

Tecnológico Nacional de México

Centro Nacional de Investigación
y Desarrollo Tecnológico

Tesis de Maestría

Seguimiento y caracterización automática de la
trayectoria de una persona, en un entorno multi-
cámaras de videovigilancia

presentado por

Ing. Damián Uriel Rosado Castellanos

como requisito para la obtención del grado de
Maestro en Ciencias de la Computación

Director de tesis

Dra. Andrea Magadán Salazar

Codirector de tesis

Dra. Daniela Alejandra Moctezuma Ochoa

Cuernavaca, Morelos, México. Enero de 2019.



SEP
SECRETARÍA DE
EDUCACIÓN PÚBLICA



TECNOLÓGICO NACIONAL DE MEXICO

Centro Nacional de Investigación y Desarrollo Tecnológico

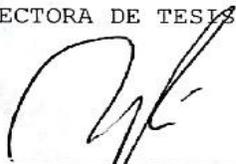
Cuernavaca, Morelos, **17/Diciembre/2018**
OFICIO No. DCC/290/2018

Asunto: Aceptación de documento de tesis

DR. GERARDO V. GUERRERO RAMÍREZ
SUBDIRECTOR ACADÉMICO
PRESENTE

Por este conducto, los integrantes de Comité Tutorial del Ing. **Damián Uriel Rosado Castellanos**, con número de control M17CE046, de la Maestría en Ciencias de la Computación, le informamos que hemos revisado el trabajo de tesis profesional titulado "**Seguimiento y caracterización automática de la trayectoria de una persona, en un entorno multi-cámaras de videovigilancia**" y hemos encontrado que se han realizado todas las correcciones y observaciones que se le indicaron, por lo que hemos acordado aceptar el documento de tesis y le solicitamos la autorización de impresión definitiva.

DIRECTORA DE TESIS



Dra. Andrea Magadán Salazar
Doctorado en Ciencias
Computacionales
10654097

CO-DIRECTOR DE TESIS



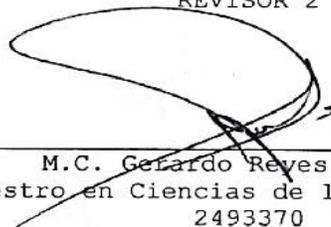
Dra. Daniela Alejandra Moctezuma Ochoa
Doctorado en tecnologías de la
información y sistemas informáticos

REVISOR 1



Dr. José Ruiz Ascencio
Doctor en Ciencias
5009035

REVISOR 2



M.C. Gerardo Reyes Salgado
Maestro en Ciencias de la Computación
2493370

C.p. M.T.I. María Elena Gómez Torres - Jefa del Departamento de Servicios Escolares.

Estudiante
Expediente

NACS/lmz

cenidet

Centro Nacional de Investigación
y Desarrollo Tecnológico

Interior Internado Palmira S/N, Col. Palmira, C. P. 62490, Cuernavaca, Morelos
Tel. (01) 777 3 62 77 70, ext. 4106, e-mail: dir_cenidet@tecnm.mx

www.tecnm.mx | www.cenidet.edu.mx





SEP
SECRETARÍA DE
EDUCACIÓN PÚBLICA



TECNOLÓGICO NACIONAL DE MEXICO

Centro Nacional de Investigación y Desarrollo Tecnológico

Cuernavaca, Mor., 8 de enero de 2019
OFICIO No. SAC/004/2019

Asunto: Autorización de impresión de tesis

ING. DAMIÁN URIEL ROSADO CASTELLANOS
CANDIDATO AL GRADO DE MAESTRO EN CIENCIAS
DE LA COMPUTACIÓN
PRESENTE

Por este conducto, tengo el agrado de comunicarle que el Comité Tutorial asignado a su trabajo de tesis titulado "Seguimiento y Caracterización Automática de la Trayectoria de una Persona, en un Entorno Multi-Cámaras de Videovigilancia", ha informado a esta Subdirección Académica, que están de acuerdo con el trabajo presentado. Por lo anterior, se le autoriza a que proceda con la impresión definitiva de su trabajo de tesis.

Esperando que el logro del mismo sea acorde con sus aspiraciones profesionales, reciba un cordial saludo.

ATENTAMENTE

Excelencia en Educación Tecnológica®
"Conocimiento y tecnología al servicio de México"

DR. GERARDO VICENTE GUERRERO RAMÍREZ
SUBDIRECTOR ACADÉMICO



SEP TecNM
CENTRO NACIONAL
DE INVESTIGACIÓN
Y DESARROLLO
TECNOLÓGICO
SUBDIRECCIÓN
ACADÉMICA

C.p. M.T.I. María Elena Gómez Torres.- Jefa del Departamento de Servicios Escolares.
Expediente

GVGR/mcr

cenidet[®]
Centro Nacional de Investigación
y Desarrollo Tecnológico

Interior Internado Palmira S/N, Col. Palmira, C. P. 62490, Cuernavaca, Morelos.
Tel. (01) 777 3 62 77 70, ext. 4106, e-mail: dir_cenidet@tecnm.mx
www.tecnm.mx | www.cenidet.edu.mx



Dedicatoria



A mi madre.

A ti que me diste todo, sin pedir nada.

A ti que dejaste todo por mí.

A ti que entregaste todo por mí...

Gracias.

Agradecimientos

Un agradecimiento al Consejo Nacional de Ciencia y Tecnología, (CONACYT). El apoyo brindado me permitió realizar los estudios de postgrado, así como realizar este trabajo de investigación.

Un agradecimiento al Centro Nacional de Investigación y Desarrollo Tecnológico, (CENIDET). Ser parte de esta institución fue un gran honor para mí, los conocimientos adquiridos por parte de los doctores y maestros que laboran en la institución me ayudaron a consolidar mi preparación personal y profesional.

Un agradecimiento a la Dra. Andrea Magadán, directora de esta tesis, por la paciencia, tiempo y entusiasmo dedicado durante la realización de este trabajo. De igual forma, un agradecimiento a la Dra. Daniela Moctezuma, co-directora de esta tesis, por todas las atenciones brindadas durante la estancia en el Centro de Investigación en Ciencias de Información Geoespacial, (CENTROGEO).

Un agradecimiento a mis revisores de tesis: Dr. José Ruiz y Dr. Gerardo Reyes. Todas las observaciones, sugerencias y tiempo dedicado fueron parte importante para la elaboración de este trabajo.

Un agradecimiento a todos los doctores de la línea de investigación de inteligencia artificial que me transmitieron su conocimiento en el aula de clases: Dr. Dante Mujica, Dr. Manuel Mejía, Dra. Andrea Magadán y Dr. Gerardo Reyes.

Un agradecimiento a mis amigos de CENIDET: Luis Montalvo, Axel Rojas, Gerardo Romero, Iván Fuentes, Oswaldo Barrera, Pablo Padilla. Todas las vivencias con ustedes fueron grandiosas en todo este periodo de estudios.

Seguimiento y caracterización automática de la trayectoria de una persona, en un entorno multi-cámaras de videovigilancia

Resumen

En esta investigación se presenta un sistema de videovigilancia multi-cámaras en el cual, el usuario específico la persona de interés a seguir, la cual es re-identificada en las diferentes cámaras que conforma el sistema, independientemente de los cambios de iluminación, forma, escala, y traslación de las personas. De igual forma, el sistema tiene la capacidad de clasificar la acción que realiza la persona (*caminar derecho, girar izquierda, girar derecha, quedarse parado*) mientras pueda ser re-identificada. Para el desarrollo de este sistema se propuso una estructura compuesta por siete módulos: adquisición de imágenes, selección de persona, descripción de la persona, detección de personas, re-identificación, seguimiento y análisis de trayectorias.

Este trabajo de tesis precisa dos campos de investigación relacionados a los sistemas de videovigilancia inteligente: re-identificación de personas y análisis de trayectorias. Para la re-identificación de personas se proponen tres métodos diferentes de re-identificación (Re-identificación mediante la descripción del torso, re-identificación mediante descriptores de color y re-identificación mediante descriptores de color de segmentos del cuerpo humano.) y para el análisis de trayectorias un método (Determinación de acciones mediante orientaciones y pendientes).

La experimentación realizada se hizo mediante una base de datos propia y se compone de 16 personas. Los resultados obtenidos demuestran que los métodos propuestos son lo suficientemente robustos para las tareas de re-identificación de personas y análisis de trayectorias.

Monitoring and automatic characterization of the trajectory from a person, in a multi-camera video surveillance environment

Abstract

In this work we present an intelligent video surveillance system in which, the user of the system has to select a person where is a suspicious person and then the system is able to do the re-identification task in the different cameras. This task is done independently of the changes in the lighting, form, scale, and translation of persons. In the same way, the system has the capacity to classify the action that the person performs (*walk straight, turn left, turn right, stand still*) as long as it can be re-identified. The development of this system has a structure composed of seven modules: image acquisition, person selection, descriptor of the person, person detection, re-identification, tracking and finally trajectory analysis.

This thesis requires two fields of research: person re-identification and trajectory analysis. Where for the person re-identification three different methods are proposed and for the trajectory analysis one method is proposed.

The experimentation was done through an own database composed from 16 people. The results obtained show that the methods proposed are sufficiently robust for the tasks of re-identification and trajectory analysis

Contenido

Resumen.....	i
Abstract.....	ii
Contenido.....	iii
Índice de figuras.....	vi
Índice de tablas.....	viii
Capítulo I.....	1
Planteamiento del problema.....	1
1.1 Introducción.....	1
1.2 Antecedentes del proyecto.....	2
1.3 Descripción del problema.....	3
1.4 Propuesta de solución.....	4
1.5 Organización de la tesis.....	7
Capítulo II.....	8
Estado del arte.....	8
2.1 Videovigilancia inteligente.....	8
2.2 Sustracción de fondo.....	12
2.3 Re-identificación de personas.....	21
2.4 Análisis de trayectorias.....	40
Capítulo III.....	48
Marco teórico.....	48
3.1 Modelo de mezclas gaussianas.....	48
3.2 Filtro gaussiano.....	49
3.3 Operadores morfológicos.....	50
3.4 Histograma de gradientes orientados.....	50
3.5 Máquinas Vector Soporte.....	51
3.6 Promedio de gradiente.....	53
3.7 Medidas de distancia.....	54
3.8 Medidas estadísticas.....	55
3.9 Espacios de color.....	56
3.10 Transformaciones de los espacios de color.....	58
3.11 RandomForest.....	60
3.12 Métricas de evaluación.....	61

Capítulo IV	64
Análisis del problema y propuesta de solución	64
4.1 Problemáticas de un sistema de videovigilancia inteligente	64
4.2 Propuesta de solución.....	66
Capítulo V	68
Análisis, diseño e implementación del sistema	68
5.1 Análisis y diseño del sistema	68
5.2 Base de datos.....	69
5.3 Adquisición de imágenes	71
5.4 Selección de la persona	73
5.5 Descripción de la persona	74
5.6 Detección de personas.....	75
5.7 Re-identificación	78
5.8 Seguimiento.....	86
5.9 Análisis de trayectorias	87
Capítulo VI.....	90
Experimentación y resultados	90
6.1 Re-identificación	90
6.2 Análisis de trayectorias	96
6.3 Conclusión.....	99
Capítulo VII	100
Conclusiones y trabajos futuros	100
7.1 Objetivos alcanzados.....	100
7.2 Conclusiones	101
7.3 Aportaciones	102
7.4 Trabajo futuro.....	103
7.5 Productos adicionales y participaciones.....	103
Referencias	106
<i>Bibliográficas</i>	106
<i>Electrónicas</i>	110
<i>Tesis de maestría</i>	110
Anexos	112
Anexo 1	113
Anexo 2	114

Anexo 3.....	115
Anexo 4.....	116
Anexo 5.....	117
Anexo 6.....	120
Anexo 7.....	121
Anexo 8.....	122
Anexo 9.....	123
Anexo 10.....	124
Anexo 11.....	125
Anexo 12.....	126
Anexo 13.....	127
Anexo 14.....	128
Anexo 15.....	129

Índice de figuras

Figura 1.1. Esquema del método de solución.	5
Figura 2.1. Diferentes áreas de investigación de un sistema de videovigilancia multi-cámaras inteligente. Las flechas indican el flujo de información en los diferentes módulos. (Wang X. , 2013).....	8
Figura 2.2. Aspectos relevantes que se relacionan con los sistemas de video.	9
Figura 2.3. Tareas que puede realizar un ISV.	9
Figura 2.4. Métodos para dotar de Inteligencia un sistema de videovigilancia.	10
Figura 2.5. Método propuesto para la re-identificación. Las características de apariencia primero son extraídas de un conjunto de imágenes de entrenamiento y proyectadas dentro de un sub-espacio RCCA con matrices de proyección aprendidas. Una instancia de descriptores de referencia es generada y concatenando sus puntajes de similitud con respecto al conjunto de referencia. La coincidencia inicial es realizada usando los descriptores de referencia y posteriormente se utiliza un esquema de rasgo sobresaliente para mejorar la precisión de re-identificación.	21
Figura 2.6. Muestra de la detección de rasgo sobresaliente en dos vistas de cámaras.	22
Figura 2.7. Esquema general de HOGG.....	23
Figura 2.8. Diagrama de flujo del aprendizaje para el mapeo y la codificación. Las entradas son muestras aleatorias extraídas de los patches x^1 y x^2 corresponden a una muestra de pixeles procedentes de las partes correspondientes de los pares de imágenes coincidentes que deben ser del mismo color. Mapping: el autocodificador mapea las entradas en un espacio de 60 dimensiones que da las representaciones de x^1 y x^2 . Con base en el diccionario “D”, los dos conjuntos en el espacio dimensional son codificados y se obtiene dos conjuntos de códigos dispersos.	24
Figura 2.9. Vista de una imagen A y B donde se extrae una región que debería corresponder al mismo color.	24
Figura 2.10. Diagrama general de un sistema basado en modelo de apariencias softbiométricas.....	25
Figura 2.11. Ejemplo de la división de un tracklet	26
Figura 2.12. Ejemplo de drásticas variaciones de apariencia con el seguimiento de un mismo objetivo.	27
Figura 2.13. Diagrama para la detección de la longitud del cabello.	28
Figura 2.14. Videos de la detección de cabello.	28
Figura 2.15. Sistema general propuesto.	28
Figura 2.16. Bolsa de características softbiométricas.	29
Figura 2.17. Esquema general del sistema propuesto.	29
Figura 2.18. Esquema general del sistema.	30
Figura 2.19. Modelo de Partición Jerárquica.	30
Figura 2.20. Esquema general del primer descriptor.	31
Figura 2.21. Esquema para la generación del descriptor.	32
Figura 2.22. Ejemplo de restricción de regiones, donde no es posible detectar una ruta mínima cuando una persona se mueva a través del pasto.	50
Figura 2.23. Ejemplo del método Fixed potential. El sistema detecta como comportamiento anormal las rutas que tienen accesos restringidos.	51
Figura 2.24. Ejemplo de los Route sets.	41
Figura 2.25. Método propuesto para la clasificación de trayectorias.....	52
Figura 2.26. Diagrama del método propuesto.	42
Figura 2.27. Comportamientos anormales a evaluar.	42
Figura 2.28. Diagrama general del sistema propuesto.	53
Figura 2.29. Donde (a) es el frame original, (b) es la detección de los objetivos y (c) es la extracción de la trayectoria.	43
Figura 2.30. Donde (a) es la secuencia distancia euclidea entre posiciones adyacente y (b) es la representación simbólica correspondiente.....	54
Figura 2.31. Esquema general del análisis de trayectorias con Data Mining.....	55

Figura 3.1. a) es la imagen actual. b) es el modelo. c) es la detección de objetos de primer plano y d) es la información superpuesta.	49
Figura 3.2. Elemento estructurante.	50
Figura 3.3. a) histograma de las orientaciones. b) imagen del histograma acumulado. (Ernest Valveny, 2018)	51
Figura 3.4. Concatenación de los histogramas de los bloques. b) descriptor resultante, (Ernest Valveny, 2018).	51
Figura 3.5. Ejemplos de entrada linealmente separables.	52
Figura 3.6. Representación de márgenes y vectores soporte.	52
Figura 3.7. Histograma del promedio de gradiente de una persona.	53
Figura 3.8. a) Histograma de gradiente, b) Segmentación del cuerpo.	54
Figura 3.9. Representación del espacio de color RGB.	57
Figura 3.10. Representación del espacio de color HLS.	57
Figura 3.11. Representación del espacio de color HSV.	58
Figura 3.12. Ejemplo de un árbol de decisión.	60
Figura 4.1. Arquitectura general del sistema propuesto.	66
Figura 5.1. Arquitectura del sistema.	68
Figura 5.2. Interfaz del sistema.	69
Figura 5.3. Cámara de videovigilancia.	70
Figura 5.4. Vista de cámaras. Lado izquierdo vista 1, medio vista 2, derecho vista 3.	70
Figura 5.5. Trayectoria de una persona, en diferentes vistas de cámara.	71
Figura 5.6. Estructura del almacenamiento de trayectorias.	71
Figura 5.7. rojo) Canales de reproducción RGB azul) Canales de reproducción de modelo de fondo.	72
Figura 5.8. Topología de cámaras.	72
Figura 5.9. Flujo de trabajo del módulo de selección de persona.	73
Figura 5.10. Selección de peatones.	74
Figura 5.11. Representación de la persona mediante características softbiométricas.	74
Figura 5.12. Esquema general del módulo de detección de personas.	75
Figura 5.13. Modelo de fondo sin procesamiento.	76
Figura 5.14. Procesamiento del modelo de fondo.	76
Figura 5.15. Modelo de fondo procesado.	76
Figura 5.16. Imágenes que pertenecen a la clase no persona.	77
Figura 5.17. Esquema de funcionamiento del detector de objetos.	77
Figura 5.18. Información superpuesta del modelo de fondo y la imagen RGB de un objeto en movimiento que pertenece a la clase persona.	78
Figura 5.19. a) Imagen de la persona. b) Imagen binaria.	78
Figura 5.20. Segmentación de partes de la persona.	79
Figura 5.21. a) Cuerpo completo. b) Extracción del torso.	79
Figura 5.22. Histograma del torso y generación del descriptor.	79
Figura 6.1. Personas de la base de datos de re-identificación.	90
Figura 6.2. Gráfica de detección de personas.	94
Figura 6.3. Curva ROC de los métodos.	96
Figura 6.4. Trayectorias de la secuencia de video.	97
Figura 6.5. Segmentos de recta.	97
Figura 6.6. Sistema adicional creado.	104

Índice de tablas

Tabla 2.1. Aportaciones de los artículos de videovigilancia inteligente.....	11
Tabla 2.2. Comparativa de los métodos.....	15
Tabla 2.3. Evaluación de desempeño de algoritmos.....	16
Tabla 2.4. Métodos de la literatura clasificados.	17
Tabla 2.5. Modelos Estadísticos Categorizados.	17
Tabla 2.6. Comparativa de los métodos estadísticos.	18
Tabla 2.7. Aportaciones de los artículos de sustracción de fondo.	19
Tabla 2.8. Características Softbiométricas utilizadas para la detección.	26
Tabla 2.9. Aportaciones de los artículos de re-identificación de personas.	36
Tabla 2.10. Modelos de la metodología propuesta.	40
Tabla 2.11. Aportaciones de los artículos de análisis de trayectorias.	47
Tabla 3.1. Definición de términos.	66
Tabla 3.2. Matriz de confusión.	71
Tabla 5.1. Vector de características de la persona seleccionada.	74
Tabla 5.2. Vector de características.	80
Tabla 5.3. Evaluadores y métodos de búsqueda utilizados para la selección de variables.	82
Tabla 5.4. Variables con mayor discriminación de los métodos.	82
Tabla 5.5. Vector de características.	83
Tabla 5.6. Selección de variables de la cabeza.	84
Tabla 5.7. Selección de variables del torso.	84
Tabla 5.8. Selección de variables de las piernas.	85
Tabla 5.9. Vector de características de la cabeza.	85
Tabla 5.10. Vector de características del torso.	85
Tabla 5.11. Vector de características de las piernas.	86
Tabla 5.12. Ejemplo de la lista de almacenamiento.	87
Tabla 6.1. Resultados del método 1.....	91
Tabla 6.2. Resultados del método 2.....	92
Tabla 6.3. Resultados del método 3.....	93
Tabla 6.4 Tiempos de ejecución de los métodos.	94
Tabla 6.5. Resultados del método 2 con RandomForest.....	95
Tabla 6.6. Resultados con un ángulo de 22.5 grados.....	98
Tabla 6.7. Resultados con un ángulo de 12 grados.....	98
Tabla 7.1. Comentarios de los objetivos específicos.	98
Tabla A1.1. Diferentes tasas de TPR y FPR de acuerdo al umbral.	120
Tabla A2.1. Evaluación de segmentos del cuerpo.	114
Tabla A3.1. Matriz de confusión del modelo generado.....	115
Tabla A4.1. Variación del color del valor H.	116
Tabla A5.1. Vector de características extendido.	117
Tabla A6.1. Resultado de los algoritmos de selección de variables y métodos de búsqueda con sus pesos. .	120
Tabla A7.1. Vector final de descripción de la persona para las variantes del método 2.....	121
Tabla A8.1. Resultados de re-identificación.....	122
Tabla A9.1. Resultados de re-identificación.....	123

Planteamiento del problema

En este capítulo se presentan la descripción del problema, la propuesta de solución, los antecedentes del trabajo de tesis, así como la estructura del presente documento de tesis.

1.1 Introducción

El abaratamiento de dispositivos visuales como lo son las cámaras de video, ha permitido que hoy en día, existan gran número de sistemas de videovigilancia tradicionales instalados tanto en espacios públicos, como privados. Un sistema de videovigilancia tradicional consiste en monitorear el comportamiento, actividades y otros cambios en el entorno, por medio de operadores visuales (humanos) que intentan proporcionar seguridad a la zona vigilada, (Moctezuma, D., 2016).

La supervisión humana en sistemas de videovigilancia resulta, en la mayoría de los casos, ineficiente y en costos elevados para las empresas o particulares. El simple hecho de procesar diferentes pantallas con una cantidad masiva de información por los seres humanos resulta una tarea difícil de realizar. Esto se debe principalmente a las necesidades fisiológicas de los mismos, que limitan la disponibilidad de supervisión continua; además, de la incapacidad de atención, en periodos largos, causada por un agotamiento visual y de entornos repetitivos visualizados en las diferentes pantallas del sistema de videovigilancia.

La videovigilancia inteligente ha sido una de las áreas de investigación más activas que están relacionadas a la visión artificial y que todavía no cuenta con un sistema genérico que logre hacer todas las tareas que involucra. Esta consiste en la extracción de información útil de diferentes vistas de cámaras de videovigilancia para la detección automática, el seguimiento y reconocimiento de objetos de interés, así como el entendimiento de sus actividades, (Wang, X., 2013). En la última década, esta área se ha convertido en una importante temática, debido a la necesidad de reemplazar la videovigilancia tradicional por sistemas automáticos de análisis de video y el reconocimiento automático de las actividades humanas.

La identificación de personas es una de las principales tareas de un sistema de videovigilancia inteligente, en donde, generalmente se utiliza información biométrica que ayude a distinguir o identificar a la persona bajo análisis, a pesar de la variabilidad en aspectos como la pose, las condiciones de iluminación, vistas, etc., debido al sujeto, ambiente y a las cámaras. Sin embargo, también existen otras características menos complejas de adquirir denominadas softbiométricas que son de gran ayuda para definir la identidad de las personas bajo observación. Algunas de estas características son, por ejemplo: el color de piel, de cabello, la ropa, la altura, la forma de andar, etc.

El análisis de las trayectorias es una valiosa fuente de información para identificar, de manera automática, comportamientos específicos o sospechosos, llevado a cabo por los objetos de interés. No obstante, el caracterizar y entender que sucede en un video, es un problema abierto

de investigación, ya que el ambiente puede ser muy dinámico, complejo y requiere de un sistema de múltiples cámaras para monitorear a las personas de forma constante.

Por lo anterior mencionado, existe la necesidad de proponer tecnologías mediante algoritmos que logren identificar y re-identificar personas, así como realizar un análisis de sus actividades, de manera automática, en los diferentes entornos en que los sistemas de videovigilancia pueden ser instalados, así como en las diferentes vistas que estas puedan tener.

En este trabajo de tesis se desarrolló e implementó un sistema de visión artificial que logra realizar la tarea de re-identificación de personas en un entorno multi-cámaras de videovigilancia, así como definir actividades (*quedarse parado, caminar derecho, girar a la derecha, girar a la izquierda*) mediante un análisis de trayectorias.

1.2 Antecedentes del proyecto

Esta sección presenta una breve descripción de los trabajos de investigación desarrollados por parte del departamento de ciencias computacionales del Centro Nacional de Investigación y Desarrollo Tecnológico (CENIDET), los cuales son los antecedentes de este trabajo actual, ya que tienen una relación de seguimiento. En total se cuenta con cinco investigaciones de maestría.

1. **Caracterización de Imágenes en Movimiento: Caminar y Correr, (Saldaña García, 2007).** Este trabajo de investigación diseño e implemento un sistema que realiza la caracterización automática de dos acciones: correr y caminar, considerando una vista lateral del sujeto. Para lo cual se lleva a cabo un procesamiento y análisis de una secuencia de imágenes en dos dimensiones, aquí se creó una base de videos que contiene la información para caracterizar dichas acciones y esto le permitió lograr patrones de posición de las partes del cuerpo al realizarlos.
2. **Caracterización Automática de una Muestra de Acciones de Video, (Dávila García, 2009).** La investigación consistió en el desarrollo de un sistema capaz de detectar, seguir, caracterizar y cuantificar de manera automática los movimientos presentados por distintas partes del cuerpo humano (cabeza, manos y brazos) en movimiento al realizar las acciones de visualizar, aproximarse y tocar una zona específica con el fin de obtener una descripción numérica de las características más importantes de estas acciones.
3. **Seguimiento y Caracterización del Comportamiento de Varios Objetos en Movimiento, (Toscano Martínez, 2011).** En esta investigación se realizó un sistema que permite la detección, seguimiento y caracterización del comportamiento de personas en movimiento. Estos movimientos son: caminar, correr y levantar las manos.
4. **Agrupamiento de Trayectorias para la Generación de un Modelo de Actividad Adaptativo, (de León Hernández, 2012).** La investigación se enfocó en proponer un modelo de aprendizaje no supervisado para la identificación automática de comportamientos anormales de personas.

5. **Caracterización Visual de Movimientos Sospechosos de Personas en Estacionamientos**, (Alcantar Juárez, 2012). El trabajo de investigación desarrollo un sistema de seguimiento de personas y caracterización de cinco movimientos: correr, levantar las manos, merodear, agacharse y acostarse.

Los trabajos de investigación anteriores, presentan interesantes metodologías para realizar la detección y seguimiento de personas, así como la caracterización de movimientos, Sin embargo, estos no son capaces de identificar a la misma persona ante diferentes vistas. De igual forma la caracterización de acciones está relacionada a la acción de la persona (ejemplo, levantar las manos) y no realizan un análisis de la trayectoria, salvo en la investigación de (de León Hernández, 2012), pero este enfocado al agrupamiento de trayectorias. Adicionalmente todas las tesis listadas trabajan en un ambiente mono-cámara.

El presente trabajo a diferencia de los anteriores, se enfoca en un entorno multi-cámaras de videovigilancia, así como la identificación y re-identificación de una persona en diferentes escenarios, y la definición de una acción basado en el análisis de trayectorias.

1.3 Descripción del problema

Los sistemas de videovigilancia tradicionales dependen del monitoreo de seres humanos para reportar diferentes actos delictivos. Sin embargo, para el personal de vigilancia, monitorear de forma precisa y eficaz diferentes escenarios del sistema las 24 horas del día, es una tarea complicada. Esto se debe principalmente a las necesidades fisiológicas que se requieren cubrir (alimentación, descanso, entre otros), así como los entornos repetitivos y la cantidad de diferentes vistas de cámaras que existen en un sistema de videovigilancia que tienen que ser analizados de forma continua, los cuales conllevan un cansancio tanto visual como mental al operador. Todo lo mencionado en conjunto reduce la alta disponibilidad de supervisión del sistema, situación que puede llevar a descuidos y reducir la tasa de efectividad de prevención y corrección de actos delictivos.

Adicionalmente, la videovigilancia inteligente es un tema abierto de investigación en la visión por computadora, ya que existen diferentes factores (*cambios de iluminación, oclusión, traslape entre objetos, variaciones en la forma y escala, así como la rotación y translación de los mismos*) que dificultan que exista un sistema general inteligente que logre hacer un análisis automático de diferentes vistas de cámaras en los diferentes entornos de instalación.

Por lo anterior, es necesario contar con un sistema de videovigilancia inteligente que pueda re-identificar personas y además que determine la acción que realiza la persona en su trayectoria mientras pueda ser detectada en las diferentes vistas. Esto con la finalidad de auxiliar a los operadores del sistema que vigilan las cámaras y aumentar la disponibilidad de supervisión, lo que se traduce en un aumento en la calidad de la tarea de videovigilancia.

1.4 Propuesta de solución

Esta sección describe los objetivos, alcances, limitaciones y metodología de solución propuestos para el desarrollo de este trabajo de investigación, que tiene como fin un sistema de videovigilancia inteligente.

1.4.1 Objetivos

Objetivo General

Proponer e implementar un sistema de visión artificial que sea capaz de realizar el seguimiento y caracterización automática de la trayectoria de una persona, en un entorno multicámaras de videovigilancia.

Objetivos Específicos

- Implementar un entorno de usuario que permita seleccionar a una persona de interés en las vistas de las cámaras.
- Implementar un módulo de descripción que permita obtener las características de la persona seleccionada.
- Implementar un módulo de detección de personas en movimiento que permita segmentar y obtener la región de interés.
- Implementar un módulo de identificación y re-identificación de la persona que determine si la región de interés corresponde a la persona seleccionada.
- Implementar un módulo de seguimiento de la persona que obtendrá la trayectoria de la persona seleccionada.
- Implementar un módulo de análisis de trayectorias que permita clasificar la trayectoria de la persona seleccionada.

1.4.2 Alcances y Limitaciones

Alcances

- El sistema proporciona una interfaz amigable que permita a un usuario seleccionar a una persona.
- El sistema tiene la capacidad de describir a una persona a través de sus características softbiométricas.
- El sistema tiene la capacidad de segmentar las diferentes vistas y obtener regiones de interés.
- El sistema evalúa las regiones de interés y determina si la región de la imagen corresponde a la persona seleccionada inicialmente o no.

- El sistema tiene la capacidad de obtener las coordenadas de la persona seleccionada y posteriormente generar su trayectoria.
- El sistema tiene la capacidad de caracterizar la trayectoria de la persona con una actividad de desplazamiento en particular.

Limitaciones

- Los procesos de descripción, identificación y re-identificación se realizan por medio de características softbiométricas.
- Los procesos de descripción, identificación y re-identificación sólo se realizan mientras la persona de interés pueda ser detectada en alguna de las vistas disponibles del ambiente de video vigilancia.
- La localización inicial de la persona a seguir se hace de manera semi-automática mediante la especificación del usuario con el ratón.
- La caracterización de la trayectoria sólo involucra las siguientes actividades: caminar recto, girar a la izquierda, girar a la derecha y quedarse parado.
- La caracterización de la trayectoria se hace únicamente con conocimiento a posteriori.
- Las vistas de las personas a seguir podrán ser de frente, posterior y lateral.
- En el seguimiento de la persona no se considera la oclusión.
- El sistema de seguimiento no realiza un reconocimiento facial sobre la identidad de la persona.
- El sistema de videovigilancia trabaja con un ambiente de multicámaras de al menos tres vistas.

1.4.3 Metodología de solución

Esta sección describe un método que permite resolver el objetivo planteado en este trabajo de investigación. Este se compone de 7 fases que comprenden los módulos del sistema y dos fases adicionales relacionados a la integración de los módulos y la experimentación, ver figura 1.1

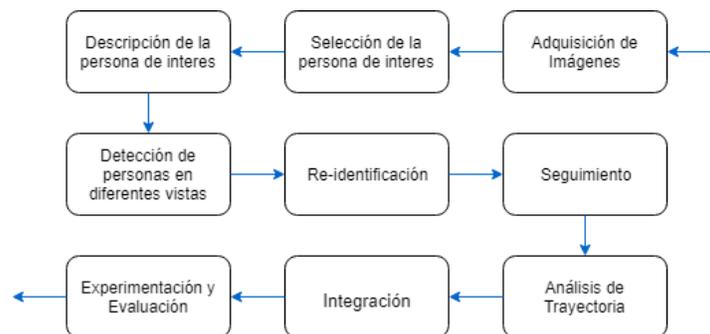


Figura 1.1. Esquema del método de solución.

A) Adquisición de Imágenes

Este apartado describe la adquisición de las secuencias de imágenes(video) que serán utilizadas en el método propuesto. La adquisición de imágenes consiste en la creación de una base de datos personal de un sistema de cámaras de videovigilancia instalados en una casa, así como la lectura de estas imágenes por parte del sistema.

B) Selección de la persona de interés

Este apartado describe la selección de una persona de interés que está bajo estudio. Es decir, se realiza la selección de una persona que aparece en las secuencias de imágenes, la cual posteriormente será re-identificada y evaluada con base en su trayectoria.

C) Descripción de la persona de interés

Este apartado es el encargado de describir a la persona seleccionada en término de sus características softbiométricas.

D) Detección de personas en diferentes vistas

Este apartado es el encargado de realizar la segmentación de objetos en movimiento. Todos los objetos que se detecten en contraste con el fondo son segmentados. Es importante resaltar que los objetos en movimiento son considerados exclusivamente personas.

E) Re-identificación

Este apartado es el responsable de poder determinar si el objeto segmentado corresponde a la persona seleccionada con base en sus características softbiométricas. En otras palabras, el sistema decide si es la misma persona o no.

F) Seguimiento

Este apartado es el encargado de almacenar las posiciones del centroide del área segmentada sólo si en el apartado de re-identificación existe la coincidencia de la persona. De igual forma una vez que las posiciones estén almacenadas se genera la trayectoria de la persona.

G) Análisis de Trayectoria

Este apartado determina el movimiento que se realizó en la trayectoria. En otras palabras, evalúa el comportamiento de la trayectoria: caminar recto, girar a la izquierda, girar a la derecha y quedarse parado.

H) Integración

Este apartado consiste en la integración de cada fase del sistema desarrollado. La integración aborda la fase A hasta la fase G para obtener un sistema completo de análisis de trayectorias.

D) *Experimentación y Evaluación*

Este apartado consiste en realizar experimentos con la base de datos generada en la fase A y evaluar el rendimiento del sistema con base en los resultados obtenidos a partir de diferentes medidas de evaluación.

1.5 Organización de la tesis

El presente documento de tesis, se encuentra estructurado en 8 capítulos. De igual forma se tiene un apartado de anexos. A continuación, se describen los capítulos restantes brevemente:

Capítulo 2. Estado del arte

El capítulo 2 muestra los trabajos relacionados sobre videovigilancia, re-identificación de personas y análisis de trayectorias, así como un análisis general de cada artículo.

Capítulo 3. Análisis del problema y propuesta de solución

El capítulo 3 muestra un análisis y un planteamiento de la metodología de solución, la cual es la base del desarrollo de un sistema de videovigilancia inteligente cuyo fin es la re-identificación y el análisis de trayectorias.

Capítulo 4. Marco teórico

El capítulo 4 presenta los diferentes métodos, utilizados en el sistema de videovigilancia inteligente.

Capítulo 5. Análisis, diseño e implementación del sistema

El capítulo 5 detalla todo el trabajo realizado para el desarrollo del sistema. Es importante mencionar que el desarrollo se basa en el planteamiento de solución documentado en el capítulo 3.

Capítulo 6. Experimentación y resultados

El capítulo 6 presenta la experimentación realiza para las tareas de re-identificación y análisis de trayectorias, pertinentes al sistema de videovigilancia inteligente. Así como los resultados obtenidos que demuestran la efectividad de los métodos propuestos para las tareas.

