro fases las cuales serán descritas a continuación:

#### 4.4.3 Fase 1: Entrenamiento de la RPN usando en Dataset específico

En esta fase del entrenamiento, el modelo AlexNet es reentrenado dentro de la RPN, usando un Dataset imágenes creado específicamente para el desarrollo de esta tesis. Este Dataset fue creado con videos proveniente del sistema de video vigilancia ciudadana de la Policía Nacional de Colombia.

Este Dataset tiene tres clases de interés para ser identificadas y reportadas al sistema de información de mando y control las cuales son: armas blancas, armas de fuero cortas y hurtos callejeros.

Como resultado de esta fase de entrenamiento se obtuvo un mapa de características de las tres clases mencionadas con el cual la RPN es capaz de identificar las regiones de interés con posibles armas blancas, armas de fuero cortas y hurtos callejeros en las imágenes a ser analizadas.

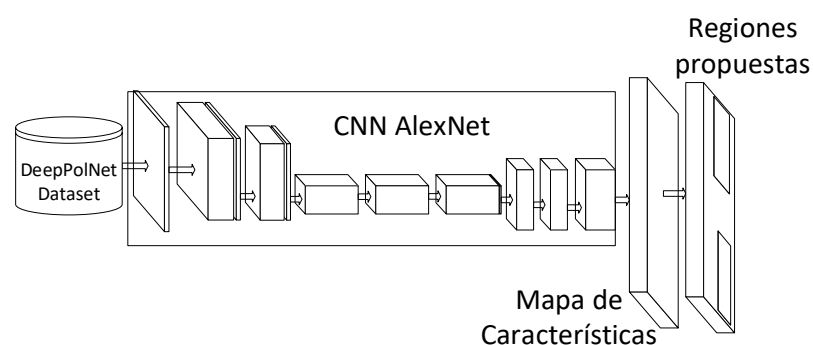


Figura 40: Fase 1: entrenamiento de la RPN



#### 4.4.4 Fase 2: Entrenamiento de una Fast R-CNN como detector de objetos.

En esta segunda fase de entrenamiento, se procedió a entrenar un detector de objetos basado en Fast R-CNN. Este detector de objetos usa AlexNet como modelo base, usando el Dataset específico para realizar el entrenamiento, además de las propuestas de regiones de interés obtenidas de la primera fase de entrenamiento para la detección de armas blancas, armas de fuego y hurtos callejeros.

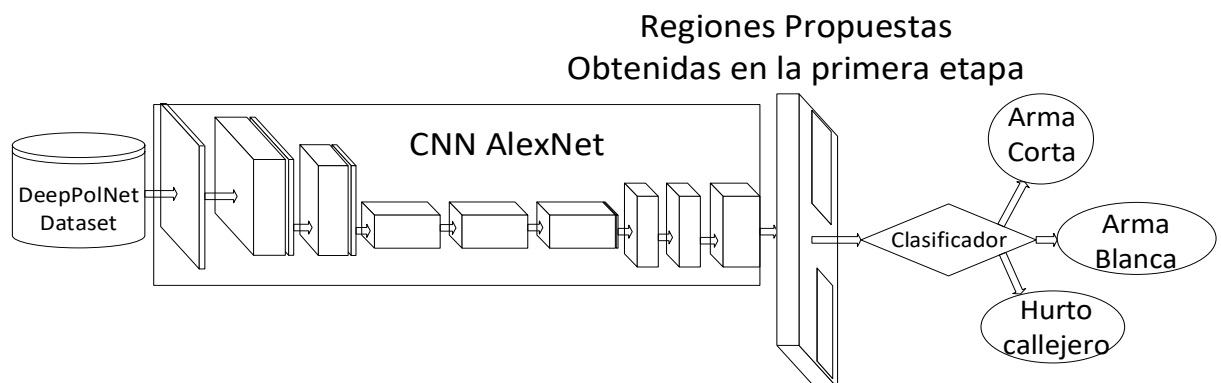


Figura 41: Fase 1: entrenamiento del detector de objetos Fast R-CNN

#### 4.4.5 Fase 3: Reentrenamiento fino de la RPN

Con el objetivo de mejorar el desempeño de la RPN, en la tercera fase se procede a realizar un entrenamiento fino de la misma, pero en esta oportunidad se incluirán los pesos de las neuronas

artificiales de la CNN AlexNet obtenidos en la segunda fase de entrenamiento como se muestra en la figura 40.

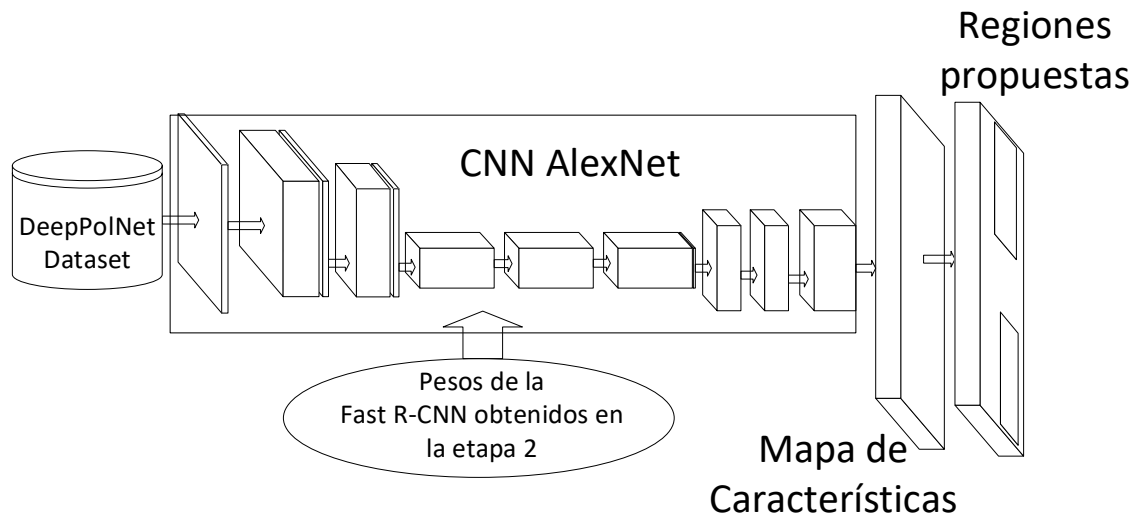


Figura 42: Fase 3: entrenamiento fino de la RPN

#### 4.4.6 Fase 4: Entrenamiento fino de la Fast R-CNN

Como fase final del entrenamiento se procede a realizar un entrenamiento fino de la Fast R-CNN usando la RPN entrenada en la tercera fase como se ve en la figura 41, eso con el objeto de mejorar el rendimiento de la Fast R-CNN.

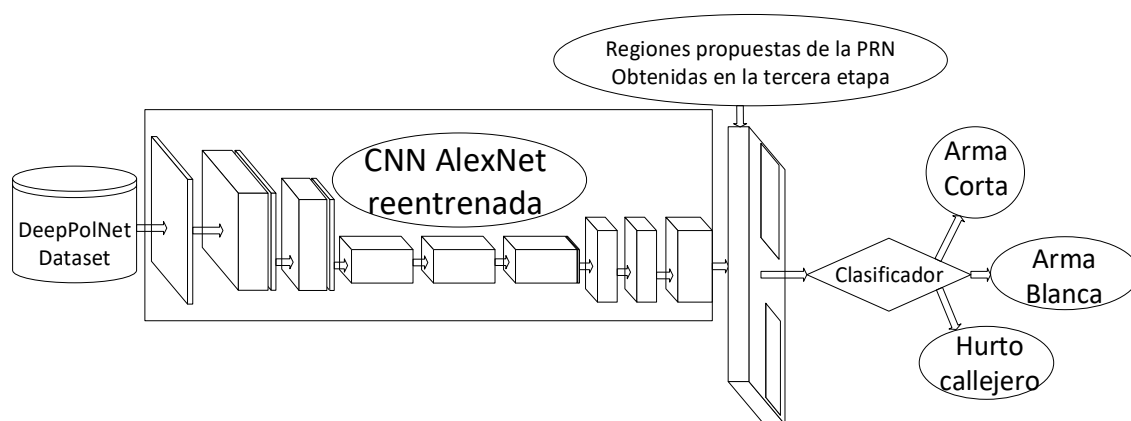


Figura 43: Fase 4: Entrenamiento fino del detecto de objetos Fast R-CNN

#### 4.5 Resultados de la detección de actividades criminales analizar video de tiempo real

Durante el desarrollo y entrenamiento del detector de actividades criminales en video se realizaron una serie de pruebas en el computador portátil MSI GT62VR-7RE [107] con un Intel Core I7 7700HQ [108], 16 GB de RAM tipo DDR4, con una GPU NVIDIA GeForce GTX 1070 [106] con 8 GB DDR5 VRAM.

Las pruebas realizadas con videos de tiempo real fueron realizadas con dos fuentes de video, las cuales fueron videos obtenidos del sistema de video vigilancia ciudadana de la Policía Nacional de Colombia.

En estos escenarios se obtuvieron muy buenos resultados con respecto al tiempo de procesamiento de cada cuadro de vídeo el cual estuvo entre 0.03 segundos y 0.05 segundos con lo cual se puede procesar video en

tiempo real con 20 a 33 cuadros por segundo, lo que es suficiente para la mayoría de los sistemas de mando y control de seguridad ciudadana pues, por ejemplo, las cámaras de la Policía Nacional de Colombia generan 30 imágenes por segundo.

Con respecto a la tasa de aciertos de detección, el detector siempre deberá contar con supervisión humana pues debido a situaciones comunes como obstrucciones de imagen, cambios de iluminación, árboles y cableado eléctrico el sistema puede presentar fallos en la detección.

Sin embargo, en las pruebas realizadas con el sistema tiene una tasa de acierto promedio del 69 % con un tiempo de procesamiento de 0.03 segundos obteniendo una tasa de 33 cuadros por segundo.

#### **4.6 Arquitectura final de detección de actividades criminales basada en análisis de vídeo en tiempo real**

Una vez entrenado el sistema basado en Faster R-CNN, según las fases descritas anteriormente, la arquitectura genérica para la detección de actividades criminales aculará de la siguiente manera:

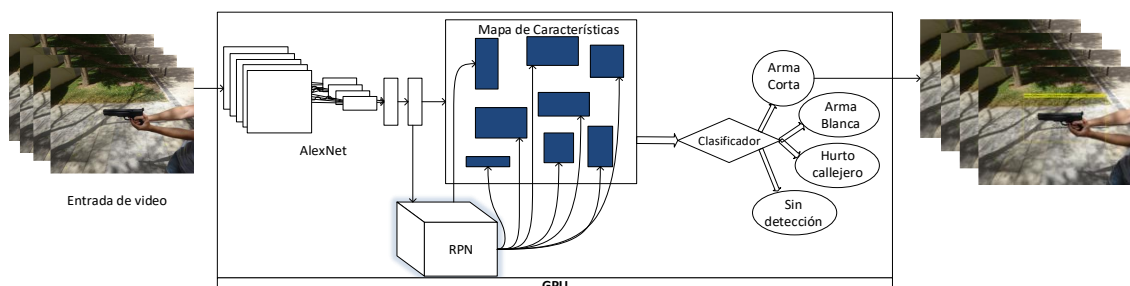


Figura 44: Funcionamiento del sistema genérico de detección de actividades criminales

Entrada de video: al sistema debe contar con una entrada de video el cual puede provenir de una cámara o de un video pregrabado, para esta carga de video se utilizan toolbox de Matlab para el procesamiento de video las cuales se encargan del control de las interfaces de hardware y de la decodificación del video.

Procesamiento de video (cuadro a cuadro): para procesar el video el sistema procede a separarlo en una serie de cuadros (*frames*) los cuales serán analizados independientemente usando el procesamiento en paralelo de la GPU del sistema, sin realizar ningún análisis de correlación, filtrado o predicción entre cuadros.

Cada imagen proveniente del video es analizada por la Faster R-CNN que usa como base AlexNET y fue entrenada como se indica anteriormente, como resultado de este análisis en cada imagen se realizará la detección de armas blancas, armas cortas, hurtos callejeros o detección negativa (sin detección) según sea el caso.

Visualización de detecciones: como fase final del proceso, el sistema de detección dibuja un recuadro en cual encierra los resultados de las detecciones positivas e indica que tipo de detección es, este resultado es mostrado en pantalla para que el operador humano pueda verificar la detección realizada por el sistema.



## **5 RENDIMIENTO DEL SISTEMA GENÉRICO DE DETECCIÓN Y CLASIFICACIÓN DE ACCIONES CRIMINALES**

Para considerar un despliegue a gran escala del sistema propuesto anteriormente, es necesario considerar el coste computacional del sistema genérico de detección y clasificación propuesta en el capítulo 4.5 y 4.6, para lo cual se procederá a comparar el coste computacional del sistema, usando los modelos de CNN revisados anteriormente VGG-16, VGG-19 y AlexNet, además de ser comparado con otro modelo para la detección de acciones como Tube Convolutional Neural Network, publicado en [109], en el cual se propone un sistema de detecciones de acciones basado en el procesamiento de múltiples imágenes para la detección de la acción.

En [109] el modelo Tube Convolutional Neural Network es implementado y probado usando Hardware con características similares al usado en este trabajo lo que permite hacer comparaciones de rendimiento, las cuales se resumen en la siguiente tabla:



**Tabla 1.** Comparación del Coste Computacional

<b>Modelo</b>	<b>Tiempo de procesamiento promedio</b>	<b>GPU</b>	<b>Desempeño de la GPU (Float 32)</b>	<b>Resolución (Píxeles)</b>
Faster R-CNN (AlexNET)	0.03 Segundos	Nvidia GTX 1070 MXM	6.738 TFLOPS	704 x 544
Faster R-CNN (VGG-16)	0.23 Segundos	Nvidia GTX 1070 MXM	6.738 TFLOPS	704 x 544
Faster R-CNN (VGG-19)	0.28 Segundos	Nvidia GTX 1070 MXM	6.738 TFLOPS	704 x 544
T-CNN	0.9 Segundos	Nvidia GTX Titan X	6.691 TFLOPS	300 x 400

Como se ve en la tabla anterior el Modelo propuesto basado en AlexNet tiene ventajas con respecto al tiempo de procesamiento de cada imagen individual, lo que tiene gran importancia a la hora de un despliegue a gran escala.

Como se mostró en el capítulo 2, el sistema de sistema genérico de detección y clasificación de acciones criminales está centrado en un posible despliegue en los sistemas de comando y control de seguridad ciudadana la Policía Nacional de Colombia, para lo cual se debe pensar que el sistema debería se desplegado a gran escala.

Teniendo en cuenta esto, se realiza un análisis de coste computacional en un escenario de despliegue a gran escala en una ciudad como Bogotá D.C. la cual en 2019 cuenta con alrededor de 2880 cámaras tipo Domo PZT instaladas en lugares estratégicos de la ciudad.

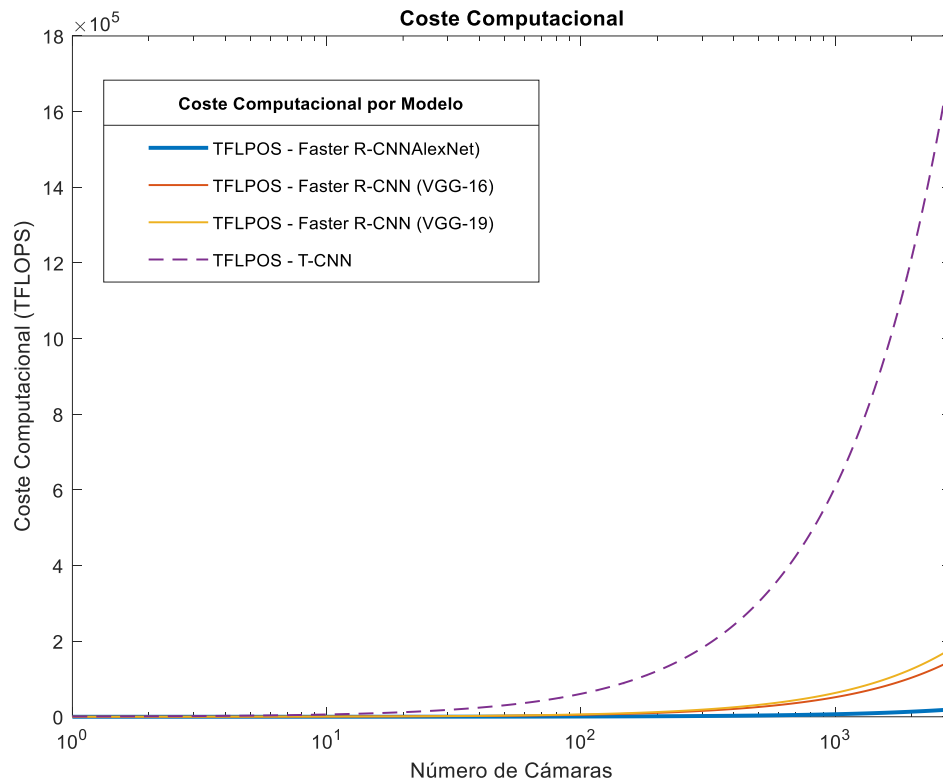


Figura 45: Comparación del coste computacional en función del número de cámaras desplegadas.

En la figura anterior, se muestra una comparación del coste computacional del sistema genérico de detección y clasificación de acciones criminales, usando AlexNet, VGG-16, VGG-19 y además es comparado con T-CNN, para un hipotético despliegue con 2880 cámaras, como se ve en la gráfica el costo computacional del sistema propuesto usando AlexNet es mucho más bajo que las opciones con VGG-16, VGG-19 y T-CNN lo que facilitaría su despliegue a gran escala en el sistemas de comando y control de seguridad ciudadana la Policía Nacional de Colombia.

Otro análisis importante para realzar es el coste económico que conllevaría un el procesamiento del video producido por las cámaras de video vigilancia ciudadana en un posible despliegue masivo.

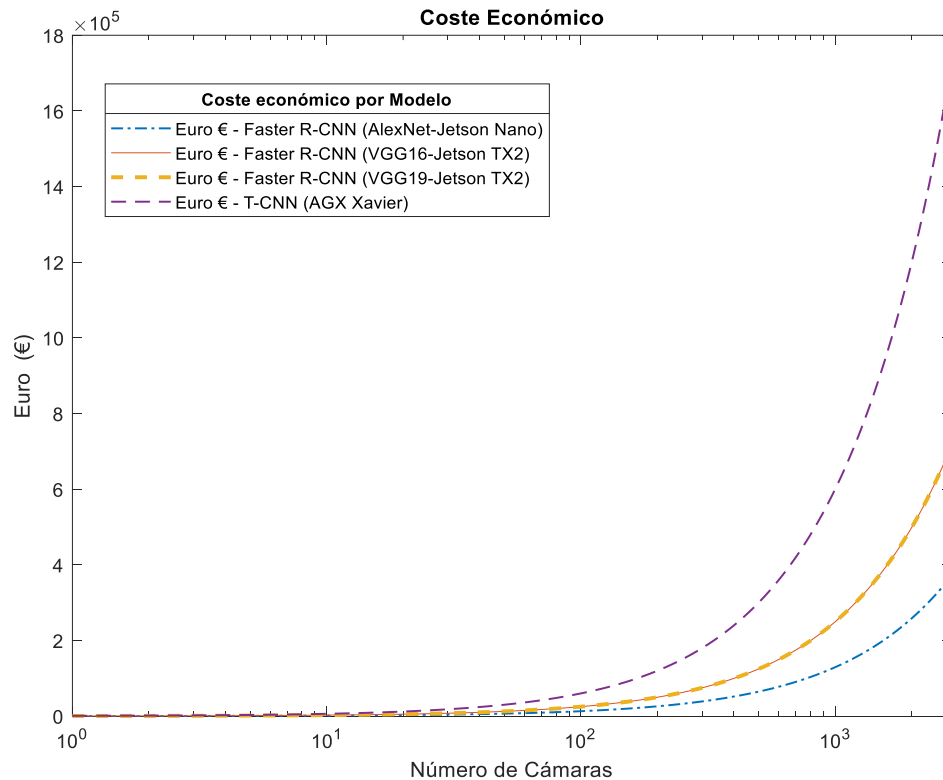


Figura 46: Comparación del coste económico en función del número de cámaras desplegadas.

En la figura anterior se muestra un cálculo del costo aproximado del despliegue masivo del despliegue a gran escala del sistema genérico de detección y clasificación de acciones criminales en el sistema de comando y control de seguridad ciudadana la Policía Nacional de Colombia, este cálculo se realizó basado en la premisa de usar sistemas embebidos Nvidia, según el coste computacional de cada modelo implementado (este

arquitectura de despliegue será analizada con más profundidad en el capítulo 6). Como se puede observar el coste económico está directamente relacionado con el coste computacional y la opción con mejor rendimiento económico es desplegar el sistema genérico de detección y clasificación de acciones criminales con AlexNet según lo propuesto en el capítulo anterior.

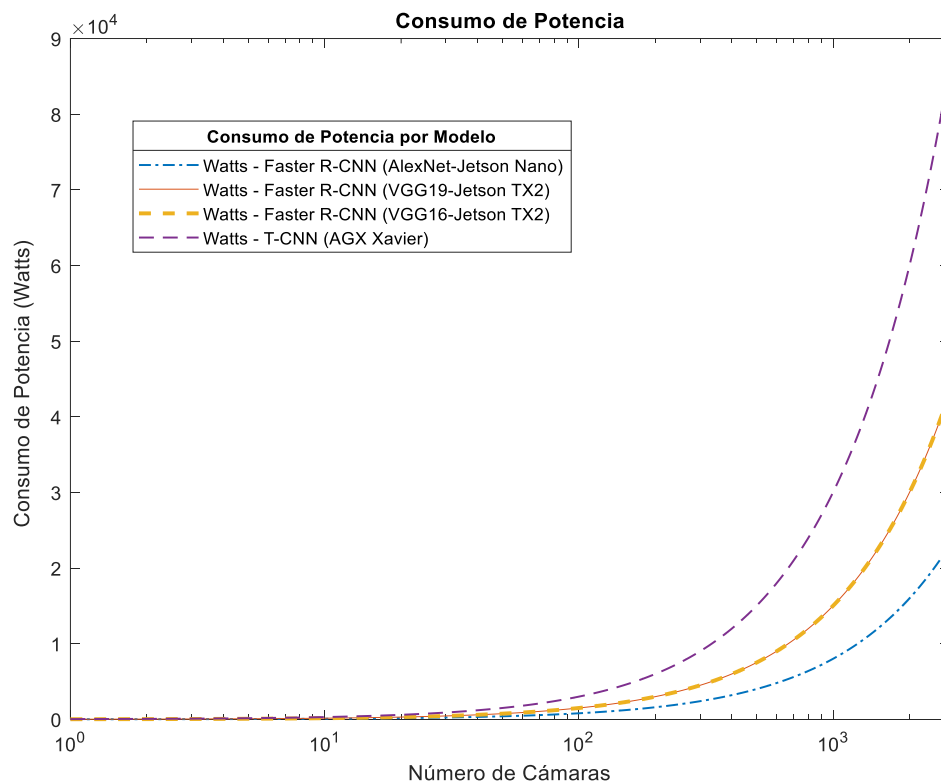


Figura 47: Comparación del Consumo Energético en función del número de cámaras desplegadas.

Otro parámetro para tener en cuenta con para un posible despliegue es el consumo eléctrico asociado al procesamiento de video en el sistema genérico de detección y clasificación de acciones criminales, al igual que

antes se asume un despliegue del sistema usando sistemas embebidos con capacidad de procesamiento en paralelo mediante GPU Nvidia.

Como se muestra en la gráfica anterior la potencia eléctrica requerida por el modelo utilizado está directamente relacionada con el coste computacional del mismo y en este caso también se muestra que el consumo de energético es menor al usar AlexNet en el sistema genérico de detección y clasificación de acciones criminales, lo que haría más factible, económica y eficiente energéticamente un despliegue a gran escala del sistema propuesto.

## **6 PROPUESTA DE APLICACIÓN DE LA ARQUITECTURA**

Una vez definida la arquitectura genérica de detección y clasificación de actividades criminales se proponen dos casos de uso para la aplicación en los sistemas de comando y control de seguridad ciudadana la Policía Nacional de Colombia.

En seguida se propondrán dos distintos tipos de casos de implementación, que serían: procesamiento distribuido de video por medio de sistemas embebidos, procesamiento centralizado de video por medio en un centro de procesamiento de datos (CPD).

### **6.1 Sistemas embebidos**

En el primer caso de uso, se plantea implementar la arquitectura de software genérica de detección y clasificación de actividades criminales propuesta en el capítulo 4 en sistemas embebidos que cuenten con capacidad de procesamiento en paralelo soportada por GPU.

### 6.1.1 Particularización de la solución genérica en sistemas embebidos

Dada la limitada potencia de cómputo con la que cuenta un sistema embebido para la implementación en este tipo de plataformas, deberá ser optimizado usando herramientas con mayor eficiencia que Matlab como TensorFlow [110] y usando librerías optimizadas por Nvidia para Deep Learning como cuDNN [111], que en su versión 7.5 tiene un rendimiento 3 veces mayor que en su sexta versión, con lo que seguramente se reduciría el costo computacional en la detección de actividades criminales en video.

Al reducir el costo computacional sería posible implementar el del sistema de detección de actividades criminales en video, en sistemas embebidos como Nvidia Jetson [112] lo que permitiría hacer una implementación de bajo costo directamente en las cámaras de seguridad ciudadana.

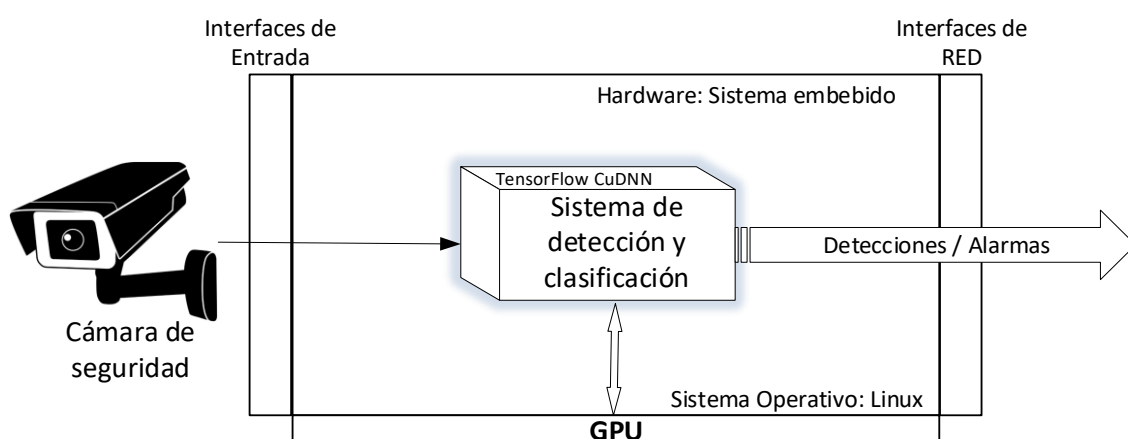


Figura 48: Arquitectura de detección de actividades criminales en un sistema embebido

Esta implementación con sistemas embebidos haría que cada cámara procese su propio video y en caso de detección genere una alarma la cual pueda ser procesada por el operador del centro de comando y control de seguridad ciudadana, pues es de gran importancia la interacción humana con las alarmas generadas con el sistema de detección de actividades criminales pues como cualquier tecnología basada en Deep Learning no es infalible en un cien por ciento.