Capítulo 7. Conclusiones y trabajo futuro.

El capítulo 7 brinda información detallada de los resultados obtenidos en la experimentación sobre los métodos propuestos. De igual forma se hace mención del trabajo futuro que se puede desarrollar a partir de este trabajo de investigación, así como los productos adicionales creados durante el periodo de estudios de la maestría.

Estado del arte

En este capítulo se detalla la recopilación y estudio del estado del arte de los trabajos relacionados recientes que tienen relación directa o parcial con este trabajo de investigación. En él se plantean los conocimientos actuales sobre los temas de videovigilancia inteligente, sustracción de fondo, re-identificación de personas y análisis de trayectorias.

2.1 Videovigilancia inteligente

2.1.1 Intelligent multi-camera video surveillance: A review, (Wang X., 2013)

El presente artículo muestra una reseña de las recientes áreas de investigación desde la perspectiva de visión artificial y reconocimiento de patrones, que se están aplicando en la videovigilancia inteligente.

De manera general, la figura 2.1 presenta un esquema de las diferentes áreas de investigación y como estas pueden relacionarse entre sí para poder generar un sistema de videovigilancia multicámaras inteligente robusto.

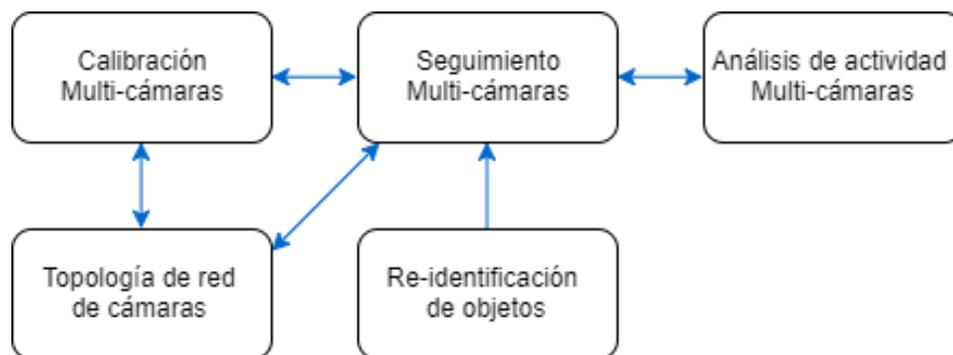


Figura 2.1. Diferentes áreas de investigación de un sistema de videovigilancia multi-cámaras inteligente. Las flechas indican el flujo de información en los diferentes módulos. (Wang X., 2013).

Calibración multi-cámara: Traza los mapas de las diferentes vistas a un sólo sistema de coordenadas.

Topología de la red de cámaras: Identifica si las vistas de las cámaras están sobrepuestas o son espacialmente adyacentes y describe la transición de tiempo de los objetos en las vistas de las cámaras.

Re-identificación de objetos: Observa dos regiones en diferentes vistas de cámara y determina si estas pertenecen al mismo objeto o no. Utiliza exclusivamente la información de apariencia y no de espacio-tiempo.

Seguimiento multi-cámara: Sigue a los objetos a través de las vistas de las cámaras.

Análisis de las actividades multi-cámara: Reconoce automáticamente las actividades de diferentes categorías y detecta actividades anormales utilizando la información de las diferentes vistas de las cámaras.

2.2.2 Intelligent Video Systems and Analytics: A Survey, (Liu, Chen, & Kubota, 2013)

El presente artículo provee un entorno general de las aplicaciones de los sistemas de videovigilancia inteligente como lo son las áreas de administración, salud y transporte. De igual forma se destaca la arquitectura de los sistemas, las tareas del mismo y métodos de análisis. De manera general se abordan tres aspectos importantes en los sistemas de videos inteligentes (*ISV, por sus siglas en ingles*).

Sistemas de Video: Este apartado aborda los aspectos relevantes que se deben tener en cuenta en los sistemas de video vigilancia como se muestra en la figura 2.2.

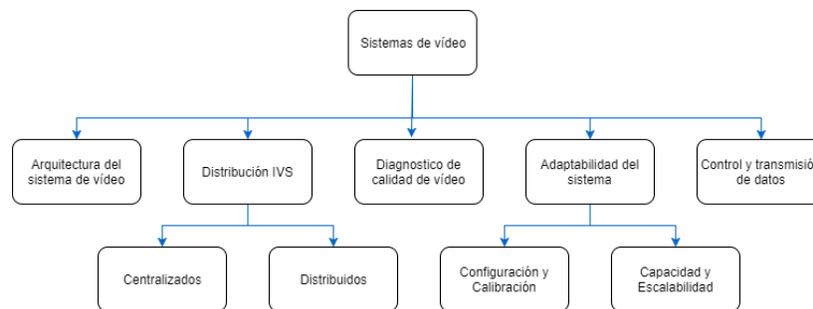


Figura 2.2. Aspectos relevantes que se relacionan con los sistemas de video.

Tareas analíticas en los ISV: Este apartado menciona las tareas que puede llegar a realizar un ISV, como se muestra en la figura 2.3.

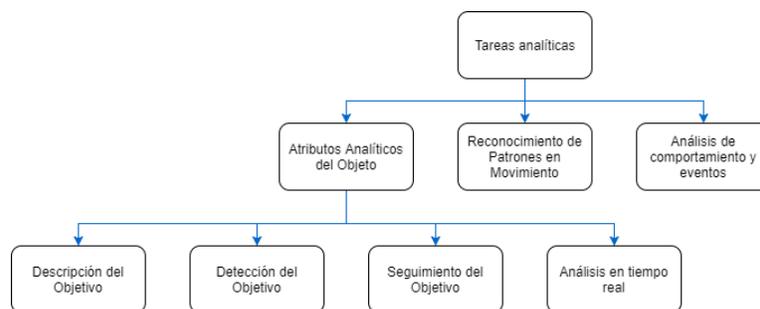


Figura 2.3. Tareas que puede realizar un ISV.

Métodos de análisis en los ISV: Este apartado menciona los principales métodos que utilizan los videos de videovigilancia para ser inteligentes, como se muestra en la figura 2.4.

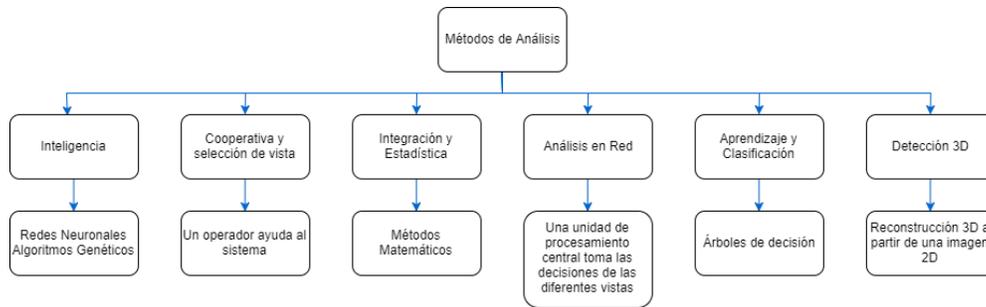


Figura 2.4. *Métodos para dotar de Inteligencia un sistema de videovigilancia.*

2.1.3 Aportaciones de videovigilancia inteligente

Tabla 2.1. Aportaciones de los artículos de videovigilancia inteligente.

Artículo	Objetivo	Técnicas o áreas	Resultados o conclusiones	Comentarios
Intelligent Video Systems and Analytics: A Survey, (Liu , Chen, & Kubota, 2013)	Brindar al lector un entorno general de las aplicaciones y áreas de investigación de los sistemas de videovigilancia inteligente.	Sistemas de Video Tareas analíticas Métodos analíticos.	Una gran cantidad de métodos son investigados actualmente para la solución de análisis de video, incluyendo el diseño del sistema, análisis en tiempo real, detección de objetos, seguimiento, aprendizaje de sustracción de fondo, clasificación, reconstrucción 3D y análisis de comportamiento.	Áreas de investigación relacionados al tema de tesis.
Intelligent multi-camera video surveillance: A review, (Wang X. 2013)	Brindar al lector un panorama general de las diferentes áreas de investigación relacionados a la videovigilancia inteligente.	Calibración de Cámaras Topología de la red de Cámaras Seguimiento multi-cámara. Re-identificación de objetos. Análisis de actividades multi-cámara.	La re-identificación de objetos es relativamente reciente y la solución de diferentes propuestas está lejos de ser completamente satisfactorio. No se ha realizado demasiada investigación en el diseño de topologías de red de cámaras y en la oclusión de cámaras híbridas cooperativas en entornos muy concurridos. Futuras investigaciones deberán usar una red de gran escala de cámaras.	Áreas de investigación relacionados al tema de tesis.

2.2 Sustracción de fondo

2.2.1 Background Subtraction Techniques: Systematic Evaluation and Comparative Analysis, (Herrero, S., & Bescós, J. 2009)

El artículo presenta una comparativa de las técnicas de sustracción de fondo más conocidas de diferentes escenarios. Cada técnica es evaluada con los mejores parámetros de inicialización de los escenarios para posteriormente evaluar su costo computacional y desempeño.

Las técnicas son clasificadas de acuerdo al modelo de representación utilizado: modelos básicos, modelos paramétricos y modelos no paramétricos. En los modelos básicos se evalúan las técnicas de diferencia de fotogramas (FD) y filtrado de la mediana (MF). En los modelos paramétricos se evalúan las técnicas de modelo simple gaussiano (SG), modelo de mezclas gaussianas (MoG) y modelo por método gamma (G). En los modelos no paramétricos se evalúan el modelo basado en el histograma (Hb) y el método de estimación de densidad de kernel (KDE).

Las evaluaciones realizadas en términos de costo computacional indican que KDE, MoG y Hb son los que tienen alto costo debido a las estimaciones que realizan en cada pixel. Técnicas simples como FD y MF presentan bajo costo computacional. G es el que presenta menor costo, pero ha sido implementado con una variante en el análisis basado en ventanas espaciales. De igual forma que tenga un costo computacional bajo no indica que sea el mejor algoritmo, razón por lo cual también se mide el desempeño.

Los resultados de la evaluación de desempeño muestran que de manera general el modelo G obtuvo los mejores resultados. Por tanto, cabe la posibilidad de considerar ventanas espaciales en lugar de estimaciones de píxeles para mejorar los resultados. Los métodos de MF y MoG funcionan de manera eficiente en los diferentes experimentos, mientras que los otros métodos tuvieron una media F baja.

Finalmente, el estudio comparativo muestra que las técnicas paramétricas tienen mayor precisión en la detección de movimiento en escenarios de interiores y exteriores de las secuencias de imagen siempre y cuando exista un bajo porcentaje de fondo dinámico. Las desventajas de estas técnicas son la dificultad de ajustar los parámetros dado el entorno en el que se trabaje, así como la complejidad de los modelos matemáticos utilizados con respecto al costo computacional. Por otro lado, las técnicas no paramétricas tienen muy baja precisión y son sensibles a estimar una buena función de densidad del pixel, además de tener un alto costo computacional.

2.2.2 Review: Moving Object Detection Techniques, (Amandeep & Goyal, 2015)

El artículo se enfoca una comparativa teórica de los algoritmos de detección de objetos en movimiento que han sido propuestos en años recientes como lo son la sustracción de fondo, flujo óptico, diferencias temporales y otros. En esta comparativa se mencionan las ventajas y desventajas de cada algoritmo.

Los algoritmos comparados para la detección de objetos en movimiento, así como sus ventajas y desventajas, son las siguientes:

Diferencias Temporales

Utiliza dos fotogramas adyacentes basados en una serie de tiempo para sustraer y obtener una diferencia de fotogramas. Tiene fácil implementación y es adaptable a escenarios con cambios dinámicos. La detección de píxeles correspondientes a objetos de primer plano que tienen cierto tipo de movimiento en ocasiones falla.

Métodos de Sustracción de Fondo

Detecta regiones en movimiento sustrayendo el fotograma actual pixel por pixel de un fotograma de fondo de referencia, la región es creada por el promedio de fotogramas sobre el tiempo. Es de fácil implementación y la precisión de detección de objetos en movimiento es buena. Es sensible a cambios externos en el entorno.

Flujo Óptico

Utiliza un vector de características, cuyos valores cambian con el tiempo para detectar áreas en movimiento de la secuencia de video. De manera general el desempeño en la detección de los objetos en movimiento es muy preciso. Tiene un alto costo computacional y no puede ser aplicado en tiempo real en cualquier tipo de hardware.

Diferencia de fotogramas

El fotograma actual es extraído de un fotograma N que representa el modelo. Usualmente es el fotograma adyacente en una serie de tiempo. Es un método simple de implementar, pero es muy sensible a cambios de iluminación y sólo tiene buen desempeño con fondos estáticos.

Mediana Aproximada

El modelo de fondo es calculado a partir de la mediana de la secuencia de imágenes, posteriormente se realiza la extracción con respecto al fotograma actual. Tiene un buen desempeño en general. Necesita de un buffer para almacenar secuencias de imágenes previas y presenta problemas con cambios de iluminación rápidos.

Modelo de mezclas gaussianas

Utiliza una función de densidad de probabilidad la cual determina el valor del pixel para generar un modelo. Tiene un costo computacional moderado, un buen desempeño y da buen resultado ante cambios de iluminación. No soporta fondos multimodales.

2.2.3 A Review of Detection and Tracking of Object from Image and Video Sequences, (Tiwari & Singhai, 2017)

El objetivo de este artículo es un análisis teórico basado en las conclusiones de diversos autores relacionados en la detección de objetos en movimiento y seguimiento de objetos. Esto permite conocer el principio básico, costo computacional y precisión de diferentes técnicas existentes.

Los métodos evaluados para la detección de objetos en movimiento son diferencias temporales, diferencias de fotogramas, mediana aproximada, promedio gaussiano, modelo de mezclas gaussianas y flujo óptico. Los métodos evaluados en el seguimiento de objetos son el filtro kalman, filtro de partículas, seguimiento de múltiples hipótesis, simple coincidencia de plantilla, método de cambio medio, máquinas de vector soporte, seguimiento basado en capas, ajuste de contornos y ajuste de formas.

El método de diferencias temporales se basa en la resta de un pixel actual con respecto a un modelo de fondo, tiene un costo computacional bajo y una precisión alta. (Chate et al., 2012) y (Mohan and Resmi, 2014) concluyen que es de fácil implementación; (Haritaoglu et al., 2000) afirma que es sensible a cambios dinámicos y (Mohan and Resmi, 2014) indica que para un buen desempeño necesita que el modelo de fondo no tenga objetos en movimiento.

La resta de fotogramas consiste en una resta del fotograma actual con respecto a un fotograma de fondo en el instante de tiempo t . El costo computacional es bajo y la precisión es moderada. (Aldhaheri and Edirisinghe, 2014) y (Haritaoglu et al., 2000) indican que es un modelo de sustracción de fondo muy simple. (Mohan and Resmi, 2014) concluye que es un modelo que no debe ser usado en aplicaciones de tiempo real debido a que es sensible a cambios de iluminación.

El método de mediana aproximada realiza una simple resta entre el fotograma actual y el modelo mediano de fotogramas. Tiene un costo computacional de bajo a moderado y su precisión es aceptable. (Aldhaheri and Edirisinghe, 2014) indica que necesita un modelo de fondo adecuado y requiere un buffer con valores de píxeles recientes de fotogramas anteriores.

El promedio gaussiano genera un modelo basado en una función de densidad de probabilidad gaussiana de los píxeles. El costo computacional es de moderado a alto y su precisión es aceptable. (Aldhaheri and Edirisinghe, 2014) determina que es muy recomendable para aplicaciones en tiempo real y el costo computacional dependerá del número de cálculos estadísticos.

El modelo de mezclas gaussianas se basa en una distribución multimodal de fotogramas y utiliza una función de densidad de probabilidad. El costo computacional es de moderado a alto y su precisión es alta. (Zhiqiang et al., 2006) indica que el consumo de memoria es bajo. (Tao Zhang et al., 2010) concluye que no puede hacer frente al solapamiento de objetos y es sensible al ruido.

El método de flujo óptico consiste en utilizar las características de distribución de los píxeles de un objeto. Tiene una complejidad computacional alta y su precisión es alta. (Krishna et al., 2011) concluye que se obtienen datos de movimiento de los objetos muy completos, pero requiere de mucho cálculo matemático.

La detección de objetos en movimiento en secuencias de video juega un papel importante en muchas aplicaciones. Específicamente en tareas de videovigilancia. La importancia de conocer las bondades que brindan los métodos existentes en la literatura permite determinar

el uso de un método en específico de acuerdo a las necesidades de la aplicación conociendo su costo computacional y la precisión.

2.2.4 Review and evaluation of commonly-implemented background subtraction algorithms, (Benezeth, Y., Jodoin, P. M., Emile, B., Laurent, H., & Rosenberger, C. 2008)

El artículo presenta los métodos comúnmente utilizados en el estado del arte para la sustracción de fondo. Estos son evaluados de forma cuantitativa. Los algoritmos evaluados son detección básica de movimiento (Basic), gaussiano (1-G), modelo de mezclas gaussianas (GMM), estimación de densidad de kernel (KDE) y diferencias máximas y mínimas de fotogramas (MinMax).

Basic consiste en imágenes usualmente en escala de grises donde se crea un modelo temporal con el filtro de la mediana. El método 1-G consiste en que cada pixel tiene una función de densidad de probabilidad y esta es evaluada por medio de una medida de distancia, usualmente Mahalanobis. El GMM es una extensión del modelo gaussiano en donde cada pixel tiene una mezcla de K gaussianos, dotándolo así de una probabilidad de ocurrencia. La KDE utiliza una estimación de ventana de Parzen de cada pixel de primer plano y es necesario tener un número de fotogramas anteriores que permita estimar la ventana. Un pixel es etiquetado como de primer plano si este pertenece a la distribución. Finalmente, MinMax utiliza un modelo de fondo hecho por los valores máximos y de los valores mínimos de una secuencia de fotogramas, un pixel pertenece a un pixel de fondo si este es menor a un umbral establecido. La tabla 2.2 muestra los resultados de la comparativa.

Tabla 2.2. Comparativa de los métodos

	Basic	1-G	KDE	MinMax	GMM
Prueba 1	0.92	0.92	0.88	0.88	0.93
Prueba 2	0.55	0.76	0.84	0.48	0.79
Prueba 3	0.62	0.77	0.75	0.28	0.76
Promedio	0.69	0.81	0.82	0.54	0.82

Los métodos 1-G, GMM y KDE son los más confiables, aunque son sensibles al ruido. El método de MinMax funciona bien solamente cuando se tiene movimiento significativo, el umbral penaliza su desempeño. Los métodos más robustos para la detección de movimiento son GMM y KDE.

2.2.5 Statistical Background Modeling for Foreground Detection: A Survey, (Bouwmans, 2010)

El artículo provee un conjunto de métodos estadísticos utilizados en la detección de objetos en movimiento de acuerdo al estado del arte, de igual forma estos son englobados en generaciones. Estos son clasificados de acuerdo a su fecha de publicación.

La primera generación comprende de 1997-2001 y se compone de los métodos de simple gaussiano (GS), modelo de mezclas gaussianas (MOG), estimación de densidad de kernel (KDE) y aprendizaje de espacio utilizando PCA (SL-PCA).

La segunda generación data de 2002 a 2006 y se compone de los métodos de máquinas de vector soporte (SVM), regresión de vector soporte (SVR) y descripción de datos de vector soporte (SVDD).

Finalmente, la tercera generación inicia en el año 2007 hasta la actualidad y se compone de los métodos de simple general gaussiano (SGG), modelo general de mezclas gaussianas (MOGG), aprendizaje de espacios utilizando ICA (SL-ICA), aprendizaje de espacios utilizando INMF (SL-INMF) y aprendizaje de espacios utilizando IRT (SL-IRT).

Los métodos de la primera generación fueron evaluados, ya que son los más estables y más utilizados de acuerdo a la literatura. La evaluación de estos métodos se hizo planteando ciertos tipos de problemas que son mencionados a continuación. Objeto en movimiento (MO) en donde una persona entra a un cuarto, contesta el teléfono y sale del cuarto. Hora del día (TOD) la iluminación cambia gradualmente de oscuro a blanco. Interruptor de luz (LS) un cuarto tiene la luz prendida, una persona entra y apaga el interruptor, posteriormente este camina en el cuarto y después prende la luz. Agitando arboles (WT) un árbol se está sacudiendo y una persona pasa enfrente del árbol, esto representa fondos dinámicos. Camuflaje una persona camina frente un fondo similar a la de la ropa de la persona. Ocupado (B) muestra secuencia de imágenes de una cafetería llena. Apertura de primer plano (FA) una persona camina lentamente. La tabla 2.3 muestra el desempeño de la primera generación en donde FN son falsos negativos y FP son falsos positivos.

Tabla 2.3. Evaluación de desempeño de algoritmos

Algoritmo	Error	Tipo de Problema							Total
		MO	TD	LS	WT	C	B	FA	
SG	FN	0	949	1857	3110	4101	2215	3464	15696
	FP	0	535	15123	357	2040	92	1290	19437
MOG	FN	0	1008	1633	1323	398	1874	2442	8678
	FP	0	20	14169	341	3098	217	530	18375
KDE	FN	0	1298	760	170	238	1755	2413	6634
	FP	0	125	14153	589	3392	933	624	19816
SL-PCA	FN	0	879	962	1027	350	304	2441	5963
	FP	0	16	362	2057	1548	6129	537	10649

Los datos demuestran que SL-PCA presenta globalmente el mejor desempeño, aunque no es el mejor algoritmo en situaciones de MO y WT. Para WT el que mejor desempeño tiene es KDE y MOG. Esta comparación no da una evaluación definitiva ya que el desempeño siempre va a depender de 3 factores importantes: la tarea, la secuencia de imágenes y su aplicación, aunque proporciona un panorama general de los métodos.

2.2.6 Recent Advanced Statistical Background Modeling for Foreground Detection - A Systematic Survey, (Bouwmans, 2011)

El artículo propone una clasificación categórica de los métodos existentes en el estado del arte enfocados a la sustracción de fondo; posteriormente vuelven a ser una clasificación categórica tomando como base los métodos de modelos estadísticos de fondo y finalmente, hacen una comparativa de estos métodos con respecto a diferentes condiciones en las secuencias de video para determinar sus bondades.

La primera clasificación determino 7 categorías de acuerdo a la estrategia utilizada como se muestra en la tabla 2.4. En esta se describen diferentes métodos de la literatura que son de uso común.

Tabla 2.4. Métodos de la literatura clasificados.

Categoría	Métodos
Básicos	Promedio
	Mediana
	Histograma
Estadísticos	Simple Gaussiano
	Modelo de Mezclas Gaussianas
	Estimación de Densidad de Kernel
Difusos	Promedio difuso
	Mezclas Gaussianas difusas
Agrupamiento	K-means
	CodeBook
Redes Neuronales	Redes Neuronales de Regresión
	Redes Neuronales auto-organizadas
Wavelets	Transformada Wavelet Discreta
Estimaciones	Filtro Weiner
	Filtro Kalman
	Filtro TchebyChev

Los autores, vuelven a categorizar los métodos de acuerdo a sus propiedades. Proponen 3 categorías las cuales se muestran en la tabla 2.5. La categoría uno representa estadísticamente el fondo para asumir la intensidad del pixel del modelo. La segunda categoría utiliza modelos de vector soporte. La tercera son una extensión de los métodos de la primera categoría.

Tabla 2.5. Modelos Estadísticos Categorizados.

Categoría	Métodos	Abreviatura
Estadísticos	Simple Gaussiano	SG
	Modelo de Mezclas Gaussianas	MOG
	Estimación de Densidad de Kernel	KDE
	Aprendizaje de espacio utilizando	SL-PCA
Vectores de Soporte	Máquinas de Vector Soporte	SVM
	Regresión de Vector Soporte	SVR
	Descripción de Datos de Vector Soporte	SVDD
Estadísticos extendidos	Simple General Gaussiano	SGG
	Mezclas Generales Gaussianas	MOGG
	Aprendizaje de espacio utilizando ICA	SL-ICA
	Aprendizaje de espacio utilizando INMF	SL-INMF
	Aprendizaje de espacio utilizando IRT	SL-IRT

Finalmente, estos métodos estadísticos son evaluados de acuerdo a fondos dinámicos y a cambios de iluminación. Con esto se puede determinar si son robustos en escenarios de interiores y exteriores, y determinar sus potenciales aplicaciones. La tabla 2.6 muestra las bondades de los algoritmos, donde SM significa movimiento lento, SC cambios lentos, IS escenas de interiores, OS escenas de exteriores.

Tabla 2.6. Comparativa de los métodos estadísticos.

Método	Fondos Dinámicos	Cambios de Iluminación	Escenarios de interiores o exteriores	Aplicaciones
SG	NO	SC	IS	Captura de movimiento
MOG	SM	SC	OS	Videovigilancia
KDE	SI	SC	OS	Videovigilancia
SL-PCA	NO	SI	OS	Videovigilancia
SVM	NO	SC	OS	Videovigilancia
SVR	SM	SC	OS	Videovigilancia
SVDD	SI	SI	OS	Videovigilancia
SGG	NO	SC	IS	Captura de movimiento
MOGG	SM	SC	OS	Videovigilancia
SL-ICA	NO	SI	OS	Videovigilancia
SL-INMF	NO	SI	OS	Videovigilancia
SL-IRT	NO	SI	OS	Videovigilancia

Las categorizaciones de los diferentes métodos de sustracción de fondo permiten tener un panorama general de las investigaciones actuales sobre el tema para poder fácilmente abordar nuevas investigaciones. De igual forma conocer las bondades de los métodos estadísticos permite poder seleccionar el que mejor se ajuste a la necesidad de trabajo.

2.2.7 Aportaciones de sustracción de fondo

Tabla 2.7 Aportaciones de los artículos de sustracción de fondo.

Artículo	Objetivo	Técnicas	Conclusiones	Comentarios
Background Subtraction Techniques: Systematic Evaluation and Comparative Analysis, (Herrero, S., & Bescós, J. 2009)	Brindar una comparativa en relación de costo computacional y desempeño de técnicas paramétricas y no paramétricas para la sustracción de fondo.	<ol style="list-style-type: none"> 1. Diferencia de fotogramas (FD) 2. Filtrado de la mediana (MF) 3. Modelo Simple Gaussiano (SG) 4. Modelo por el método gamma (G) 5. Modelo de Mezclas Gaussianas (MoG) 6. Modelo basado en histograma(Hb) 7. Método de estimación de densidad (KDE). 	Las técnicas son medidas en relación al costo computacional y desempeño. Los modelos simples como FD, MF, SG y G tienen un menor costo de tiempo de procesamiento, pero su desempeño es bajo. Los modelos complejos como MoG, KDE y Hb tienen un mejor desempeño pero a mayor costo computacional.	El modelo de mezclas gaussianas con respecto a las otras técnicas tiene un costo computacional medio y con buena precisión.
Review: Moving Object Detection Techniques, (Amandeep & Goyal, 2015)	Brindar una lista de algoritmos para la detección de objetos en movimiento propuestos en recientemente, así como brindar sus ventajas y desventajas de cada algoritmo.	<ol style="list-style-type: none"> 1. Diferencias Temporales. 2. Sustracción de Fondo. 3. Flujo Óptico. 4. Diferencia de fotogramas. 5. Mediana Aproximada. 6. Modelo de mezclas gaussianas. 	Los algoritmos son comparados teóricamente. Los métodos de diferencia de fotogramas, mediana aproximada y mezclas gaussianas son de fácil implementación, pueden ser utilizados en aplicaciones de tiempo real y no requieren hardware especializado.	El modelo de mezclas gaussianas tiene como ventaja buen desempeño, aunque no logra diferenciar entre fondos multimodales.
A Review of Detection and Tracking of Object from Image and Video Sequences, (Tiwari & Singhai, 2017)	El artículo realiza un estudio teórico comparativo de métodos de detección y seguimiento del estado del arte.	<ol style="list-style-type: none"> 1. Diferencias Temporales. 2. Diferencia de Fotogramas. 3. Mediana aproximada. 4. Promedio Gaussiano. 5. Modelo de mezclas gaussianas. 6. Flujo Óptico. 	La detección de objetos en movimiento en videos juega un papel importante en tareas de videovigilancia. Conocer las bondades que brindan los métodos existentes en la literatura permite determinar el uso de un método en específico de acuerdo a las necesidades de la aplicación.	El modelo de mezclas gaussianas tiene un costo computacional moderado, con precisión alta en comparación a otros métodos que tiene costo alto y precisión alta o caso contrario.

Artículo	Objetivo	Técnicas	Conclusiones	Comentarios
Review and evaluation of commonly-implemented background subtraction algorithms, (Benezeth, Y., Jodoin, P. M., Emile, B., Laurent, H., & Rosenberger, C. 2008)	Brinda una comparativa de la precisión de diferentes métodos de detección de movimiento.	<ol style="list-style-type: none"> 1. Detección de movimiento básico (Basic) 2. Gaussiano (1-G) 3. Modelo de mezclas gaussianas (GMM) 4. Estimación de densidad de kernel (KDE) 5. Diferencias máximas y mínimas de fotogramas (MinMax). 	Los métodos 1-G, GMM y KDE son los más confiables, aunque son sensibles al ruido. El método de MinMax funciona bien solamente cuando se tiene movimiento significativo, el umbral penaliza su desempeño. Los métodos más robustos para la detección de movimiento son GMM y KDE.	El método de mezclas gaussianas sigue demostrando resultados superiores ante otros métodos de sustracción de fondo. Otro método que vale la pena evaluar es el de estimación de densidad de kernel.
Statistical Background Modeling for Foreground Detection: A Survey, (Bouwman, 2010)	Brindarle al lector una clasificación de los métodos estadísticos de sustracción de fondo, así como una clasificación de acuerdo a su desempeño de la primera generación.	<ol style="list-style-type: none"> 1. Modelo Gaussiano. 2. Modelo de Mezclas Gaussianas. 3. Estimación de Densidad de Kernel. 4. Aprendizaje de Sub-espacio utilizando PCA. 	Los métodos de modelos de mezclas gaussianas y estimación de densidad de kernel son los que tienen un mejor desempeño de las diferentes situaciones evaluadas. La comparación no da una evaluación definitiva ya que el desempeño siempre va a depender de 3 factores importantes: la tarea, la secuencia de imágenes y su aplicación. La ideal principal es tener un panorama general de los métodos.	El método de mezclas gaussianas, así como el de estimación de densidad de kernel son robustos para la sustracción de fondo. Se proponen métodos de segunda y tercera generación que podrían brindar mejores resultados.
Recent Advanced Statistical Background Modeling for Foreground Detection - A Systematic Survey, (Bouwman, 2011)	Brindarle al lector una lista de los métodos de sustracción de fondo del estado del arte, así como una evaluación de los métodos de modelos estadísticos de fondo en escenas de interiores y exteriores, y aplicaciones.	<ol style="list-style-type: none"> 1. Modelo Gaussiano. 2. Modelo de Mezclas Gaussianas. 3. Estimación de Densidad de Kernel. 4. Aprendizaje de Sub-espacio utilizando PCA. 5. Máquinas de Vector Soporte. 6. Regresión de Vector Soporte. 7. Datos de Vector Soporte. 8. Modelo General Gaussiano. 9. Modelo General de Mezclas Gaussianas. 10. Análisis de Componentes Independientes. 11. Factorización de matrices no negativas incrementales. 12. Tensores de rangos incrementales. 	Conocer las propiedades de los algoritmos en fondos dinámicos y cambios de iluminación permiten saber que tan sensibles. Esto determina el tipo de escena en el que pueden ser utilizados, así como posibles aplicaciones.	Los modelos de mezclas gaussianas son buenos en fondos de cambios dinámicos lentos y los cambios de iluminación no deben ser abruptos ya que esto genera una gran diferencia entre el fotograma actual y el modelo. Es útil en escenarios exteriores y su principal aplicación en la videovigilancia.

2.3 Re-identificación de personas

2.3.1 Person Reidentification With Reference Descriptor, (An, L., Kafai, M., Yang, S., & Bhanu, B. 2016)

El presente artículo propone un método para la re-identificación de personas a través de diferentes vistas de cámaras. En lugar de manejar similitudes de apariencias por medio de descriptores, el método propuesto primeramente aprende una subregión de la imagen en donde las correlaciones de los datos de diferentes cámaras son maximizadas usando un análisis de correlación cónica regularizada (*RCCA*, por sus siglas en inglés) como se describe en la figura 2.5.

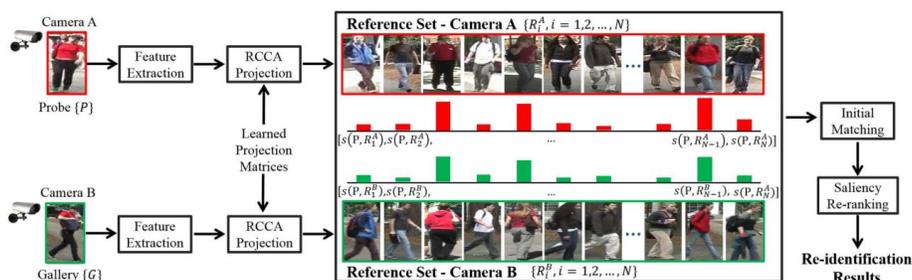


Figura 2.5. Método propuesto para la re-identificación. Las características de apariencia primero son extraídas de un conjunto de imágenes de entrenamiento y proyectadas dentro de un sub-espacio RCCA con matrices de proyección aprendidas. Una instancia de descriptores de referencia es generada y concatenando sus puntajes de similitud con respecto al conjunto de referencia. La coincidencia inicial es realizada usando los descriptores de referencia y posteriormente se utiliza un esquema de rasgo sobresaliente para mejorar la precisión de re-identificación.

El método propuesto es dividido en dos procesos: “*offline re-identification*” y “*online re-identification*”.

Offline re-identification.

En este proceso las matrices de proyección de RCCA son aprendidas y las referencias de descriptores son generados (RD, por sus siglas en inglés).

La técnica utilizada es el análisis de correlaciones canónicas (CCA, por sus siglas en inglés.). El objetivo principal es explorar las relaciones entre los dos conjuntos de variables aleatorias de diferentes observaciones en el mismo ejemplo de datos.

CCA encuentra tales proyecciones que tienen relación entre los dos conjuntos y estos son maximizados. Hablando matemáticamente dado dos conjuntos de datos observados X y Y , se obtienen dos conjuntos de vectores tales que los coeficientes de la correlación son maximizados.

Las características del conjunto de entrenamiento extraídas son los histogramas de color y descriptores de textura. Estos son almacenados en dos conjuntos de características diferentes.

Una vez que los conjuntos de características son definidos primeramente se aprende una subregión RCCA en el cual las correlaciones entre el conjunto de características proyectados son maximizadas y con esto las matrices de proyección de RCCA son aprendidos.

Online re-identification.

Los descriptores de referencia son generados y son comparados con el conjunto de descriptores de referencias generados en el proceso offline.

Las características de una subregión correspondientes a un sujeto son extraídas. Los descriptores de referencia del sujeto son proyectados por RCCA. La identidad del sujeto es determinada por la coincidencia entre la subregión y los descriptores de referencias generados, entonces se toma la mayor coincidencia encontrada.

Para mejorar la precisión de la re-identificación se usa un rasgo sobresaliente de la subregión como se muestra en la figura 2.6. Esto brinda una característica visual discriminativa que hace coincidir al mismo sujeto en diferentes vistas.

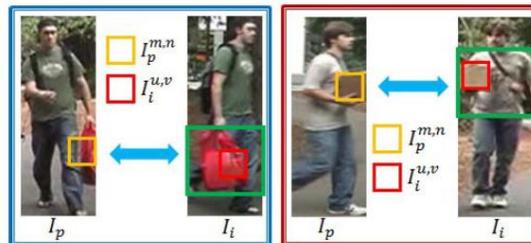


Figura 2.6. Muestra de la detección de rasgo sobresaliente en dos vistas de cámaras.

2.3.2 Re-identificación de personas a través de sus características softbiométricas en un entorno multi-cámara de video-vigilancia, (Moctezuma, D., 2016)

El presente artículo propone un sistema completo de videovigilancia inteligente. El sistema propuesto se compone de dos módulos: La detección de personas y la re-identificación de personas por características softbiométricas. Estos conjuntos proporcionan un sistema de visión artificial para la re-identificación de personas en un entorno multi-cámaras de videovigilancia.

En el módulo de detección de personas proponen un algoritmo, combinando los filtros de Gabor y utilizando el algoritmo de Histogram Of Gradients (HOG).

El método propuesto en el módulo de detección de personas se resume de la siguiente forma:

1. Detectar las regiones en movimiento.
2. Aplicar filtros de Gabor para resaltar las características de las regiones.
3. Aplicar el algoritmo de HOG para extraer las características de las regiones
4. Utilizar un clasificador, en este caso se utiliza una máquina de vector soporte para determinar si la región pertenece a una persona o no.

Lo mencionado antes se resume gráficamente como se muestra en la figura 2.7:

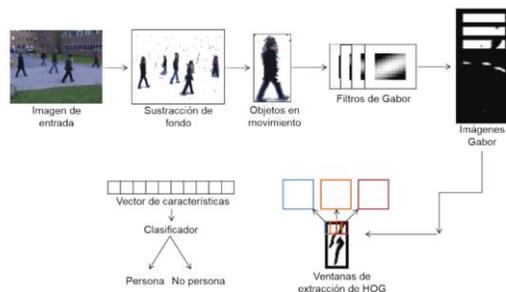


Figura 2.7. Esquema general de HOGG

En el módulo de re-identificación de personas se propone el uso de un conjunto de características softbiométricas. Estas son evaluadas para medir su relevancia mediante la posición y la ponderación utilizando diferentes técnicas.

Las características softbiométricas son aquellas que proveen cierta información sobre las personas, pero que carecen de suficiente permanencia y de un alto nivel distintivo para diferenciar a un individuo de otro como lo define (Jain, A. K., Dass, S. C., & Nandakumar, K., 2004). Las características utilizadas son: Estadísticas de color, estadísticas de histograma, geometría y textura.

Las técnicas para medir la relevancia fueron: Análisis de Componentes Principales (*PCA por sus siglas en inglés.*) y diferentes variantes propuestas a partir de esta. Las características softbiométricas ordenadas por la técnica son ponderadas y ayudan a determinar si las regiones corresponden a la misma persona o no.

2.3.3 Learning Invariant Color Features for Person Reidentification, (Varior, R. R., Wang, G., Lu, J., & Liu, T. 2016)

El presente artículo propone un sistema de aprendizaje de patrones de colores de píxeles de dos diferentes vistas de cámara. El objetivo principal es que independiente de las variaciones de las condiciones de luz que pueden presentarse en las diferentes vistas y que afectan los valores de los píxeles de una región que deberían ser del mismo color, la representación final del píxel de un color en particular será estable a las posibles variaciones. La figura 2.8 muestra un esquema general del sistema propuesto. La metodología propuesta para lograr esto se divide en cinco etapas:

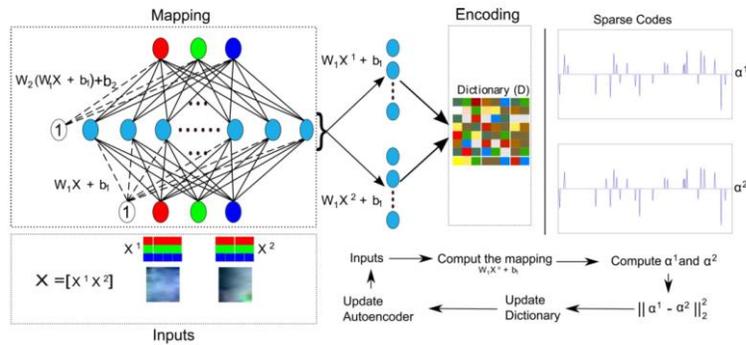


Figura 2.8. Diagrama de flujo del aprendizaje para el mapeo y la codificación. Las entradas son muestras aleatorias extraídas de los patches x^1 y x^2 corresponden a una muestra de píxeles procedentes de las partes correspondientes de los pares de imágenes coincidentes que deben ser del mismo color. Mapping: el autocodificador mapea las entradas en un espacio de 60 dimensiones que da las representaciones de x^1 y x^2 . Con base en el diccionario “D”, los dos conjuntos en el espacio dimensional son codificados y se obtiene dos conjuntos de códigos dispersos.

1.- Conjunto de entrenamiento:

En esta etapa se define manualmente las parejas de “patch” extraídos de pares de imágenes del conjunto de entrenamiento como se muestra en la figura 2.9.



Figura 2.9. Vista de una imagen A y B donde se extrae una región que debería corresponder al mismo color.

2.- Función objetivo

Una vez que los datos de entrenamiento están preparados, la función objetivo para aprender las características invariantes es formulada con una restricción de invariancia para la representación de características obtenidas de cada pareja de píxeles de las dos vistas. La función objetivo es formulada con un codificador automático.

El codificador automático con restricciones de invariancia es capaz de aprender filtros que pueden ser capturados en los patrones de datos. De igual forma pueden ser aplicados a los descriptores de representación.

3.- Optimización

La optimización es hecha a los valores de α , D , w_1 , b_1 , w_2 , b_2 . Para este caso es utilizado el gradiente de L-BFGS para actualizar sus valores.

El uso del gradiente basado en la optimización es simple para las funciones objetivos. Para una convergencia rápida y optima se requiere un poco de finura. Una buena inicialización es requerida para los valores de D , w_1 , w_2 .

4.- Estructura multi-capa.

Considerando la complejidad del enfoque sólo se emplean dos capas en el modelo propuesto.

La representación final de los ejemplos de las muestras de *patch* de los pares de imágenes son igualados los unos a los otros. Todos los pixeles del conjunto de ejemplos son codificados usando la función objetivo y la optimización. Es necesario utilizar un esquema máxima agrupación sobre las regiones de 2×2 para obtener la representación de un *patch* en la primera capa. Una vez que esta representación es obtenida la segunda capa repite el esquema de máxima agrupación, pero sobre regiones de 4×4 .

5.- Parámetros

La formulación contiene varios parámetros, tales como penalizaciones en los pesos en términos de la función de costo, las dimensiones de los espacios transformados linealmente y el número de átomos del diccionario. Todos estos son determinados por un *2-fold cross-validation* en los datos de entrenamiento de la base de datos (ViPeR dataset).

2.3.4 Soft Biometrics Integrated Multi-target Tracking, (Chen, X., & Bhanu, B. 2014)

El presente artículo presenta un sistema basado en un modelo de apariencia de características softbiométricas para el seguimiento de personas en una sola vista de cámara. El objetivo principal es poder reubicar a las personas que contienen las mismas características a lo largo del desplazamiento. La propuesta es fusionar las características softbiométricas generado por *tracklets* para aprender un modelo de apariencia discriminativo. Un *tracklet* es definido como secuencias cortas de imágenes de un mismo objeto de interés.

La metodología propuesta tiene como base el flujo general de modelos de apariencia como se muestra en la figura 2.10.

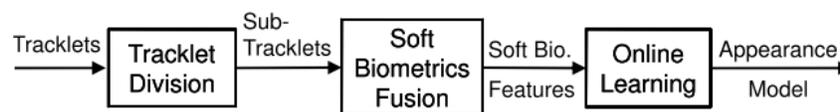


Figura 2.10. Diagrama general de un sistema basado en modelo de apariencias softbiométricas.

Dado un conjunto de *tracklets*, el objetivo del seguimiento basado en la asociación es encontrar el mejor conjunto de asociaciones con la máxima probabilidad de similitud. El modelo de apariencia va estar dado por:

1. Generación de sub-tracklets

En un *tracklet*, la apariencia del objetivo puede variar con el tiempo, pero la detección responde a que estos son obtenidos en *frames* consecutivos que poseen alta similitud. Para crear una representación concisa, el *tracklet* es dividido en *sub-tracklets* que representan instancias de apariencia de un objetivo como se muestra en la figura 2.11.

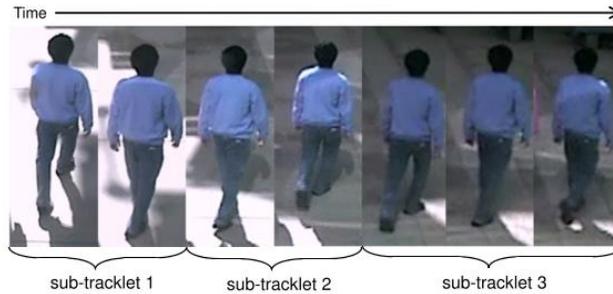


Figura 2.11. Ejemplo de la división de un tracklet

2. La fusión de características softbiométricas

Las características utilizadas para la detección son extraídas por diferentes técnicas y son categorizadas en tres tipos: simbólicas, valor escalar y valor de vector. La tabla 2.8 describe estas características.

Tabla 2.8. Características softbiométricas utilizadas para la detección.

Name	Value	Type
Gender	Male, Female	Symbolic
HairColor	Light, Dark	Symbolic
SkinColor	Caucasian, Non_Caucasian	Symbolic
Height	Centimeters	Scalar
Weight	Kilograms	Scalar
BodyColor	1-D probability distribution	Vector
TorsoColor	1-D probability distribution	Vector
LegsColor	1-D probability distribution	Vector

3. Similitud de características softbiométricas

Cada *sub-tracklet* es representado en un sólo conjunto de características softbiométricas. Para las características simbólicas, valores escalares y vectores de similitud se calcula el valor de máxima similitud.

4. Aprendizaje del modelo de apariencia.

El objetivo es diseñar un modelo de apariencia discriminativo que de la similitud por medio de pares de *tracklets*. Si estos contienen el mismo objetivo (alta similitud) se definen como pares positivos, mientras que los que tengan objetivos diferentes (baja similitud) son definidos como pares negativos. El algoritmo utilizado para el aprendizaje de los modelos de apariencias es AdaBoost.

5. Colección de ejemplos de entrenamiento.

Las muestras positivas y negativas son obtenidas con criterios de selección de similitud. Un par de *tracklets* es negativo si cumple los siguientes criterios:

- Dos *tracklets* tienen solapamiento
- Dos *tracklets* tienen una pequeña separación en los *frames* pero una gran diferencia espacial.

La primera restricción se basa en la observación de que un mismo objetivo no puede ser seguido simultáneamente en dos diferentes *tracklets* y la segunda en que un objetivo no puede realizar un movimiento de distancia significativo dentro de un corto periodo de tiempo.

El valor de una muestra de entrenamiento es un vector de 8 dimensiones, donde cada elemento del vector es la similitud entre las características softbiométricas de dos *subtracklets*.

Con todo lo anterior mencionado es posible realizar el seguimiento de personas utilizando características softbiométricas. La figura 2.12 muestra un ejemplo del seguimiento.



Figura 2.12. Ejemplo de drásticas variaciones de apariencia con el seguimiento de un mismo objetivo.

2.3.5 Human Hair Segmentation and Length Detection for Human Appearance Model, (Wang, Y., Zhou, Z., Teoh, E. K., & Su, B. 2014)

El presente artículo propone un método para la detección de la longitud del cabello humano. Este método no requiere la detección de caras y puede segmentar el cabello de cabezas humanas en diferentes vistas con baja resolución.

La metodología utilizada en el sistema propuesto como se muestra en la figura 2.13 es hacer una sustracción de fondo para detectar personas. Una vez que se cuenta con las regiones de la imagen que contienen personas, lo siguiente es detectar las cabezas humanas con un detector entrenado. Posteriormente se realiza un análisis del histograma en la región de la cabeza detectada para segmentar la región del cabello con el algoritmo de agrupamiento K-means. La longitud del cabello se determina por un escaneo lineal en la región del cabello segmentado. La ventaja del método es que es robusto ante la variación de las vistas, la baja resolución y las vistas de pequeños cuerpos humanos.

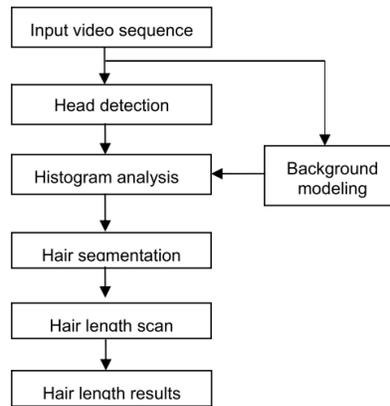


Figura 2.13. Diagrama para la detección de la longitud del cabello.

El método propuesto fue evaluado con una base de datos propia. Esta se compone de 5 videos donde cada uno contiene entre 5063 a 9060 cuadros. La figura 2.14 es un ejemplo de lo descrito.

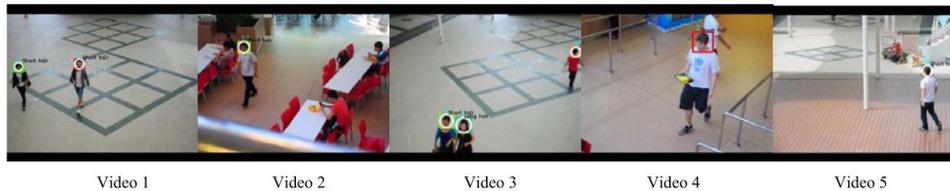


Figura 2.14. Videos de la detección de cabello.

2.3.6 A database for person re-identification in multi-camera surveillance networks, (Bialkowski, Denman, Sridharan, Fookes, & Lucey, 2012)

El presente artículo propone 3 tipos de características softbiométricas para un sistema de re-identificación. Estas características son de espacios de color, altura y textura. El sistema propuesto se muestra en la figura 2.15.

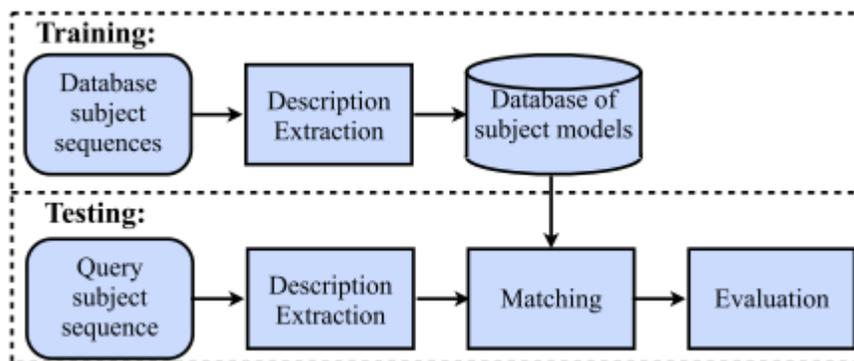


Figura 2.15. Sistema general propuesto.

La etapa de descripción y extracción se basa en 3 tipos de características softbiométricas y es necesario que el cuerpo humano este segmentado en cabeza, torso y piernas. La característica de altura es calculada a partir de cada segmento del cuerpo las cuales son

convertidas a valores del mundo real a partir de la calibración de cámaras. La característica de textura utiliza el algoritmo de *Local Binary Patterns* ya que es invariante a iluminación y rotación. La característica de color es calculada a partir de la distribución de color *HSL* generando dos modelos de color de características softbiométricas. Finalmente, cada característica softbiométrica es sumada a través de un peso ponderado y se obtiene el descriptor final.

2.3.7 Soft-biometrics evaluation for people re-identification in uncontrolled multi-camera environments, (Moctezuma, Conde, De Diego, & Cabello, 2015)

En este artículo se presenta un sistema de re-identificación de personas mediante características softbiométricas. Primeramente, se propone una bolsa de características de tamaño 23 que se relacionan al color, textura, características locales y geometría. Cada una de estas características es extraída de los individuos. La figura 2.16 muestra las características calculadas.

Number	Category	Description
1	RGB Color	Mean of Red color channel
2	RGB Color	Mean of Green color channel
3	RGB Color	Mean of Blue color channel
4	RGB Color	Mean of three RGB channels
5	RGB Color	Standard Deviation of three RGB channels
22	RGB Color	Brightness
6	Grayscale	Mean
7	Grayscale	Standard Deviation
8	Geometry	Eccentricity
9	Grayscale Histogram	Entropy
10	Grayscale Histogram	Dispersion
11	Grayscale Histogram	Mean
12	Grayscale Histogram	Standard Deviation
15	Grayscale Histogram	Energy
16	Grayscale Histogram	Kurtosis
13	HSV Color	Mean
14	HSV Color	Standard Deviation
17	Statistics of Co-occurrence Matrix	Energy
18	Statistics of Co-occurrence Matrix	Maximum Probability
19	Statistics of Co-occurrence Matrix	Entropy
20	Statistics of Co-occurrence Matrix	Inertia
23	Statistics of Co-occurrence Matrix	Homogeneity
21	Local Binary Pattern (LBP)	Simple LBP

Figura 2.16. Bolsa de características softbiométricas.

Una vez que las características son calculadas estas son medidas por su relevancia por diferentes métodos, como análisis de componentes principales. Dependiendo de su relevancia, estas son ordenadas según el rango de posición y posteriormente se le asigna un peso con respecto al rango obtenido. Cada característica fue evaluada en escenarios mono-cámara y multi-cámara. El esquema general del sistema se muestra en la figura 2.17.

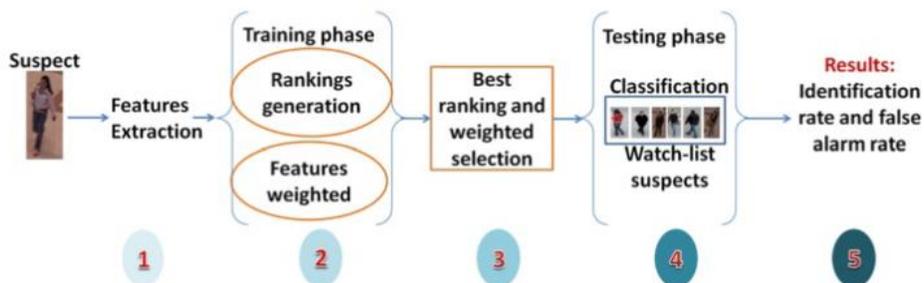


Figura 2.17. Esquema general del sistema propuesto.

2.3.8 Pedestrian re-identification using color feature in multi surveillance video, (Zhang, Matsumoto, Yoshida, & Kita, 2016)

El artículo propone un sistema que trata problemáticas de cambios de iluminación, posición y oclusión. El sistema propuesto consta de 3 etapas principales: La extracción del fotograma donde detecta a los objetos en primer plano y el fondo; la detección de objetos donde se marcan las regiones que sólo corresponden a personas y por último la extracción de características y la medida de similitud, donde se calcula las características de la imagen y se determina la re-identificación. La figura 2.18 muestra estructura del sistema.

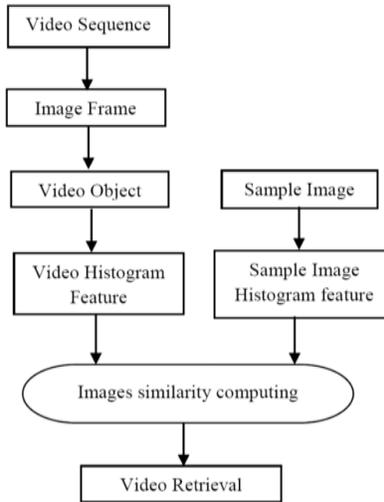


Figura 2.18. Esquema general del Sistema.

El sistema utiliza el histograma como descriptor en los espacios de color *RGB*, *SHV* y *UVW*. El cálculo del descriptor a partir del histograma se hace por medio de un modelo de partición jerárquica, como se muestra en la figura 2.19, que permite conservar la estructura de la información del histograma, ya que los métodos tradicionales no lo permiten.

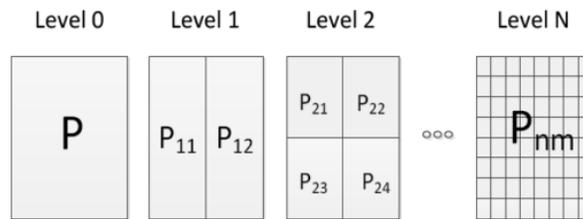


Figura 2.19. Modelo de Partición Jerárquica.

Una vez que el descriptor es obtenido la similitud es calculada en los niveles correspondientes del modelo de partición jerárquica utilizando la ecuación 2.1, en donde $p_{i,j}$ es la región p de la imagen en cierta posición. $d(p_{i,j}, q_{i,j})$ es el grado de similitud de la característica de la imagen P y Q . $k_{i,j}$ es un peso de ponderación.

$$d(P, Q) = \sum k_{ij}d(p_{ij}, q_{ij}) \quad \text{Ec. (2.1)}$$

2.3.9 Discriminative Image Descriptors for Person Re-identification, (Ma, Su, & Jurie, 2014)

Este artículo propone dos descriptores para la re-identificación de personas. El primero es la combinación de características biológicamente inspiradas (BIF, por sus siglas en inglés) y descriptores de covarianza como lo son los filtros de Gabor y el operador MAX. El segundo descriptor se basa en los vectores Fisher.

El primer descriptor se calcula de manera que se tiene una imagen de entrada, esta es dividida en sus 3 canales en el espacio de color *HSV*. Posteriormente se calculan los filtros de Gabor en diferentes escalas. Las parejas de escalas vecinas son agrupadas en una banda. Las magnitudes de las imágenes son producidas aplicando el operador MAX y se divide en pequeñas celdas donde cada celda es un descriptor de covarianza. La figura 2.20 muestra el esquema general del descriptor.

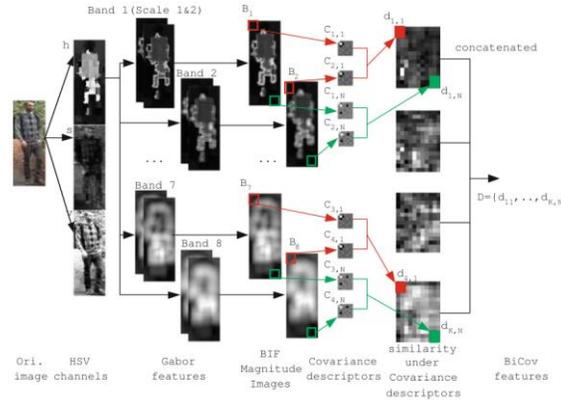


Figura 2.20. Esquema general del primer descriptor.

El segundo descriptor codifica las estadísticas de orden más alto de las características locales en un vector de 7 características, las cuales capturan la intensidad, la textura y la forma. La ecuación 2.2 representa el vector, donde (x, y) son la posición de los píxeles y B es la información del vector.

$$f_i(x, y) = [x, y, B_i(x, y), B_{ix}(x, y), B_{iy}(x, y), B_{ixx}(x, y), B_{iyy}(x, y)] \quad \text{Ec. (2.2)}$$

Finalmente se calcula el descriptor de covarianza por cada región, para posteriormente obtener la diferencia de covarianza entre dos bandas, posteriormente estas son concatenadas para la representación de la imagen y la distancia euclídea, entre dos imágenes, es utilizada para determinar la similitud.

2.3.10 Ensemble of different approaches for a reliable person re-identification system, (Nanni, Munaro, Ghidoni, Menegatti, & Brahnam, 2016)

El artículo propone un sistema de re-identificación de personas usando diferentes espacios de color y algunas variantes de los enfoques del estado del arte. Estos se combinan y prometen ser un descriptor robusto.

Los descriptores utilizados en este artículo son gBiCov que es una representación inspirada biológicamente y en una matriz de covarianza, Symmetry-Driven Accumulation of Local Features (SDALF). Custom Pictorial Structures (CPS), chromatic content and color displacement (CCD), Color Invariants (CI) y Skeleton-based Person Signature (SPS), Weber's Law Descriptor proposed (WLD), Local Phase Quantization from Three Orthogonal Plane (LPQT) y Volume Local Phase Quantization (VLPQ).

De manera general la forma en que trabaja el enfoque es primeramente cambiar del espacio de color *RGB* a *XYZ*, se extraen las estructuras pictóricas (PS) de ambas imágenes. El siguiente paso es extraer las articulaciones del esqueleto de la imagen del espacio *RGB* en 3 dimensiones, de igual forma se extraen los valores de gBiCov y SDALF en el espacio RGB. Finalmente, las imágenes se igualan usando una medida de distancia apropiada y se combinan de acuerdo a la coincidencia con una regla de suma. La figura 2.21 muestra el esquema de cálculo del descriptor.

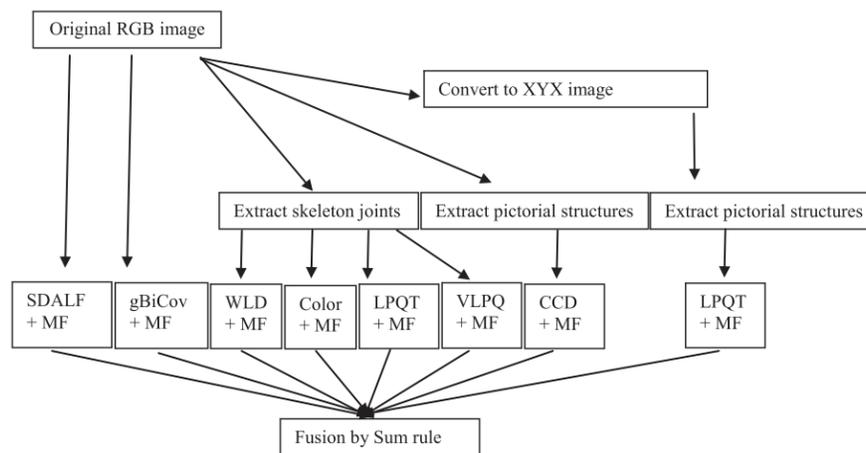


Figura 2.21. Esquema para la generación del descriptor.

Los resultados en los experimentos demuestran que el enfoque propuesto tiene mejor desempeño en comparación a los del estado del arte. Este método puede ser programado de forma paralela y tiene la desventaja que presenta este método es que no puede ser implementado en tiempo real.

2.3.11 Survey on colour, texture and shape features for person re-identification, (Poongothai & Suruliandi, 2016)

El artículo se basa en hacer una comparativa teórica de diferentes descriptores utilizados para la re-identificación de personas. Estos descriptores se relacionan directamente con el color, la textura y la forma. Del análisis general obtenido se sabe que los que tienen mejor desempeño son las características de color y de forma.

Los espacios de color evaluados fueron *RGB*, *HSV*, *YCbCr* y *CIE L*a*b**. La exactitud de la identificación se ve afectada por el espacio de color seleccionado. La combinación eficaz y

eficiente de componentes espectrales sobre varios espacios de color mejoran significativamente el rendimiento de la precisión de identificación.

El espacio de color *RGB* no requiere transformaciones para mostrar la información del color en diferentes pantallas y es considerado el espacio de color base para diferentes aplicaciones. Uno de los problemas que presentan sus componentes es que consideran el brillo y las propiedades de color de manera conjunta, así como la visión humana es diferente al color y brillo que presenta. No suele ser útil para el reconocimiento de colores. Es muy difícil que se pueda determinar el color específico en este modelo. Este espacio de color es usado comúnmente en gráficos de computadora, detección de objetos, análisis de video y detección de personas.

El espacio de color $L^*a^*b^*$ se basa en 3 componentes donde L es el brillo del color, a^* es un valor de posición entre magenta y verde y b^* es un valor de posición entre amarillo y azul. Está diseñado para aproximar la visión humana y se inspira en la percepción del color uniforme. El problema que suele presentar es que la gama de colores es tan grande que la mayoría de dispositivos no puede reproducirlos. Este espacio de color es usado comúnmente en gráficos de computadora, detección de objetos, análisis de video y detección de personas.

El espacio de color *YCbCr* está basado en *RGB*, la idea es separar los componentes en Y el cual representa la variación de luminosidad, Cb representa la variación de azul y Cr representa la variación de rojo. Separar la luminosidad en un componente independiente aproxima directamente al ojo y cerebro a ser más sensibles a los cambios de la información del color. Este espacio tiene una dimensión reducida debido a un proceso de compresión. La información del color siempre va a depender del espacio de color *RGB* y siempre es requerido la información de compresión para mostrar el color. Este espacio de color es usado comúnmente en gráficos de computadora, detección de objetos, análisis de video y detección de personas.

El espacio de color *HSV* descompone sus componentes en H que representa una medida del brillo de un color de la imagen, S que es el porcentaje del valor de color y V el cual indica la variación en la percepción de un color. Este espacio de color es usado comúnmente en visión robótica, reconocimiento de objetos, recuperación de imágenes basada en contenido y análisis de imágenes médicas.

Las características de textura y forma que se evaluaron fueron Local Binary Pattern (LBP), Scale Invariant Feature Representation (SIFT), Speeded Up Robust Feature (SURF), Gradient Location and Orientation Histogram (GLOH), Gabor Filters, Schmid y Histogram of Oriented Gradient (HOG).

LBP describe una superficie de texturas de dos dimensiones por patrones espaciales locales y contrastes de la imagen en una escala de grises. Es tolerante a cambios de iluminación y tiene una simplicidad computacional lo que permite que sea rápido de calcular. No requiere parámetros de inicio. Los problemas que suele presentar es que la velocidad de

reconocimiento es lenta en grandes bases de datos, en ciertas circunstancias puede perder la estructura local y es muy sensible al ruido. Es usado en visión por computadora, reconocimiento facial, sustracción de fondo de imagen, biometría y clasificación facial de la edad.

SIFT es un extractor de características basado en la magnitud y orientación del gradiente. Es muy robusto ante cambios de iluminación intensos, escala y rotación y los píxeles de interés son localizados fácilmente. El detalle que suele presentar es que leves cambios de iluminación producen puntos de interés similares. Es usado en reconocimiento de objetos, detección de puntos destacados, modelado 3D y seguimiento en secuencias de imágenes.

SURF es un detector y descriptor de puntos de interés en imágenes, en donde la imagen se transforma en coordenadas usando la técnica de pirámide de multi-resolución. La extracción de puntos de interés es muy rápida, tiene un mejor desempeño en escala y variaciones de iluminación, es fácil localizar las correspondencias entre fotogramas y es robusto ante movimientos ligeros imprevistos de un fotograma. El método tiene una dependencia total de la vista del fotograma, la extracción de puntos de interés es difícil en entornos de 3 dimensiones, el costo computacional es alto y la apariencia de los puntos de interés puede cambiar entre fotogramas. Es usado en reconocimiento y seguimiento de objetos, mapeo y localización robótica, registro de imágenes y recuperación de imágenes.

GLOH es una extensión del descriptor SIFT, pero está basado en el uso de histogramas de la orientación de gradiente. La descripción de la estructura de la escena es precisa y es muy robusto a cambios de iluminación. Presenta la desventaja que los bordes suelen desaparecer en caso de existir desenfoque en la imagen. Es usado en reconocimiento y seguimiento de objetos, registro de imágenes y recuperación de imágenes.

Los filtros de Gabor son un descriptor que brindan la frecuencia y orientación y es un método aproximado al sistema visual humano. Es robusto ante variaciones de brillo, iluminación, rotación y es invariante al contraste de imágenes. El detalle que presenta es que mientras mayor sea el número de filtros aplicados, mayor será la dimensión del vector de características, por tanto, el costo del almacenamiento computacional es muy elevado. Es usado en reconocimiento facial, medición, control y videovigilancia del tráfico.

Schmid es un descriptor inspirado en el sistema de visión humano y tiene una baja complejidad computacional. El problema que presenta es que realizar seguimientos incorrectos debido a un error de distribución calculado. Es usado en videovigilancia de tráfico y detección de peatones.

HOG es un descriptor que brinda información de la apariencia y forma de un objeto. Está basado en el uso de gradientes y es particularmente utilizado en la detección de humanos. Es robusto ante variaciones locales de transformaciones fotométricas y geométricas. Tiene la habilidad de detectar partes del cuerpo, manejar oclusión y solapamientos. El problema que

suele tener es que es sensible a rotación y translación siempre y cuando estos sean menores al tamaño espacial local definido de la celda.

De manera general se puede observar que el espacio de color RGB no es un descriptor robusto ya que contempla variaciones de iluminación en su canal. Los otros espacios de color presentados demuestran que pueden tener un mayor grado de precisión ya que estos logran separar la iluminación de la información de color y darle un grado de invariancia. Descriptores como SIFT y SURF podrían ser omitidos en la re-identificación ya que trabajan principalmente con puntos destacados. Un buen descriptor que valdría la pena evaluar es LBP el cual es muy utilizado en la visión por computadora, así como también lo es HOG para la detección de personas. Tener un resumen de las ventajas, desventajas y áreas de aplicación ayuda a entender de manera general que tan preciso pueden ser estos descriptores para la re-identificación.

2.3.12 Aportaciones de re-identificación

Tabla 2.9. Aportaciones de los artículos de re-identificación de personas.

Artículo	Objetivo	Técnicas	Resultados	Comentarios
Person Reidentification With Reference Descriptor (An, L., Kafai, M., Yang, S., & Bhanu, B. 2016)	Un sistema para la re-identificación de personas utilizando RCCA.	Canonical correlation analysis	El sistema es robusto ante la baja resolución, posiciones arbitrarias, cambios de iluminación y de oclusión.	Base de Datos VIPeR CUHK Campus
Re-identificación de personas a través de sus características softbiométricas en un entorno multi-cámara de video-vigilancia, (Moctezuma, D., 2016)	Un sistema de detección de personas y la re-identificación de personas por medio de características softbiométricas.	Filtros de Gabor. Histograma de Orientación Gradiente. Análisis de Componentes principales. Maquinas Vector Soporte.	El sistema tiene una gran precisión en la detección de personas a comparación de otros métodos. De igual forma se da en la re-identificación.	Características softbiométricas Filtros de realce de características. Base de datos PETS 2006 PETS 2007 PETS 2009. CAVIAR. MUBA.
Learning Invariant Color Features for Person Reidentification (Varior, R. R., Wang, G., Lu, J., & Liu, T. 2016)	Un sistema de re-identificación de personas utilizando un aprendizaje de patrones de colores de pixeles de dos diferentes vistas.	Diagonal Matrix Transform (DMT). Brightness Transfer Function (BTF).	El sistema es robusto ante las variaciones de iluminación de las vistas de las dos cámaras.	Método de consistencia de color. Base de Datos VIPeR. Person Re-ID 2011. CAVIAR4REID.
Soft Biometrics Integrated Multi-target Tracking (Chen, X., & Bhanu, B. 2014)	Un sistema basado en el modelo de apariencia de características softbiométricas.	Algoritmo Hungarian. AdaBoost.	Sistema robusto para la re-identificación y seguimiento de personas.	Características softbiométricas Base de datos propia
Human Hair Segmentation and Length Detection for Human Appearance Model (Wang, Y., Zhou, Z., Teoh, E. K., & Su, B. 2014)	Un sistema de re-identificación de personas utilizando el cabello.	Histograma de Orientación de Gradiente. Local Binary Pattern. Gaussian mixture model.	El sistema tiene una precisión promedio mayor al 80% de re-identificación utilizando solamente un descriptor.	Base de datos propia.