### **6.1.2 Arquitectura de implementación en sistemas embebidos**

Para un posible despliegue a gran escala de la solución basada en sistemas embebidos se implementará uno de estos sistemas en cada cámara de video vigilancia y como se ha mencionado anteriormente la arquitectura de software para la detección de actividades criminales en video deberá ser optimizada para esta plataforma.

El video será analizado en tiempo real, generando alarmas según las detecciones y clasificaciones generadas; estas alarmas serán enviadas por la red de telecomunicaciones del operador local hasta el centro de comando y control de seguridad ciudadana.

Una vez las alarmas están en el centro de comando y control, estas deberán ser contrastadas con el video en tiempo real proveniente de la



misma cámara para descartar falsas alarmas y proceder con las acciones correspondientes.

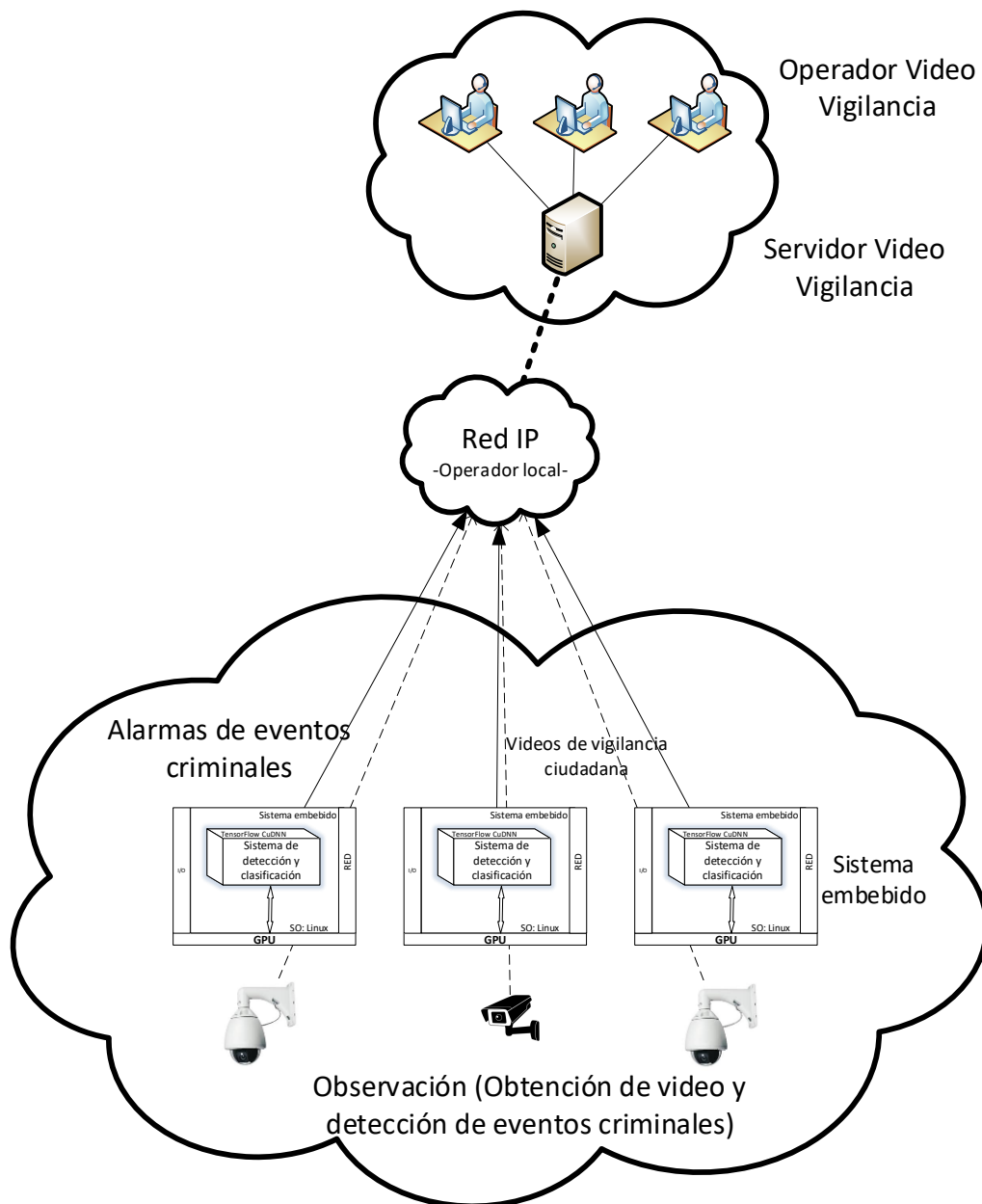


Figura 49: Arquitectura propuesta para la Policía Nacional de Colombia basada en sistemas embebidos.

## 6.2 Procesamiento de video centralizado

Otra posible implementación del sistema de detección de actividades criminales en video en la Policía Nacional de Colombia es realizar el procesamiento del video en un centro de procesamiento de datos, con capacidad de procesamiento en paralelo mediante GPUs.

Una implementación centralizada reduciría los puntos de falla y gestión de la red de cámaras de seguridad ciudadana, sin embargo los equipos de cómputo deben contar además con gran capacidad de procesamiento en paralelo, implementando GPU de gran potencia de cómputo.

Alternativas de procesamiento en paralelo de gran poder en la nube pública como NVIDIA GPU Cloud [113], podrían ser una solución sin embargo por el carácter de seguridad nacional, de la información manejada por la Policía Nacional de Colombia hace que esta solución no sería viable para la policía.

Otro gran inconveniente es que los proveedores de soluciones Cloud como NVIDIA GPU Cloud [113], Google Cloud Platform [114] o Amazon Web Services [115] están físicamente ubicados en Estados Unidos lo que implicaría un retardo importante, pues el vídeo proveniente de las cámaras desplegadas en las ciudades colombianas debería viajar a Estados Unidos para ser procesados, recorriendo alrededor de 3000 km de fibras ópticas

y pasando por una gran variedad de redes y equipos de Telecomunicaciones.

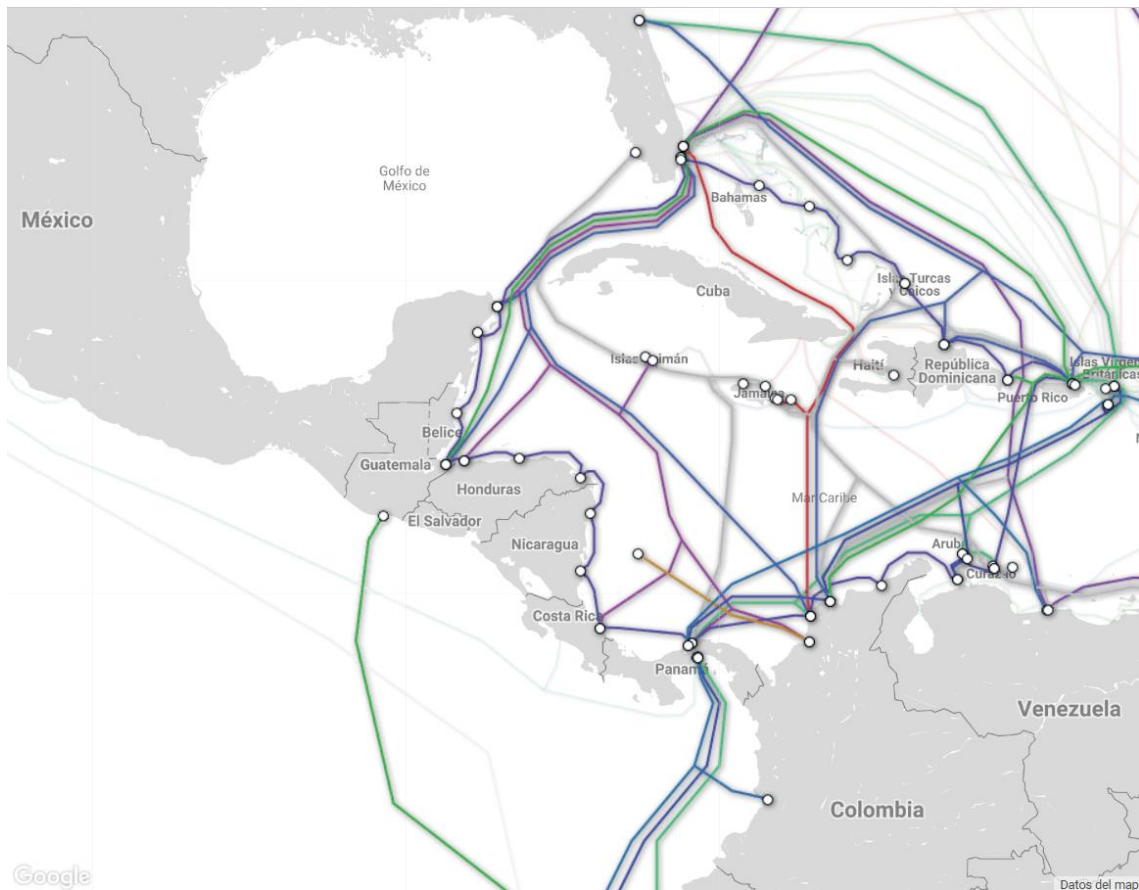


Figura 50: Arquitectura de Red Óptica internacional disponible en Colombia [116]

Otro inconveniente con esta solución en la capacidad de los enlaces internacionales para transportar esta cantidad de video lo que aumentaría en gran medida el costo de la solución; por estas razones no se considera viable el procesamiento del video hasta que no se realicen despliegues de infraestructura cloud con capacidad de procesamiento basada en GPU.

## 6.2.1 Implementación centralizada un centro de procesamiento de datos CPD

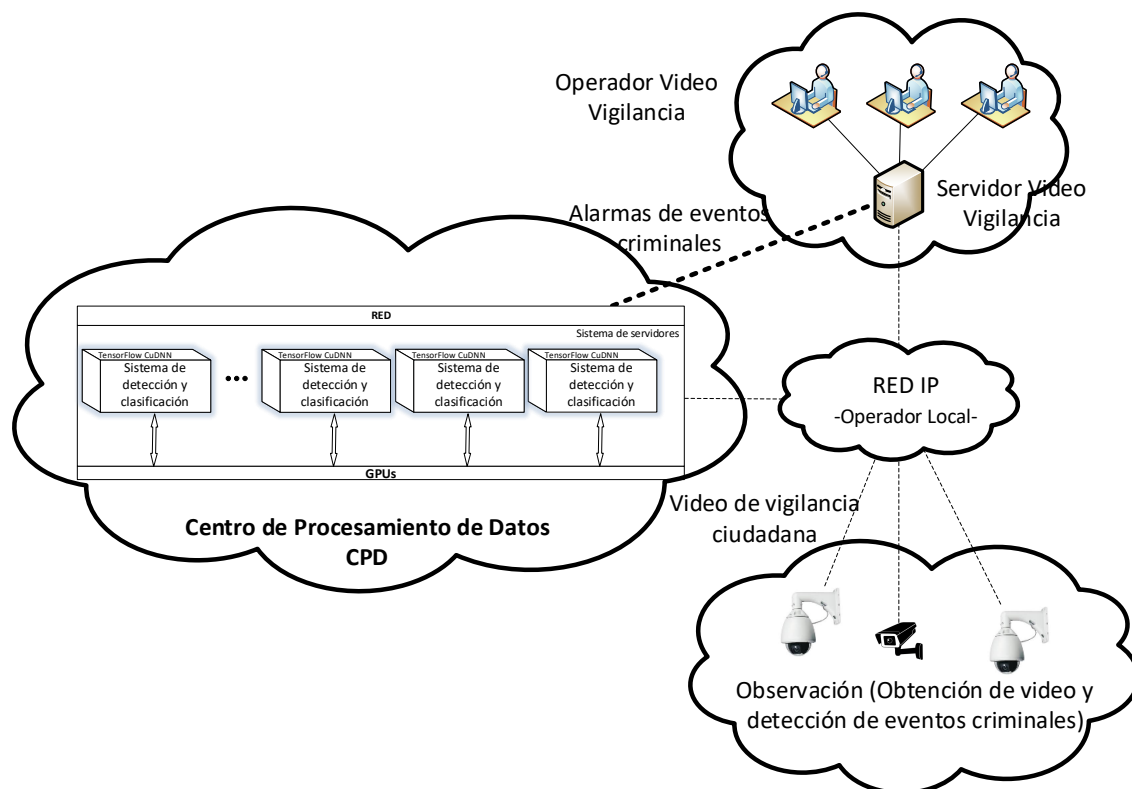


Figura 51: Arquitectura propuesta para la Policía Nacional de Colombia basada en procesamiento centralizado

Como se muestra en la figura, en esta propuesta de implementación el sistema de detección de actividades criminales se implementaría de manera centralizada en una serie de servidores con capacidad de cómputo en paralelo soportada en GPU de gran potencia. Sus resultados son entregados a los operadores del subsistema de video vigilancia los cuales procesarían los crímenes detectados por el

sistema, descartando las posibles detecciones falsas generadas por las siguientes limitaciones del sistema:

- **Calidad del video a analizar y el tamaño en pixeles de las armas a detectar:** al aplicar el sistema de detección en situaciones reales su efectividad se ve limitada en gran medida por estos factores restringiendo su aplicabilidad a las partes del Sistema de Video Vigilancia con buena calidad de video.
- **Condiciones de iluminación:** si el video producido por el Sistema de Video Vigilancia tiene condiciones tenues de iluminación y los eventos criminales no se distinguen muy bien del fondo (arma oscura en fondo oscuro), la efectividad del sistema se reduce en gran medida.
- **Obstrucciones parciales:** las obstrucciones parciales generadas por la infraestructura eléctrica, semáforos, árboles y demás elementos comunes en un ambiente urbano pueden hacer que el prototipo realice detecciones erróneas o que no detecte un arma presente en el entorno.