Artículo	Objetivo	Técnicas	Resultados	Comentarios
A database for person re-identification in multi-camera surveillance networks, (Bialkowski, Denman, Sridharan, Fookes, & Lucey, 2012)	Un sistema de re-identificación utilizando descriptores de color, textura y altura.	<ol style="list-style-type: none"> 1. Modelo de Mezclas Gaussianas. 2. Característica de Altura. 3. Característica de color HSL. 4. Local Binary Patterns. 	La base de datos propuesta es útil para la re-identificación de personas. Los modelos de color re-identifican mejor en diferentes ángulos y vistas en comparación de la altura y textura.	La base de datos puede ser útil para evaluación. La característica de textura puede ser útil para la re-identificación.
Soft-biometrics evaluation for people re-identification in uncontrolled multi-camera environments, (Moctezuma, Conde, De Diego, & Cabello, 2015)	Propone un sistema de re-identificación por medio de una comparativa de características softbiométricas, en un ambiente no controlado de multi-cámaras.	<ol style="list-style-type: none"> 1. Diferencia de Fotogramas 2. Filtros de Gabor 3. Histograma de Gradientes 4. Máquinas de Vector Soporte 5. Características en el espacio de color RGB 6. Características en escala de grises 7. Características en el espacio de color HSV 8. Matriz de Co-ocurrencia 9. Local Binary Pattern 10. Excentricidad 11. Curvas ROC 	La bolsa de las características softbiométricas propuestas tienen resultados prometedores en la re-identificación ya que son superiores al 90%.	Las diferentes características softbiométricas pueden ser utilizadas para la re-identificación. El único detalle es que necesita saber cuál es la característica softbiométrica más discriminante por secuencia de video.
Pedestrian re-identification using color feature in multi surveillance video, (Zhang, Matsumoto, Yoshida, & Kita, 2016)	Propone un sistema de re-identificación que sea robusto ante cambios de iluminación, posición y oclusión.	<ol style="list-style-type: none"> 1. Características en el espacio de color UVW 2. Partition-hierarchical model (SPM) 3. Modelo de mezclas gaussianas 	El espacio de color utilizado mejora las debilidades con respecto al histograma y el SPM mejora la falta de información espacial. Esto mejora la precisión de computo de la imagen para la re-identificación.	El SPM puede ser útil como descriptor espacial para la re-identificación.

Artículo	Objetivo	Técnicas	Resultados	Comentarios
Discriminative Image Descriptors for Person Re-identification, (Ma, Su, & Jurie, 2014)	Propone un sistema de re-identificación utilizando la combinación de dos descriptores llamado BiCov.	<ol style="list-style-type: none"> 1. Low-Level Biologically Inspired Features (BIF) 2. Filtros de Gabor 3. MAX pooling operation 4. Modelo de Mezclas Gaussianas. 	El descriptor propuesto es más robusto a cambios de iluminación, escala, variaciones de fondo para la re-identificación que los enfoques del estado del arte.	Los descriptores propuestos pueden ser útiles para la re-identificación. El modelo de mezclas gaussianas se sigue utilizando para la sustracción de fondo.
Ensemble of different approaches for a reliable person re-identification system, (Nanni, Munaro, Ghidoni, Menegatti, & Brahnam, 2016)	Propone un sistema de re-identificación utilizando diferentes espacios de color y variantes de técnicas del estado del arte.	<ol style="list-style-type: none"> 1. gBiCov 2. Symmetry-Driven Accumulation of Local Features (SDALF) 3. Custom Pictorial Structures (CPS) 4. chromatic content and color displacement (CCD) 5. Color Invariants (CI) 6. Skeleton-based Person Signature (SPS) 7. Weber's Law Descriptor proposed (WLD), Local Phase Quantization from Three Orthogonal Plane (LPQT) 8. Volume Local Phase Quantization (VLPQ). 	El sistema no trabaja en tiempo real pero produce un descriptor general para la re-identificación de personas que puede ser generalizado en diferentes secuencias de imágenes.	El artículo propone diferentes características para la re-identificación aunque debido a la complejidad computacional puede no ser útil.

Artículo	Objetivo	Técnicas	Resultados	Comentarios
<p>Survey on colour, texture and shape features for person re-identification, (Poongothai & Suruliandi, 2016)</p>	<p>Brindar al lector un análisis de las diferentes características existentes para la re-identificación de personas en el estado del arte.</p>	<ol style="list-style-type: none"> 1. Características en el espacio de color RGB. 2. Características en el espacio de color L*a*b*. 3. Características en el espacio de color YCbCr. 4. Características en el espacio de color HSV. 5. Local Binary Pattern. 6. Scale Invariant Feature 7. Representation. 8. Speeded Up Robust Feature. 9. Gradient Location and Orientation Histogram. 10. Gabor. 11. Schmid. 12. Histogram of Oriented Gradient. 	<p>El artículo presenta un resumen de las ventajas y desventajas de cada característica, así como diferentes aplicaciones. Esto ayuda a entender las condiciones de la re-identificación y generar descriptores más robustos.</p>	<p>La evaluación de las características del espacio de color puede ser útil para la generación de un descriptor robusto.</p>

2.4 Análisis de trayectorias

2.4.1 On the use of a minimal path approach for target trajectory analysis (Cancela, B., Ortega, M., Penedo, M. G., Novo, J., & Barreira, N. 2013).

El presente artículo muestra una estrategia para el análisis de trayectorias en escenarios multi-objetivos y está basado en el uso de un algoritmo de rutas mínimas utilizando contornos geodésicos activos. El principal objetivo es evitar la necesidad de calcular y almacenar rutas típicas.

La metodología propuesta se basa en la aplicación de un algoritmo para obtener la trayectoria y en el uso de técnicas de ruta mínima para el análisis de la trayectoria.

El método de rutas mínimas con contornos geodésicos es “Sethian Fast Marching”. Este fue adaptado restringiendo las posibilidades de que un objetivo pasé en ciertas regiones asignando un valor de pixel alto. La figura 2.22 muestra la restricción de regiones.

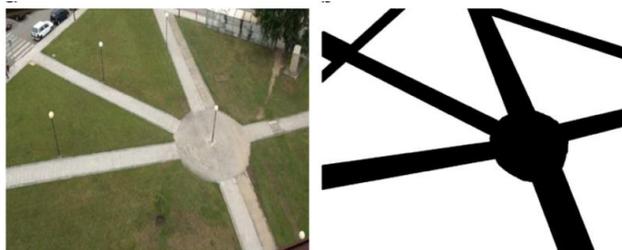


Figura 2.22. Ejemplo de restricción de regiones, donde no es posible detectar una ruta mínima cuando una persona se mueva a través del pasto.

Un punto crucial de la metodología es como obtener la ruta mínima correcta. Esto va depender de los diferentes escenarios. Por tanto, se utilizan diferentes modelos dependiendo de las condiciones, como se muestra en la tabla 2.10.

Tabla 2.10 Modelos de la metodología propuesta.

Método	Descripción
Fixed potential	Cuando es posible incluir información a priori del entorno de las imágenes. Por ejemplo regiones prohibidas de acceso.
Online potential	Cuando no se tiene información de la escena. Se utiliza el movimiento de los objetos para crear una imagen potencial. Es decir se almacena la posición de los objetos a lo largo del tiempo para poder obtener la ruta.
Pool of potential images	Cuando tienes diferentes puntos de inicio y fin por diferentes objetos es necesario definir diferentes rutas mínimas que se activan dependiendo del nivel del conocimiento.

Una vez que se tiene la ruta mínima asociado a cada objetivo se utiliza una métrica para determinar la ruta del objetivo como comportamiento normal o anormal como se muestra en la figura 2.23.



Figura 2.23. *Ejemplo del método Fixed potential. El sistema detecta como comportamiento anormal las rutas que tienen accesos restringidos.*

2.4.2 Visual abnormal behavior detection based on trajectory sparse reconstruction analysis (Li, C., Han, Z., Ye, Q., & Jiao, J. 2013).

El presente artículo propone un sistema de detección de comportamiento anormal basado en el análisis de trayectorias.

La metodología propuesta se basa en un diccionario de trayectorias predefinido de un escenario de video estático. Primeramente, se reúnen un conjunto de trayectorias de comportamientos normales de una secuencia de video por medio de un algoritmo de seguimiento de objetos o de detección de movimiento. Cuando el conjunto de trayectorias está completo, se realiza una observación de las trayectorias y son manualmente clasificadas en diferentes sub-conjuntos a partir de su apariencia, llamados “Route sets” como se muestra en la figura 2.24.

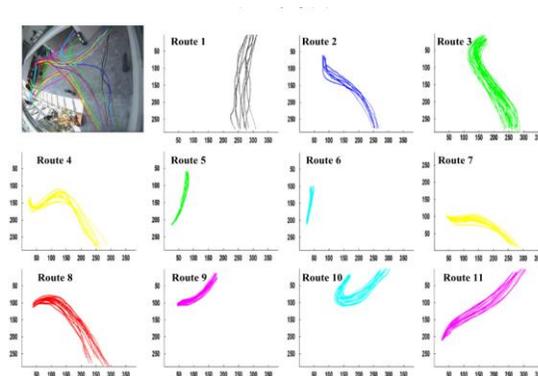


Figura 2.24. *Ejemplo de los Route sets.*

Para toda la colección de trayectorias, las características son representadas con “Least-squares Cubic Spline Curves Approximation” (LCSCA, por sus siglas en ingles) y son extraídas para la construcción del diccionario.

Al realizar la detección del comportamiento anormal, es necesario utilizar el análisis de reconstrucción disperso en el diccionario para clasificar las trayectorias de los objetos, donde el objetivo es reconstruir la trayectoria con las muestras de los diccionarios. De igual forma

la minimización L1-norm es utilizado para resolver la reconstrucción de los coeficientes, en los cuales los residuos de construcción de cada Route set también puede ser calculado. El residuo de reconstrucción mínimo es usado para clasificar la trayectoria como un comportamiento normal o anormal con un umbral empíricamente definido. El método propuesto es mostrado en la figura 2.25.

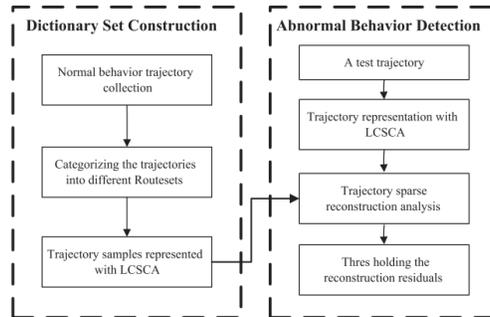


Figura 2.25. Método propuesto para la clasificación de trayectorias

2.4.3 A Method of Abnormal Pedestrian Behavior Detection Based on the Trajectory Model (Li, X., Lu, L., Yu, J., & Wang, K. 2013)

El presente artículo propone un método para la detección de 5 comportamientos anormales de peatones para la reducción de accidentes de tráfico. El modelo de detección está compuesto en relación con la trayectoria del peatón y la carretera. El método propuesto está compuesto del reconocimiento de peatones, seguimiento del peatón y la detección de comportamiento anormal como se muestra en la figura 2.26.

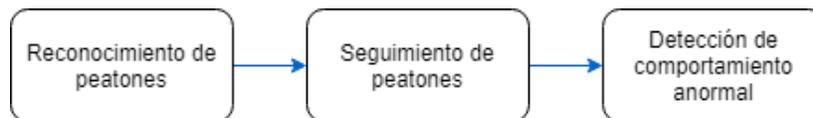


Figura 2.26. Diagrama del método propuesto.

Los comportamientos anormales están catalogados en 5 categorías: cruce de fronteras por carretera, paradas ilegales, cruce de caminos, caminar a lo largo de la frontera de la carretera, y entrar en la región de carretera, como se muestra en la figura 2.27.

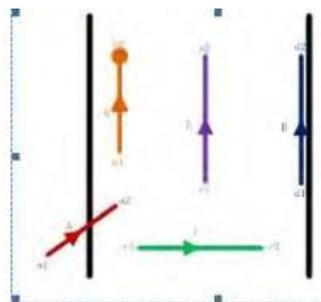


Figura 2.27. Comportamientos anormales a evaluar.

El método propuesto utiliza una cámara estática, donde realiza la detección de movimiento por medio de diferencias de fondos basado en el modelo de mezclas gaussianas. Posteriormente se obtienen los fondos de primer plano de los peatones y son reconocidos con base en la relación de aspecto de peatones y vehículos. La trayectoria de la persona es obtenida por medio de un modelo de predicción lineal con base en la posición actual. El modelo para la detección del comportamiento es construido a partir de las características de los 5 comportamientos anormales y la estructura del camino, donde se obtiene modelos matemáticos por cada comportamiento anormal que son utilizados para detectar y clasificar si una trayectoria corresponde a un comportamiento anormal del peatón.

2.4.4 Abnormal Behavior Detection Using Trajectory Analysis in Camera Sensor Networks (Wang, Y., Wang, D., & Chen, F. 2013).

El presente artículo propone un sistema para la detección de comportamiento anormal utilizando una red de cámaras. El análisis de trayectoria y el modelado de anomalías se llevan a cabo mediante el procesamiento de un sólo nodo, mientras que la detección de anomalías se realiza mediante votación multi-nodo. Las trayectorias son construidas y presentadas como una secuencia de símbolos. Las secuencias utilizan el modelo de Markov para la construcción de la matriz de transición de probabilidad la cual es utilizada para automáticamente analizar un comportamiento anormal. El sistema esta descrito en la figura 2.28.

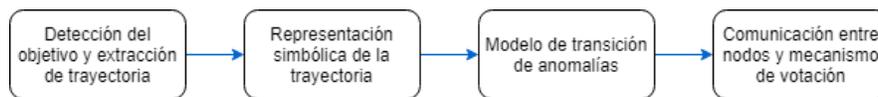


Figura 2.28. Diagrama general del sistema propuesto.

Detección del objetivo y extracción de trayectoria:

En este apartado se utiliza un método de modelado y actualización de fondo basado en una adaptación de mezclas gaussianas para la detección de objetivos en movimiento y la extracción de su posición, como se muestra en la figura 2.29. La posición del objetivo es definida como el centro del rectángulo de etiquetado y la trayectoria está compuesta por una secuencia de posiciones.

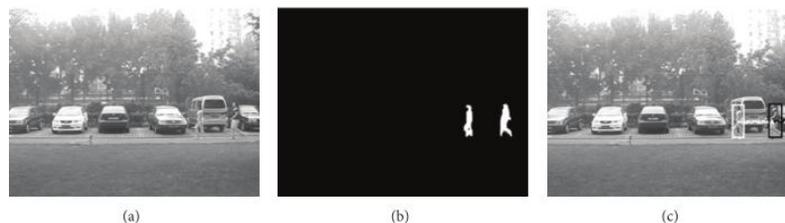


Figura 2.29. Donde (a) es el frame original, (b) es la detección de los objetivos y (c) es la extracción de la trayectoria.

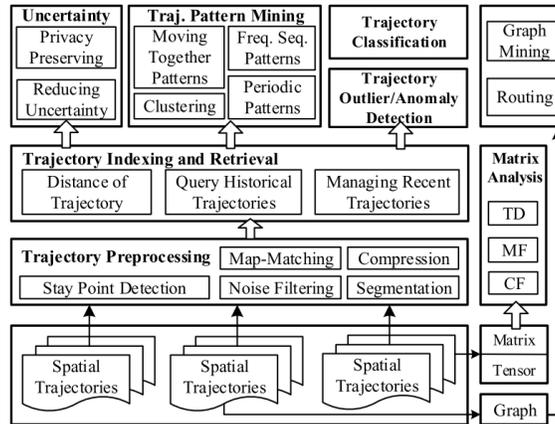


Figura 2.31. Esquema general del análisis de trayectorias con Data Mining.

Spatial Trajectories (Trayectorias espaciales): Define la representación de trayectorias. Generalmente se representan como personas, vehículos, animales y fenómenos naturales.

Trajectory Preprocessing (Preprocesamiento de trayectorias): Etapa fundamental para el análisis de trayectorias. Generalmente engloban diferentes tareas mejoran la trayectoria para procesos futuros. Este preprocesamiento está dado por:

- *Map-Matching*: Proyecta cada punto de una trayectoria en un segmento de camino donde realmente fue generado.
- *Compression*: Reduce el tamaño de la trayectoria manteniendo la información útil.
- *Stay Point Detection*: Algoritmo que identifica la localización de un objeto en movimiento que se ha quedado estático durante un periodo de tiempo.
- *Noise Filtering*: Remueve puntos de ruido en la trayectoria causados por una mala localización en los sistemas de posición.
- *Segmentation*: Divide la trayectoria en fragmentos definidos por un intervalo de tiempo o forma espacial para futuros procesos como agrupamiento o clasificación.

Trajectory Indexing and Retrieval (Trayectoria indexada y recuperada): El análisis de trayectorias es muy lenta, ya que se acceden a muchas muestras o partes de estas en un gran número de veces. Para solucionar este problema es necesario aplicar alguna de las técnicas mostradas en la figura 2.31.

Trajectory Uncertainty (Trayectoria incierta): Los objetos se mueven continuamente, pero sus localizaciones pueden ser sólo actualizados en tiempos discretos.

Trajectory Pattern Mining (Trayectoria de patrones): El alto volumen de trayectorias espaciales da la oportunidad de analizar los patrones de movilidad de los objetos.

Trajectory Classification (Clasificación de trayectorias): Usando métodos de aprendizaje supervisados, es posible clasificar las trayectorias o segmentos de estas en clases, las cuales representan actividades.

Trajectory Outlier Detection (Detección de trayectorias atípicas): Detecta trayectorias que pueden ser anormales definidos en términos de métricas de similitud.

Las trayectorias pueden ser convertidas en otros formatos como matrices, gráficas o “tensor”. Las nuevas representaciones expanden los enfoques para la extracción de datos de trayectorias permitiendo aprovechar otras técnicas como filtrado colaborativo (CF), factorización de matriz (MF) y descomposición tensorial (TD).

2.4.6 Aportaciones de análisis de trayectorias

Tabla 2.11. Aportaciones de los artículos de análisis de trayectorias.

Artículo	Objetivo	Técnicas / Áreas	Resultados	Comentarios
On the use of a minimal path approach for target trajectory analysis (Cancela, Ortega, Penedo, Novo, & Barreira, 2013)	Un sistema de análisis de trayectorias basado en el uso de un algoritmo de rutas mínimas.	Sethian Fast Marching Algoritmo minimal path Geodesic active contours	El análisis de trayectorias arroja buenos resultados en la interpretación pero es necesario condicionar ciertos escenarios.	Delimitación de áreas prohibidas Base de datos CANDELA BARD
Visual abnormal behavior detection based on trajectory sparse reconstruction analysis (Li, Han, Ye, & Jiao, 2013)	Un sistema de detección de comportamiento anormal basado en el análisis de trayectorias	Least-squares Cubic Spline Curves Approximation (LCSCA) Sparse Reconstruction Analysis (SRA) L1-norm minimization	El método propuesto logra buenos resultados detectando trayectorias anormales con un porcentaje mayor al 85% en comparación con otros métodos propuestos	Construcción de un diccionario para la representación de datos LCSSA. Base de Datos Caviar NGSIM
A Method of Abnormal Pedestrian Behavior Detection Based on the Trajectory Model (Li, Lin, Yu, & Wang, 2013)	Un sistema de detección de 5 comportamientos de peatones anormales para la reducción de accidentes de tráfico.	Background difference Gaussian Mixture Model Transformación R	La detección de los cinco comportamientos se hace con precisión aunque está limitado a que no exista oclusión.	Base de Datos Guangzhou Road Monitoring System
Abnormal Behavior Detection Using Trajectory Analysis in Camera Sensor Networks (Wang, Wang, & Chen, 2014)	Un sistema de análisis de trayectorias utilizando una red de cámaras, en donde un nodo principal es el detector de trayectorias anormales	Red Bayesiana Hidden Markov model (HMM) Symbolic aggregate approximation(SAX)	Los resultados experimentales muestran que la representación SAX de la trayectoria objetivo es efectiva y el método propuesto posee una fiabilidad relativamente alta de detección de comportamiento anormal.	Representación de datos simbólicos de las trayectorias. Base de Datos Propia
Trajectory Data Mining : An Overview (Yu, 2015)	Brindar al lector un panorama general de las diferentes áreas técnicas para el análisis de trayectorias.	Trayectorias espaciales Preprocesamiento de trayectorias Trayectoria indexada y recuperada Trayectoria incierta Trayectoria de patrones Clasificación de trayectorias Detección de trayectorias atípicas	Ofrece una visión general sobre cómo obtener conocimiento desde las trayectorias.	Preprocesamiento de trayectorias Trayectoria de patrones Clasificación de trayectorias Detección de trayectorias atípicas.

Marco teórico

El presente capítulo habla del marco teórico, en donde se detallan diferentes teorías y métodos utilizadas en el desarrollo de este sistema de videovigilancia inteligente. A continuación, se detallan los métodos utilizados para los módulos del sistema con base en la arquitectura general del sistema propuesto, es importante notar que el apartado 4.12 sólo se utiliza para la evaluación de la experimentación y no corresponde a ningún módulo del sistema.

- *Módulo de detección de personas*
 - ✓ Modelo de mezclas gaussianas
 - ✓ Filtro gaussiano
 - ✓ Operadores morfológicos
 - ✓ Histograma de gradientes orientados
 - ✓ Máquinas Vector Soporte

- *Módulo de descripción de personas*
 - ✓ Promedio de gradiente
 - ✓ Medidas estadísticas
 - ✓ Espacios de color
 - ✓ Transformaciones de los espacios de color

- *Módulo de re-identificación*
 - ✓ Medidas de distancia
 - ✓ RandomForest

3.1 Modelo de mezclas gaussianas

El modelo de mezclas gaussianas fue inicialmente propuesto por (Friedman & Russell, 1997) en un sistema de videovigilancia de tráfico y mejorado por (Stauffer & Grimson, 1999) utilizando una matriz de covariancia.

La técnica analiza una imagen en un tiempo t , genera una imagen compuesta del promedio de los píxeles con mayor probabilidad gaussiana como modelo de fondo. Para finalmente obtener los píxeles que pertenecen a objetos en primer plano y generar la imagen con la información superpuesta. La figura 3.1 muestra de manera general la técnica.

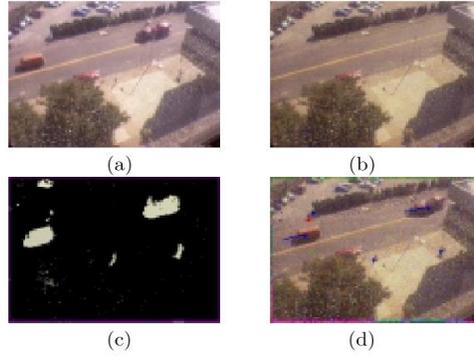


Figura 3.1. a) es la imagen actual. b) es el modelo. c) es la detección de objetos de primer plano y d) es la información superpuesta.

Esta técnica consiste en que cada pixel de la imagen es caracterizado por su intensidad en el espacio de color *RGB*. La probabilidad de observar el pixel se obtiene dada la ecuación 3.1, donde K es el número de distribuciones. $w_{i,t}$ es un peso asociado a la muestra en el tiempo i con una $\mu_{u,t}$. $\Sigma_{i,t}$ es la desviación estándar de la matriz de covariancia.

$$P(X_t) = \sum_{i=1}^K w_{i,t} \eta(X_t, \mu_{u,t}, \Sigma_{i,t}) \text{ donde } \Sigma_{i,t} = \sigma_{i,t}^2 I \quad \text{Ec. (3.1)}$$

Para poder obtener la probabilidad de observación es necesario contar con una función de densidad de probabilidad gaussiana. Esta función se describe en la ecuación 3.2.

$$\eta(X, \mu, \Sigma) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(X-\mu)\Sigma^{-1}(X-\mu)} \quad \text{Ec. (3.2)}$$

3.2 Filtro gaussiano

Un filtro gaussiano es la simulación de una distribución gaussiana bivalente. El valor máximo aparece en el pixel central y disminuye hacia los extremos tanto más rápido cuanto menor sea el parámetro de desviación típica. El resultado será un conjunto de valores entre 0 y 1. Para transformar la matriz resultante a una matriz de números enteros se divide toda la matriz por el menor de los valores obtenidos. La ecuación 3.3 representa la convolución del filtro gaussiano, (D'Haeyer, J. P. F, 1989), en donde I_s es la imagen suavizada, I es la imagen original y \otimes es la operación de convolución.

$$I_s = \frac{1}{273} \begin{bmatrix} 1 & 4 & 7 & 4 & 1 \\ 4 & 16 & 26 & 16 & 4 \\ 7 & 26 & 41 & 26 & 7 \\ 4 & 16 & 26 & 16 & 4 \\ 1 & 4 & 7 & 4 & 1 \end{bmatrix} \otimes I \quad \text{Ec. (3.3)}$$

3.3 Operadores morfológicos

El procesamiento morfológico es una técnica de la visión por computador que analiza las imágenes basándose en propiedades de la forma, empleando álgebra de conjuntos, (Alegre, Pajares, de la Escalera 2016). Estas se encargan de obtener la forma y estructura de los objetos de la imagen. Además, permiten separar los objetos unos de otros, obtener contornos, descomponer formas complejas en formas más simples, reconstruir elementos distorsionados entre otros. Existen dos tipos de operadores elementales, la erosión y la dilatación. La dilatación utiliza el operador lógico *OR* y la erosión utiliza el operador lógico *AND*. A partir de ellos se crean otros operadores denominados de apertura y cerradura. Para poder aplicar los operadores es necesario definir un elemento estructurante, el cual determina los píxeles que se utilizan o no en la operación. Las ecuaciones 3.4, 3.5, 3.6 y 3.7 representan los operadores morfológicos, donde M es una imagen binaria y E es el elemento estructurante.

$$\text{dilatación} = M \oplus E \quad \text{Ec. (3.4)}$$

$$\text{erosión} = M \otimes E \quad \text{Ec. (3.5)}$$

$$M_{\text{apertura}} = (M \otimes E) \oplus E \quad \text{Ec. (3.6)}$$

$$M_{\text{cerradura}} = (M \oplus E) \otimes E \quad \text{Ec. (3.7)}$$

El elemento estructurante utilizado en la implementación fue un rectángulo de tamaño 5. La figura 3.2 muestra el elemento estructurante.

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Figura 3.2. Elemento estructurante.

3.4 Histograma de gradientes orientados

El Histograma de Gradientes Orientados fue propuesto por (Dalal & Triggs, 2005). La idea general es obtener un descriptor de la imagen dividida en celdas por el cálculo de su gradiente y la distribución del histograma de la celda. Dada una imagen el primer paso es computar el gradiente de la imagen en orientación vertical y horizontal realizando una convolución con una máscara de Sobel u otra similar. Posteriormente se divide la imagen en celdas de tamaño 8 y se calcula el histograma de gradientes de cada una. El histograma acumula el número de orientaciones de 0 a 180 grados. La figura 3.3 muestra el histograma acumulado del gradiente.



Figura 3.3. a) *histograma de las orientaciones.* b) *imagen del histograma acumulado.* (Ernest Valveny, 2018)

El siguiente paso es el agrupamiento de las celdas en bloques de tamaño 16 para la normalización. El bloque contiene un único valor y es la combinación de los histogramas de las celdas. La normalización del bloque se utiliza partiendo de la norma L2. La ecuación 3.8 es el cálculo de la normalización, donde v es el vector del histograma acumulado del bloque. ε valor decimal pequeño que evita divisiones entre 0.

$$v' = \frac{v}{\sqrt{\|v\|_2^2 + \varepsilon}} \quad \text{Ec. (3.8)}$$

Una vez cada bloque se ha normalizado, el siguiente paso es realizar una concatenación de los vectores de los bloques de la imagen. Esta concatenación genera un vector resultante de tamaño N y es el descriptor de la imagen. La figura 3.4 da una idea general del descriptor.

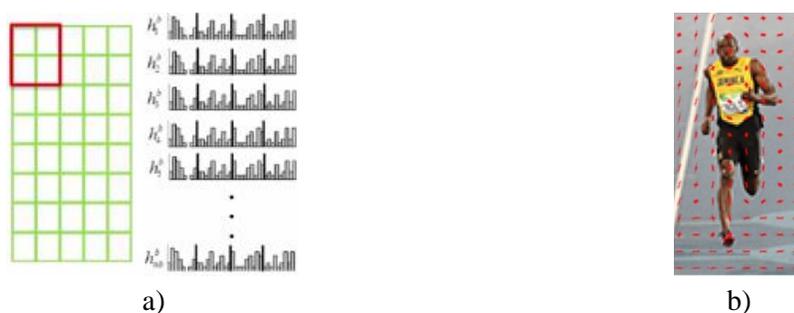


Figura 3.4. *Concatenación de los histogramas de los bloques.* b) *descriptor resultante,* (Ernest Valveny, 2018).

3.5 Máquinas Vector Soporte

Las máquinas de vector soporte (*SVM*) se basan en aprendizaje estadístico. Fue introducido por (Boser, Guyon, & Vapnik, 1992) y fueron pensados inicialmente para resolver problemas de clasificación binaria. Actualmente también son usadas para otros problemas de regresión y agrupamiento.

Las SVM se categorizan como clasificadores lineales, ya que inducen separadores lineales/hiperplanos en el espacio de los ejemplos de entrada, si éstos son separables entre si, ver figura 3.5. Además se definen como problemas de optimización.

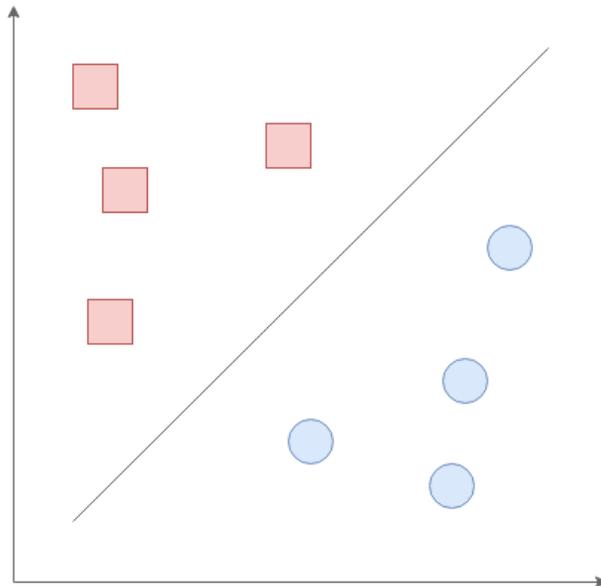


Figura 3.5. Ejemplos de entrada linealmente separables.

La idea principal es seleccionar un hiperplano de separación que equidiste de los ejemplos más cercanos para cada clase de tal forma que se consiga la distancia máxima (*margen máximo*) a cada lado del hiperplano. Es importante saber que para definir el hiperplano de separación, sólo se consideran los ejemplos de entrenamiento de cada clase que se encuentren en la frontera de los márgenes (*vectores soporte*), ver figura 3.6. Con lo anterior mencionado se obtiene buena capacidad de generalización de los datos y se evita el problema de sobreajuste.

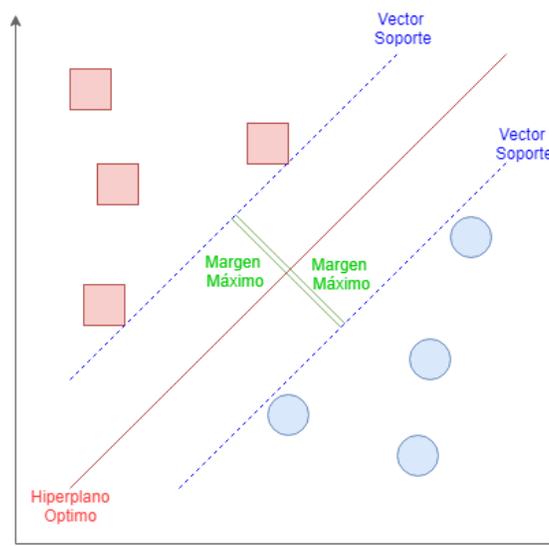


Figura 3.6. Representación de márgenes y vectores soporte.

Dado un conjunto de ejemplos $S = \{(x_1, y_1), \dots, (x_n, y_n)\}$, donde x representa la ubicación espacial e y representa la clase perteneciente. Se define un hiperplano de separación como una ecuación lineal que separa dicho conjunto de manera óptima, ver ecuación 3.9, en donde w y x son coeficientes reales.

$$h(x) = \langle w, x \rangle + b \quad \text{Ec. (3.9)}$$

Un hiperplano será óptimo cuando su margen sea de tamaño máximo entre las fronteras de las clases de la muestra de datos. Adicionalmente las máquinas de vector pueden ser extendidas a casos linealmente no separables, con n cantidad de clases. Para fines prácticos sólo se utilizó la clasificación binaria.

3.6 Promedio de gradiente

(Denman, Fookes, Bialkowski, & Sridharan, 2009) propone un método de segmentación de partes del cuerpo humano (*piernas, torso y cabeza*). Se asume inicialmente que los personas siempre se van a encontrar en orientación vertical.

Esta técnica consiste en dividir en 3 partes el cuerpo humano de tal forma que se tenga la cabeza, el torso y las piernas. El cálculo del gradiente se obtiene mediante la ecuación 3.10. Donde $V_g(i)$ es el vector promedio de gradiente, i es la posición de la fila. N es el ancho de la imagen. $I(i, j)$ es el valor del pixel y $M(i, j)$ es una imagen binaria donde 1 representa al objeto en primer plano y 0 representa fondo.

$$V_g(i) = \frac{\sum_{j=1}^{N-1} |I(i, j) - I(i, j - 1)| M(i, j)}{\sum_{j=0}^{N-1} M(i, j)} \quad \text{Ec. (3.10)}$$

Una vez aplicada la ecuación se busca obtener un histograma del promedio de gradiente en la cual se localizan dos puntos: cuello y cintura. Estos dan la referencia para la segmentación de las 3 partes del cuerpo humano. La figura 3.7 muestra un ejemplo del histograma del promedio de gradiente.

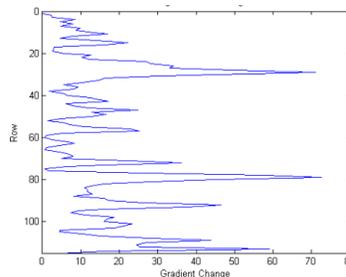


Figura 3.7. Histograma del promedio de gradiente de una persona.

Posteriormente se detectan los puntos de división de las regiones. El primer punto a detectar es el cuello y este se ubica entre el 10% al 30% de la altura de la persona, (Denman, Fookes,

Bialkowski, & Sridharan, 2009). El cuello se encuentra mediante la ecuación 3.11, donde P_n es la fila donde se encuentra el cuello y H es la altura de la persona en píxeles.

$$P_n = \operatorname{argmax}_{0.1(H)}^{0.3(H)} v_g(i) \quad \text{Ec. (3.11)}$$

El segundo punto a detectar es la cintura y este se ubica entre el 35% al 70% de la altura de la persona, (Denman, Fookes, Bialkowski, & Sridharan, 2009). La cintura se encuentra mediante la ecuación 3.12, donde P_w es la fila donde se encuentra la cintura y H es la altura de la persona en píxeles.

$$P_w = \operatorname{argmax}_{0.35(H)}^{0.7(H)} v_g(i) \quad \text{Ec. (3.12)}$$

El resultado de aplicar las ecuaciones 4, 5 y 6 genera los puntos de referencia para la segmentación del cuerpo de la persona. En la figura 3.8 se puede observar el histograma y la segmentación.

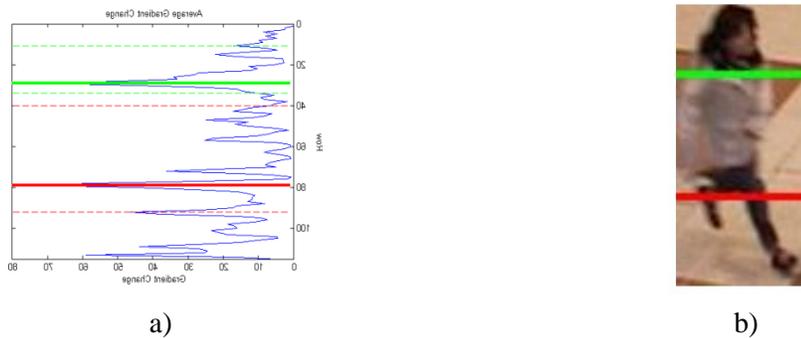


Figura 3.8. a) Histograma de gradiente, b) Segmentación del cuerpo.

3.7 Medidas de distancia

Las medidas de distancia se utilizan para medir el grado de similaridad o disimilaridad entre dos objetos por medio de una representación simbólica, (Huang, 2008). Esto se interpreta en una representación geométrica, en donde los objetos son puntos en un espacio definido.

Hablando matemáticamente la distancia se define como la representación de dos vectores de los objetos, ver ecuación 3.13. En donde se obtiene un valor numérico que indica la similaridad/disimilaridad entre los objetos. El número de distancia indica que mientras menor sea existe mayor similitud en los objetos, caso contrario, mientras más grande sea indica menor similitud, es decir mayor disimilaridad.

$$D(\vec{x}, \vec{x}') \quad \text{Ec. (3.13)}$$

Existen varias medidas de distancia, en este trabajo sólo hablaré de las distancias cuantitativas. Se definen dos vectores numéricos del mismo tamaño de dos objetos diferentes: $\vec{x} = \{x_1, x_2, x_3 \dots x_n\}$ y $\vec{x}' = \{x'_1, x'_2, x'_3 \dots x'_n\}$ en donde el resultado de medir

las n variables indican la similaridad/disimilaridad. Para esto se aplican las siguientes distancias:

- Distancia Euclídea, ver ecuación 3.14. Es una de las distancias más comunes y suele ser sensible a las medidas, es decir valores altos pueden generar mucha diferencia de similitud entre los objetos, una solución es normalizar los datos.

$$D(\vec{x}, \vec{x}') = \sqrt{\sum_{i=1}^n (x_i - x'_i)^2} \quad \text{Ec. (3.14)}$$

- Distancia Manhattan, ver ecuación 3.15. Esta se representa como la sumatoria de las diferencias absolutas entre las coordenadas que componen al vector.

$$D(\vec{x}, \vec{x}') = |x_i - x'_i| \quad \text{Ec. (3.15)}$$

- Distancia Chebyshev, ver ecuación 3.16. Esta se representa como la mayor diferencia en cualquiera de las coordenadas del espacio.

$$D(\vec{x}, \vec{x}') = \max(|x_i - x'_i|) \quad \text{Ec. (3.16)}$$

- Distancia del Coseno, ver ecuación 3.17. Este se representa como el producto escalar de dos vectores y un ángulo definido como un criterio de similaridad.

$$D(\vec{x}, \vec{x}') = \|\vec{x}\| \|\vec{x}'\| \cos(\theta) \quad \text{Ec. (3.17)}$$

3.8 Medidas estadísticas

El histograma de una imagen es una función discreta que contabiliza el número de ocurrencias de los canales de una imagen. Representa una gráfica en donde la abscisa es el nivel de gris o color y la ordenada es la frecuencia de aparición de los niveles. A simple vista un histograma brinda una idea aproximada de la distribución de los niveles de grises/color de los canales, (Moctezuma, D., 2016). Esto permite extraer información estadística de la imagen.

La *media*, ver ecuación 3.18, en donde i es el nivel de gris/color y $p(i)$ es la probabilidad de ocurrencia de ese color, L representa el tamaño del histograma. Esta medida estadística Indica el valor medio de la imagen que coincide con el valor medio del histograma.

$$\mu = \sum_{i=0}^{L-1} i \cdot p(i) \quad \text{Ec. (3.18)}$$

La *desviación estándar*, ver ecuación 3.19. Es una medida de dispersión que indica que tan distante están los niveles dispersión entre los niveles de grises/color de la imagen los unos a los otros con respecto a la media.

$$\sigma = \sqrt{\sum_{i=0}^{L-1} (i - \mu)^2 \cdot p(i)} \quad \text{Ec. (3.19)}$$

La *energía*, ver ecuación 3.20. Indica el grado de dispersión de los niveles de grises/color de la imagen. Si existe un único nivel de color (por ejemplo, sólo 255) el valor de la energía sería 1, por tanto, es una imagen homogénea.

$$\text{energía} = \sum_{i=0}^{L-1} p(i)^2 \quad \text{Ec. (3.20)}$$

La *entropía*, ver ecuación 3.21. Es la función inversa de la energía, este indica que mientras mayor sea el valor, existe mayor desorden en los niveles de grises/color de la imagen. Es decir, existe mayor participación de los niveles de grises/color.

$$\text{entropía} = \sum_{i=0}^{L-1} p(i) \cdot \log(p(i)) \quad \text{Ec. (3.21)}$$

3.9 Espacios de color

Un espacio de color es un método de representación que especifica, crea o visualiza cualquier color, (Alegre, Pajares, & Escalera, 2016). La especificación numérica de un color se efectúa mediante 3 componentes que lo definen de forma cualitativa y cuantitativa. Por ejemplo, el televisor trabaja con los componentes rojo, verde y azul, mientras que las impresoras actuales utilizan cian, magenta, amarillo y negro. Existe una terminología que es imperante conocer antes de profundizar los diferentes espacios de color, ver tabla 3.1.

Tabla 3.1. Definición de términos.

Termino	Definición
Brillo	Indica el grado de iluminación de una imagen
Tono	Indica la similitud de colores (rojo, amarillo, verde, azul o porciones entre dos de ellos)
Color	Grado de intensidad de los colores
Luminosidad	Brillo de una zona respecto a otra en la imagen
Croma	Color de un área respecto al brillo de un blanco de referencia
Saturación	Relación entre el color y brillo.

- *Escala de Grises*

Es un espacio cromático que representa valores grises. Este puede ir de un rango de [0,255] valores. Esta escala es de las más comunes usadas en el procesamiento digital de imágenes.

- *RGB*

Es el espacio de color más común, se compone de 3 colores primarios *rojo (R)*, *verde (G)* y *azul (B)*. La suma aritmética de estos colores primarios, ver ecuación 3.22, se traduce como la mezcla que resulta en un x color.

$$x = R + G + B \quad \text{Ec. (3.22)}$$

Los colores primarios que componen este espacio se representan mediante 3 componentes elementales y define un espacio de color básico, ver figura 3.9. En donde cada color se define como un punto tridimensional perteneciente a un cubo.

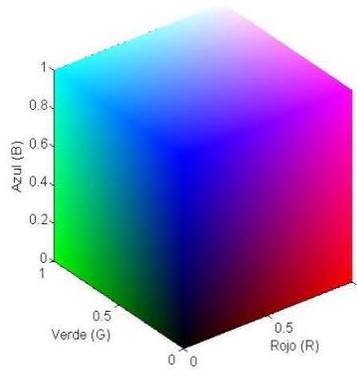


Figura 3.9. Representación del espacio de color RGB.

- *HLS*

Este espacio de color está basado en la forma de percepción humana. Se caracteriza en términos del *tono (H)*, *iluminación (L)*, *saturación (S)*. Su representación espacial se define como un doble cono, ver figura 3.10. Es importante notar que el rango de valores de H , es de $[0,359]$

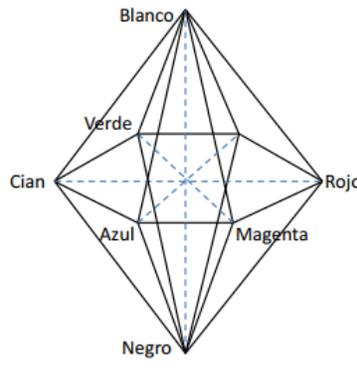


Figura 3.10. Representación del espacio de color HLS

- *HSV*

Este espacio de color, al igual que el HLS, está basado en la percepción humana. Se caracteriza en términos del tono (H), saturación (S) y el valor (V). Su representación espacial se define como un cono invertido, ver figura 3.11.

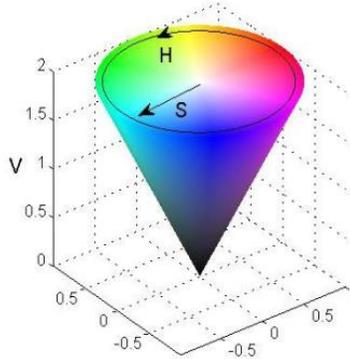


Figura 3.11. Representación del espacio de color HSV.

- *Cie-Lab*

Los espacios de color anteriormente mencionados pueden tener coeficientes negativos en determinadas longitudes de onda, lo cual es un error importante en los cálculos colorimétricos. Por tanto, este espacio de color busca evitar ese tipo de errores.

Lab se caracteriza en términos de *luminosidad (L)*, *a* es una referencia a la relación rojo/azul y *b* es una referencia a la relación amarillo/azul.

- *YCrCb*

Este espacio de color se basa en la luminosidad y cromaticidad. *Y* indica la luminancia, *Cb* es una representación de color en una escala entre azul y amarillo y *Cr* es una representación de color en una escala entre rojo y verde.

3.10 Transformaciones de los espacios de color

Ecuaciones obtenidas de (OpenCV: “Color conversions document”, 2018).

- *RGB a Escala de Grises*

$$gris = \frac{R + G + B}{3} \quad Ec. (3.23)$$

- *RGB a HLS*

$$H = \begin{cases} \frac{60(G - B)}{(\max(R, G, B) - \min(R, G, B))} \\ \frac{180(B - R)}{\max(R, G, B) - \min(R, G, B)} \\ \frac{300(R - G)}{\max(R, G, B) - \min(R, G, B)} \end{cases} \quad \text{Ec. (3.24)}$$

$$L = \frac{\max(R, G, B) - \min(R, G, B)}{2} \quad \text{Ec. (2.25)}$$

$$S = \begin{cases} \frac{\max(R, G, B) - \min(R, G, B)}{\max(R, G, B) + \min(R, G, B)} & L < 0.5 \\ \frac{\max(R, G, B) - \min(R, G, B)}{2 - \max(R, G, B) + \min(R, G, B)} & L \geq 0.5 \end{cases} \quad \text{Ec. (3.26)}$$

- *RGB a HSV*

$$H = \begin{cases} \frac{60(G - B)}{(\max(R, G, B) - \min(R, G, B))} \\ \frac{180(B - R)}{\max(R, G, B) - \min(R, G, B)} \\ \frac{300(R - G)}{\max(R, G, B) - \min(R, G, B)} \end{cases} \quad \text{Ec. (3.27)}$$

$$S = \begin{cases} \frac{\max(R, G, B) - \min(R, G, B)}{\max(R, G, B)} & \max(R, G, B) \neq 0 \\ 0 & \text{otro caso} \end{cases} \quad \text{Ec. (3.28)}$$

$$V = \max(R, G, B) \quad \text{Ec. (3.29)}$$

- *RGB a XYZ*

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad \text{Ec. (3.30)}$$

- *RGB a Lab*

$$L = \begin{cases} 116 \cdot Y^{1/3} - 16 & Y > 0.008856 \\ 903.3 \cdot Y & Y \leq 0.008856 \end{cases} \quad \text{Ec. (3.31)}$$

$$a = 500(f(X) - f(Y) + \Delta) \quad \text{Ec. (3.32)}$$

$$b = 500(f(Y) - f(Z) + \Delta) \quad \text{Ec. (3.33)}$$

$$\text{en donde, } \begin{cases} f(t) = \begin{cases} t^{1/3} & t > 0.008856 \\ 7.787t + 16/116 & t \leq 0.008856 \end{cases} \\ \Delta = \begin{cases} 128 & \text{imágenes 8 - bits} \\ 0 & \text{imágenes puntos flotante} \end{cases} \end{cases} \quad \text{Ec. (3.34)}$$

- *RGB a YCrCb*

$$\begin{bmatrix} Y \\ Cr \\ Cb \end{bmatrix} = \begin{bmatrix} 0 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 0.299 & 0.578 & 0.114 \\ 0.500 & -0.419 & -0.081 \\ -0.169 & -0.331 & -0.081 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad \text{Ec. (3.35)}$$

3.11 RandomForest

Los árboles de decisión se encuentran entre los métodos de clasificación más comunes, (Breiman, J.H., Olshen, & Stone, 1984). Se compone de *nodos* que validan las características, *ramificaciones*, que son las características evaluadas de la muestra de entrenamiento; y finalmente *hojas* que representan la clase, ver figura 3.12.

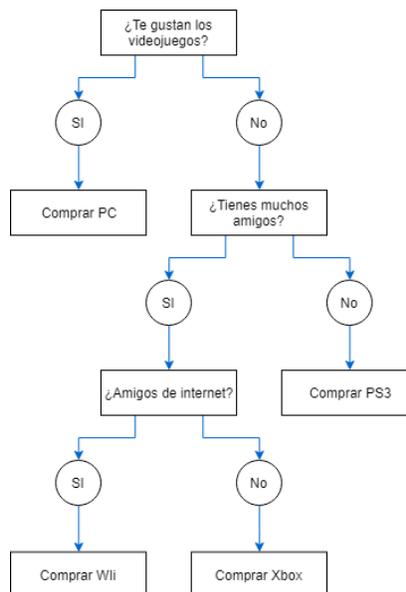


Figura 3.12. Ejemplo de un árbol de decisión.

Un árbol no tiene complejidad computacional siempre y cuando el número de características sea razonable, además son de fácil interpretación, aunque tiene la problemática de ser

sensible a variaciones en los datos de entrada, ya que pueden encaminar al árbol por un camino diferente, dando una clasificación errónea.

La forma en que estos trabajan se base en ramificaciones sucesivas del espacio de características, en donde siempre se busca la característica y el umbral que logran maximizar la homogeneidad del árbol.

Una de las formas más comunes de medir la homogeneidad es con el índice de *Gini*, ver ecuación 3.36, donde c son las clases que pertenecen a la ramificación, C es el total de clases y f_c es la porción de casos en la ramificación que pertenecen a la clase c . Básicamente el índice toma la clase con mayor ocurrencia y distingue la característica que mejor la discrimina.

$$G = \sum_{c=1}^c f_c \cdot (1 - f_c) \quad \text{Ec. (3.36)}$$

RandomForest es un algoritmo de múltiples arboles de decisión de aprendizaje supervisado propuesto por (Ho, 1995). Este algoritmo es usado para tareas de clasificación y regresión. La idea general es tener un bosque conformado de árboles de decisión, en donde cada árbol se entrena con un subconjunto aleatorio de la muestra de datos (*tree bagging*) y las características (*feature sampling*), ver ecuación 37.

La construcción del *tree bagging* consiste en la construcción de N árboles de decisión. Para esta tarea se toman de manera aleatoria de la muestra de entrenamiento un subconjunto de datos con remplazo definidos como (x, y) en donde x representan las características y y la clase. Por tanto se tiene que para el subconjunto (x, y) denotamos (x_n, y_n) donde $(n \in \{1 \dots N\})$. Finalmente, se entrena un árbol sobre cada par (x, y) . Lo anterior hace que no se sobreajuste. Para el *feature sampling*, se tiene un total de x características, en donde cada árbol sólo tiene acceso a una parte, se define como \sqrt{x} . Esto ayuda a bajar la correlación (ρ) existente entre los árboles.

Finalmente es necesario saber la independencia entre los árboles, por tanto, se calcula la varianza del bosque, ver ecuación 3.37, en donde N es el promedio de las variables aleatorias y σ^2 es la varianza de los árboles.

$$\sigma_{bosque}^2 = \rho\sigma^2 + \frac{\sigma^2(1 - \rho)}{N} \quad \text{Ec. (3.37)}$$

3.12 Métricas de evaluación

3.12.1 Matriz de Confusión

Una matriz de confusión es una tabla que permite ver el rendimiento de un algoritmo, (Stehman, 1997). Contiene información sobre las clasificaciones reales y predichas hechas

por un sistema de clasificación, ver tabla 3.2. Esta evalúa la predicción del algoritmo contra la clase real del objeto. Se definen los términos *verdaderos positivos (VP)*, *falsos positivos (FP)*, *falsos negativos (FN)*, *verdaderos negativos (VN)*.

Tabla 3.2. Matriz de confusión.

		Clase Real	
		V	F
Predicción	Resultado		
	V	VP	FP
	F	FN	VN

Los *VP* indican los valores positivos correctamente clasificados, es decir el valor real de la clase es *verdadero* y la predicción de la clase es *verdadero*. Los *VN* son los valores negativos correctamente clasificados, es decir el valor real de la *falso* y la predicción de la clase es *falso*. Los *FP* son los valores positivos incorrectamente clasificados, es decir el valor real de la *verdadero* y la predicción de la clase es *falso*. Los *FN* son los valores negativos incorrectamente clasificados, es decir el valor real de la *falso* y la predicción de la clase es *verdadero*.

3.12.2 Medidas Estadísticas

Dada la matriz de confusión, existen diferentes métricas que se pueden utilizar para medir los resultados. Suponiendo que se tiene un conjunto de clases definidas como $\{X, Y\}$. (Powers D. , 2011) estas medidas ayudan a saber que tan discriminante es el algoritmo.

La *tasa de verdaderos positivos (TPR)* indica la porción de instancias que fueron clasificadas correctamente, ver ecuación 3.38.

$$TPR = \frac{VP}{VP + FN} \quad Ec. (3.38)$$

La *tasa de falsos positivos (FPR)* indica la porción de instancias que fueron clasificadas incorrectamente, ver ecuación 3.39.

$$FPR = \frac{FP}{FP + VN} \quad Ec. (3.39)$$

El *accuracy* es una medida simple que indica la relación las predicciones correctas y las observaciones totales, ver ecuación 3.40.

$$accuracy = \frac{VP + VN}{VP + VN + FP + FN} \quad Ec. (3.40)$$

La *precisión* indica la relación entre las observaciones positivas predichas correctamente y las observaciones predichas totales, es decir indica que tan preciso es el algoritmo para clasificar, ver ecuación 3.41.

$$precisión = \frac{VP}{VP + FP} \quad Ec. (3.41)$$

El *recall* indica la relación entre las observaciones positivas correctamente predichas y el total de las observaciones de la clase real, en otras palabras indica que tanto aprendió el sistema, ver ecuación 3.42.

$$recall = \frac{VP}{VP + FN} \quad Ec. (3.42)$$

La *media-F* es la relación ponderada entre la *precisión* y el *recall*, indica que tan exacto es el sistema en clasificar, ya que toma en cuenta los *FP* y *FN*, ver ecuación 3.43.

$$media - F = \frac{2 * Precisión * Recall}{Precisión + Recall} \quad Ec. (3.43)$$

Análisis del problema y propuesta de solución

En este capítulo se hace un estudio de las diferentes problemáticas que deben ser resueltas para realizar el sistema de videovigilancia inteligente. De igual forma se detalla la metodología de solución que permite resolver los objetivos planteados en este trabajo de investigación.

4.1 Problemáticas de un sistema de videovigilancia inteligente

Como bien se ha mencionado anteriormente, la visión por computadora sigue siendo un tema de investigación vigente, esto debido a la ausencia de un sistema de visión artificial genérico que resuelva diferentes problemáticas como lo hace el ser humano.

En este trabajo se tienen detectado 5 principales problemas que dificultan la creación de un sistema de videovigilancia inteligente genérico. Estos son listados a continuación.

4.1.1 *Sustracción de fondo*

El primer paso de un sistema de videovigilancia inteligente es saber diferenciar entre los objetos de primer plano de interés de estudio, así como el fondo presente en la imagen que no proporciona información relevante.

Para la sustracción de fondo se tienen en cuenta 3 factores importantes que pueden influir directamente en el desarrollo del sistema:

1. *Variaciones de iluminación:* Cambios drásticos en las variaciones de iluminación afectan gravemente el desempeño de los algoritmos de sustracción de fondo. Esto se precisa para entornos exteriores, en donde la intensidad de luz solar es variante.
2. *Translación de los objetos:* Un objeto que tenga una translación lenta, influye directamente en el modelo de fondo que se genera. Los pixeles que lo conforman tienen mayor probabilidad de ser parte del modelo de fondo, esto se debe que, al aparecer en el mismo lugar, los algoritmos dejan de detectarlo como un objeto en movimiento. Por ejemplo, si un objeto en movimiento se queda estático, este pasara a ser parte del modelo de fondo, por tanto, si la translación es lenta, tendrá menos probabilidad de ser detectado.
3. *Sombras de los objetos:* Dependiendo de las condiciones de iluminación puede existir un crecimiento de regiones de los objetos debido a la proyección de la sombra de estos.

En el estado del arte se presentan diferentes métodos que logran realizar esta tarea de manera efectiva siempre y cuando existan ciertas limitantes. Existen métodos básicos como lo son la diferencia de fotogramas, métodos estadísticos como el modelo de mezclas gaussianas, así

como métodos con mayor complejidad como modelo de mezclas gaussianas difusas y segmentación del fondo mediante algoritmos de agrupamiento, entre otros.

Los métodos básicos, son los menos efectivos ya que son demasiado sensible a variaciones de iluminación y en fondos dinámicos no tienen precisión. Los métodos estadísticos son los más utilizados en el estado del arte y reportan buenos resultados. Finalmente, los métodos restantes, según la literatura, mejoran los resultados de los anteriores, pero al tener mayor grado de complejidad, estos difícilmente pueden ser implementados en un sistema de videovigilancia inteligente que funcione en tiempo real.

4.1.2 *Detección de personas*

Un punto importante para un sistema de videovigilancia inteligente, es que debe tener la capacidad de diferenciar objetos (vehículos, animales, humanos, etc.). En este trabajo el objetivo es lograr detectar personas. La literatura ofrece diferentes representaciones mediante descriptores, muchos de ellos están basados en descriptores a nivel local, a nivel región y a nivel global. El descriptor que ha tenido mayor reporte en el estado del arte es el *Histograma de Gradientes Orientados*.

Las problemáticas que se consideran para este trabajo en la detección de personas son:

1. *Oclusión*: Si una persona no aparece de cuerpo completo en las diferentes vistas de cámara esta difícilmente pueda ser detectada como un objeto de la clase persona.
2. *Forma*: El principal problema de la detección de personas, es que esta no tiene una forma constante, dependiendo del instante de tiempo la forma del objeto puede variar, sobre todo si se consideran diferentes posiciones como agacharse o llevar objetos consigo.

4.1.3 *Re-identificación de personas*

La re-identificación de personas, es la tarea principal de este trabajo de investigación para el sistema de videovigilancia inteligente. Se toman a consideración diferentes factores que pueden influir negativamente en esta tarea.

1. *Cambios de iluminación*: Los cambios de iluminación de la persona pueden hacer esta tarea complicada ya que el color de este, va a variar con el tiempo.
2. *Forma*: La forma de la persona dificulta esta tarea. Como bien se ha mencionado anteriormente, la variación de la posición de la persona es un factor influyente en la tarea de re-identificación.
3. *Escala*: El tamaño de una persona en una imagen de una vista específica, puede ser diferente en alguna otra vista. La variación de proporciones representa un reto en la re-identificación

4. *Translación:* Una persona que aparece en una secuencia de video, se espera que este en constante movimiento. Que un sistema tenga la capacidad de poder detectarlo en cualquier instante de tiempo, es importante.
5. *Oclusión:* La ausencia de segmentos del cuerpo de la persona genera mayor complejidad en esta tarea. Es muy diferente tener a una persona de cuerpo completo, a tener sólo la mitad de su cuerpo.

4.1.4 Seguimiento de personas

Un sistema ideal de videovigilancia debería poder re-identificar a las personas bajo estudio en un 100%, desafortunadamente esto aún está fuera de la realidad. El principal problema que se presenta en el seguimiento de personas, es que el sistema no logre re-identificar a la persona de interés, o que este se confunda con otra persona. Esto genera confusión en el seguimiento, así como en la generación de su trayectoria.

4.1.5 Caracterización de la trayectoria

La caracterización de la trayectoria de una persona, es un tema poco estudiado en el estado del arte. La literatura reporta diferentes métodos para la detección de trayectorias normales y anormales. Para lograr esta tarea la principal problemática que se tiene es que el sistema pueda hacer un seguimiento correcto de las personas, caso contrario, esto puede generar confusión en caracterización.

4.2 Propuesta de solución

Con base en lo planteado en la sección 1.4, se cuenta con un sistema compuesto de 7 módulos, ver figura 4.1, que permiten el funcionamiento del sistema de videovigilancia inteligente. Estos tienen una relación secuencial y dependen del módulo anterior para funcionar. Es importante saber que este proceso es paralelo y por cámara, siempre y cuando existan imágenes en la secuencia.

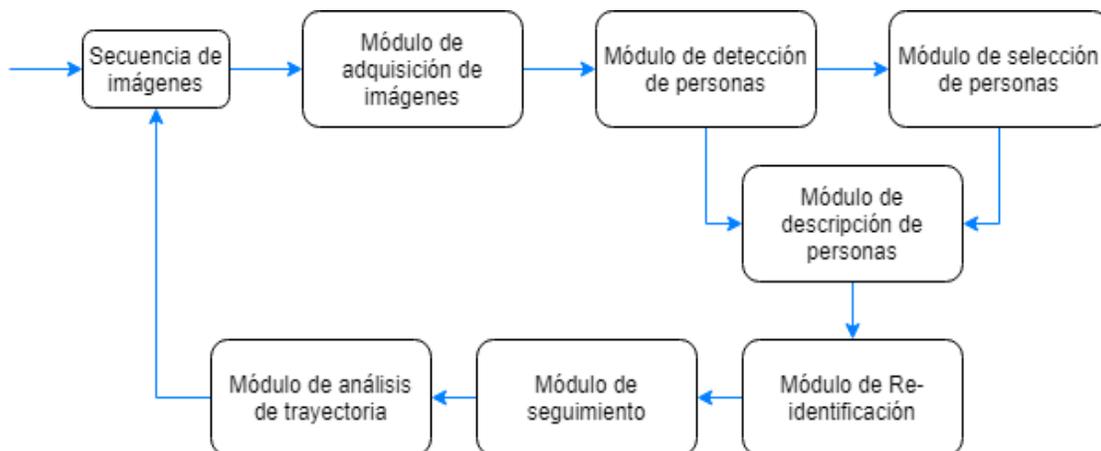


Figura 4.1. Arquitectura general del sistema propuesto.

1. *Módulo de adquisición de imágenes.*

Se recibe como entrada una secuencia de imágenes, este módulo se encarga de la lectura de cada una de las imágenes de las diferentes cámaras del sistema.

2. *Módulo de detección de personas.*

Dado la imagen de entrada, se genera un modelo de sustracción de fondo. Esto permite obtener los objetos en primer plano que se encuentran en movimiento en la secuencia de imágenes. Una vez que estos objetos son detectados, es necesario relacionarlos a una clase, precisamente para este caso, sólo se maneja la clase *persona* y la clase *no persona*.

3. *Módulo de selección de personas.*

Cuando los objetos en primer plano son clasificados en la clase *persona*, el sistema permite al usuario seleccionar a una persona denominada *sospechoso*, este será re-identificado en las diferentes cámaras, así como se realizará el análisis de trayectorias. Es importante saber que este módulo sólo va a funcionar si el usuario no ha seleccionado a una persona de interés, en caso contrario, este módulo es omitido en la arquitectura del sistema.

4. *Módulo descripción.*

Este módulo recibe como entrada todos los objetos en movimiento pertenecientes a la clase *persona*, cada uno de ellos es descrito en términos de características softbiométricas.

5. *Módulo de re-identificación.*

Cuando el usuario ha seleccionado a un *sospechoso* y el módulo de descripción ha creado los descriptores de estos, este módulo los recibe como entrada. Cada descriptor de las personas que aparecen en las secuencias de imágenes es comparado con el descriptor del *sospechoso*. El sistema evalúa si existe coincidencia entre estos y si la respuesta es *verdadera* se dice que son la misma persona, caso contrario son personas diferentes.

6. *Módulo de seguimiento.*

Para el seguimiento de una persona, es necesario recibir solamente como entrada una respuesta *verdadera* por parte del módulo de re-identificación, si la respuesta es *falsa*, el seguimiento no se realiza. El seguimiento de la persona obtiene el centroide de la sub-región correspondiente a la persona y esta es almacenada para generar su trayectoria.

7. *Módulo de análisis de trayectoria.*

Este es el último módulo del sistema, para que este funcione se tiene como entrada dos segmentos de trayectorias, las cuales fueron generadas por el módulo de seguimiento. Mediante el cálculo de orientaciones y pendientes de la recta de las trayectorias, este módulo logra clasificar cuatro acciones durante el trayecto de aparición de la persona: quedarse parado, caminar derecho, girar a la derecha, girar a la izquierda.

Análisis, diseño e implementación del sistema

En este capítulo se describe el análisis, diseño e implementación del sistema. El apartado de análisis y diseño corresponde al lenguaje de programación utilizado, entorno de desarrollo e interfaz gráfica del sistema. El apartado de implementación está implícito en los 7 módulos del sistema, los cuales son descritos en el transcurso del capítulo.

5.1 Análisis y diseño del sistema

La finalidad del sistema es lograr la re-identificación de una persona en un ambiente de multi-cámaras. Para lograr este cometido se parte de los módulos descritos en la propuesta de solución, sección 4.2. Es importante tener en cuenta que los módulos propuestos son secuenciales y no pueden trabajar de forma independiente.

La arquitectura del sistema, ver figura 5.1, trabaja con modelo de mezclas gaussianas, operaciones morfológicas, histograma de gradientes orientados, maquinas vector soporte, características softbiométricas y cálculos de rectas. Cada una de estas etapas es descrita en las secciones posteriores. De igual forma es importante saber que la arquitectura expuesta es por cada canal de cámara de video del sistema, ya que estos trabajan de forma independiente y paralela.

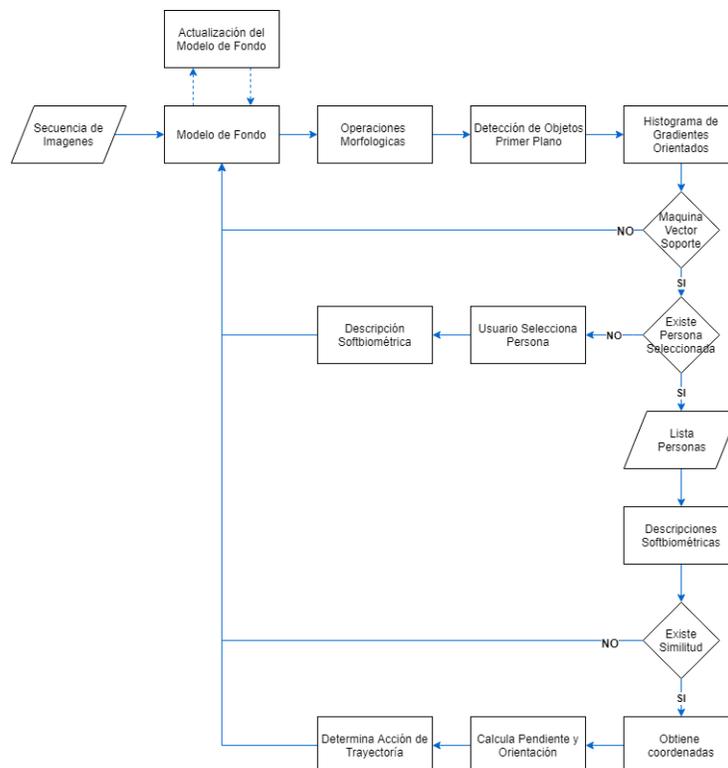


Figura 5.1. Arquitectura del sistema.

El sistema se desarrolló utilizando el lenguaje de programación JavaFX en su versión 10.0, (disponible en [1]). Este lenguaje es orientado a objetos y brinda portabilidad entre sistemas operativos. El entorno de desarrollo usado fue Eclipse en su versión Oxygen, (disponible en [2]). Para la creación de la interfaz del sistema se utilizó Scene Builder en su versión 10.0, (disponible en [3]).

La interfaz del sistema se estableció de forma que sea fácil e intuitiva para el usuario. Esta se compone de 3 secciones, ver figura 5.2. La primera sección se conforma de un menú de operaciones que permite al usuario cargar, reproducir y detener las secuencias de imágenes, así como la opción de habilitar y deshabilitar la selección de la persona de interés. La segunda sección se compone de los canales de reproducción y la información que brinda el sistema al usuario. La última sección consiste en un modelo de fondo que detecta objetos en movimiento. La estructura de la interfaz brinda un entorno amigable al usuario.



Figura 5.2. Interfaz del sistema.

5.2 Base de datos

Este apartado se compone de dos bases de datos. La primera es un conjunto de imágenes utilizadas para la re-identificación. La segunda es un conjunto de segmentos de rectas utilizados para el análisis de trayectorias.

5.2.1 Base de datos para re-identificación

Para generar la base de datos para la tarea de re-identificación se utilizaron 3 cámaras del modelo E7TURBO de la marca EPCOM. Estas cuentan con un sensor CMOS 1/3, un ángulo de visión de 2.9 mm. y con una resolución de 1280x720px. Las características físicas de las cámaras son: 89x59.1 mm. de tamaño, 400 gr. de peso. y una alimentación de 12 V. La cámara se muestra en la figura 5.3.

[1] <https://www.oracle.com/technetwork/es/java/javafx/overview/index.html>

[2] <https://www.eclipse.org/oxygen/>

[3] <https://gluonhq.com/products/scene-builder/>



Figura 5.3. Cámara de videovigilancia.

Para la adquisición de los videos se utilizaron 3 cámaras, cada una de ellas cuenta con una vista diferente en un ambiente no controlado, con variaciones de intensidad de iluminación, cambios de translación, perspectiva y forma. De estas grabaciones se obtienen secuencias cortas de personas de interés. Las imágenes se encuentran en un formato de color *RGB* y tienen una dimensión de 854x480px. La figura 5.4 muestra las vistas de la base de datos.



Figura 5.4. Vista de cámaras. Lado izquierdo vista 1, medio vista 2, derecho vista 3.

La base de datos se compone de 9 secuencias de video, en donde existen un total de 14 personas. Estas fueron creadas en fechas y horarios diferentes. El primer conjunto de datos se creó el día 10 de febrero de 2018, en un horario de 17:00 a 18:00 horas. De esta se obtuvieron un total de 4 personas diferentes. La segunda base de datos se generó el día 4 de abril de 2018 en un horario de 19:00 a 20:00 horas, de esta se obtuvieron las imágenes restantes.

5.2.2 Base de datos para trayectorias

La generación de la base de datos de trayectorias se obtuvo mediante las secuencias de imágenes de los peatones, ver figura 5.5. Así como la generación de trayectorias aleatorias. De la secuencia de imágenes se obtuvieron un total de 1708 segmentos de trayectorias, un número pequeño para la evaluación. Por tanto, se generaron segmentos de rectas aleatorias pertenecientes a la clase: *parado*, *camina derecho*, *gira izquierda* y *gira derecha*, en donde por cada acción se tienen 9673 trayectorias.



Figura 5.5. Trayectoria de una persona, en diferentes vistas de cámara.

Cada trayectoria tanto como la de la secuencia de imágenes y los generados aleatoriamente se almacenan en un archivo de texto, ver figura 5.6. La estructura del archivo se compone del total de la trayectoria y cada línea es un segmento en donde se describen los puntos espaciales, la cámara y el peatón.

```

trayectoria: Bloc de notas
Archivo Edición Formato Ver Ayuda
1078/
-116.0/-126.0/-71.0/-80.0/2/0/
-71.0/-80.0/-31.0/-45.0/2/2/
-31.0/-45.0/-2.0/-16.0/2/2/
-2.0/-16.0/25.0/31.0/2/2/
25.0/31.0/53.0/33.0/2/2/
53.0/33.0/74.0/49.0/2/2/
74.0/49.0/102.0/60.0/2/2/
102.0/60.0/112.0/67.0/2/2/
112.0/67.0/116.0/68.0/2/1/
116.0/68.0/118.0/66.0/2/1/
118.0/66.0/114.0/67.0/2/1/
0.0/0.0/-221.0/43.0/0/0/
-221.0/43.0/-212.0/16.0/0/2/
-212.0/16.0/-210.0/16.0/0/1/
-210.0/16.0/-210.0/16.0/0/1/
-210.0/16.0/-209.0/21.0/0/1/

```

Figura 5.6. Estructura del almacenamiento de trayectorias.

5.3 Adquisición de imágenes

Esta etapa consiste en la lectura de las secuencias de imágenes obtenidas en los videos. Para esta tarea se establecen los canales de reproducción mediante componentes visuales del lenguaje de programación JavaFX.