- **Detecciones erróneas:** en situaciones reales se presentan direcciones erróneas lo que dispararía falsas alarmas, por lo que el sistema siempre debe funcionar bajo la supervisión humana.

Finalmente es importante destacar que la potencia de cómputo disponible en el CPD debe ser suficiente para asignar recursos de hardware y software individuales para cada señal de video que se deba analizar.



## 7 INTEGRACIÓN EN EL SISTEMA DE MANDO Y CONTROL DE SEGURIDAD CIUDADANA

Para una posible integración de la arquitectura de detección de actividades criminales con los sistemas de la Policía Nacional, se debe describir cómo se estructura el servicio de Policía y su relación con el centro de comando y control.

### 7.1 Estructura del servicio de Policía

Según la Policía Nacional de Colombia el servicio de policía prestado en el marco del Modelo Nacional de Vigilancia Comunitaria por Cuadrantes tiene una estructura que clasifica el grado de complejidad del problema de seguridad en tres grados llamados: «*Fenómenos y problemáticas complejas*»; «*Fenómenos y problemáticas estructuradas*» y finalmente «*Puntos críticos/motivos de policía y requerimientos ciudadanos*».

Estos tres niveles de complejidad clasifican el servicio de policía integrado, los niveles de planeación y definen las direcciones o departamentos de policía encargados de ejecutar las operaciones policiales para el cumplimiento de los objetivos estratégicos.



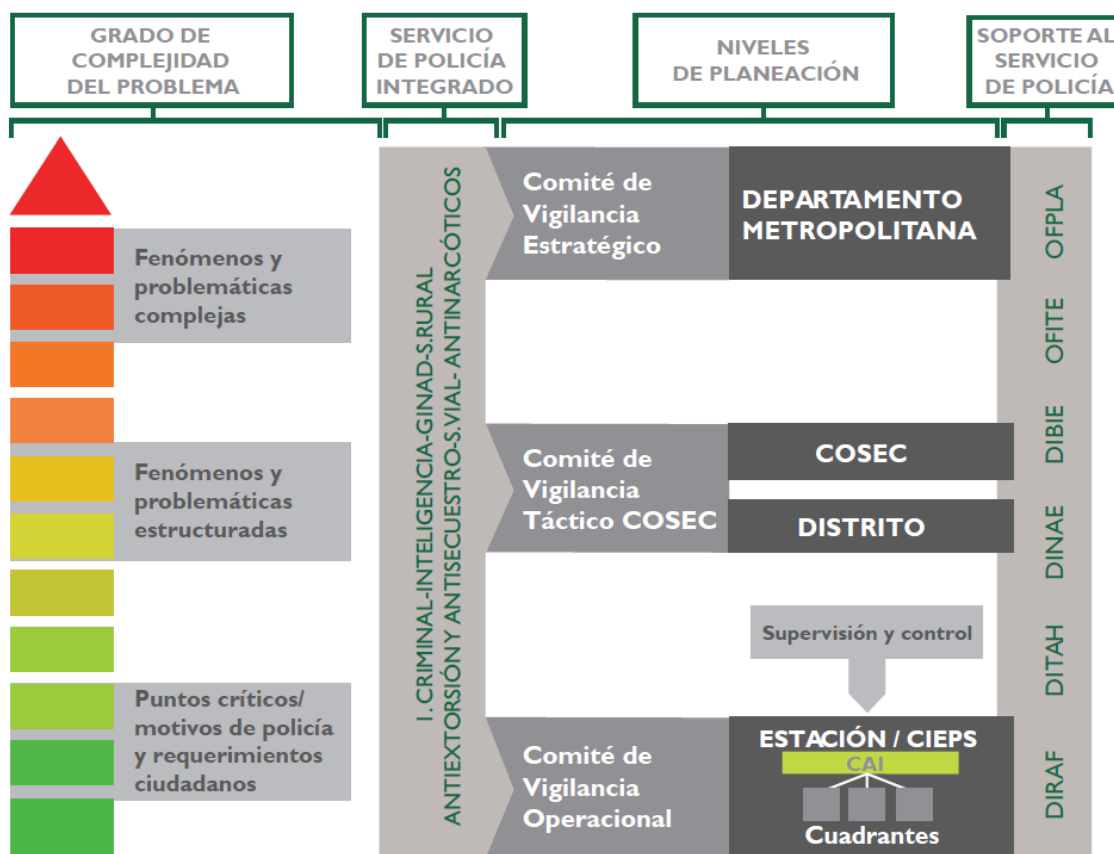


Figura 52: Estructura del servicio de Policía en Colombia [87]

## 7.2 Estrategias operativas Policía Nacional de Colombia

Según los patrones delictivos presentes en Colombia la policía Nacional de Colombia cuenta con el Plan Integral Policial Para La Seguridad del Ciudadano “Corazón Verde” (PIPSC-CV) [117], en la cual se definen ejes para plantear estrategias operativas para prevenir y mitigar las acciones delictivas en los siguientes ejes:

- Narcotráfico
- Bandas criminales

- Infancia y Adolescencia
- Contrabando
- Secuestro
- Extorción / Microextorción
- Seguridad Vial
- Terrorismo
- Población vulnerable
- Delitos informáticos
- Microtráfico
- Hurto a Celulares
- Restitución de Tierras
- Seguridad Rural
- Minería ilegal

Para el despliegue de estas estrategias operativas y articulación con el Modelo Nacional de Vigilancia Comunitaria por Cuadrantes, la Policía Nacional Plantea la Estrategia Institucional de Seguridad y Convivencia Ciudadana (EICOS) la cual está orientada a proteger a los ciudadanos de los factores que atentan contra la vida, integridad, libertad y patrimonio económico, diezmando el accionar delincuencia, reducir los delitos de impacto y fortalecer la gobernabilidad.

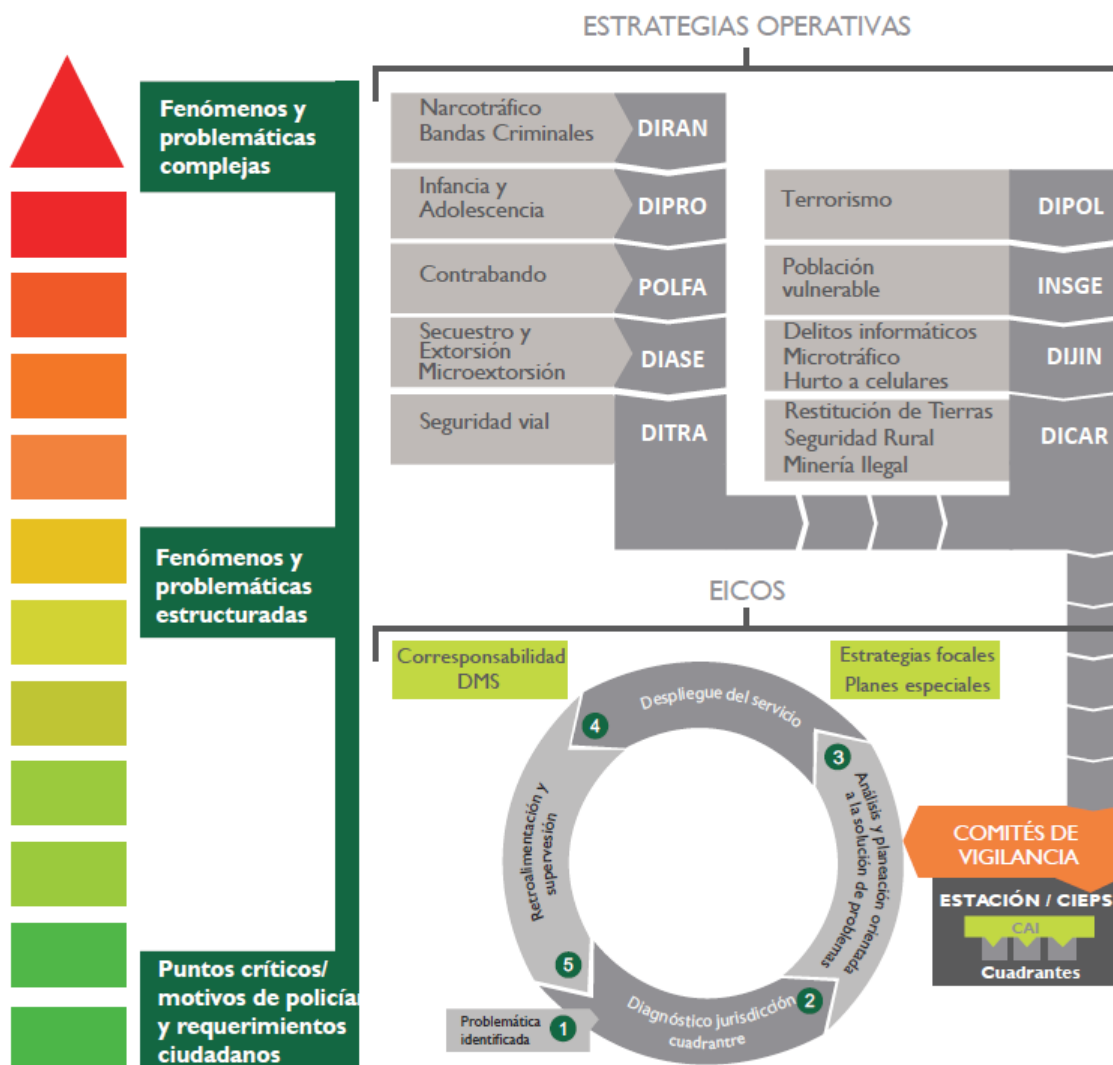


Figura 53: Estrategias Operativas y Estrategia Institucional de Seguridad y Convivencia Ciudadana (EICOS) -Policía Nacional de Colombia [87]

### 7.3 Operaciones de Policía en Colombia

La planeación, preparación y ejecución de operaciones policía en Colombia siempre está lineada al servicio de Policía, a las estrategias

operáticas y a la Estrategia Institucional de Seguridad y Convivencia Ciudadana mencionadas anteriormente.

En este sentido los centros de comando y control de la Policía Nacional de Colombia son los puntos neurálgicos para la asegurar la seguridad ciudadana en el territorio colombiano, según lo planteado en el Modelo Nacional de Vigilancia Comunitaria por Cuadrantes, en el Plan Integral Policial para La seguridad del ciudadano y en la estrategia institucional de Seguridad y Convivencia Ciudadana.



Figura 54: Esquema de ejecución en operaciones Policiales - Policía Nacional de Colombia

Por ejemplo, si en determinada zona geográfica (cuadrante) se han detectado actividades delictivas de bandas criminales dedicadas al

Narcotráfico, la Policía Nacional de Colombia en cumplimiento de sus objetivos estratégicos, planeará, alistará y ejecutará una operación Policial, desde el Centro de Comando Control con el objeto de desarticular la banda delictiva, capturar a sus miembros y ponerlos a disposición de la fiscalía general de la nación.

#### **7.4 Infraestructura de Centros de comando y control Policía Nacional de Colombia**

Como se ha mostrado los centros de comando y control son fundamentales para la planeación, preparación y ejecución de las operaciones de policía que permiten que la institución cumpla con sus objetivos misiones según los planes y estrategias mencionados anteriormente.

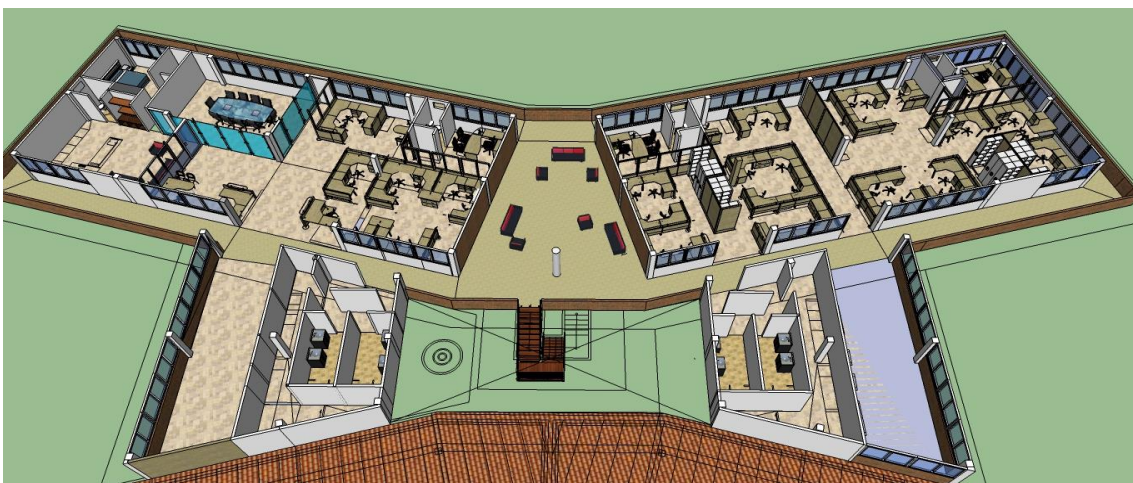


Figura 55: Centro de comando y control (Planta Baja) - Policía Nacional de Colombia

Los centros de comando y control de la Policía Nacional de Colombia cuentan con infraestructura física y tecnológica suficiente para la planeación, preparación y ejecución de operaciones de policía, por ejemplo:

- Sistema unificado de llamadas de Emergencia (123)
- Sistema de recepción y gestión de casos de policía.
- Sistema de recepción y gestión de contravenciones.
- Visualización y control del sistema de Video Vigilancia
- Sistema de comunicación en tiempo real con agentes y patrullas en el campo.
- Sistema de información geográfico alimentado en tiempo real con ubicación de agentes, patrullas y casos de policía
- Salas de crisis / comando y salas de juntas.