5.3.1 Interfaz gráfica

Se diseña una interfaz gráfica compuesta de 6 canales completamente independientes, ver figura 5.7. Cada canal se compone de un componen visual de la clase *ImageView*, el cual es el encargado las imágenes. Los primeros 3 canales serán utilizados para mostrar las secuencias de imágenes en el espacio de color *RGB*, los 3 canales restantes, son auxiliares y muestran una imagen binaria, correspondiente al modelo de fondo que detecta los objetos en movimiento.



Figura 5.7. rojo) Canales de reproducción RGB azul) Canales de reproducción de modelo de fondo.

5.3.2 Topología de cámaras

La topología existente entre las cámaras es completamente independiente. Es decir, los objetos que se encuentran en la vista de la cámara uno, pueden aparecer en la vista de la cámara 2 o la cámara 3. La figura 5.8 muestra la independencia entre las vistas.

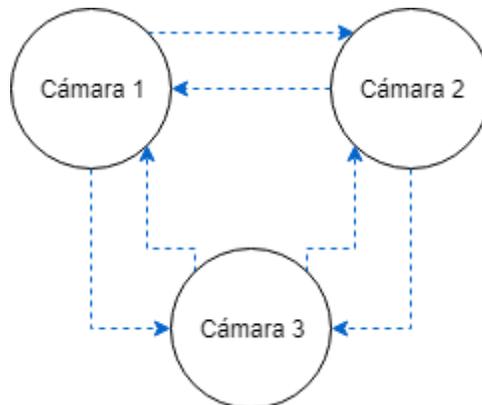


Figura 5.8. Topología de cámaras.

5.3.3 Lectura

La lectura de la secuencia de imágenes utiliza la clase *BufferedImage*. Esta clase permite obtener las imágenes para que puedan ser procesadas, así como para mostrarlo en las pantallas de la interfaz del sistema. La lectura y reproducción de la secuencia de imágenes es paralela e independiente en cada canal.

5.4 Selección de la persona

El objetivo de esta etapa es dotar al sistema con la habilidad de sustraer regiones, de la imagen completa, que el usuario seleccione con el ratón. Para esto, el sistema utiliza un componente visual de la clase *ImageView* y su evento *OnMousePressed* el cual devuelve las coordenadas de la posición del ratón con respecto al *ImageView*. En general la forma de funcionamiento de este modulo es, dada las diferentes vistas de las cámaras, se obtiene una lista de personas posibles a seleccionar, el usuario selecciona a una persona de la lista y el sistema la almacena para usarla posteriormente, ver figura 5.9.

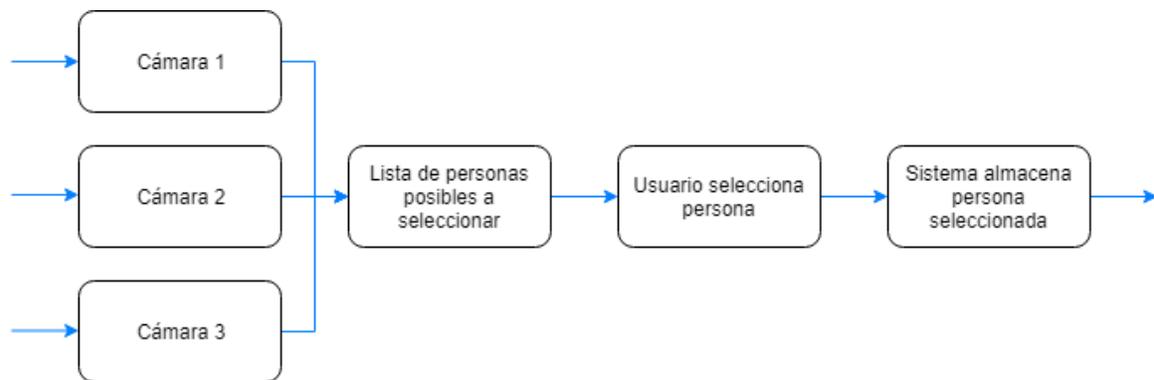


Figura 5.9. Flujo de trabajo del módulo de selección de persona.

De manera automática el sistema realiza una detección de movimiento, en las tres vistas del sistema multi-cámaras, posteriormente se determina si estas regiones en movimiento corresponden a la clase persona o no persona. En caso de que la región en movimiento sea etiquetada en la clase persona, la región es pintada con un cuadro amarillo y es posible seleccionarla como un peatón, el cual será re-identificado posteriormente. El módulo de selección cuenta con un botón que permite habilitar o deshabilitar la selección. La figura 5.10 muestra la detección de la región correspondiente a un peatón posible a seleccionar.



Figura 5.10. Selección de peatones.

5.5 Descripción de la persona

Esta etapa consiste, en la descripción de la persona por medio de características softbiométricas de la persona seleccionada, ver figura 5.11, y que va a ser re-identificada en el sistema.

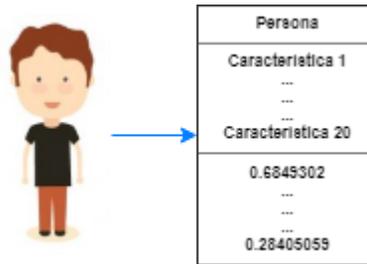


Figura 5.11. Representación de la persona mediante características softbiométricas.

Para este modulo se implementa un descriptor formado por un vector de características de 4 medidas estadísticas (*media*, *desviación estandar*, *energia* y *entropía*) en los espacios de color *HSV*, *LAB*, *YCBR*, *HLS* y *RGB*. Ver tabla 5.1. La descripción completa de este descriptor, se menciona mas adelante en el método 2, en el apartado de re-identificación.

Tabla 5.1. Vector de características de la persona seleccionada.

Estadística	Canal	Espacio de Color	Valor
Media	HSV	HSV	0.528450
Desviación Estándar	S	HSV	0.040349
Media	A	LAB	0.230494
Media	B	LAB	0.495069
Desviación Estándar	A	LAB	0.039840
Desviación Estándar	B	LAB	0.098272
Desviación Estándar	LAB	LAB	0.004856
Entropía	A	LAB	0.950912
Desviación Estándar	CR	YCRCB	0.029012

Estadística	Canal	Espacio de Color	Valor
Media	H	HLS	0.704930
Media	H	HSV	0.193049
Entropía	LAB	LAB	0.089493
Media	S	HLS	0.940201
Energía	A	LAB	0.049850
Media	CB	YCRCB	0.794930
Desviación Estándar	CB	YCRCB	0.091981
Desviación Estándar	YCRCB	YCRCB	0.012345
Media	HLS	HLS	0.863004
Desviación Estándar	S	HLS	0.019293
Media	B	RGB	0.193950

5.6 Detección de personas

Este módulo tiene como objetivo detectar regiones en movimiento e identificar si la región pertenece a la clase persona o no persona. Para esta etapa las técnicas seleccionadas e implementadas, por su buen resultado reportado de acuerdo a la revisión del estado del arte es el modelo de mezclas gaussianas (*MOG*) para la segmentación de regiones en movimiento y el histograma de gradientes orientados (*HOG*) con máquinas de vector soporte (*SVM*) para la detección de personas. La figura 5.12 muestra el orden para la detección de personas.

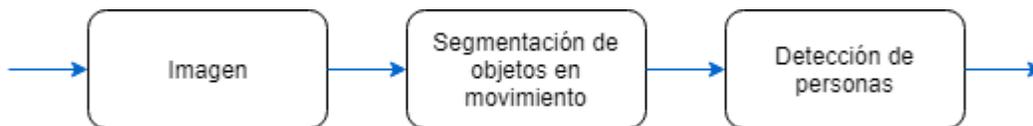


Figura 5.12. Esquema general del módulo de detección de personas.

La literatura dice que para la detección de regiones en movimiento (Afsar, Cortez, & Santos, 2015) el *MOG* es una de las técnicas estándar en la sustracción de fondo de regiones en movimiento. (Bouwman, El Baf, & Vachon, 2010) afirman que el *MOG* ha sido utilizado en múltiples tareas de videovigilancia para la detección del movimiento de objetos en escenas de videovigilancia. (Denman, Chandran, & Sridharan, 2007) recopilan diferentes investigaciones propuestas para el mejoramiento del *MOG*. La tabla comparativa que realiza (Tiwari & Singhai, 2017) muestra que el *MOG* tiene buena precisión en la detección de movimiento y un costo computacional moderado en comparación con otras técnicas. En (Benezeth, Jodoin, Emile, & Rosenberger, 2008) la comparativa muestra que el *MOG* es lo suficientemente confiable. (Bouwman, 2011) afirma en su comparativa que esta técnica, es utilizada principalmente en aplicaciones de videovigilancia y es robusto en escenarios exteriores.

Por tanto para la detección de objetos en movimiento se utiliza *MOG*, ver sección 3.1. El modelo de fondo generado necesita ser preprocesado y posprocesado para mejorar los resultados. Esto se debe a que existen regiones pequeñas que corresponden a cambios de iluminación que generan ruido al modelo de fondo, ver figura 5.13.



Figura 5.13. Modelo de fondo sin procesamiento.

Para la generación del modelo de fondo, ver figura 5.14, el preprocesamiento consiste en aplicar un filtro de suavizado, ver sección 3.2, a la imagen entrante, el posprocesamiento consiste en aplicar operadores morfológicos, ver sección 3.3, a la imagen resultante del modelo de fondo.

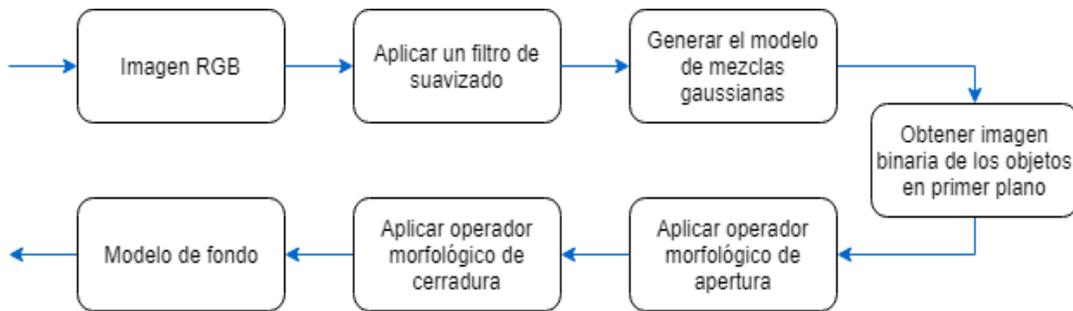


Figura 5.14. Procesamiento del modelo de fondo.

El resultado del esquema anterior mencionado es eliminar las pequeñas perturbaciones que corresponden a ruido existente en el modelo de fondo, así como mejorar la forma de los objetos en movimiento, ver figura 5.15.



Figura 5.15. Modelo de fondo procesado.

Dado que *MOG* detecta todas aquellas regiones en movimiento, existen imágenes en la secuencia de video que corresponden a ruido por algún cambio abrupto en la iluminación, presencia de objetos que pertenecen a la clase no persona (perros, árboles, vehículos y otros) y regiones que pertenecen a la persona, pero no se encuentra completas debido a la oclusión, ver figura 5.16.



Figura 5.16. Imágenes que pertenecen a la clase no persona.

Este tipo de regiones mencionadas deben ser omitidas. Estas se categorizan como la clase *no persona*. Por tanto, se implementa un sub-módulo de detector de objetos que consiste en identificar las regiones que corresponden a la clase persona y las que no.

De acuerdo a la literatura existen diferentes tipos de descriptores para representar un objeto. Una de las técnicas más comunes son bordes a nivel local, los cuales detectan los bordes del objeto, entre ellos se encuentra sobel, canny y otros; Bordes a nivel región, en donde el más utilizado es el histograma de gradientes orientados; Finalmente se encuentran los descriptores de forma donde se pueden encontrar técnicas como local binary pattern. Estos atributos pueden ser combinados para generar un descriptor más robusto. (Zhang, Matsumoto, Yoshida, & Kita, 2016) describe que en la actualidad una de las técnicas que tiene mayor preferencia y con mejores resultados para describir a una persona es el *HOG*, debido a su robustez con respecto a cambios de iluminación, rotación y translación. En diferentes artículos relacionados en videovigilancia como lo son (Moctezuma, Conde, De Diego, & Cabello, 2015), (Chen, An, & Bhanu, 2015) y (Li, Liu, & Yan, 2014) se puede apreciar el uso de *HOG* y *SVM* para la detección de personas.

Con lo anterior mencionado el sub-módulo se compone de *HOG*, ver sección 3.4, como descriptor y *SVM*, ver sección 3.5, como clasificador, ver figura 5.17. El proceso que se sigue es dada una imagen de entrada correspondiente a una región en movimiento, este se describe mediante *HOG*. La *SVM* recibe el descriptor de la región de la imagen y determina si esta pertenece a la clase *persona* o *no persona*.

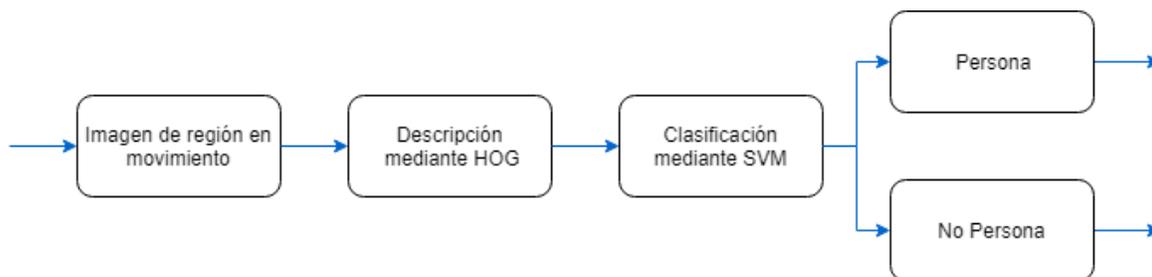


Figura 5.17. Esquema de funcionamiento del detector de objetos.

El sub-módulo detector de objetos, dota al sistema con la habilidad de descartar, de manera automática, regiones en movimiento que simplemente no son de interés y permite quedarse sólo con regiones que pertenecen a la clase persona. Esto ayuda en el tiempo de ejecución porque se describen mediante características softbiométricas a las regiones de interés, mientras las otras son omitidas.

Una vez que las perturbaciones del modelo de fondo son eliminadas, se localizan las regiones que corresponden a objetos en movimiento en contraste con el fondo. Las regiones son evaluadas con el detector de objetos y si son etiquetados en la clase *persona*, la información obtenida es superpuesta en la imagen *RGB* actual en el tiempo t , ver figura 5.18.



Figura 5.18. Información superpuesta del modelo de fondo y la imagen *RGB* de un objeto en movimiento que pertenece a la clase *persona*..

5.7 Re-identificación

Para la tarea de re-identificación se proponen 3 métodos. Estos métodos trabajan con diferentes espacios de color: *Escala de grises*, *HSV*, *HSL*, *RGB*, *YCRCB*, *LAB*, ver sección 3.9, en combinación con método de segmentación del cuerpo.

El primer método consiste en una característica softbiométrica, el cual describe el color del torso. El segundo método utiliza un vector de características softbiométricas, el cual inicialmente tiene un tamaño de 84 y con algoritmos de selección de variables y métodos de búsqueda se reduce la dimensionalidad. Finalmente, el tercer método, es una extensión del segundo, utiliza un vector de características softbiométricas que describe el torso y las piernas de una persona.

5.7.1 Re-identificación mediante la descripción del torso (MÉTODO 1)

1. El módulo recibe como entrada una imagen, que contiene a una persona, ver figura 5.19, a). A la imagen se le aplica el método conocido como modelos de mezclas gaussianas para detectar a la persona, separándola del fondo. La imagen resultante es una imagen binaria, en donde la zona blanca es la región que pertenece a la categoría persona, ver figura 5.19, b).



Figura 5.19. a) Imagen de la persona. b) Imagen binaria.

2. Una vez que la imagen binaria es definida, se procede a segmentar el cuerpo humano mediante el método de promedio de gradiente, (Denman S. , Fookes, Bialkowski, & Sridharan, 2009), ver sección 3.6. Esto permite obtener las regiones de la cabeza, el torso y las piernas, ver figura 5.20.



Figura 5.20. Segmentación de partes de la persona.

3. La siguiente tarea es la extracción del torso, para ello:

- a. La imagen de entrada (imagen que contiene el cuerpo completo), ver figura 5.21, a), es transformada del espacio de color RGB a HLS.
- b. De esta nueva imagen, se calcula el histograma del torso, ver figura 5.21, b), en el canal H, tomando en cuenta sólo los píxeles que pertenecen a la región del objeto del primer plano de la imagen binaria. Los otros canales no se consideran porque sólo dan información relacionada con la iluminación y saturación.



Figura 5.21. a) Cuerpo completo. b) Extracción del torso.

4. El histograma proporciona información sobre la distribución del color del torso. En este caso, el nivel de color que tenga mayor frecuencia, es el que define el color de torso, ver figura 5.22. Por ejemplo, si el valor con mayor frecuencia es el nivel de color 350, indica que el torso es de color rojo. Para determinar el nivel con mayor frecuencia se hace un barrido simple del histograma.



Figura 5.22. Histograma del torso y generación del descriptor.

5. Finalmente, la decisión de re-identificación mediante el color se define en la ecuación 5.1 en donde teniendo en cuenta que el intervalo de valores del histograma en el canal H del espacio de color HLS es de $[0,359]$, se define como x_s el color del torso de la primera persona seleccionada y x_r el color de torso de alguna persona que aparece en la secuencia de video. El valor del umbral se explica en el anexo 4.

$$re - identificación = \begin{cases} true & \text{si } x_s - 5 \leq x_r \leq x_s + 5 \\ false & \text{caso contrario} \end{cases} \quad \text{Ec. (5.1)}$$

5.7.2 Re-identificación mediante descriptores de color (MÉTODO 2)

Este método, también utiliza la imagen de entrada y la binaria que contiene en blanco el objeto persona.

1. La imagen de entrada, se transforma en los espacios de color: *Escala de Grises, RGB, HSV, HLS, YCRCB, LAB*, ver sección 3.10.
2. Se crea un vector de características, integrado de 4 medidas estadísticas (*Media, Desviación Estándar, Energía, Entropía*), ver sección 3.8, calculadas en cada espacio de color, así como por canal individual. Es decir, cada espacio de color se compone de 3 canales y un canal adicional de combinación, salvo el espacio de escala de grises. El tamaño del vector se define: $4 \times 4 \times 5 + 4$, donde hay 4 estadísticas, 4 canales, 5 espacios de color y las estadísticas de un sólo canal (*niveles de grises*). Por tanto, se obtiene un vector de 84 características, ver tabla 5.2.

Tabla 5.2. Vector de características.

Atributo	Estadística	Canal	Espacio de Color
0	Media	Escala de Grises	Escala de Grises
1	Desviación Estándar	Escala de Grises	Escala de Grises
2	Energía	Escala de Grises	Escala de Grises
3	Entropía	Escala de Grises	Escala de Grises
4	Media	R	RGB
5	Media	G	RGB
6	Media	B	RGB
7	Media	RGB	RGB
8	Desviación Estándar	R	RGB
9	Desviación Estándar	G	RGB
10	Desviación Estándar	B	RGB
11	Desviación Estándar	RGB	RGB
12	Energía	R	RGB
13	Energía	G	RGB
14	Energía	B	RGB
15	Energía	RGB	RGB
16	Entropía	R	RGB
17	Entropía	G	RGB
18	Entropía	B	RGB
19	Entropía	RGB	RGB
20	Media	H	HLS
21	Media	L	HLS
22	Media	S	HLS

Atributo	Estadística	Canal	Espacio de Color
23	Media	HLS	HLS
24	Desviación Estándar	H	HLS
25	Desviación Estándar	L	HLS
26	Desviación Estándar	S	HLS
27	Desviación Estándar	HLS	HLS
28	Energía	H	HLS
29	Energía	L	HLS
30	Energía	S	HLS
31	Energía	HLS	HLS
32	Entropía	H	HLS
33	Entropía	L	HLS
34	Entropía	S	HLS
35	Entropía	HLS	HLS
36	Media	H	HSV
37	Media	S	HSV
38	Media	V	HSV
39	Media	HSV	HSV
40	Desviación Estándar	H	HSV
41	Desviación Estándar	S	HSV
42	Desviación Estándar	V	HSV
43	Desviación Estándar	HSV	HSV
44	Energía	H	HSV
45	Energía	S	HSV
46	Energía	V	HSV
47	Energía	HSV	HSV
48	Entropía	H	HSV
49	Entropía	S	HSV
50	Entropía	V	HSV
51	Entropía	HSV	HSV
52	Media	L	LAB
53	Media	A	LAB
54	Media	B	LAB
55	Media	LAB	LAB
56	Desviación Estándar	L	LAB
57	Desviación Estándar	A	LAB
58	Desviación Estándar	B	LAB
59	Desviación Estándar	LAB	LAB
60	Energía	L	LAB
61	Energía	A	LAB
62	Energía	B	LAB
63	Energía	LAB	LAB
64	Entropía	L	LAB
65	Entropía	A	LAB
66	Entropía	B	LAB
67	Entropía	LAB	LAB
68	Media	Y	YCRCB
69	Media	CR	YCRCB
70	Media	CB	YCRCB
71	Media	YCRCB	YCRCB
72	Desviación Estándar	Y	YCRCB
73	Desviación Estándar	CR	YCRCB
74	Desviación Estándar	CB	YCRCB
75	Desviación Estándar	YCRCB	YCRCB

Atributo	Estadística	Canal	Espacio de Color
76	Energía	Y	YCRCB
77	Energía	CR	YCRCB
78	Energía	CB	YCRCB
79	Energía	YCRCB	YCRCB
80	Entropía	Y	YCRCB
81	Entropía	CR	YCRCB
82	Entropía	CB	YCRCB
83	Entropía	YCRCB	YCRCB

La dimensionalidad de este vector genera complejidad de cálculo computacional. Se utilizaron métodos de selección de variables para reducir la dimensionalidad del vector y obtener solamente características lo suficiente discriminantes.

Para esta tarea se utilizó el 20% del total de la base de datos y el otro 80% se quedó para evaluación. Se escogieron diferentes evaluadores de atributos, así como métodos de búsqueda, ver tabla 5.3. Esto con la finalidad de tener varios criterios de selección de variables.

Tabla 5.3. Evaluadores y métodos de búsqueda utilizados para la selección de variables.

Método	Evaluador de Atributos	Método de Búsqueda
A	CfsSubsetEval	BestFirst
B	CfsSubsetEval	GreedyStepWise
C	CorrelationAttributeEval	Ranker
D	GainRatioAttribute	Ranker
E	InfoGainAttribute	Ranker
F	OneAttribute	Ranker
G	Relief	Ranker
H	Symmetrical	Ranker

Cada evaluador con su método de búsqueda asigna diferente peso a cada característica del vector definido inicialmente, ver tabla 5.4. Por tanto, cada uno devuelve un vector reducido diferente y es necesario definir las variables más discriminantes.

Tabla 5.4. Variables con mayor discriminación de los métodos.

A	B	C	D	E	F	G	H
5	5	59	70	57	53	58	57
6	6	73	57	54	57	57	70
10	10	58	54	53	54	26	54
20	20	54	73	70	70	41	53
27	27	70	75	73	65	73	73
36	36	41	74	65	73	54	74
39	39	57	53	74	20	74	75
40	40	75	65	75	36	59	65
41	41	74	58	58	74	20	58
42	42	26	36	59	75	36	41
43	43	65	20	41	39	65	36
53	53	67	41	67	58	70	20
54	54	42	22	20	59	75	59

A	B	C	D	E	F	G	H
55	55	39	39	36	41	53	67
57	57	66	67	39	22	69	39
58	58	43	61	61	61	23	61
59	59	53	59	22	23	39	22
61	61	34	26	37	67	22	26
65	65	23	66	26	43	6	66
67	67	22	23	63	69	55	23
73	73	6	63	66	37	37	63

3. Reducir la dimensionalidad del vector. Con lo anterior mencionado, se definió un vector de tamaño 20, ver tabla 5.5. Esto se definió como las características que tuvieran mayor frecuencia de aparición en los diferentes métodos. Esto permite tener un vector robusto, ya que no se considera sólo un criterio de selección de variables, sino varios criterios que definen un conjunto de las variables, las cuales tienen mayor capacidad de discriminación.

Tabla 5.5 Vector de características.

Característica	Frecuencia	Estadística	Canal	Espacio de Color
39	8	Media	HSV	HSV
41	8	Desviación Estándar	S	HSV
53	8	Media	A	LAB
54	8	Media	B	LAB
57	8	Desviación Estándar	A	LAB
58	8	Desviación Estándar	B	LAB
59	8	Desviación Estándar	LAB	LAB
65	8	Entropía	A	LAB
73	8	Desviación Estándar	CR	YCRCB
20	7	Media	H	HLS
36	7	Media	H	HSV
67	7	Entropía	LAB	LAB
22	6	Media	S	HLS
61	6	Energía	A	LAB
70	6	Media	CB	YCRCB
74	6	Desviación Estándar	CB	YCRCB
75	6	Desviación Estándar	YCRCB	YCRCB
23	5	Media	HLS	HLS
26	5	Desviación Estándar	S	HLS
6	4	Media	B	RGB

4. Finalmente, dado dos vectores la distancia Manhattan α es calculada. Si esta distancia es menor o igual a un umbral establecido de 1.6, ver anexo 1, el sistema determina que es la misma persona, caso contrario son personas diferentes, ver ecuación 5.2.

$$re - identificación = \begin{cases} true & \alpha \leq 1.6 \\ false & \text{caso contrario} \end{cases} \quad Ec. (5.2)$$

5.7.3 Re-identificación mediante descriptores de color de segmentos del cuerpo humano (MÉTODO 3)

Este método es la combinación del método 1 y 2.

1. A partir de las dos imágenes iniciales (la imagen recibida como entrada y la binaria) se aplica la segmentación utilizando el método propuesto en (Denman S. , Fookes, Bialkowski, & Sridharan, 2009) y se obtiene las partes del cuerpo (*cabeza, torso y piernas*) . Cada parte de cuerpo es definido con un vector de características iniciales definidos anteriormente en la tabla 5.2.
2. Se sigue la metodología de la selección de variables, llevada a cabo en el método 2. Se utiliza el 20% de la base de datos para la selección de variables y el 80% para la evaluación. Los métodos utilizados para la selección de variables son los que listan en la tabla 5.3.

Como se mencionó, cada parte de la imagen segmentada en cabeza, torso y piernas tienen su propio vector de características. Esto da independencia a cada parte del cuerpo y se puede determinar qué tan discriminante es cada uno de estos. Los resultados de los métodos de selección de variables se definen en las tablas 5.6 para la cabeza, 5.7 para el torso, 5.8 para las piernas; y el vector de características resultante se define en las tablas 5.9 para la cabeza, 5.10 para el torso, 5.11 para las piernas.

Tabla 5.6. Selección de variables de la cabeza.

A	B	C	D	E	F	G	H
6	6	73	54	54	54	58	54
26	26	57	59	73	70	74	70
39	39	74	70	70	57	73	57
53	53	58	58	57	73	57	75
54	54	54	74	75	39	26	73
57	57	70	57	53	75	39	74
58	58	75	65	74	74	41	58
59	59	65	26	58	53	23	59
65	65	22	67	67	22	40	67
66	66	39	73	39	67	24	65
69	69	59	53	66	58	65	66
70	70	53	61	59	65	22	26
75	75	41	62	65	41	54	39
		67	41	22	66	20	41
		26	23	23	23	36	61
		37	39	41	26	66	23
		23	22	63	37	70	22
		6	63	26	59	75	63
		66	37	61	61	53	62
		38	6	62	36	6	37
		34	34	37	20	59	6

Tabla 5.7. Selección de variables del torso.

A	B	C	D	E	F	G	H
5	5	73	57	53	53	74	57
6	6	75	70	54	57	54	70
13	13	54	54	70	20	57	54
20	20	70	74	57	36	53	74
21	24	57	75	73	54	73	53
24	37	74	53	74	73	70	75
36	43	58	73	36	70	22	73
37	47	65	65	20	37	75	65
43	52	59	20	65	74	20	20
47	53	37	36	75	5	36	36
52	54	22	39	37	69	37	39
53	55	53	37	39	65	58	37

A	B	C	D	E	F	G	H
54	57	41	5	55	39	6	5
55	58	55	69	22	22	39	69
57	65	4	58	5	55	65	58
58	69	39	52	69	58	23	55
59	70	6	61	58	75	55	22
61	75	67	55	61	6	4	52
65		26	22	52	52	5	61
69		66	23	0	61	69	0
70		61	68	68	0	26	68

Tabla 5.8. Selección de variables de las piernas.

A	B	C	D	E	F	G	H
6	6	56	54	54	70	20	54
14	14	72	70	70	54	36	70
20	20	42	23	69	69	23	69
23	23	43	69	36	36	69	36
26	26	36	36	20	20	56	20
27	27	20	20	23	75	72	23
36	36	59	39	39	59	10	39
39	39	27	59	59	39	39	59
42	42	10	56	75	23	42	75
43	43	11	8	56	43	41	56
53	53	8	72	27	27	9	72
54	54	25	75	53	26	8	53
55	55	23	53	42	42	11	42
56	56	1	42	43	56	1	8
58	58	69	43	10	72	25	43
59	59	39	10	27	10	59	10
60	60	9	11	8	6	40	11
62	62	54	1	11	8	24	1
69	69	70	25	9	11	43	27
70	70	75	27	1	53	26	25
75	75	41	9	25	25	27	9

Tabla 5.9. Vector de características de la cabeza.

Característica	Frecuencia	Estadística	Canal	Espacio de Color
26	8	Desviación Estándar	S	HLS
39	8	Media	HSV	HSV
54	8	Media	B	LAB
57	8	Desviación Estándar	A	LAB
58	8	Desviación Estándar	B	LAB
59	8	Desviación Estándar	LAB	LAB
65	8	Entropía	A	LAB
70	8	Media	CB	YCRCB
53	7	Media	A	LAB
66	7	Entropía	B	LAB
75	7	Desviación Estándar	YCRCB	YCRCB
6	6	Media	B	RGB
22	6	Media	S	HLS
23	6	Media	HLS	HLS
41	6	Desviación Estándar	S	HSV
73	6	Desviación Estándar	CR	YCRCB
74	6	Desviación Estándar	CB	YCRCB
37	5	Media	S	HSV
67	5	Entropía	LAB	LAB
61	4	Energía	A	LAB

Tabla 5.10. Vector de características del torso.

Característica	Frecuencia	Estadística	Canal	Espacio de Color
20	8	Media	H	HLS
37	8	Media	S	HSV
53	8	Media	A	LAB
54	8	Media	B	LAB
55	8	Media	LAB	LAB
57	8	Desviación Estándar	A	LAB
58	8	Desviación Estándar	B	LAB
65	8	Entropía	A	LAB
70	8	Media	CB	YCRCB
75	8	Desviación Estándar	YCRCB	YCRCB
5	7	Media	G	RGB
69	7	Media	CR	YCRCB
74	7	Desviación Estándar	CB	YCRCB
6	6	Media	B	RGB
22	6	Media	S	HLS

Característica	Frecuencia	Estadística	Canal	Espacio de Color
36	6	Media	H	HSV
39	6	Media	HSV	HSV
52	6	Media	L	LAB
61	6	Energía	A	LAB
73	6	Desviación Estándar	CR	YCRCB

Tabla 5.11. Vector de características de las piernas.

Característica	Frecuencia	Estadística	Canal	Espacio de Color
20	8	Media	H	HLS
23	8	Media	HLS	HLS
27	8	Desviación Estándar	HLS	HLS
36	8	Media	H	HSV
39	8	Media	HSV	HSV
42	8	Desviación Estándar	V	HSV
43	8	Desviación Estándar	HSV	HSV
56	8	Desviación Estándar	L	LAB
59	8	Desviación Estándar	LAB	LAB
69	8	Media	CR	YCRCB
54	7	Media	B	LAB
70	7	Media	CB	YCRCB
75	7	Desviación Estándar	YCRCB	YCRCB
8	6	Desviación Estándar	R	RGB
10	6	Desviación Estándar	B	RGB
11	6	Desviación Estándar	RGB	RGB
25	6	Desviación Estándar	L	HLS
53	6	Media	A	LAB
72	6	Desviación Estándar	Y	YCRCB
1	5	Desviación Estándar	ESCALA GRIS	ESCALA GRIS

3. Del cuerpo se selecciona el torso y las piernas. Por tanto, se obtiene un total de dos vectores por la imagen del objeto persona. Dado dos personas se utiliza la distancia Manhattan α como medida de similitud, esto significa que se tienen 4 vectores, 2 correspondientes al torso de la persona x y la persona y , así como 2 correspondientes a las piernas de las personas. Los vectores del torso y las piernas son comparadores y se obtiene un valor numérico que indica la similitud de los segmentos del cuerpo. Posteriormente se realiza una suma ponderada de estas distancias, ver ecuación 5.3. Estos devuelven un valor de similitud.

$$\delta = 0.80 \times \alpha_{torso} + 0.20 \times \alpha_{piernas} \quad \text{Ec. (5.3)}$$

4. El valor de similitud tiene que ser menor o igual a un umbral establecido de 1.6, ver ecuación 5.4. El sistema determina que es la misma persona en caso de que se cumpla la condición, caso contrario determina que son personas diferentes.

$$re - identificación = \begin{cases} true & \delta \leq 1.6 \\ false & \text{caso contrario} \end{cases} \quad \text{Ec. (5.4)}$$

5.8 Seguimiento

Este módulo se encarga de obtener la posición de la persona re-identificada. Dado una región de imagen que pertenece a la clase persona se ubica la posición central de la región definido en un punto (x, y) , ver ecuación 5.5, en dónde $R_{(x,y)}$ son los puntos donde inicia la región de la imagen segmentada e I es la imagen de la secuencia de video.

$$\begin{aligned}
 x &= R_x + \frac{I_{ancho}}{2} \\
 y &= R_y + \frac{I_{alto}}{2}
 \end{aligned}
 \tag{Ec. (5.5)}$$

Una vez que el punto es definido, se almacena en una lista, ver tabla 5.12. En esta se guarda el punto, la cámara y la persona mediante un id. Este seguimiento, representa la entrada para el módulo de análisis de trayectorias.

Tabla 5.12. Ejemplo de la lista de almacenamiento.

Punto	Cámara	Sospechoso
(200,200),	1	1
(205,205),	1	1
(210,215),	1	1
(210,220),	1	1
(230,220),	1	1
(235,235),	1	1
(235,235),	1	1
(400,25),	2	1
(400,321),	3	1

5.9 Análisis de trayectorias

Para esta tarea se propone un método denominado *determinación de acciones mediante orientaciones y pendientes*, cuyo principio está basado en el cálculo de la orientaciones y pendientes de segmentos de rectas. Esto permite determina la acción de la persona cuando aparece en la secuencia de video.

Dado una imagen, cuyas propiedades tienen un alto y ancho definido. Es necesario convertir el espacio de la imagen a un espacio de sistema de coordenadas cartesianas mediante la ecuación 5.6, donde dado un punto en el espacio de la imagen $P(x, y)$ este será definido en un plano cartesiano como $C(f(x), (y))$. Esto facilita los cálculos matemáticos respecto a las ecuaciones de la recta.

$$\begin{aligned}
 f(x) &= \frac{x + ancho}{2} - ancho \\
 f(y) &= alto - \frac{y + alto}{2}
 \end{aligned}
 \tag{Ec. (5.6)}$$

Una vez definido el plano cartesiano, el segundo paso es dado dos puntos cualesquiera, es necesario definir la orientación (δ). Esto se realiza obteniendo las diferenciales en el eje X (ΔX) y en el eje Y (ΔY). Las orientaciones definidas toman base de la rosa de los vientos y son las siguientes: *Norte (N)*, *Noroeste (NO)*, *Oeste (O)*, *Suroeste (SO)*, *Sur (S)*, *Sureste (SE)*, *Este (E)*, *Noreste (NE)*. Dado ΔX y ΔY se establecen una serie de condiciones que permiten saber la orientación, ver ecuación 5.7.

$$\delta = \begin{cases} N & \Delta X = 0, \Delta Y > 0 \\ NO & \Delta X < 0, \Delta Y > 0 \\ O & \Delta X < 0, \Delta Y = 0 \\ SO & \Delta X < 0, \Delta Y < 0 \\ S & \Delta X = 0, \Delta Y < 0 \\ SE & \Delta X > 0, \Delta Y < 0 \\ E & \Delta X > 0, \Delta Y = 0 \\ NE & \Delta X > 0, \Delta Y > 0 \end{cases} \quad Ec. (5.7)$$

Cuando la orientación es definida el siguiente paso es el cálculo de la pendiente (m), ver ecuación 5.8, la constante de multiplicación se utiliza para convertir de radianes a grados. Es necesario recordar que la pendiente de una recta de 90 grados es igual a infinito, así como el intervalo de valores que toma $[0,179]$ grados.

$$m = 57.2958 \arctg \frac{\Delta Y}{\Delta X} \quad Ec. (5.8)$$

La importancia de la orientación es que permite manejar un intervalo de valores de $[0,360]$ grados. Una vez con la pendiente definida es necesario determinar el valor real del ángulo de la pendiente de la recta. Para esto es necesario definir un conjunto de reglas descritas en la ecuación 5.9. Esto permite que el sistema pueda manejar direcciones en todos los sentidos.

$$m = \begin{cases} 90 & \delta = N \\ m + 180 & \delta = NO \\ 180 & \delta = O \\ m + 180 & \delta = SO \\ 270 & \delta = S \\ m + 360 & \delta = SE \\ 0 & \delta = E \\ m & \delta = NE \end{cases} \quad Ec. (5.9)$$

Para definir el estado de la trayectoria de la persona es necesario obtener dos segmentos de rectas. Estos son obtenidos conforme la secuencia de video. Es importante notar que, dado que la información es obtenida al momento, existe un pequeño retraso del estado de la trayectoria de la persona, es decir da la acción inmediata anterior a la actual.

Dado los dos segmentos de recta definidas como R_1 y R_2 con sus respectivas orientaciones δ_1 y δ_2 y pendientes m_1 y m_2 . Se establece un conjunto de reglas que permiten devolver la acción de la trayectoria de la persona, ver ecuación 5.10, en donde *dis* es la distancia entre dos puntos de alguna de las rectas. La idea general es comparar los dos segmentos obtenidos con base en el conjunto de reglas, esto dota al sistema con la capacidad de determinar si la persona se queda parada (*p*), camina derecho (*cd*), gira izquierda (*gi*) o gira derecha (*gd*). Adicionalmente se agrega un estado de indefinido (*i*) ya que pueden presentarse situaciones, por ejemplo donde retroceda.

$$\text{acción} = \begin{cases} p & \text{dis} \leq 5 \\ cd & \delta_1 = \delta_2 \quad || \quad m_2 - 22.5 \leq m_1 \leq m_2 + 22.5 \\ gi & \delta_2 > \delta_1 \\ gd & \delta_2 < \delta_1 \\ i & \text{otro caso} \end{cases} \quad \text{Ec. (5.10)}$$

Experimentación y resultados

En este capítulo se describe la experimentación, así como los resultados obtenidos en la re-identificación de las personas y el análisis de trayectorias. Para el apartado de re-identificación se proponen 3 métodos, estos son evaluados y comparados. Para el apartado de análisis de trayectoria se propone un método. Las métricas de evaluación utilizadas son estadísticas como Tasa de Verdaderos Positivos (TPR), Tasa de Falsos Positivos (FPR) que tienen base en la matriz de confusión: Verdaderos Positivos (VP), Verdaderos Negativos (VN), Falsos Positivos (FP) y Falsos Negativos (FN), ver sección 3.12.

Los experimentos se diseñaron de tal forma que, especificada la persona a seguir a partir de este momento denominada como “*sospechoso*”, este pueda ser re-identificado sin importar la vista de la cámara, el tiempo o el orden de aparición. Se obtiene una lista de personas de las secuencias de video las cuales son comparadas con la persona sospechosa. Así como dado el sospechoso identificado y re-identificado el sistema tenga la capacidad de definir la acción de la persona: *parado*, *caminar derecho*, *girar izquierda*, *girar derecha*.

6.1 Re-identificación

El banco de imágenes de re-identificación cuenta con tres diferentes vistas de cámaras en un ambiente no controlado, variaciones de intensidad de luz, de perspectiva y forma. Se compone de un total de 14 personas (P1, P2, P3, P4, P5, P6, P7, P8, P9, P10, P11, P12, P13, P14), ver figura 6.1, en donde en algunas imágenes existen similitud de vestimenta, dando mayor complejidad a la re-identificación. Para este apartado se tiene un total de 13305 imágenes.



Figura 6.1. Personas de la base de datos de re-identificación.

6.1.1 Re-identificación mediante la descripción del torso (MÉTODO 1)

El método uno, es evaluado con base en medidas estadísticas que se obtienen a partir de la matriz de confusión, ver tabla 6.1. Este método está basado en la teoría del color HLS y utiliza el color del torso de la persona como descriptor, realizando un cálculo de máxima frecuencia del histograma en el canal H.

Tabla 6.1. Resultados del método 1.

PEATON	TPR	FPR	VP	VN	FP	FN
1	40.4761905	20.104337	187	10261	2582	275
2	96.8328841	2.46172067	1437	11530	291	47
3	63.997061	1.21399866	871	11799	145	490
4	91.0845588	11.4295983	1982	9857	1272	194
5	95.3521127	11.8132099	1354	10481	1404	66
6	55.5693069	2.48096501	898	11399	290	718
7	52.1178637	20.3126279	566	9737	2482	520
8	62.1776504	1.98363692	217	12699	257	132
9	2.38095238	7.3316479	6	12096	957	246
10	95.1830443	6.10824339	494	12005	781	25
11	91.5492958	11.7170791	195	11558	1534	18
12	41.6666667	19.6571429	75	10545	2580	105
13	83.1215971	17.2808531	458	10550	2204	93
14	92.7261614	1.53397892	1517	11490	179	119
MEDIA	68.8739533	9.67350283				

La tabla muestra valores altos en *TPR* en las personas donde el torso se compone de un sólo color, como se ve en el peatón 2. Si el torso se contiene dos o más colores y presenta texturas, como es el caso del peatón 9 (compuesto de azul y gris) y el peatón 1 (blanco y azul), se obtienen pésimos resultados en *TPR*. Por esta razón la media general de *TPR* presenta un valor bajo del 67% para la re-identificación. Otro factor influyente es el canal H, por sí sólo, no logra interpretar de manera correcta el color gris, blanco y negro, es necesario extender la definición del color con los canales *L* y *S*, razón por la cual en el peatón 7 tiene un alto valor en los *FPR*, ya que estos canales si consideran la variabilidad en la iluminación.

El método en general, dado las propiedades del canal *H*, del espacio de color *HLS*. Es invariante a cambios en la intensidad de iluminación, escala y forma. El método siempre brindará buenos resultados en la re-identificación siempre y cuando las condiciones se ajusten a que el torso de la persona se componga de un sólo color y no exista la textura, esto dado que la principal característica del método propuesto se basa en el nivel de color con mayor frecuencia en el histograma.

6.1.2 Re-identificación mediante descriptores de color (MÉTODO 2)

El método dos, es evaluado con base en las medidas de la matriz de confusión, ver tabla 6.2. Este método está considera toda la información de la región de la imagen de la persona y lo representa mediante un vector de características de tamaño 20. Se utiliza la distancia Manhattan, por tanto, se establece un umbral de 1.6, ver anexo 1.

Tabla 6.2. Resultados del método 2.

PEATON	TPR	FPR	VP	VN	FP	FN
1	72.7272727	31.0052168	336	8861	3982	126
2	90.4986523	7.1059978	1343	10981	840	141
3	95.7384276	22.5720027	1303	9248	2696	58
4	88.0055147	6.54146824	1915	10401	728	261
5	96.7605634	19.4615061	1374	9572	2313	46
6	77.7846535	13.0464539	1257	10164	1525	359
7	45.4880295	11.4657501	494	10818	1401	592
8	77.6504298	13.4686632	271	11211	1745	78
9	62.3015873	25.6645982	157	9703	3350	95
10	65.5105973	27.2016268	340	9308	3478	179
11	96.2441315	21.8606783	205	10230	2862	8
12	96.1111111	8.34285714	173	12030	1095	7
13	95.0998185	17.719931	524	10494	2260	27
14	89.1809291	15.5968806	1459	9849	1820	177
MEDIA	82.0786942	17.2181165				

Los resultados demuestran que un vector de características softbiométricas, que considera todo el cuerpo de la persona, hace un mejor manejo en los múltiples colores y textura que componen a la persona, como se ve en el aumento de *TPR* del peatón 1 y 9. Así como brindar mayor capacidad de discriminación en la re-identificación. Se sigue presentando la problemática de baja detección en los colores negros, como se ve en el peatón 7. En este caso se considera que al tomar toda la región del cuerpo genera mayor grado de confusión, esto bajo la premisa de que algunas regiones del cuerpo como la cabeza, en la mayoría de los casos son similares. Por tanto, se tiene la hipótesis que descarta la región de la cabeza mejorará los valores de *TPR* y *FPR*.

Un vector de características compuesto por diferentes medidas estadísticas y espacios de color, que describen el color de la persona, son lo suficientemente discriminante para la tarea de re-identificación. Además, se conserva la propiedad de ser invariante a cambios de intensidad de iluminación, translación, escala y forma.

Adicionalmente este método fue extendido en dos variantes, con la hipótesis de mejorar los resultados con un vector mayor y asignando pesos a las características del vector final. Dado que los resultados fueron similares y no generan un aumento significativo en la tasa de *TPR* y un decremento en la tasa de *FPR* la experimentación sólo será mencionada.

1. Extensión 1: El vector inicial se compone de 6 medidas estadísticas, en los mismos espacios de color y tiene un tamaño de 128 características, ver anexo 5. El paso de selección de variables reduce la dimensionalidad a 20 características, ver anexo 6 y 7. Posteriormente se evalúa con el umbral 1.6, sin considerar los pesos, ver anexo 8.
2. Extensión 2: El vector inicial es igual a la extensión 1, se realiza la selección de variables y se reduce la dimensionalidad. Adicionalmente se considera un peso de discriminación de los atributos clasificados, ver anexo 7, en donde los pesos son

multiplicados por la diferencia de las características de dos vectores. Finalmente se evalúa con el mismo umbral, ver anexo 9.

6.1.3 Re-identificación mediante descriptores de color de segmentos del cuerpo humano (MÉTODO 3)

El método tres, es evaluado con base a las estadísticas de matriz de confusión, ver tabla 6.3. Este método segmenta las partes de la persona: *cabeza*, *torso* y *piernas*. Dado la hipótesis planteada en el método 2, se hizo una evaluación de la capacidad de discriminación de cada parte del cuerpo, ver anexo 2. Considerando solamente las regiones del torso y las piernas, cada región tiene un vector de características de tamaño 20 diferentes. Se utiliza la distancia Manhattan, por tanto, se establece un umbral de 1.6.

Tabla 6.3. Resultados del método 3.

PEATON	TPR	FPR	VP	VN	FP	FN
1	87.012987	15.6894807	402	10828	2015	60
2	92.2506739	6.90296929	1369	11005	816	115
3	92.3585599	10.4236437	1257	10699	1245	104
4	60.8915441	8.73393836	1325	10157	972	851
5	86.971831	16.4408919	1235	9931	1954	185
6	75.990099	6.32218325	1228	10950	739	388
7	79.46593	3.9855962	863	11732	487	223
8	81.6618911	15.0355048	285	11008	1948	64
9	72.2222222	13.2076917	182	11329	1724	70
10	82.4662813	12.0131394	428	11250	1536	91
11	94.8356808	4.59822793	202	12490	602	11
12	82.7777778	7.36	149	12159	966	31
13	99.4555354	7.11149443	548	11847	907	3
14	95.7212714	14.1828777	1566	10014	1655	70
MEDIA	84.5773061	10.1434028				

Como muestran los resultados la omisión de la cabeza mejora la precisión de la re-identificación, como se ve en el peatón 13. Además, la tasa TPR no es menor al 60%, el peatón 4 es un claro ejemplo, en comparación con los métodos anteriores y la tasa de FPR no es mayor al 16%, como se ve en el peatón 5. Esto indica que el método es preciso para la tarea de re-identificación.

La segmentación del cuerpo humano y la descripción de cada una de estas, mediante vectores de características softbiométricas de color demuestran que tienen la suficiente capacidad de discriminación para re-identificar personas con precisión. Al igual que los anteriores métodos, se conservan las propiedades de ser invariante a cambios de intensidad de iluminación, translación, escala y forma. El único detalle de este método es que al ser la combinación del método uno y dos, requiere más pasos para su ejecución, adicionalmente la descripción de la persona se hace mediante dos vectores de características, todo lo anterior se resume en que existirá mayor complejidad computacional.

Para concluir de los métodos 3 métodos propuestos, los que brindan mejores resultados son el método 2 y el método 3. En la figura 6.2, se puede ver como el método uno presenta caídas drásticas en la tarea de re-identificación, lo cual se traduce en baja detección de las personas, el método 2 minimiza estas caídas tan drásticas, pero siguen siendo inferiores al 50%. El método 3 presenta una gráfica regular en donde no hay caídas drásticas en la tarea de re-identificación, esto se debe a que utiliza descriptores por cada parte del cuerpo, logrando así mejor precisión de la persona.

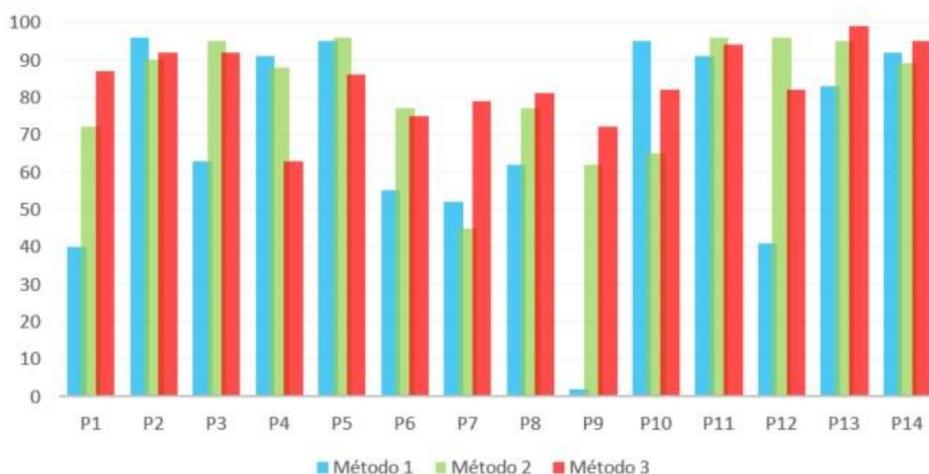


Figura 6.2. Gráfica de detección de personas.

6.1.4 Tiempos de ejecución

Este apartado tiene como objetivo mostrar el tiempo computacional que se requiere para la re-identificación de dos personas de los métodos anteriormente mencionados, ver tabla 6.4. Para esta medición se consideran los 14 peatones, en donde se toman dos muestras aleatorias de imágenes. Los tiempos consideran la generación de los descriptores de las dos imágenes, así como el tiempo que requiere el sistema para determinar si son la misma persona o son personas diferentes.

Tabla 6.4. Tiempos de ejecución de los métodos.

Peatón	Método 1		Método 2		Método 3	
	Milisegundos	Segundos	Milisegundos	Segundos	Milisegundos	Segundos
1	261	2.61	256	2.56	310	3.1
2	109	1.09	33	0.33	204	2.04
3	69	0.69	33	0.33	138	1.38
4	119	1.19	34	0.34	181	1.81
5	98	0.98	32	0.32	157	1.57
6	98	0.98	39	0.39	136	1.36
7	66	0.66	31	0.31	104	1.04
8	30	0.3	32	0.32	53	0.53
9	146	1.46	28	0.28	285	2.85
10	28	0.28	30	0.3	72	0.72

11	37	0.37	40	0.4	62	0.62
12	34	0.34	31	0.31	43	0.43
13	46	0.46	29	0.29	69	0.69
14	70	0.7	27	0.27	125	1.25
MEDIA	86.5	0.865	48.21428571	0.48214285	138.5	1.385

6.1.5 *RandomForest*

La experimentación anteriormente realizada fue evaluada utilizando medidas de similaridad, ver sección 3.7, particularmente con la distancia Manhattan. El detalle de esta evaluación es que, en la representación espacial, en ocasiones un elemento a cierta distancia no necesariamente indica que pertenecen a la misma clase. Además, dependiendo del valor de umbral pueden variar los resultados. Ante la carencia de una estructura de datos que logre determinar qué tan discriminante es cada característica del vector, se hizo la experimentación con el clasificador RandomForest, ver sección 3.11.

De los 3 métodos propuestos, se tomó el método dos, ya que este tiene buenos resultados en la re-identificación en comparación del método uno; y tiene menor complejidad computacional que el método 3. La generación del modelo del clasificador utiliza el 20% de la base de datos, mientras que, el otro 80% es usado para evaluación. Los resultados se pueden ver en la tabla 6.5 y la matriz de confusión en el anexo 3.

Tabla 6.5. Resultados del método 2 con RandomForest.

PEATÓN	TPR	FPR	VP	VN	FP	FN
1	82.8125	0.14619883	318	10245	15	66
2	99.494524	0.07401925	1181	9450	7	6
3	96.9359331	0.22995714	1044	9545	22	33
4	99.5970063	0.12349837	1730	8896	11	7
5	100	0	1122	8221	0	0
6	96.8085106	1.09348199	1274	9226	102	42
7	98.2638889	0.26584867	849	9754	26	15
8	99.6441281	0.02894915	280	10360	3	1
9	98.5	0.08617388	197	10435	9	3
10	97.0873786	0.0488663	400	10227	5	12
11	95.8823529	0	163	10474	0	7
12	96.6666667	0.02175568	145	9191	2	5
13	99.7716895	4.17381361	437	10056	438	1
14	99.6171516	0	1301	9338	0	5
MEDIA	97.2201236	0.44946878				

Los resultados en general demuestran que utilizar un clasificador, el cual se caracteriza por tener una estructura de datos, en lugar de sólo una medida de similaridad, aumenta drásticamente la precisión de la re-identificación. El caso 5 por ejemplo logra ser detectado en todos los casos, mientras que la menor *TPR*, no es inferior al 80%, en el peatón 1. Adicionalmente el valor máximo de *FPR* no excede el 5%.

Por tanto, utilizar un clasificador (*RandomForest*) con un vector de características que representan el color de la persona (*método 2*) tiene gran capacidad de distinción entre peatones independientemente de la vista de la cámara, la iluminación, la translación, escala y forma del objeto, ya que presenta mayor área bajo la curva, ver figura 6.3, en comparación con los otros métodos evaluados con la *distancia Manhattan*. Razón por la cual fue implementado para el sistema final.

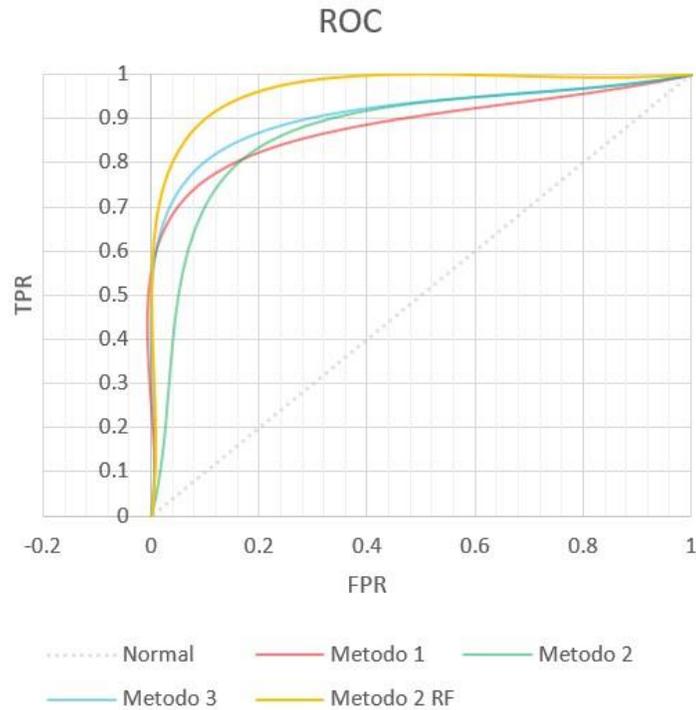


Figura 6.3. Curva ROC de los métodos.

6.2 Análisis de trayectorias

La experimentación consiste en la clasificación de cuatro acciones de la trayectoria de la persona: *camina derecho (CD, verde)*, *gira derecha (GD, magenta)*, *gira izquierda (GI, azul)* y *parado (P, parado)*, ver figura 6.4, adicionalmente se agregan dos estados cuando el análisis de trayectorias se inicia (*rojo*) y la trayectoria no puede ser definida (*negro*). Estos fueron diseñados de forma que dado dos segmentos de recta se pueda identificar la acción realizada por la persona en la secuencia de video.

ángulos de experimentación, de 22.5 grados, ver tabla 6.6 y un ángulo de 12 grados, ver tabla 6.7. Estos son evaluados con base en las estadísticas de matriz de confusión.

Tabla 6.6. Resultados con un ángulo de 22.5 grados.

ACCION	TPR	FPR	VP	VN	FP
CD	98.01	1.06333333	9801	29681	319
GD	96.53	0.22333333	9653	29933	67
GI	97.43	0.31666667	9743	29905	95
P	100	0.49333333	10000	29852	148
MEDIA	97.9925	0.52416667			

Tabla 6.7. Resultados con un ángulo de 12 grados.

ACCION	TPR	FPR	VP	VN	FP
CD	97.93	0.34758197	9793	29817	104
GD	97.69	0.2439758	9769	29848	73
GI	99.0625945	0.32333333	9828	29903	97
P	100	0.49463587	10000	29773	148
MEDIA	98.6706486	0.35238174			

Los resultados obtenidos, demuestran que el método logra clasificar con gran precisión la acción de la persona. Las acciones de camina derecho y gira a la derecha presentan valores bajos en comparación a las acciones restantes, esto se debe a la definición del umbral. Si la pendiente de la recta se encuentra en cierto grado de inclinación cercano al umbral, existe un cierto grado de confusión al interpretar la acción, pero no representa un error significativo, ya que si logra precisar la acción en la mayor parte de los casos.

El umbral en este caso, se interpreta como un rango de ángulo en grados que permite clasificar la acción. Un ángulo en grados muy alto dará mayor probabilidad de que la acción se detecte como *camina derecho*. Un ángulo con valor bajo dará mayor probabilidad a las acciones: *giro de izquierda o derecha*. Para el caso de la acción *quedarse parado* sólo se define una medida de distancia, si esta es menor o igual a 5 significa que el sospechoso está estático, caso contrario, se encuentra en movimiento. La definición de esta distancia dependerá de la velocidad de translación del sospechoso en la vista de la cámara, así como la obtención de los puntos cada determinado número de imágenes.

Es importante saber que cada acción realizada es por cámara y estas son completamente independientes. En caso de la translación de una vista a otra vista de las cámaras, después de 15 cuadros de imagen, al no existir información, se reinicia la información almacenada de la trayectoria de la cámara donde dejó de aparecer el sospechoso y se vuelve a obtener el estado de iniciando análisis de la trayectoria. En caso de existir trayectorias, en donde el sospechoso retroceda o se realice otra acción, este entra en un estado de acción denominado *no definido*, el cual es establecido en el conjunto de reglas.

Finalmente, un ángulo de 12 grados da buenos resultados y logra discriminar bien la acción del sospechoso con el método propuesto, así como una distancia 5 unidades logra definir

correctamente el estado *quedarse parado*. Por tanto, para el sistema de videovigilancia se establecen estos valores de umbral.

6.3 Conclusión

Para el apartado de re-identificación, los resultados de la experimentación demuestran que el método dos propuesto en combinación con el clasificador *RandomForest* es lo suficientemente discriminante para re-identificar a una persona, pese a las variaciones existentes de cambios de iluminación, translación, oclusión y forma. Para el apartado de análisis de trayectorias, el método tiene buena tasa de reconocimiento de las acciones de la persona. Los métodos propuestos en combinación brindan la terminación del sistema de videovigilancia inteligente, el cual logra realizar con precisión las tareas anteriormente mencionadas en un sistema multicámaras y en tiempo real.

Conclusiones y trabajos futuros

En este capítulo se presentan las conclusiones obtenidas en el desarrollo de este trabajo de investigación, así como las aportaciones que este brinda y los trabajos futuros que pueden mejorar el sistema de videovigilancia inteligente. (Agregar cumplimiento de objetivos)

7.1 Objetivos alcanzados

El objetivo general de este trabajo consistió en el desarrollo de un sistema de videovigilancia inteligente que pueda realizar la re-identificación de una persona en un entorno multicámaras, así como la clasificación de cuatro comportamientos en su trayectoria (caminar derecho, quedarse parado, girar a la izquierda, girar a la derecha). El desarrollo de este trabajo de investigación se ha desarrollado de manera satisfactoria y cada uno de los objetivos específicos se cumplieron de acuerdo a los establecido, ver tabla 7.1.

Tabla 7.1. Comentarios de los objetivos específicos.

Objetivo específico	Comentario
Implementar un entorno de usuario que permita seleccionar a una persona de interés en las vistas de las cámaras	El entorno de desarrollo se creó mediante el lenguaje JavaFX y utilizando el API de SceneBuilder.
Implementar un módulo de descripción que permita obtener las características de la persona seleccionada.	El modulo genera un vector de características utilizando diferentes medidas estadísticas, así como en distintos espacios de color.
Implementar un módulo de detección de personas en movimiento que permita segmentar y obtener la región de interés.	El modulo utiliza el <i>MOG</i> para detectar objetos en movimiento, posteriormente la imagen se representa con <i>HOG</i> y el clasificador <i>SVM</i> determina si la región es de interés, es decir es una persona.
Implementar un módulo de identificación y re-identificación de la persona que determine si la región de interés corresponde a la persona seleccionada.	El módulo utiliza los vectores de características y el clasificador RandomForest determina si existe coincidencia entre la lista de personas de las secuencias de videos y la persona seleccionada.
Implementar un módulo de seguimiento de la persona que obtenga la trayectoria de la persona seleccionada	Este módulo depende de que exista la re-identificación. En caso de existir se obtiene el centro de la región de la imagen que corresponde a la persona re-identificada. Posteriormente es almacenado en coordenadas (x,y) y utilizado para el análisis de trayectorias,.
Implementar un módulo de análisis de trayectorias que permita clasificar la trayectoria de la persona seleccionada.	El modulo utiliza dos segmentos de rectas, se hace el cálculo de la orientación y la pendiente de estas. Mediante un sistema de reglas se determina la acción a clasificar (caminar derecho, quedarse parado, girar izquierda y girar derecha).

7.2 Conclusiones

En este trabajo se presentó la metodología usada para las tareas de re-identificación y análisis de trayectorias, así como la experimentación y los resultados. Estas dos tareas brindan la esencia del sistema de videovigilancia inteligente.

Se presenta tres métodos para la re-identificación, evaluados con clasificadores y distancias. Cada uno presenta sus propias bondades en esta tarea. De igual forma, se propone un método para la tarea de análisis de trayectorias, el cual es suficientemente discriminante para clasificar acciones.

La experimentación demuestra que utilizar la distancia Manhattan como criterio de evaluación para la re-identificación, brinda resultados aceptables. Esto se debe a que cada característica del descriptor es considerada con la misma capacidad de discriminación, lo cual ocasiona mayor grado de confusión. Utilizar un clasificador, en este caso RandomForest, el cual se compone de una estructura de datos. Permite que el vector por cada característica tenga mayor grado de discriminación y por tanto el descriptor sea robusto en la tarea de re-identificación, en comparación de aplicar medidas de similaridad.

El método uno presenta debilidades ante el reconocimiento de texturas y definición del color de la persona, cuando este se compone de dos o más. El método tres tiene mayor grado de discriminación, pero tienen un mayor costo computacional debido a que es necesario generar un vector de características por cada parte del cuerpo. Finalmente, el método dos solventa las debilidades del método uno y tres, ya que tiene menor costo computacional al ser sólo un vector y considera todo el color que compone el cuerpo de la persona.

La evaluación de los resultados demuestra que el método dos propuesto, con un clasificador, brinda buenos resultados en la tarea de re-identificación. Este método discrimina mejor a las personas en comparación del método uno y tres, razón por la cual se implementó para esta tarea.

Utilizar un vector de características con diferentes espacios de color y estadísticas, logra que la tarea de re-identificación sea lo suficientemente precisa cuando existan casos de cambios de iluminación, forma, escala, rotación y traslación.

De igual forma se presentó un método que permite describir la acción de la trayectoria de una persona en tiempo real. Utilizar un sistema de rectas y ángulos brinda buenos resultados definiendo el estado de la persona en: *parado*, *camina derecho*, *gira izquierda*, *gira derecha*. Esto brinda al usuario la facilidad de conocer la acción que realiza la persona en un entorno de videovigilancia.

Finalmente, se elimina ciertas limitaciones del sistema que fueron propuestas inicialmente. Estas son listadas a continuación:

- Se logra la re-identificación cuando existe oclusión en algunos casos.
- El sistema logra trabajar en modo online.

- La re-identificación se logra con más de una persona por secuencia de video.
- La localización inicial de las personas se hace de manera automática.

7.3 Aportaciones

Las aportaciones que se obtuvieron en la realización de este proyecto de investigación relacionado en la videovigilancia inteligente son:

1. *Algoritmo de re-identificación de personas.*

Se propone un nuevo algoritmo de re-identificación que genera un descriptor de una persona, basándose en su color. El descriptor se compone de un conjunto de características que permite re-identificar a una persona, independientemente de las variaciones de iluminación, translación, escala, forma y oclusión. En el estado del arte se reportan varios trabajos relacionados a la re-identificación, muchos de ellos presentan la problemática de no poder ser implementados en tiempo real, cuestión que el algoritmo propuesto logra manejar sin problemas.

2. *Algoritmo de análisis de trayectorias*

Se propone novedoso algoritmo de análisis de trayectorias. Este permite saber cuatro acciones de la persona: *caminar derecho*, *girar izquierda*, *girar derecha* y *quedarse parado*. La revisión del estado del arte reporta que la mayoría de los artículos reportan trabajos sobre clasificar la trayectoria como típica o atípica, pero caracterizar la acción de la trayectoria han sido poco tratados.

3. *Respuesta en tiempo real*

Los algoritmos mencionados fueron diseñados de tal forma que puedan ser implementados en tiempo real, esto ofrece una gran ventaja al sistema, ya que no tiene mayor grado de complejidad computacional.

4. *Sistema de Videovigilancia Inteligente*

Se propone un novedoso sistema en donde con una interfaz amigable hacia el usuario. El sistema tiene varias opciones, en donde el usuario puede interactuar directamente, por ejemplo, en la selección de un sospechoso, así como la opción de seleccionar nuevamente a otra persona en la secuencia de imágenes. De igual forma el sistema cuenta con la capacidad de detectar al sospechoso en las diferentes vistas de la cámara y determinar la acción que está realizando.

7.4 Trabajo futuro

Los trabajos futuros que se plantean realizar sobre este trabajo son:

1. Mejorar el modelo de fondo, de tal forma que en las imágenes se logre obtener mejor las regiones del cuerpo, ya que en algunas ocasiones la imagen binaria presenta huecos en el cuerpo o regiones cortadas. Otro aspecto a considerar es que logre diferenciar entre dos regiones de personas diferentes, ya que si existe solapamiento estas se montan.
2. Lograr la detección de personas completamente estáticas, que no pertenecen realmente al fondo del entorno. Al no existir variación entre los píxeles de la persona quieta, estos tienen baja probabilidad de aparición, lo cual representa un problema para la re-identificación de la persona.
3. Realizar un estudio profundo del método de segmentación utilizado en el método 1 de re-identificación, esto con la finalidad de lograr segmentar el cuerpo humano cuando existan variaciones en la forma, ya que el método original está condicionado a que sólo se encuentre la persona en vertical.
4. Extender el algoritmo del método 2, en donde no sólo se consideren espacios de color y medidas estadísticas, sino también de otras posibles características que puedan mejorar la precisión de los resultados y reducir la dimensionalidad del vector.
5. Extender el algoritmo de análisis de trayectorias de tal forma que pueda clasificar otro tipo de trayectorias, por ejemplo: cuando retrocede.
6. Generar un vector de características con base a los segmentos de trayectorias y medir si este, pudiera ser útil para la tarea de re-identificación.
7. Extender la interfaz del sistema, de tal modo que no esté limitado a la selección de personas, sino que dada la secuencia de imágenes, este pueda seleccionar N personas de manera automática.
8. Combinar los métodos propuestos de este trabajo con información espacio-temporal de las personas que aparecen en la secuencia de video para aumentar la precisión en la tarea de re-identificación.

7.5 Productos adicionales y participaciones

En esta sección se presentan las diferentes participaciones realizadas durante el periodo de estudios de la maestría, así como los productos adicionales creados en el desarrollo del sistema de videovigilancia inteligente.

- Productos adicionales

En este trabajo se produjo como producto adicional, ver figura 6.6, una extensión del sistema original que no sólo permita la re-identificación de una persona seleccionada. En él se carga una lista de personas, las cuales son buscadas en las diferentes vistas de las cámaras y definen las trayectorias.

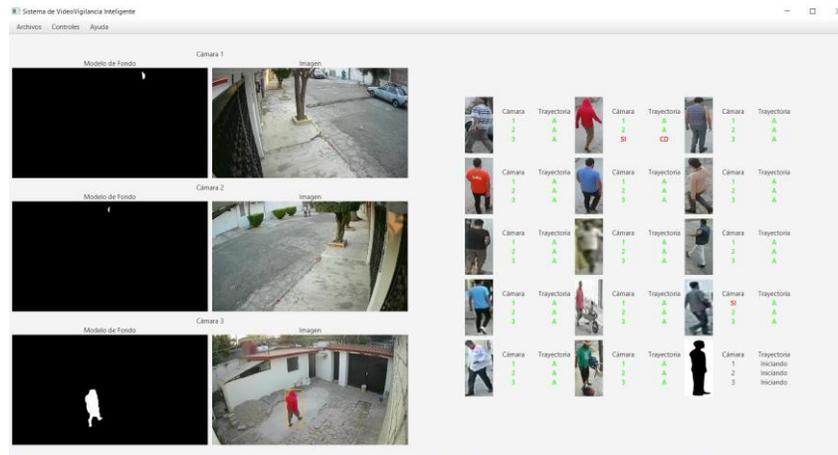


Figura 6.6. Sistema adicional creado.

- Participaciones

Las participaciones realizadas durante el transcurso de la maestría son listadas a continuación:

- ✓ Participación como asistente en las actividades de la escuela de inteligencia artificial y robótica de la Universidad Tecnológica Emiliano Zapata llevado a cabo el 19 y 20 de octubre de 2017, anexo 10.
- ✓ Participación como asistente en las actividades del Congreso Internacional de Ingeniería Mecatrónica, Electrónica y Automotriz, organizado por la Universidad Autónoma del Estado de Morelos y llevado a cabo del 21 al 24 de Noviembre de 2017, anexo 11.
- ✓ Participación como conferencista en el Instituto Tecnológico de Campeche con el tema “Introducción al procesamiento digital de imágenes”, llevado a cabo el día 22 de marzo de 2018, anexo 12.
- ✓ Participación como conferencista en el simposio de la Universidad Autónoma de Campeche con el tema de “Mi tema de tesis de maestría en CENIDET”, llevado a cabo del 20 al 22 de Marzo de 2018, anexo 13.