Sin embargo, los centros de comando y control en Colombia no cuentan con herramientas automáticas para el procesamiento de información proveniente de los diversos los sistemas que entregan información en tiempo al centro, y todo el procesamiento debe ser realizado por agentes de policía, lo que en ocasiones supera las capacidades operativas de la Policía Nacional.

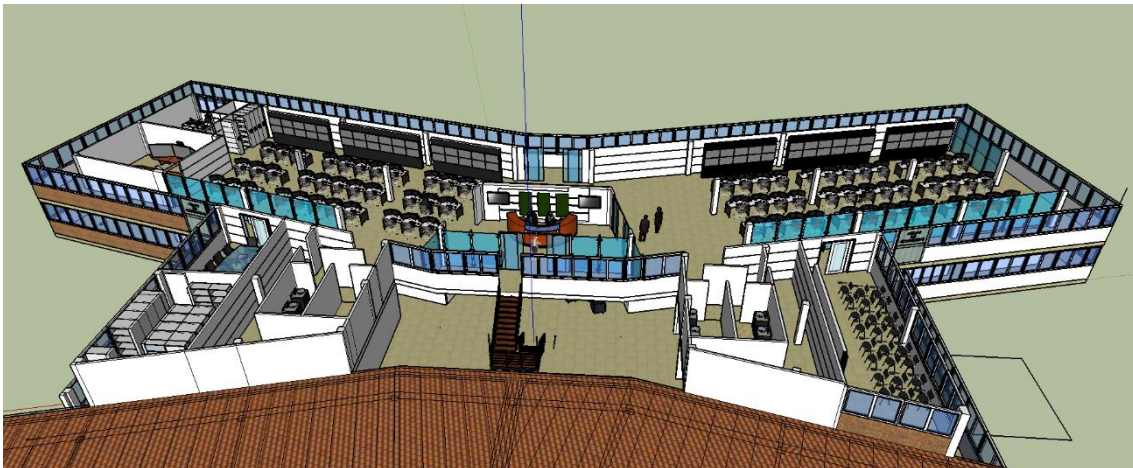


Figura 56: Centro de comando y control (primera planta) - Policía Nacional de Colombia

## 7.5 Arquitectura genérica del centro de comando y control Policía Nacional de Colombia

Las arquitecturas detalladas de los centros de comando y control desplegados por en la república de Colombia son de carácter confidencial por sus implicaciones en la seguridad ciudadana, sin embargo, todos los centros de comando y control tienen una arquitectura genérica que sigue la siguiente estructura lógica.

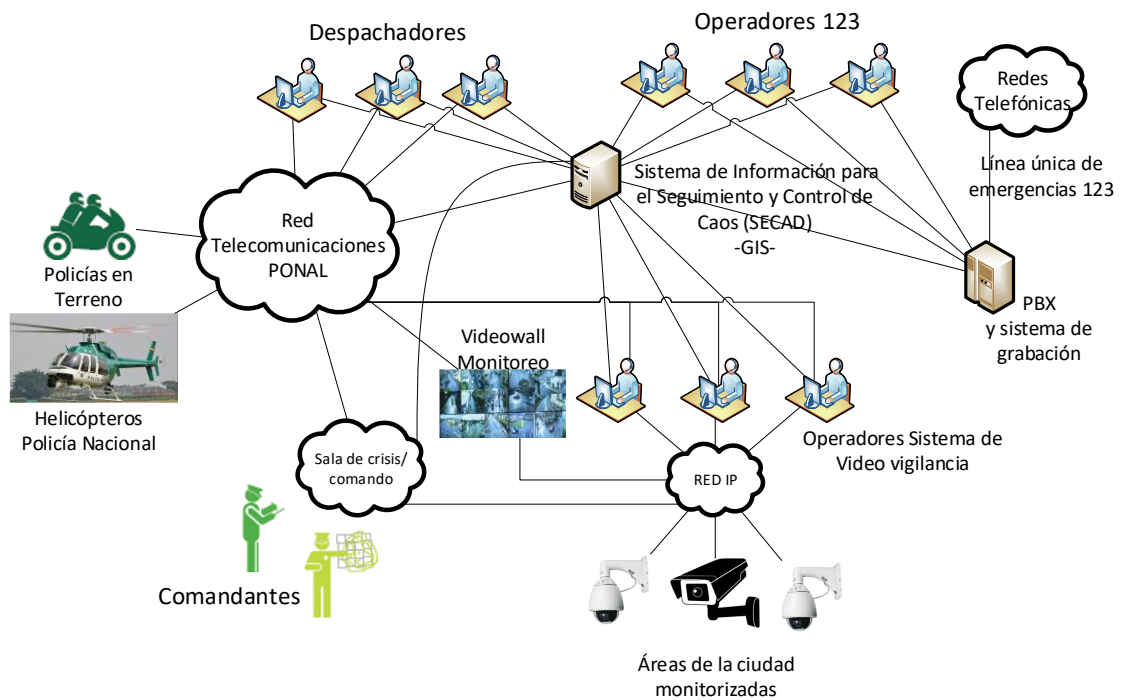


Figura 57: Arquitectura genérica del centro de comando y control de seguridad ciudadana - Policía Nacional de Colombia

En forma general, los centros de comando y control de seguridad ciudadana desplegados en Colombia están configurados de la siguiente manera:



### **7.5.1 SISTEMA DE INFORMACIÓN PARA EL SEGUIMIENTO Y CONTROL DE CASOS (SECAD)**

El SECAD es el sistema de información desarrollado por la oficina de telemática de la policía nacional, encargado de gestionar los casos de policía, reportados por la ciudadanía o por las fuerzas policiales.

En Colombia se entiende como caso de policía a cualquier incidente que requiere la atención y el seguimiento de las autoridades policiales, como reportes de hurto, lesiones personales, riñas, tentativas de homicidio, reportes de por alto ruido en zonas residenciales, etc; estos casos pueden ser clasificados según su urgencia, gravedad y diferentes parámetros definidos por la policía nacional.

El SECAD está soportado en un sistema de información geográfica (GIS por su sigla en inglés) el cual le permite manejar y visualizar información georreferenciada mediante mapas interactivos que se actualizan en tiempo real de los casos de policía y de las patrullas en terreno en tiempo real, aumentando así la conciencia situacional (*situational awareness*) en el centro de comando y control de seguridad ciudadana.

Las coordenadas geográficas de los casos reportados por la ciudadanía por medio de llamadas de emergencia son entregadas por los operadores de telefonía móvil celular y son cargadas al SECAD. Por otro lado, las coordenadas de las patrullas de policía en terreno son capturadas por

receptores GPS y enviadas al SECAD por medio de la red de Telecomunicaciones de la Policía Nacional.

Finalmente, el SECAD cuenta con una base de datos en la cual se almacenan los casos de policía y se registra su seguimiento y la gestión que se hace sobre ellos.

El SECAD tienen diferentes tipos de usuarios como son los: operadores 123 quienes atienden las llamas de emergencia de la ciudadanía y crean los casos en el SECAD. Despachadores, encargados de asignar agentes de policía en terreno para resolver los distintos casos reportados en el SECAD; operadores del sistema de Video vigilancia, son los encargados de monitorear y controlar las cámaras móviles desplegadas por la ciudad, en caso de detectar alguna actividad delictiva tienen la capacidad de crear el caso de policía y remitirlo a los despachadores.

Al SECAD también tiene acceso los comandantes del centro de comando y control para seguridad ciudadana, obteniendo un reporte en tiempo real de la situación de seguridad ciudadana.

## **7.5.2 SUBSISTEMA DE ATENCIÓN A LLAMADAS DE EMERGENCIA**

### **123**

Este subsistema [118] es está encargado de recibir y gestionar las llamadas de emergencia realizadas por la ciudadanía, estas llamadas de

emergencia llegan al sistema de comando y control por medio de troncales IP bajo el protocolo SIP, estas llamadas se centralizan en una IP-PBX y de allí son repartidas a los operadores del subsistema. Integrado con la IP-PBX el subsistema cuenta con un sistema de grabación de llamadas el cual está activo para cualquier llamada que curse por este sistema.

Dado que en ocasiones las llamadas de emergencia tienen carácter crítico, puede tratarse de situaciones de vida o muerte, los operadores cuentan con capacitación especial para tratar este tipo de situaciones.

En ciudades como Bogotá y Medellín existe personal especializado para la atención psicológica de casos específicos, por ejemplo, existe atención especializada a las mujeres y así brindar vigilancia y protección a las mujeres víctimas de violencia.

Este subsistema también clasifica casos que no solo deben ser atendidos por Policía Nacional, sino que requieren participación de otras agencias como Bomberos o defensa civil, y en estos casos se realizan las gestiones con estas agencias.

### **7.5.3 SUBSISTEMA DE VIDEO VIGILANCIA CIUDADANA**

El subsistema de Video Vigilancia está compuesto por una extensa red de cámaras de vigilancia instaladas en sectores estratégicos de la ciudad que están interconectadas con un servidor de almacenamiento y monitoreo de video ubicado en el centro de comando y control de seguridad ciudadana.

Esta interconexión es suministrada por un operador de telecomunicaciones local quien entrega enlaces IP entre cada una de las cámaras instaladas en la ciudad con el servidor de almacenamiento y monitoreo de video.

El servidor cuenta con un software propietario del fabricante de las cámaras con el cual se almacena el video proveniente de las cámaras, además este permite el control centralizado de las cámaras, específicamente se pueden controlar parámetros como dirección de observación de las cámaras móviles, el zoom y programar movimientos automáticos en cámaras específicas.

Los operadores del subsistema de video vigilancia tienen acceso a clientes de este servidor, los cuales son configurados para el monitoreo y control de un grupo de cámaras específico de una zona de la ciudad en concreto.

Finalmente, el sistema cuenta con una conexión con el Videowall del centro de comando y control para mostrar el video proveniente de las cámaras críticas en la ciudad.

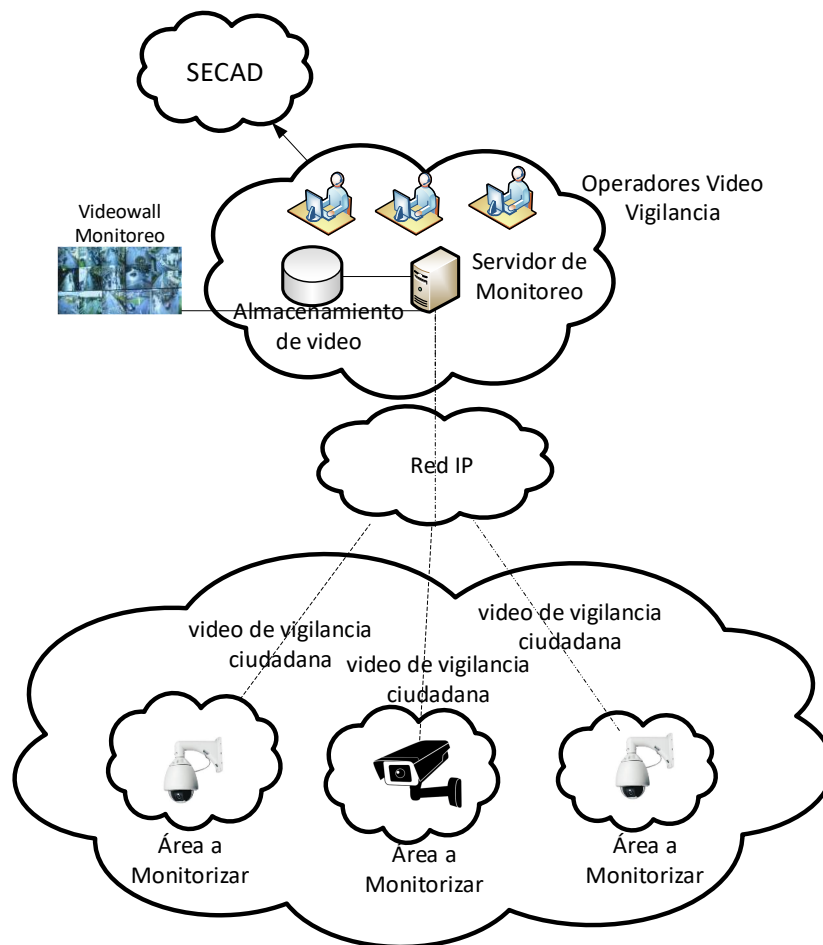


Figura 58: Arquitectura genérica del subsistema de video vigilancia

#### 7.5.4 SUBSISTEMA DE DESPACHO

El subsistema de despacho es el encargado de la comunicación en tiempo real entre los policías en campo y el centro de comando y control.

Los agentes encargados de este sistema son llamados despachadores y son responsables del control y seguimiento de los casos de policía que se encuentren activos.

Los despachadores cuentan con un número terminado de casos para hacer seguimiento y la comunicación con los agentes de policía en terreno se realiza por una red de telecomunicaciones inalámbrica propiedad de la policía nacional, en casos críticos los despachadores pueden coordinar sus operaciones con los operadores del sistema de video vigilancia para enviar información y coordinar la ejecución de ordenes en tiempo real con los agentes de policía desplegados en la ciudad.

#### **7.5.5 SALAS DE CRISIS / COMANDO**

El último componente de los centros de comando y control de seguridad ciudadana son las salas de crisis / comando, en donde los comandantes tienen acceso a la información de todos los subsistemas en tiempo real.

En estas salas se planea, prepara y coordina la ejecución de operaciones policiales como capturas y desarticulaciones de bandas criminales, incautación de drogas a narcotraficantes, etc.

Cuando hay una condición de orden público excepcional, como un atentado terrorista, protestas ciudadanas o cualquier citación que requiera atención especial la coordinación se realiza desde las salas de crisis / comando.



Figura 59: Sala de crisis Dirección General de la Policía Nacional de Colombia - DIPON

## **7.6 AUTOMATIZACIÓN DE ALARMAS EN POSIBLES CASOS POLICIALES DETECTADOS POR VIDEO**

Como se ha mencionado, en determinadas oportunidades la capacidad de información que llega al centro de comando y control supera a la capacidad de procesamiento con la que cuentan los agentes de policía que operan el sistema de comando y control.

Por esta situación a continuación se evaluará la posible aplicación del sistema de detección de actividades criminales en video aplicado en Sistemas de Mando y Control para seguridad ciudadana, se procederá a

estudiar el caso de una posible implementación en la Policía Nacional de Colombia quienes apoyaron el desarrollo de esta tesis doctoral.

Lo primero que se debe tener en cuenta es el número de cámaras de seguridad ciudadana desplegadas en Colombia, por ejemplo la capital del país cuenta a diciembre de 2018 con 3300 cámaras instaladas y tiene planeado contar con 4000 cámaras [119], en las demás ciudades de Colombia se cuenta con despliegue de cientos de cámaras, por ejemplo Medellín cuenta con alrededor de 900 cámaras [120] y Santiago de Cali alrededor de 800 cámaras [121].

Teniendo en esto en cuenta, se tendría que procesar una cantidad importante de video generado por segundo que ascendería a tener que procesar alrededor de 170.000 imágenes por segundo solo para estas tres ciudades, lo que tendría un costo computacional extremadamente elevado y por tanto un costo económico bastante alto.

## **7.7 INTEGRACIÓN CON EL SISTEMA DE COMANDO Y CONTROL DE SEGURIDAD CIUDADANA**

Según la disposición del sistema de comando y control de seguridad ciudadana con la que cuenta la policía nacional, el sistema detección de actividades criminales basada en análisis de vídeo en tiempo real propuesta en esta tesis se debería integrar en el subsistema de video vigilancia, entregando alarmas generadas por el sistema a los operadores



del subsistema de video vigilancia, las cuales podrán ser verificadas visualmente por los mismos procediendo a generar los casos de policía en el SECAD si este es el caso.

Teniendo en cuenta que la arquitectura de detección de actividades criminales es una arquitectura de software basada en técnicas de Deep Learning que aprovecha las capacidades de procesamiento en paralelo de las GPU compatibles con CUDA®, su integración con el sistema de comando y control de seguridad ciudadana debe estar acorde con los casos de uso y con la disponibilidad de la capacidad de computo para procesamiento en paralelo, pues como se mostró en el capítulo 4 cada señal de video proveniente de cada cámara deberá tener sus propios recursos de hardware y software.

Por tanto, la integración en el sistema al sistema de comando y control de seguridad ciudadana se tiene que dar enmarcar dentro de los dos casos de uso expuestos en el capítulo 5, siempre dentro del subsistema de video vigilancia, siendo entendidos por los comandantes y operadores como una nueva funcionalidad de procesamiento de video.

Para integrar las capacidades de la arquitectura de detección de actividades criminales que se propone en esta tesis, con el sistema de

comando y control de seguridad ciudadana de la mejor forma posible, se plantean dos maneras el procesamiento distribuido mediante sistemas embebidos y el procesamiento centralizado en un centro de procesamiento de datos en el centro de comando y control de la policía nacional. En 2019 se descarta el procesamiento de video en nubes públicas, pues en Colombia no existe infraestructura de un operador de Telecomunicaciones o proveedor de servicios de Cloud Computing con capacidad de procesamiento en paralelo por GPU, esta capacidad está instalada en Estados Unidos lo que representaría un retraso importante al mover esta gran cantidad de video por fibras ópticas internacionales.

En este sentido la primera propuesta de integración sería instalado en cada cámara un sistema embebido provisto de una GPU. A este sistema embebido ingresaría la señal de video de la cámara, dentro del sistema embebido se ejecutaría el sistema detección de actividades criminales en tiempo real el cual realizaría las detecciones generando las alarmas correspondientes. Finalmente, por medio de la interfaz de red del sistema embebido se enviaría la información de las alarmas hacia los operadores del subsistema de video vigilancia los cuales deberán verificar la validez de las alarmas generadas y en tal caso crear el caso de policía en el SECAD.

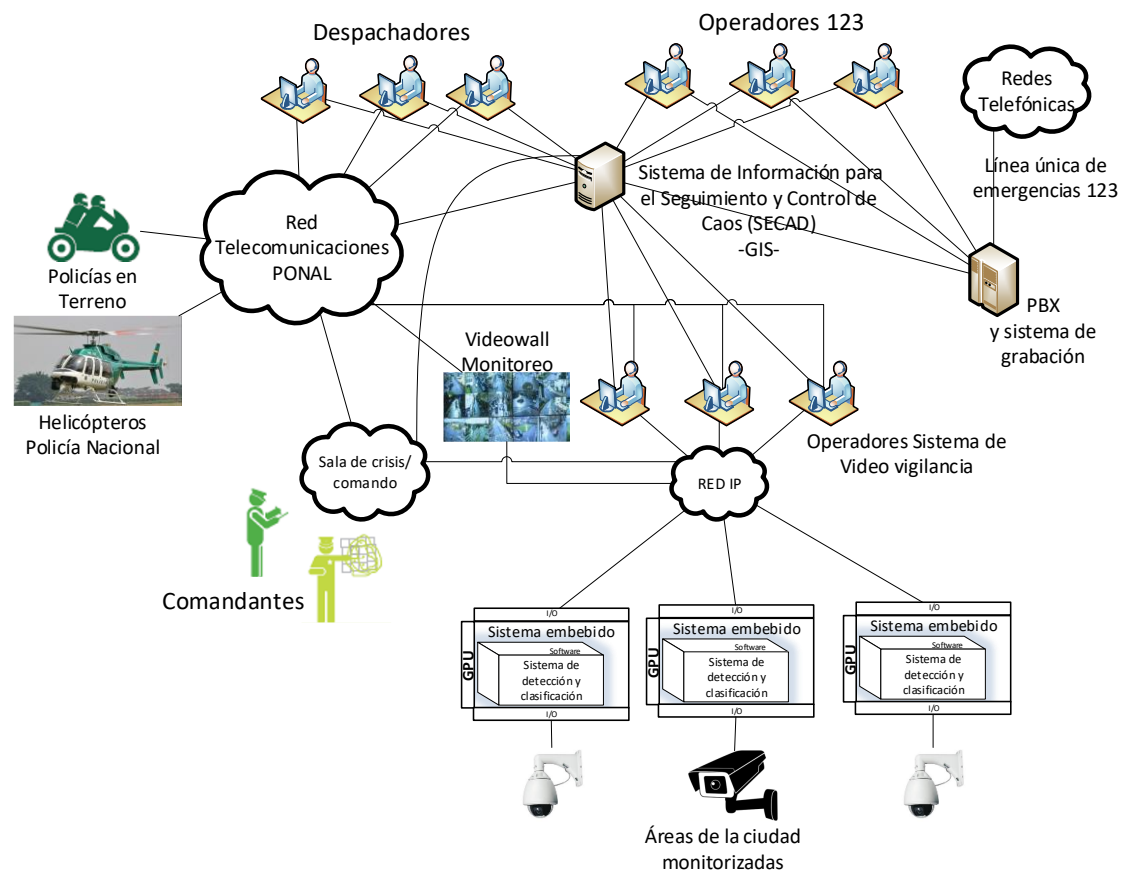


Figura 60: Integración mediante sistemas embebidos

La segunda forma de integración es ejecutando el sistema de detección de actividades criminales en un centro de procesamiento de datos de la policía nacional, este CPD deberá contar con capacidades suficientes de procesamiento paralelo basado en GPU.

El video proveniente de las cámaras desplegadas por la ciudad será llevado al CPD en donde se ejecutará el sistema de detección de actividades criminales individualmente para cada cámara usando recursos de hardware y software para cada tarea. Desde el CPD se detectarían las actividades criminales en video generando la correspondiente información

de las alarmas hacia los operadores del subsistema de video vigilancia los cuales deberán verificar la validez de las alarmas generadas y en tal caso crear el caso de policía en el SECAD.

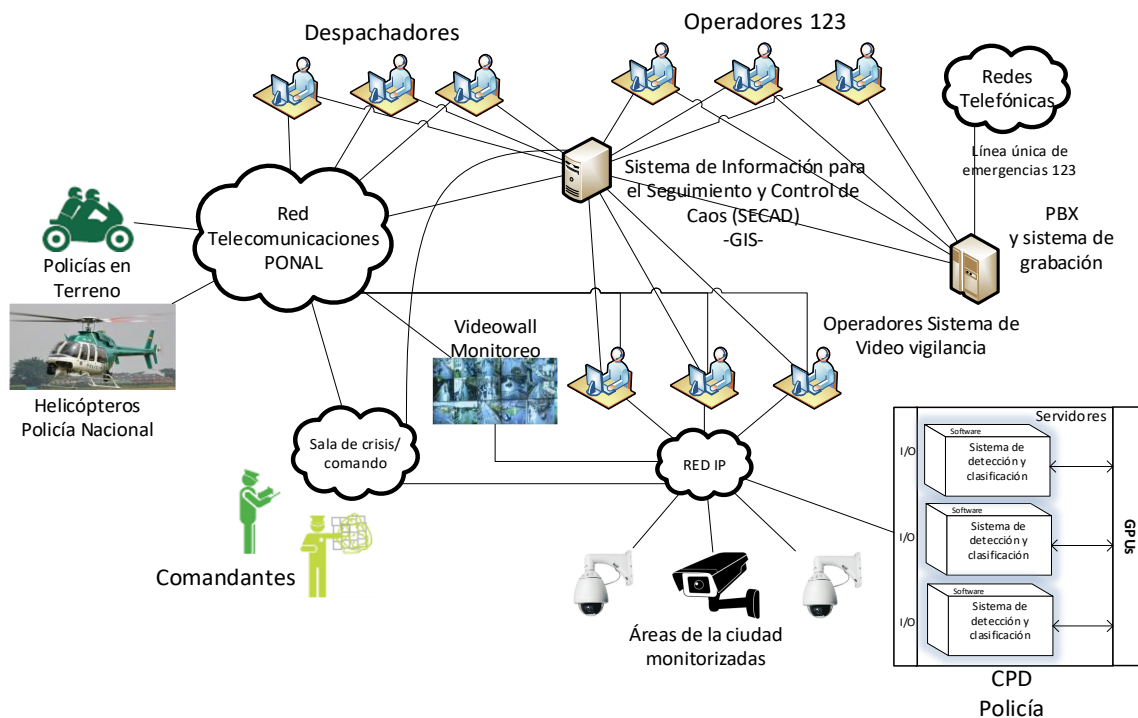


Figura 61: Integración mediante centro de procesamiento de datos CPD

Con estas dos posibles integraciones del sistema de detección de actividades criminales con el sistema de comando y control de seguridad ciudadana se tendría una mejor conciencia situacional, pues aumentaría la tasa de detección de actividades criminales desde el subsistema de video vigilancia.





## **8 CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN**

### **8.1 CONCLUSIONES**

Al desarrollar esta tesis doctoral se demostró que es posible hacer un detector de actividades criminales con basado en Deep Learning y que es posible entrenar una Red Neuronal Convolutiva para la detección de hurtos callejeros tratándolos como si fueran objetos a detectados.

También es posible mejorar la conciencia situacional en un centro de mando y control de seguridad ciudadana al generar alarmas cuando se detectan actividades criminales previa validación humana.

Sin embargo, se debe tener en cuenta que el sistema no es infalible y como todas las soluciones basadas en Deep Learning está limitada por las condiciones de captura del video como baja iluminación, obstrucciones y demás situaciones que afectan la calidad del video capturado.

Automatizar este tipo de procesos puede ayudar a las agencias de seguridad en su labor, sin embargo, siempre debe tener supervisión humana y obviamente las decisiones estratégicas deberán ser tomadas por los comandantes con la información más fiable posible.

El costo computacional es un factor decisivo, el cual puede hacer que una solución sea viable para usar en tiempo real o no lo sea, en este sentido se identificó que para aplicaciones de tiempo real basadas en Deep Learning no siempre es bueno usar modelos CNN de alta complejidad, pues su impacto en el rendimiento puede eliminar el procesamiento de video en tiempo real.

El procesamiento de video en tiempo real es una actividad computacionalmente muy demandante y aplicaciones a gran escala como el sistema de video vigilancia de la Policía Nacional de Colombia, este costo computacional se traduce en cientos de miles de imágenes a ser procesadas cada segundo, lo que a su vez implica un costo económico considerable, por tanto este tipo de soluciones deben ser diseñadas tratando de minimizar los costos al mínimo, tanto así que el procesamiento pueda realizarse en hardware de bajo costo como el disponible en ciertos sistemas embebidos.



## 8.2 FUTURAS LÍNEAS DE INVESTIGACIÓN

Esta tesis doctoral abre varias líneas de investigación a aplicar técnicas de inteligencia artificial en el campo de la seguridad ciudadana entre las cuales destacan.

- **Detectores de actividades sospechosas en sistemas de video vigilancia.**

Aunque existen varios trabajos en este aspecto, al aplicar Deep Learning usando análisis de video en tiempo real se pueden obtener mejores resultados en lo correspondiente a análisis de trayectorias e identificación de objetos sospechosos.

- **Detección de patrones de ataque y defensa en objetivos.**

Las técnicas basadas en Learning Machine pueden ser entrenada para la detección de patrones de ataque y/o defensa en situaciones

tácticas, permitiendo identificar con algún tiempo extra emboscadas y demás estrategias del enemigo.

- **Sistemas de predicción de actividades delictivas o terroristas.**

Al usar análisis de datos basado en Learning Machine sería posible identificar patrones de operación en bandas criminales o terroristas, pudiendo predecir cuales serían sus próximas acciones.

## REFERENCIAS

- [1] D. S. Alberts y R. E. Hayes, Understanding Command And Control, CCRP, 2011, p. 222.
- [2] Naciones Unidas, «SINOPSIS: SEGURIDAD CIUDADANA,» Naciones Unidas, Nueva York, 2013.
- [3] República de Colombia, «Política de Seguridad y Convivencia (PSC),» Bogotá D.C., 2014.
- [4] Gobierno de España, *ESTRATEGIA DE SEGURIDAD NACIONAL*, Madrid, 2017.
- [5] Casa editorial el tiempo, «Policía inaugura el centro de control y comando más grande del país,» *El Tiempo*, 7 11 2017.
- [6] L. G. Shapiro y G. C. Stockman, Computer vision, Prentice Hall, 2001.
- [7] L. Deng y D. Yu, Deep Learning:Methods and Applications, Now Foundations and Trends, 2014.
- [8] S. Khan, H. Rahmani, S. A. A. Shah, M. Bennamoun, G. Medioni y S. Dickinson, A Guide to Convolutional Neural Networks for Computer Vision, Morgan & Claypool, 2018.

- [9] Large Scale Visual Recognition Challenge (ILSVRC), «ImageNet Large Scale Visual Recognition Challenge (ILSVRC),» 2010-2017. [En línea]. Available: <http://www.image-net.org/challenges/LSVRC/>. [Último acceso: 18 04 2018].
- [10 E. Alpaydin, Machine Learning: The New AI, Cambridge, Massachusetts: MIT Press, 2016.
- [11 W. Qian, R. Li, F. Xie, H. Luo, Y. Wang y T. Zheng, «A machine learning based method of constructing virtual inertial measurement predictor of human body,» Dalian, China, 2017.
- [12 A. Lahouar y J. B. H. Slama, «Comparative study of learning machine predictors for half-hour and day-ahead electricity price forecast in deregulated markets,» de *Renewable Energy Congress (IREC), 2016 7th International*, Hammamet, Tunisia, 2016.
- [13 P. Juneja y U. Ojha, «Casting online votes: To predict offline results using sentiment analysis by machine learning classifiers,» de *Computing, Communication and Networking Technologies (ICCCNT), 2017 8th International Conference on*, Delhi, India, 2017.
- [14 S. K. Gonugondla, M. Kang y N. Shanbhag, «A 42pJ/decision 3.12TOPS/W robust in-memory machine learning classifier with on-

chip training,» de *Solid - State Circuits Conference - (ISSCC), 2018 IEEE International*, San Francisco, CA, USA, 2018.

[15 D. Watkins, G. Gallardo y S. Chau, «Pilot Support System: A Machine Learning Approach,» de *Semantic Computing (ICSC), 2018 IEEE 12th International Conference on*, Laguna Hills, CA, USA, USA, 2018.

[16 A. Bifet, R. Gavaldá, G. Holmes y B. Pfahringer, *Machine Learning for Data Streams: with Practical Examples in MOA*, MIT Press, 2018.

[17 H. Geng, *Internet of Things and Data Analytics Handbook*, Wiley, 2017.

[18 K. Fukushima, «Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position,» *Biological Cybernetics*, vol. 36, nº 4, pp. 193-202, 1980.

[19 K. Fukushima, S. Miyake y T. Ito, «Neocognitron: A neural network model for a mechanism of visual pattern recognition,» *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 13, nº 5, pp. 826-834, 1983.

[20 A. Krizhevsky, I. Sutskever y G. Hinton, «ImageNet Classification with Deep Convolutional,» de *Advances in Neural Information Processing Systems 25*, C. B. L. B. K. W. F. Pereira, Ed., Curran Associates, Inc., 2012, pp. 1097-1105.

- [21 W. Choi, K. Choi y J. Park, «Low Cost Convolutional Neural Network Accelerator Based on Bi-Directional Filtering and Bit-Width Reduction,» *IEEE Access*, vol. 6, pp. 14734 - 14746, 2018.
- [22 L. Xiao, Q. Yan y S. Deng, «Scene classification with improved AlexNet model,» de *Intelligent Systems and Knowledge Engineering (ISKE), 2017 12th International Conference on*, Nanjing, China, 2017.
- [23 S. Gu, L. Ding, Y. Yang y X. Chen, «A new deep learning method based on AlexNet model and SSD model for tennis ball recognition,» de *Computational Intelligence and Applications (IWCIA), 2017 IEEE 10th International Workshop on*, Hiroshima, Japan, 2017.
- [24 M. D. Zeiler y R. Fergus, «Visualizing and Understanding Convolutional Networks,» *CoRR*, vol. 1311.2901, 2013.
- [25 K. Simonyan y A. Zisserman, «VERY DEEP CONVOLUTIONAL FOR LARGE-SCALE IMAGE RECOGNITION,» de *ICLR 2015*, San Diego, USA., 2015.
- [26 C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke y A. Rabinovich, «Going deeper with convolutions,» de *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 2015.

- [27 K. He, X. Zhang, S. Ren y J. Sun, «Deep Residual Learning for Image Recognition,» de *Computer Vision and Pattern Recognition (CVPR)*, 2016 *IEEE Conference on*, Las Vegas, NV, USA, 2016.
- [28 A. Krizhevsky, I. Sutskever y G. E. Hinton, «ImageNet Classification with Deep Convolutional,» de *Advances in Neural Information Processing Systems 25*, C. B. L. B. K. W. F. Pereira, Ed., Curran Associates, Inc., 2012, pp. 1097-1105.
- [29 C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke y A. Rabinovich, «Going deeper with convolutions,» de *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 2015.
- [30 B. Y. M. Lobato y A. A. Garza, «Classification algorithm for measuring human emotion: "I like it" and "I do not like" in Neuromarketing,» *IEEE Latin America Transactions*, vol. 15, nº 11, pp. 2177-2184, 2017.
- [31 R. Girshick, J. Donahue, T. Darrell y J. Malik, «Region-Based Convolutional Networks for Accurate Object Detection and Segmentation,» *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, nº 1, pp. 142-158, Jan 2016.
- [32 R. Girshick, «Fast R-CNN,» de *Computer Vision (ICCV)*, 2015 *IEEE International Conference on*, Santiago, Chile, 2015.

- [33 S.-C. Hsu, Y.-W. Wang y C.-L. Huang, «Human Object Identification for Human-Robot Interaction by Using Fast R-CNN,» de *Robotic Computing (IRC), 2018 Second IEEE International Conference on*, Laguna Hills, CA, USA, USA, 2018.
- [34 K. Shi, H. Bao y N. Ma, «Forward Vehicle Detection Based on Incremental Learning and Fast R-CNN,» de *Computational Intelligence and Security (CIS), 2017 13th International Conference on*, Hong Kong, China, 2018.
- [35 R. Qian, Q. Liu, Y. Yue, F. Coenen y B. Zhang, «Road surface traffic sign detection with hybrid region proposal and fast R-CNN,» de *Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), 2016 12th International Conference on*, Changsha, China, 2016.
- [36 P. Burlina, «MRCNN: A stateful Fast R-CNN,» de *Pattern Recognition (ICPR), 2016 23rd International Conference on*, Cancun, Mexico, 2016.
- [37 K. Wang, Y. Dong, H. Bai, Y. Zhao y K. Hu, «Use fast R-CNN and cascade structure for face detection,» de *Visual Communications and Image Processing (VCIP), 2016*, Chengdu, China, 2016.



- [38 S. Ren, K. He, R. Girshick y J. Sun, «Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,» *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, nº 6, pp. 1137-1149, 2017.
- [39 X. L. S. S. T. X. ., J. F. a. S. Y. Jianan Li, «Scale-Aware Fast R-CNN for Pedestrian Detection,» *IEEE Transactions on Multimedia*, vol. 20, nº 4, pp. 985 - 996, 2018.
- [40 A. Dominguez-Sanchez, M. Cazorla y S. Orts-Escolano, «Pedestrian Movement Direction Recognition Using Convolutional Neural Networks,» *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, nº 12, pp. 3540 - 3548, 9 Aug 2017.
- [41 L. L. F. F. J. W. H. X. M. B. S. P. D. M. G. K. L. Xiaofei Li, «A Unified Framework for Concurrent Pedestrian and Cyclist Detection,» *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, nº 2, pp. 269 - 281, 2017.
- [42 El Colombiano, «Sicario disparó 11 veces contra una mujer en Medellín,» *El Colombiano*, 17 04 2018.
- [43 T. Liu, H. Y. Fu, Q. Wen, D. K. Zhang y L. F. Li, «Extended faster R-CNN for long distance human detection: Finding pedestrians in UAV

images,» de *Consumer Electronics (ICCE), 2018 IEEE International Conference on*, Las Vegas, NV, USA, 2018.

[44 Z. Wu, Y. Huang, L. Wang, X. Wang y T. Tan, «A Comprehensive  
] Study on Cross-View Gait Based Human Identification with Deep  
CNNs,» *IEEE Transactions on Pattern Analysis and Machine  
Intelligence*, vol. 39, nº 2, pp. 209-226, 2017.

[45 F. M. Castro, M. J. Marin-Jimenez, N. Guil, S. Lopez-Tapia y N. P. d.  
] I. Blanca, «Evaluation of Cnn Architectures for Gait Recognition Based  
on Optical Flow Maps,» de *Biometrics Special Interest Group  
(BIOSIG), 2017 International Conference of the*, Darmstadt, Germany,  
2017.

[46 M. Marín-Jiménez, F. Castro, N. Guil, F. d. I. Torre y R. Medina-  
] Carnicer, «Deep multi-task learning for gait-based biometrics,» de  
*Image Processing (ICIP), 2017 IEEE International Conference on*,  
Beijing, China, 2017.

[47 Z. Ji, W. Zheng y Y. Pang, «Deep pedestrian attribute recognition  
] based on LSTM,» de *Image Processing (ICIP), 2017 IEEE  
International Conference on*, Beijing, China, 2017.

- [48 J. Cao, Y. Pang y X. Li, «Pedestrian Detection Inspired by Appearance  
] Constancy and Shape Symmetry,» *IEEE Transactions on Image  
Processing*, vol. 25, nº 12, pp. 5538 - 5551, 2016.
- [49 Y. Zhou, L. Liu y L. Shao, «Vehicle Re-Identification by Deep Hidden  
] Multi-View Inference,» *IEEE Transactions on Image Processing*, vol.  
27, nº 7, pp. 3275 - 3287, 2018.
- [50 Y. Tang, D. Wu, Z. Jin, W. Zou y X. Li, «Multi-modal metric learning for  
] vehicle re-identification in traffic surveillance environment,» de *Image  
Processing (ICIP), 2017 IEEE International Conference on*, Beijing,  
China, 2017.
- [51 Y. Zhang, D. Liu y Z.-J. Zha, «Improving triplet-wise training of  
] convolutional neural network for vehicle re-identification,» de  
*Multimedia and Expo (ICME), 2017 IEEE International Conference on*,  
Hong Kong, China, 2017.
- [52 L. Xie, T. Ahmad, L. Jin, Y. Liu y S. Zhang, «A New CNN-Based  
] Method for Multi-Directional Car License Plate Detection,» *IEEE  
Transactions on Intelligent Transportation Systems*, vol. 18, nº 2, pp.  
507 - 517, 2018.
- [53 S. Lee, K. Son, H. Kim y J. Park, «Car plate recognition based on CNN  
] using embedded system with GPU,» de *Human System Interactions*

(HSI), 2017 10th International Conference on, Ulsan, South Korea, 2017.

[54 B. S. Prabhu, S. Kalambur y D. Sitaram, «Recognition of Indian license plate number from live stream videos,» de *Advances in Computing, Communications and Informatics (ICACCI), 2017 International Conference on*, Udupi, India, 2017.

[55 S. Chandra, M. A.-a. Nowshad, M. J. Islam y Marium-E-Jannat, «An automated system to detect and recognize vehicle license plates of Bangladesh,» de *Computer and Information Technology (ICCIT), 2017 20th International Conference of*, Dhaka, Bangladesh, 2017.

[56 W.-Y. Chang, S.-H. Hsu y J.-H. Chien, «FATAUVA-Net: An Integrated Deep Learning Framework for Facial Attribute Recognition, Action Unit Detection, and Valence-Arousal Estimation,» de *Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on*, Honolulu, HI, USA, 2017.

[57 B. Yang, J. Cao, R. Ni y Y. Zhang, «Facial Expression Recognition Using Weighted Mixture Deep Neural Network Based on Double-Channel Facial Images,» *IEEE Access*, vol. 6, pp. 4630 - 4640, 2017.

[58 P. Tzirakis, G. Trigeorgis, M. A. Nicolaou, B. W. Schuller y S. Zafeiriou, «End-to-End Multimodal Emotion Recognition Using Deep Neural

Networks,» *IEEE Journal of Selected Topics in Signal Processing* (, vol. 11, nº 8, pp. 1301-1309, 2017.

[59 A. Keçeli y A. Kaya, «Violent activity detection with transfer learning  
] method,» *Electronics Letters*, vol. 53, nº 15, pp. 1047 - 1048, 2017.

[60 S. Sudhakaran y O. Lanz, «Learning to detect violent videos using  
] convolutional long short-term memory,» de *Advanced Video and  
Signal Based Surveillance (AVSS), 2017 14th IEEE International  
Conference on*, Lecce, Italy, 2017.

[61 M. Z. Zaheer, J. Y. Kim, H.-G. Kim y S. Y. Na, «A Preliminary Study  
] on Deep-Learning Based Screaming Sound Detection,» de *IT  
Convergence and Security (ICITCS), 2015 5th International  
Conference on*, Kuala Lumpur, Malaysia, 2015.

[62 F. Colangelo, F. Battisti, M. Carl, A. Neri y F. Calabró, «Enhancing  
] audio surveillance with hierarchical recurrent neural networks,» de  
*Advanced Video and Signal Based Surveillance (AVSS), 2017 14th  
IEEE International Conference on*, Lecce, Italy, 2017.

[63 World Bank, «United Nations, Perspectives of global urbanization.,»  
] Washington D. C, 2018.

[64 United Nations Development Programme, «Colombia en el Informe de  
] Desarrollo Humano "Progreso multidimensional: bienestar más allá  
del ingreso",» United Nations, Bogotá D.C., 2016.

[65 Departamento Administrativo Nacional de Estadística, República de  
] Colombia, «Boletín técnico, Principales indicadores del mercado  
laboral,» Bogotá D.C., 2018.

[66 REDACCIÓN JUSTICIA Y NACIÓN, EL TIEMPO, «Atracos, el lado  
] oscuro del éxodo de los venezolanos,» 17 3 2017.

[67 Fiscalía General de La Nación, República de Colombia, «Informe de  
] venezolanos capturados en flagrancia en Colombia,» Bogotá D.C.,  
2018.

[68 El Espectador & Unión Europea, «Líderes sociales: los silenciados y  
] los que resisten,» *El Espectador*, 2 12 2017.

[69 El Tiempo, «Bogotá terminará el 2017 con la tasa de homicidios más  
] baja en 32 años,» *El Tiempo*, 27 12 2017.

[70 El Tiempo, «Bajan homicidios y hurtos en Bogotá: Informe,» 7 23  
] 2017.

[71 Policía Nacional de Colombia, Dirección de investigación criminal e ] INTERPOL, «Estudio criminológico 2017, Hurto a Personas, Homicidios comunes,» Policia Nacional de Colombia, 2017.

[72 Naciones Unidas, «Programa de las Naciones Unidas para el ] Desarrollo,» 2018. [En línea]. Available: <http://www.undp.org/content/undp/es/home.html>. [Último acceso: 20 4 2018].

[73 Unión Europea, «Delegación de la Unión Europea en Colombia,» ] 2018. [En línea]. Available: [https://eeas.europa.eu/delegations/colombia\\_es](https://eeas.europa.eu/delegations/colombia_es). [Último acceso: 20 04 2018].

[74 U.S. Governmen, «Colombia, U.S. Agency for International ] Development - USAIS,» 2018. [En línea]. Available: <https://www.usaid.gov/es/colombia>. [Último acceso: 20 4 2018].

[75 Gobierno de Colombia, «DNP Departamento Nacional de ] Planeación,» 2018. [En línea]. Available: <https://www.dnp.gov.co/Paginas/inicio.aspx>. [Último acceso: 20 4 2018].

[76 United Nations, UNDP, «Human Development Report 2015,» United ] Nations, 2015.

[77 Ejército Nacional de Colombia, «Ejército Nacional de Colombia,»  
] 2018. [En línea]. Available: [www.ejercito.mil.co](http://www.ejercito.mil.co). [Último acceso: 20 4  
2018].

[78 Armada Nacional, «Armada Nacional,» 2018. [En línea]. Available:  
] <https://www.armada.mil.co/>. [Último acceso: 20 4 2018].

[79 Fuerza Aérea Colombiana, «Fuerza Aérea Colombiana,» 2018. [En  
] línea]. Available: <https://www.fac.mil.co>. [Último acceso: 20 4 2018].

[80 Policía Nacional de Colombia, «Policía Nacional de Colombia,» 2018.  
] [En línea]. Available: <https://www.policia.gov.co>. [Último acceso: 20 4  
2018].

[81 N. RCN, «En 2019 crecerá el número de migrantes venezolanos que  
] llegan a Colombia,» *Noticias RCN*, 14 Diciembre 2018.

[82 El Heraldo, «Migración Colombia habilita segunda fase del Permiso  
] Especial de Permanencia para Venezolanos,» *El Heraldo*, 18 2 2018.

[83 Revista Semana, «Alerta por xenofobia en contra de los venezolanos  
] en Colombia,» 6 02 2018.

[84 El Tiempo, «Rutas legales para que un venezolano proteja sus  
] derechos en Colombia,» 07 09 2018.



- [85 Secretaría Distrital de Seguridad, Convivencia y Justicia, Bogotá D.C.,  
] «Centro de Comando de Emergencias C-4,» 07 11 2017. [En línea].  
Available: <https://scj.gov.co/es/noticias/tecnología-punta-bogotá-transformó-su-centro-comando-emergencias-c-4>. [Último acceso: 02  
02 2018].
- [86 Ministerio de Tecnologías de la Información y las Comunicaciones,  
] República de Colombia, «Gobierno Nacional inauguró Centro de  
Mando y Control de Villavicencio,» 07 Abril 2016.
- [87 Policía Nacional de Colombia, «MODELO NACIONAL DE  
] VIGILANCIA COMUNITARIA POR CUADRANTES,» Imprenta  
Nacional, Bogotá D.C., 2012.
- [88 Policía Nacional de Colombia, «Concepto Comando y Control,»  
] Bogotá D.C., 2016.
- [89 Policía Nacional de Colombia, «MODELO NACIONAL DE VIGILANCIA  
] COMUNITARIA POR CUADRANTES,» 2018 3 7. [En línea]. Available:  
<https://www.policia.gov.co/cuadrantes>.
- [90 Caracol Radio, «Bogotá está diez veces más videovigilada que hace  
] tres años,» 03 01 2019.

- [91 ALCALDÍA MAYOR DE BOGOTÁ ALCALDE MAYOR DE BOGOTÁ  
] D.C., «Análisis demográfico y proyecciones poblacionales de Bogotá,» Bogotá D.C., 2018.
- [92 Aetina, «NVIDIA® Pascal™ GeForce 1070 MXM,» [En línea].  
] Available: <https://www.aetina.com/products-detail.php?i=87>. [Último acceso: 15 05 2018].
- [93 O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg y L. Fei-Fei, «ImageNet Large Scale Visual Recognition Challenge,» *International Journal of Computer Vision (IJCV)*, vol. 115, nº 3, pp. 211-252, 2015.
- [94 Guns.com, «Gun Review: Colt Detective Special revolver in .38 Spl,» Illinois, 2017.
- [95 W. Zeng, J. Jia, Z. Zheng, C. Xie y L. Guo, «A comparison study: Support vector machines for binary classification in machine learning,» de *Biomedical Engineering and Informatics (BMEI), 2011 4th International Conference on*, Shanghai, China, 2011.
- [96 C. Silva, D. Welfer, F. P. Gioda y C. Dornelles, «Cattle Brand Recognition using Convolutional Neural Network and Support Vector Machines,» *IEEE Latin America Transactions*, vol. 15, nº 2, pp. 310-316, 2017.

- [97 M. Nieto, L. Varona, O. Senderos, P. Leskovsky y J. Garcia, «Real-time video analytics for petty crime detection,» de *7th International Conference on Imaging for Crime Detection and Prevention (ICDP 2016)*, 2016.
- [98 T. Senst, V. Eiselein, A. Kuhn y T. Sikora, «Crowd Violence Detection Using Global Motion-Compensated Lagrangian Features and Scale-Sensitive Video-Level Representation,» *IEEE Transactions on Information Forensics and Security*, vol. 12, nº 12, pp. 2945 - 2956, 2017.
- [99 V. M. Arceda, K. F. Fabian y J. C. Gutierrez, «Real time violence detection in video,» de *Pattern Recognition Systems (ICPRS-16)*, *International Conference on*, Talca, Chile, 2016.
- [10 P. Bilinski y F. Bremond, «Human violence recognition and detection in surveillance videos,» de *Advanced Video and Signal Based Surveillance (AVSS), 2016 13th IEEE International Conference on*, Colorado Springs, CO, USA, 2016.
- [10 Z. Zhang, Z. He, G. Cao y W. Cao, «Animal Detection From Highly Cluttered Natural Scenes Using Spatiotemporal Object Region Proposals and Patch Verification,» *IEEE Transactions on Multimedia*, vol. 18, nº 10, pp. 2079 - 2092, 2016.

- [10 W. Xu, J. He, H. L. Z. B. Mao y J. Cao, «Real-Time Target Detection  
2] and Recognition with Deep Convolutional Networks for Intelligent  
Visual Surveillance,» de *Utility and Cloud Computing (UCC)*, 2016  
*IEEE/ACM 9th International Conference on*, Shanghai, China, 2016.
- [10 T. H. N. Le, Y. Zheng, C. Zhu, K. Luu y M. Savvides, «Multiple Scale  
3] Faster-RCNN Approach to Driver's Cell-Phone Usage and Hands on  
Steering Wheel Detection,» de *Computer Vision and Pattern  
Recognition Workshops (CVPRW)*, 2016 *IEEE Conference on*, Las  
Vegas, NV, USA, 2016.
- [10 K. Saleh, M. Hossny, A. Hossny y S. Nahavandi, «Cyclist detection in  
4] LIDAR scans using faster R-CNN and synthetic depth images,» de  
*Intelligent Transportation Systems (ITSC)*, 2017 *IEEE 20th  
International Conference on*, Yokohama, Japan, 2017.
- [10 H. Jiang y E. Learned-Miller2017, «Face Detection with the Faster R-  
5] CNN,» de *Automatic Face & Gesture Recognition (FG 2017)*, 2017  
*12th IEEE International Conference on*, Washington, DC, USA, 2017.
- [10 Tech power up, «NVIDIA GeForce GTX 1070 Mobile,» [En línea].  
6] Available: <https://www.techpowerup.com/gpudb/2869/geforce-gtx-1070-mobile>. [Último acceso: 25 04 2018].

[10 MSI, «GT62VR 7RE,» 2017. [En línea]. Available:  
7] <https://www.msi.com/Laptop/GT62VR-7RE-Dominator-Pro.html>.

[Último acceso: 4 25 2017].

[10 Intel, «Intel® Core™ i7-7700HQ Processor,» [En línea]. Available:  
8] [https://ark.intel.com/products/97185/Intel-Core-i7-7700HQ-](https://ark.intel.com/products/97185/Intel-Core-i7-7700HQ-Processor-6M-Cache-up-to-3_80-GHz)

[Processor-6M-Cache-up-to-3\\_80-GHz](https://ark.intel.com/products/97185/Intel-Core-i7-7700HQ-Processor-6M-Cache-up-to-3_80-GHz). [Último acceso: 25 4 2018].

[10 C. C. M. S. Rui Hou, «Tube Convolutional Neural Network (T-CNN) for  
9] Action Detection in Videos,» de *IEEE International Conference on Computer Vision (ICCV) 2017, Venice, Italy, 2017*.

[11 M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.  
0] Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat y L. Goodfellow,  
*TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, Google, 2015.

[11 Nvidia Corporation, *NVIDIA CUDA® Deep Neural Network library*  
1] (*cuDNN*), 2018.

[11 NVIDIA, *Jetson developer kit*, 2018.

2]

[11 NVIDIA Corporation, «NVIDIA GPU CLOUD,» Nvidia, 2018. [En línea].

3] Available: <https://www.nvidia.com/en-us/gpu-cloud/>. [Último acceso:  
04 20 2018].

[11 Google Inc., «Google Cloud Platfotm,» 2019.

4]

[11 Amazon Web Services, «Amazon Web Services,» 2019.

5]

[11 PriMetrica, Inc., «submarinecablemap,» TeleGeography, 23 04 2019.

6] [En línea]. Available: <https://www.submarinecablemap.com>.

[11 Policía Nacional de Colombia- Comité de Revisión Estratégica e

7] Innovación Policial (CREIP), Plan Integral Policial Para La Seguridad del Ciudadano “Corazón Verde” (PIPSC-CV), Bogotá D.C.: Imprenta Nacional de Colombia, 2013.

[11 Alcaldía Mayor de Bogotá D.C., «Numero Único de Seguridad y

8] Emergencias 123,» 2018. [En línea]. Available: <http://www.123bogota.gov.co/index.php/noticias/52-el-nuevo-centro-de-comando-control-comunicaciones-y-computo-de-bogota-c4-ya-esta-en-marcha>. [Último acceso: 25 04 2018].

[11 Alcaldía Mayor de Bogotá D.C., «Secretaría Distrital de Seguridad,

9] Convivencia y Justicia,» 10 1 2018. [En línea]. Available: <https://scj.gov.co/es/noticias/bogot%C3%A1-cuenta-m%C3%A1s-1600-c%C3%A1maras-vigilancia>. [Último acceso: 26 4 2018].

[12 Alcaldía de Medellín, «Sistema Inteligente de Movilidad,» 2013. [En 0] línea]. Available: <https://www.medellin.gov.co/simm/camaras-de-circuito-cerrado>. [Último acceso: 4 23 2018].

[12 Alcaldía de Santiago de Cali, «Alcaldía en Línea,» 2015 01 29. [En 1] línea]. Available: [http://www.cali.gov.co/alcaldenlinea/publicaciones/107192/cali\\_tendra\\_camaras\\_de\\_seguridad\\_en\\_las\\_calles/](http://www.cali.gov.co/alcaldenlinea/publicaciones/107192/cali_tendra_camaras_de_seguridad_en_las_calles/). [Último acceso: 26 04 2018].

[12 R. B. Girshick, «Fast R-CNN,» *CoRR*, vol. 1504.08083, 2015. 2]