- ✓ Participación en el concurso 3 minutos sobre mi tesis, llevado a cabo el 31 de Mayo de 2018 en el Centro Nacional de Investigación y Desarrollo Tecnológico. Se obtuvo el tercer lugar, anexo 14.
- ✓ Participación en el 3er congreso nacional de informática, llevado a cabo el 25 de Octubre de 2018 en el Instituto Tecnológico del Valle del Guadiana, Durango con el tema “*Técnicas para la videovigilancia inteligente*”, anexo 15.

Referencias

Bibliográficas

- Afsar, P., Cortez, P., & Santos, H. (2015).** Automatic visual detection of human behavior: A review from 2000 to 2014. *Expert Systems with Applications*, 6935-6956.
- Aldhaheri, A.R. & Edirisinghe, E.A., 2014.** Detection and Classification of a Moving Object in a Video Stream, in: *Proc. of the Intl. Conf. on Advances in Computing and Information Technology*. Institute of Research Engineers and Doctors, Saudi Arabia, pp. 105–111.
- Alegre, E., Pajares, G., & Escalera, A. d. (2016).** *Conceptos y Métodos en Visión por Computador*. España: Grupo de Visión del Comité Español de Automática, [s.l].
- Amandeep, & Goyal, M. (2015).** Review: Moving Object Detection Techniques. *International Journal of Computer Science and Mobile Computing*, 4(9), 345 – 349.
- An, L., Kafai, M., Yang, S., & Bhanu, B. (2016).** Person reidentification with reference descriptor. *IEEE Transactions on Circuits and Systems for Video Technology*, 26(4), 776-787.
- Benezeth, Y., Jodoin, P. M., Emile, B., Laurent, H., & Rosenberger, C. (2008, December).** Review and evaluation of commonly-implemented background subtraction algorithms. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on* (pp. 1-4). IEEE.
- Bialkowski, A., Denman, S., Sridharan, S., Fookes, C., & Lucey, P. (2012, December).** A database for person re-identification in multi-camera surveillance networks. In *Digital Image Computing Techniques and Applications (DICTA), 2012 International Conference on* (pp. 1-8). IEEE.
- Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992).** A training algorithm for optima margin classifiers. *Proceedings of the Fifth Annual Workshop on Computational Learning Theory, COLT*, 144-152.
- Bouwman, Thierry & El Baf, Fida & Vachon, Bertrand. (2010).** Statistical Background Modeling for Foreground Detection: A Survey. *Handbook of Pattern Recognition and Computer Vision*, (4), 181-199.
- Bouwman, T. (2011).** Recent advanced statistical background modeling for foreground detection-a systematic survey. *Recent Patents on Computer Science*, 4(3), 147-176.
- Bradski, G. (2000).** The opencv library (2000). *Dr. Dobb's Journal of Software Tools*.
- Breiman, L., J.H., F., Olshen, R., & Stone, C. (1984).** *Classification and Regression Trees*. New York: Chapman and Hall/CRC.

- Cancela, B., Ortega, M., Penedo, M. G., Novo, J., & Barreira, N. (2013).** On the use of a minimal path approach for target trajectory analysis. *Pattern Recognition*, 46(7), 2015-2027.
- Cánovas-García, F., Alonso-Sarría, F., & Gomariz-Castillo, F. (2016).** Modificación del algoritmo RandomForest para su empleo en clasificación de imágenes de teledetección. *Aplicaciones de las Tecnologías de la Información Geográfica (TIG) para el desarrollo económico sostenible*, XVII Congreso Nacional de Tecnologías de Información Geográfica, 359-368.
- Chate, M., Amudha, S., & Gohokar, V., 2012.** Object Detection and tracking in Video Sequences. *ACEEE Int. J. Signal Image Process.* 3.
- Chen, X., An, L., & Bhanu, B. (2015).** Multitarget Tracking in Nonoverlapping Cameras Using a Reference Set. *IEEE Sensors Journal*, 2692-2704.
- Chen, X., & Bhanu, B. (2014, August).** Soft biometrics integrated multi-target tracking. In *Pattern Recognition (ICPR), 2014 22nd International Conference on* (pp. 4146-4151). IEEE.
- Cortes, C., & Vapnik, V. (1995).** Support-vector networks. *Machine Learning*, 20(3), 273-297.
- Dalal, N., & Triggs, B. (2005).** Histograms of oriented gradients for human detection. *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, 886-893.
- Denman, S., Chandran, V., & Sridharan, S. (2007).** Robust Real Time Multi-Layer Foreground Segmentation. *IAPR Conference on Machine Vision Applications*, 16-18.
- Denman, S., Fookes, C., Bialkowski, A., & Sridharan, S. (2009, December).** Soft-biometrics: unconstrained authentication in a surveillance environment. In *Digital Image Computing: Techniques and Applications, 2009. DICTA'09.* (pp. 196-203). IEEE.
- D'Haeyer, J. P. F. (1989).** Gaussian filtering of images: A regularization approach. *Signal Processing*, 18(2), 169-181.
- Eibe Frank, Mark A. Hall, and Ian H. Witten (2016).** The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016.
- Friedman, N., & Russell, S. (1997).** Image Segmentation in Video Sequences: A Probabilistic Approach. *UAI'97 Proceedings of the Thirteenth conference on Uncertainty in artificial intelligence*, 175-181.
- González, R. C., & Woods, R. E. (2002).** *Digital Image Processing*. Prentice Hall 2da Edición
- Hahnel, M., Klunder, D., & Kraiss, K. F. (2004, July).** Color and texture features for person recognition. In *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on* (Vol. 1, pp. 647-652). IEEE.

Hall, M. A. (1998). Correlation-based feature subset selection for machine learning. Thesis submitted in partial fulfillment of the requirements of the degree of Doctor of Philosophy at the University of Waikato.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA Data Mining Software: An Update. SIGKDD Explorations, Volume 11, Issue 1.

Haritaoglu, I., Harwood, D. & Davis, L.S., 2000. W/sup 4/: real-time surveillance of people and their activities. IEEE Trans. Pattern Anal. Mach. Intell. 22, 809–830.

Herrero, S., & Bescós, J. (2009, September). Background subtraction techniques: Systematic evaluation and comparative analysis. In International conference on Advanced Concepts for Intelligent Vision Systems (pp. 33-42). Springer, Berlin, Heidelberg.

Ho, T. K. (1995). Random Decision Forests. Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, 278–282.

Huang, A. (2008). Similarity measures for text document clustering. Proceedings of the sixth new zealand computer science research student conference, 49-56.

Jain, A. K., Dass, S. C., & Nandakumar, K. (2004, August). Can soft biometric traits assist user recognition?. In Biometric Technology for Human Identification (Vol. 5404, pp. 561-573). International Society for Optics and Photonics.

Jain, A. K., Duin, R. P., & Mao, J. (2000). Statistical pattern recognition a review. IEEE Transaction on pattern recognition and machine intelligence, 22(1):4–37.

Krishna, M.T.G., Ravishankar, M. & Babu, D.R.R., 2011. Automatic detection and tracking of moving objects in complex environments for video surveillance applications, in: 2011 3rd International Conference on Electronics Computer Technology. IEEE, pp. 234–238.

Li, A., Liu, L., & Yan, S. (2014). Clothes Attributes Assisted Person Re-identification. IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, 119-138.

Li, C., Han, Z., Ye, Q., & Jiao, J. (2013). Visual abnormal behavior detection based on trajectory sparse reconstruction analysis. Neurocomputing, 119, 94-100.

Li, X., Lu, L., Yu, J., & Wang, K. (2013). A method of abnormal pedestrian behavior detection based on the trajectory model. In ICTE 2013: Safety, Speediness, Intelligence, Low-Carbon, Innovation (pp. 2861-2867).

Liu, H., Chen, S., & Kubota, N. (2013). Intelligent Video Systems and Analytics: A Survey. IEEE Trans. Industrial Informatics, 9(3), 1222-1233.

Ma, B., Su, Y., & Jurie, F. (2014). Discriminative image descriptors for person re-identification. In Person Re-Identification (pp. 23-42). Springer, London.

- Moctezuma, D. (2016).** Re-identificación de personas a través de sus características soft-biométricas en un entorno multi-cámara de video-vigilancia. *Ingeniería, investigación y tecnología*, 17(2), 257-271.
- Moctezuma, D., Conde, C., De Diego, I. M., & Cabello, E. (2015).** Soft-biometrics evaluation for people re-identification in uncontrolled multi-camera environments. *EURASIP Journal on Image and Video Processing*, 2015(1), 28.
- Mohan, A.S. & Resmi, R., 2014.** Video image processing for moving object detection and segmentation using background subtraction, in: 2014 First International Conference on Computational Systems and Communications (ICCSC). IEEE, pp. 288–292.
- Nanni, L., Munaro, M., Ghidoni, S., Menegatti, E., & Brahnam, S. (2016).** Ensemble of different approaches for a reliable person re-identification system. *Applied Computing and Informatics*, 12(2), 142-153.
- Pongothai, E., & Suruliandi, A. (2016).** Survey on colour, texture and shape features for person re-identification. *Indian Journal of Science and Technology*, 9(29).
- Powers, D. (2011).** Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation. *Journal of Machine Learning Technologies*, 1-24.
- Stauffer, C., & Grimson, W. (1999).** Adaptive background mixture models for real-time tracking. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 246-252.
- Stehman, S. V. (1997).** Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment*, 62 (1): 77–89.
- Strehl, A. G. (2000).** Impact of Similarity Measures on Web-page Clustering. *Workshop on artificial intelligence for web search*, Vol. 58, p. 64.
- Tao Zhang, Zaiwen Liu, Xiaofeng Lian & Xiaoyi Wang, 2010.** Study on moving-objects detection technique in video surveillance system, in: 2010 Chinese Control and Decision Conference. IEEE, pp. 2375–2380.
- Tiwari, M., & Singhai, R. (2017).** A review of detection and tracking of object from image and video sequences. *Int. J. Comput. Intell. Res.*, 13(5), 745-765.
- Varior, R. R., Wang, G., Lu, J., & Liu, T. (2016).** Learning invariant color features for person reidentification. *IEEE Transactions on Image Processing*, 25(7), 3395-3410.
- Visa, S., Ramsay, B., R. A., & Van Der Knaap, E. (2011).** Confusion Matrix-based Feature Selection. *MAICS*, 120-127.
- Wang, X. (2013).** Intelligent multi-camera video surveillance: A review. *Pattern recognition letters*, 34(1), 3-19.

Wang, Y., Wang, D., & Chen, F. (2013). Abnormal behavior detection using trajectory analysis in camera sensor networks. *International Journal of Distributed Sensor Networks*, 10(1), 839045.

Wang, Y., Zhou, Z., Teoh, E. K., & Su, B. (2014, August). Human hair segmentation and length detection for human appearance model. In *Pattern Recognition (ICPR), 2014 22nd International Conference on* (pp. 450-454). IEEE.

Zhang, G., Matsumoto, K., Yoshida, M., & Kita, K. (2016). Pedestrian re-identification using color feature in multi surveillance video. *International Journal of Advances in Electronics and Computer Science*,(3), 11-14.

Zheng, Y. (2015). Trajectory data mining: an overview. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 6(3), 29.

Zhiqiang, W., Xiaopeng, J. & Peng, W., 2006. Real-time is moving object detection for video monitoring systems. *J. Syst. Eng. Electron.* 17: 731–736.

Electrónicas

Ernest Valveny. (20 de 09 de 2018). Coursera. Obtenido de <https://es.coursera.org/lecture/deteccion-objetos/l4-2-hog-calculo-del-gradiente-WSJ9t>

WorkWithColor.com. "HSL Color Picker." Accessed September 6, 2018. <http://www.workwithcolor.com/hsl-color-picker-01.htm>.

OpenCV: "Color conversions document", (15 de 09 de 2018). https://docs.opencv.org/3.4/de/d25/imgproc_color_conversions.html

Tesis de maestría

Alcantar Juárez, R. (2012). Caracterización Visual de Movimientos Sospechosos de Personas en Estacionamientos. Tesis de Maestría en Ciencias de la Computación, Centro Nacional de Investigación y Desarrollo Tecnológico.

Dávila García, M. L. (2009). Caracterización Automática de una Muestra de Acciones en Video. Tesis de Maestría en Ciencias de la Computación, Centro Nacional de Investigación y Desarrollo Tecnológico.

De León Hernández, G. (2012). Agrupamiento de Trayectorias para la Generación de un Modelo de Actividad Adaptativo. Tesis de Maestría en Ciencias de la Computación, Centro Nacional de Investigación y Desarrollo Tecnológico.

Saldaña Garcia, J. A. (2007). Caracterización de imágenes en movimiento: correr y caminar. Tesis de Maestría en Ciencias de la Computación, Centro Nacional de Investigación y Desarrollo Tecnológico.

Toscano Martínez, J. H. (2011). Seguimiento y Caracterización del Comportamiento de Varios Objetos en Movimiento. Tesis de Maestría en Ciencias de la Computación, Centro Nacional de Investigación y Desarrollo Tecnológico.

Anexos

1. Establecimiento de umbral 1.6
2. Comparativa de vector de características de partes del cuerpo humano
 3. Matriz de confusión del modelo RandomForest
 4. Establecimiento de umbral de 5, método 1
5. Vector de características extendido a 128 características
 6. Selección de variables y pesos de 128 características
 7. Vector reducido de 128 características
 8. Variante 1 del método 2.
 9. Variante 2 del método 2.
 10. Participación UTEZ 2017
 11. Participación ICMEAE 2017
 12. Participación ITC 2018
 13. Participación UAC 2018
 14. Participación 3MT 2018
 15. Participación ITVG 2018

Anexo 1

Este anexo comprende el estudio del establecimiento de un umbral para la distancia Manhattan, ver tabla A1.1. Dado dos vectores de características normalizados y compuestos de diferentes estadísticas y espacios de color, de tamaño 20. Se establece un umbral de valor 1.6 utilizando el 20% de la base de datos de re-identificación.

Tabla A1.1. Diferentes tasas de TPR y FPR de acuerdo al umbral.

Umbral	TPR	FPR
0.1	0.114523686	0
0.2	0.739198334	0
0.3	2.144716294	0.001735207
0.4	5.059864654	0.043380184
0.5	9.349297241	0.187402395
0.6	15.20041645	0.570883221
0.7	22.7589797	1.200763491
0.8	30.39042166	1.983342009
0.9	38.72982821	2.935970849
1	46.8401874	4.1037654
1.1	54.87766788	5.386083637
1.2	62.28006247	6.920006941
1.3	68.21447163	8.66736075
1.4	73.64914107	10.4719764
1.5	78.55283706	12.27485685
1.6	82.47787611	14.14367517
1.7	85.60124935	16.06454971
1.8	87.94377928	17.90733993
1.9	89.77615825	19.62693042
2	91.33784487	21.37428423
2.1	92.68089537	23.06090578
2.2	93.75325351	24.62085719
2.3	94.70067673	26.16172133
2.4	95.46069755	27.73381919
2.5	96.14783967	29.30765226
2.6	96.90786049	30.92833594
2.7	97.65747007	32.50910984
2.8	98.28214472	34.13846955
2.9	98.68818324	35.71056741
3	98.95887559	37.30175256

Un umbral mayor a 1.6 aumenta TPR, pero al mismo tiempo aumenta FPR. Es decir, el número de VP aumenta a costa de un decremento en los VN. Por tanto, la TPR será mayor al 80% pero alcanzando FPR superiores al 20%. Caso contrario es con un umbral inferior, el número de VP decremента, pero el número de VN incrementa. Por tanto, la TPR será inferior al 80% pero se tendrán FPR muy bajas. El intervalo de valores de un umbral ideal se define [1.5,1.9].

Anexo 2

Este anexo comprende el estudio de los vectores de características generados de los segmentos del cuerpo: *cabeza*, *torso* y *piernas*, ver tabla A2.1. Cada vector se compone de diferentes estadísticas y espacios de color y se compone de un tamaño 20. Se establece un umbral de valor 1.6 utilizando el 20% de la base de datos de re-identificación.

Tabla A2.1. Evaluación de segmentos del cuerpo.

PEATON	Cabeza		Torso		Piernas	
	TPR	FPR	TPR	FPR	TPR	FPR
1	35.0649351	14.6461107	81.81818	12.82411	99.35065	72.58429
2	5.25606469	0.40605702	79.85175	5.456391	98.98922	92.16648
3	24.761205	18.335566	84.20279	8.389149	100	79.85599
4	26.9301471	21.4574535	56.52574	7.125528	84.55882	64.85758
5	24.7887324	25.393353	73.59155	14.74127	99.78873	80.77408
6	20.7920792	15.8439559	71.84406	3.413466	98.88614	82.7188
7	38.213628	18.0784025	72.74401	2.651608	99.90792	79.66282
8	42.9799427	1.88329731	78.2235	12.31862	98.2808	64.23279
9	38.4920635	26.9823029	64.28571	10.25052	100	75.48456
10	14.8362235	3.55075864	77.64933	8.08697	100	90.73205
11	5.16431925	0.44301864	94.3662	2.268561	100	91.65139
12	24.4444444	4.07619048	72.77778	6.300952	100	74.41524
13	19.2377495	18.4569547	98.00363	4.375098	100	78.17155
14	36.7359413	20.7901277	94.8044	10.62645	97.98289	91.5674
MEDIA	25.5498197	13.5959678	78.62062	7.773477	98.41037	79.91964

Los resultados demuestran que las piernas son el segmento del cuerpo que tiene mayor TPR, pero su valor de FPR es demasiado grande. Esto se debe a que es muy probable la coincidencia de valores de color en las piernas de una persona. El torso es el segmento del cuerpo que tiene mayor discriminación, presenta valores altos en TPR, así como valores bajos en FPR. Como bien se había supuesto, la cabeza es el segmento de cuerpo menos útil para la re-identificación.

La generación de un descriptor que omita el área de la cabeza, mejorará los resultados en la re-identificación. Por tanto, el descriptor debe estar compuesto del torso y las piernas con sus respectivos vectores de características. Es importante mencionar que, dado las bondades de cada segmento, es necesario asignar un peso a cada parte del cuerpo. Esto considera que el peso de la cabeza debe ser de un valor 0, el del torso debe tener un peso [0.7,0.9] y finalmente las piernas deben tener un peso [0.1-0.3]. Esto para que el descriptor sea lo suficientemente discriminante, la combinación de las partes supondrá el aumento de TPR, así como un bajo FPR.

Anexo 3

Este anexo despliega la matriz de confusión del modelo de clasificación RandomForest generado, ver tabla A3.1. Se utiliza el 20% de la base de datos para la generación del modelo y el otro 80% para la evaluación.

Tabla A3.1. Matriz de confusión del modelo generado.

PEATON	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Evaluación	Entrenamiento	Total
1	318	0	4	0	0	58	3	0	1	0	0	0	0	0	384	78	462
2	0	1181	0	4	0	0	0	1	0	0	0	0	1	0	1187	297	1484
3	5	0	1044	0	0	28	0	0	0	0	0	0	0	0	1077	284	1361
4	0	7	0	1730	0	0	0	0	0	0	0	0	0	0	1737	439	2176
5	0	0	0	0	1122	0	0	0	0	0	0	0	0	0	1122	298	1420
6	6	0	18	0	0	1274	17	1	0	0	0	0	0	0	1316	300	1616
7	3	0	0	0	0	9	849	0	3	0	0	0	0	0	864	222	1086
8	0	0	0	0	0	0	1	280	0	0	0	0	0	0	281	68	349
9	0	0	0	0	0	0	2	1	197	0	0	0	0	0	200	52	252
10	0	0	0	0	0	2	3	0	5	400	0	2	0	0	412	107	519
11	0	0	0	7	0	0	0	0	0	0	163	0	0	0	170	43	213
12	0	0	0	0	0	5	0	0	0	0	0	145	0	0	150	30	180
13	1	0	0	0	0	0	0	0	0	0	0	0	437	0	438	113	551
14	0	0	0	0	0	0	0	0	0	5	0	0	0	1301	1306	330	1636
															10644	2661	13305

Los datos muestran gran precisión en la re-identificación y baja tasa de confusión con respecto a los otros peatones. El vector de características propuesto del método 2, es lo suficientemente discriminante para esta tarea.

Anexo 4

El canal H se representa en un intervalo de valores de 0 a 360 grados. Realizando un análisis del color con un valor de 180 en el canal H , ver tabla A4.1. Se define un rango de valores donde se logre preservar el mismo tono de color.

Tabla A4.1. Variación del color del valor H .

Valor $H = 180$			
Valor H	Color	Valor H	Color
178		182	
176		184	
174		186	
172		188	
170		190	
168		192	
166		194	
164		196	
162		198	
160		200	

La variación del color depende del valor H . A partir del valor de 174 y 186 se comienza a notar una diferencia en la tonalidad del color. Por tanto, se define un intervalo de valor 5, dado un valor H . Es decir $H - 5 \leq H \leq H + 5$. Aumentar el rango de valores causara confusión entre la distinción de colores, reducir el rango de valores hará que el criterio de similitud sea muy estricto.

Anexo 5

Este anexo muestra el vector de características extendido, ver tabla A5.1, para las variantes propuestas del método 2.

Tabla A5.1. Vector de características extendido.

Atributo	Estadística	Canal	Espacio de Color
0	Media	Escala de Grises	Escala de Grises
1	Desviación Estándar	Escala de Grises	Escala de Grises
2	Energía	Escala de Grises	Escala de Grises
3	Entropía	Escala de Grises	Escala de Grises
4	Media	R	RGB
5	Media	G	RGB
6	Media	B	RGB
7	Media	RGB	RGB
8	Desviación Estándar	R	RGB
9	Desviación Estándar	G	RGB
10	Desviación Estándar	B	RGB
11	Desviación Estándar	RGB	RGB
12	Energía	R	RGB
13	Energía	G	RGB
14	Energía	B	RGB
15	Energía	RGB	RGB
16	Entropía	R	RGB
17	Entropía	G	RGB
18	Entropía	B	RGB
19	Entropía	RGB	RGB
20	Media	H	HLS
21	Media	L	HLS
22	Media	S	HLS
23	Media	HLS	HLS
24	Desviación Estándar	H	HLS
25	Desviación Estándar	L	HLS
26	Desviación Estándar	S	HLS
27	Desviación Estándar	HLS	HLS
28	Energía	H	HLS
29	Energía	L	HLS
30	Energía	S	HLS
31	Energía	HLS	HLS
32	Entropía	H	HLS
33	Entropía	L	HLS
34	Entropía	S	HLS
35	Entropía	HLS	HLS
36	Media	H	HSV
37	Media	S	HSV
38	Media	V	HSV
39	Media	HSV	HSV
40	Desviación Estándar	H	HSV
41	Desviación Estándar	S	HSV
42	Desviación Estándar	V	HSV
43	Desviación Estándar	HSV	HSV
44	Energía	H	HSV

Atributo	Estadística	Canal	Espacio de Color
45	Energía	S	HSV
46	Energía	V	HSV
47	Energía	HSV	HSV
48	Entropía	H	HSV
49	Entropía	S	HSV
50	Entropía	V	HSV
51	Entropía	HSV	HSV
52	Media	L	LAB
53	Media	A	LAB
54	Media	B	LAB
55	Media	LAB	LAB
56	Desviación Estándar	L	LAB
57	Desviación Estándar	A	LAB
58	Desviación Estándar	B	LAB
59	Desviación Estándar	LAB	LAB
60	Energía	L	LAB
61	Energía	A	LAB
62	Energía	B	LAB
63	Energía	LAB	LAB
64	Entropía	L	LAB
65	Entropía	A	LAB
66	Entropía	B	LAB
67	Entropía	LAB	LAB
68	Media	Y	YCRCB
69	Media	CR	YCRCB
70	Media	CB	YCRCB
71	Media	YCRCB	YCRCB
72	Desviación Estándar	Y	YCRCB
73	Desviación Estándar	CR	YCRCB
74	Desviación Estándar	CB	YCRCB
75	Desviación Estándar	YCRCB	YCRCB
76	Energía	Y	YCRCB
77	Energía	CR	YCRCB
78	Energía	CB	YCRCB
79	Energía	YCRCB	YCRCB
80	Entropía	Y	YCRCB
81	Entropía	CR	YCRCB
82	Entropía	CB	YCRCB
83	Entropía	YCRCB	YCRCB
84	Dispersión	Escala de grises	Escala de Grises
85	Dispersión	R	RGB
86	Dispersión	G	RGB
87	Dispersión	B	RGB
88	Dispersión	RGB	RGB
89	Dispersión	H	HLS
90	Dispersión	L	HLS
91	Dispersión	S	HLS
92	Dispersión	HLS	HLS
93	Dispersión	H	HSV
94	Dispersión	S	HSV
95	Dispersión	V	HSV

Atributo	Estadística	Canal	Espacio de Color
96	Dispersión	HSV	HSV
97	Dispersión	L	LAB
98	Dispersión	A	LAB
99	Dispersión	B	LAB
100	Dispersión	LAB	LAB
101	Dispersión	Y	YCRCB
102	Dispersión	CR	YCRCB
103	Dispersión	CB	YCRCB
104	Dispersión	YCRCB	YCRCB
105	Homogeneidad	Escala de grises	Escala de Grises
106	Homogeneidad	R	RGB
107	Homogeneidad	G	RGB
108	Homogeneidad	B	RGB
109	Homogeneidad	RGB	RGB
110	Homogeneidad	H	HLS
111	Homogeneidad	L	HLS
112	Homogeneidad	S	HLS
113	Homogeneidad	HLS	HLS
114	Homogeneidad	H	HSV
115	Homogeneidad	S	HSV
116	Homogeneidad	V	HSV
117	Homogeneidad	HSV	HSV
118	Homogeneidad	L	LAB
119	Homogeneidad	A	LAB
120	Homogeneidad	B	LAB
121	Homogeneidad	LAB	LAB
122	Homogeneidad	Y	YCRCB
123	Homogeneidad	CR	YCRCB
124	Homogeneidad	CB	YCRCB
125	Homogeneidad	YCRCB	YCRCB
126	Media	LBP	LBP
127	Desviación Estándar	LBP	LBP
128	Homogeneidad	LBP	LBP

Anexo 6

Este anexo muestra los resultados de la selección de variables y métodos de búsqueda, con sus respectivos pesos, ver tabla A6.1, en donde C_w representa el peso de la características. Estos fueron utilizados en las variantes del método 2. Es importante notar que dado la naturaleza de los métodos A y B, estos no generan un peso de discriminación de las características. Los algoritmos y métodos utilizados ya han sido descritos anteriormente en la tabla 5.3.

Tabla A6.1. Resultado de los algoritmos de selección de variables y métodos de búsqueda con sus pesos.

A	B	C	C_w	D	D_w	E	E_w	F	F_w	G	G_w	H	H_w
6	6	59	0.3931	70	0.4246	57	1.7558	53	0.6955752	58	0.2981	57	0.5104
8	8	73	0.3894	103	0.4246	54	1.7109	98	0.6955752	57	0.274	70	0.5029
10	10	58	0.3874	57	0.4209	99	1.7109	57	0.671317	26	0.2616	103	0.5029
20	20	99	0.3871	54	0.4107	53	1.6833	54	0.6624675	41	0.2451	54	0.4979
27	27	54	0.3871	99	0.4107	98	1.6833	99	0.6624675	73	0.2339	99	0.4979
29	29	70	0.3849	73	0.4043	70	1.6695	70	0.6311296	74	0.2004	53	0.4827
36	36	103	0.3849	75	0.397	103	1.6695	103	0.6306091	59	0.1959	98	0.4827
41	41	41	0.3833	74	0.3952	73	1.5907	65	0.6291515	54	0.193	73	0.479
42	42	57	0.3825	98	0.3946	65	1.5492	73	0.605518	99	0.193	74	0.4577
43	43	75	0.3822	53	0.3946	74	1.4723	20	0.6003123	65	0.1917	75	0.454
53	53	74	0.3777	65	0.374	75	1.4355	89	0.5997918	75	0.1845	65	0.4524
54	54	26	0.3753	58	0.3737	58	1.4324	93	0.5992712	70	0.1809	58	0.438
55	55	65	0.3716	36	0.3545	59	1.3504	36	0.5986465	103	0.1809	41	0.4081
57	57	67	0.37	89	0.3521	41	1.3403	74	0.5642894	98	0.1791	36	0.4046
58	58	120	0.3623	20	0.3521	67	1.2928	75	0.5631442	53	0.1791	89	0.4033
59	59	42	0.3565	93	0.3515	20	1.2776	96	0.5627277	93	0.1628	20	0.4033
61	61	39	0.3556	41	0.3471	93	1.2775	39	0.5589797	23	0.1628	93	0.4029
65	65	96	0.3556	22	0.3364	89	1.2775	58	0.5514836	91	0.1626	59	0.3959
67	67	66	0.3521	91	0.3364	36	1.2757	59	0.5393024	69	0.1626	67	0.3924
73	73	82	0.3508	96	0.334	96	1.2432	41	0.5331598	102	0.1588	96	0.3867
81	81	43	0.3497	39	0.3337	39	1.2429	91	0.5262884	22	0.1587	39	0.3865
96	96	98	0.3438	67	0.3331	61	1.2325	22	0.5251432	96	0.1582	61	0.3831
98	98	53	0.3438	61	0.3307	22	1.1993	61	0.5181676	39	0.1579	22	0.3824
105	105	34	0.342	59	0.3283	91	1.1993	23	0.5177512	94	0.1579	91	0.3824
107	107	23	0.3417	82	0.3162	94	1.1605	67	0.5174388	37	0.1572	82	0.3609
110	110	92	0.3417	26	0.3152	37	1.1603	43	0.5162936	43	0.1497	26	0.3608
116	116	91	0.3389	66	0.3126	26	1.1422	92	0.5140031	87	0.1495	66	0.3558
120	120	22	0.3389	23	0.3071	82	1.138	69	0.5112962	6	0.1448	23	0.3483
127	127	6	0.3313	92	0.3067	63	1.1367	102	0.5111921	67	0.1443	92	0.3479

Anexo 7

Este anexo muestra el vector resultante, ver tabla A7.1, utilizado para describir a la persona, en las variantes del método 2 y con base en la tabla A6.1 del anexo 6.

Tabla A7.1. Vector final de descripción de la persona para las variantes del método 2.

Característica	w	Frecuencia	Estadística	Canal	Espacio de Color
57	0.669152833	8	Desviación Estándar	A	LAB
54	0.643677917	8	Media	B	LAB
98	0.629845867	8	Media	CB	YCRCB
53	0.629845867	8	Dispersión	A	LAB
73	0.617136333	8	Media	A	LAB
65	0.59467525	8	Desviación Estándar	CR	YCRCB
58	0.5801806	8	Entropía	A	LAB
41	0.5428433	8	Desviación Estándar	B	LAB
59	0.533817067	8	Desviación Estándar	CB	YCRCB
67	0.5083398	8	Desviación Estándar	YCRCB	YCRCB
96	0.50673795	8	Desviación Estándar	S	HSV
70	0.6323216	6	Desviación Estándar	B	LAB
74	0.577931567	6	Entropía	LAB	LAB
75	0.5693907	6	Dispersión	V	HSV
39	0.50592995	6	Media	HSV	HSV
91	0.4909814	6	Dispersión	S	HLS
22	0.490140533	6	Media	S	HLS
36	0.43890775	6	Media	H	HSV
20	0.438885383	6	Media	H	HLS
61	0.4107446	6	Energía	A	LAB

Anexo 8

Este anexo muestra los resultados obtenidos de la primera variante del método 2, ver tabla A8.1. La variación del vector de características que describe a la persona tiene un ligero decremento en la *TPR* y *FPR*, este cambio favorece solamente a la *FPR*, aunque en general no brinda un cambio significativo comparado con el método 2.

El vector inicial de 128 características con 6 medidas estadísticas reducido, tiene una capacidad de discriminación similar al vector inicial de 84 características con 4 estadísticas reducido. Esto evaluado en la descripción de la persona en los 5 espacios de color mencionados anteriormente.

Tabla A8.1. Resultados de re-identificación.

PEATON	TPR	FPR	VP	VN	FP	FN
1	70.56277056	29.3934439	326	9068	3775	136
2	89.42048518	6.67456222	1327	11032	789	157
3	93.97501837	20.98961822	1279	9437	2507	82
4	85.47794118	6.128133705	1860	10447	682	316
5	95.84507042	18.20782499	1361	9721	2164	59
6	74.19554455	12.1823937	1199	10265	1424	417
7	43.64640884	10.86013585	474	10892	1327	612
8	76.50429799	11.89410312	267	11415	1541	82
9	60.71428571	24.11706121	153	9905	3148	99
10	64.35452794	25.6765212	334	9503	3283	185
11	95.30516432	20.45523984	203	10414	2678	10
12	95.55555556	7.306666667	172	12166	959	8
13	94.1923775	16.55951074	519	10642	2112	32
14	88.56968215	14.86845488	1449	9934	1735	187
MEDIA	80.59422359	16.09383359				

Anexo 9

Este anexo muestra los resultados obtenidos de la segunda variante del segundo método, ver tabla A9.1. La asignación de pesos a las características que compone el vector, aumenta la *TPR* y *FPR*, pero no representa un cambio significativo. Los resultados son similares al método 2, pero esta variante genera mayor complejidad computacional al considerar los pesos.

Tabla A9.1. Resultados de re-identificación.

PEATON	TPR	FPR	VP	VN	FP	FN
1	72.7272727	30.5691817	336	8917	3926	126
2	91.3746631	7.20751205	1356	10969	852	128
3	95.9588538	22.1784997	1306	9295	2649	55
4	88.6488971	6.82001977	1929	10370	759	247
5	96.8309859	20.3954565	1375	9461	2424	45
6	75.6806931	12.8069125	1223	10192	1497	393
7	46.0405157	11.2611507	500	10843	1376	586
8	77.0773639	13.1367706	269	11254	1702	80
9	59.1269841	25.4500881	149	9731	3322	103
10	66.2813102	28.3200375	344	9165	3621	175
11	97.1830986	21.9065078	207	10224	2868	6
12	96.1111111	6.64380952	173	12253	872	7
13	95.6442831	17.5552768	527	10515	2239	24
14	89.7310513	16.3338761	1468	9763	1906	168

UNIVERSIDAD TECNOLÓGICA
EMILIANO ZAPATA DEL ESTADO DE MORELOS
ORGANISMO PÚBLICO DESCENTRALIZADO DEL GOBIERNO DEL ESTADO DE MORELOS

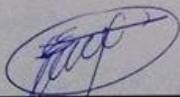
otorga el presente

RECONOCIMIENTO

A: Rosado Castellanos Damián Uriel

Por su participación, en el marco de actividades de la Escuela de Inteligencia Artificial y Robótica 2017, realizada los días 19 y 20 de octubre del presente año en las instalaciones de esta Universidad Tecnológica.

Emiliano Zapata, Mor., octubre 2017



M.C. Jaime Vázquez Colín
Director de la División Académica de
Mecánica Industrial





El Centro de Investigación en Ingeniería y Ciencias Aplicadas de la Universidad Autónoma del Estado de Morelos y el Instituto de Ingenieros Electrónicos y Eléctricos de Morelos A.C

otorgan el presente

RECONOCIMIENTO

A: **Damián Uriel Rosado Castellanos**

Por su destacada participación como:

Asistente

En el marco del Congreso Internacional de Ingeniería Mecatrónica, Electrónica y Automotriz realizado del 21 al 24 de noviembre en la ciudad de Cuernavaca, Morelos , México.


Ing. Leoncio Aguilar Negrete
Presidente del Instituto de Ingenieros Electrónicos y Eléctricos de Morelos A.C
Instituto de Ingenieros Electronicos y Eléctricos de Morelos A.C.


Dr. Pedro Antonio Márquez Aguilar
Director del Centro de Investigación en Ingeniería y Ciencias Aplicadas de la UAEM





LA RAMA ESTUDIANTIL IEEE
DE LA FACULTAD DE INGENIERÍA
DE LA UNIVERSIDAD AUTÓNOMA DE CAMPECHE
OTORGA EL PRESENTE



Reconocimiento

A
ING. DAMIÁN U. ROSADO CASTELLANOS

CONFERENCISTA

CON EL TEMA:

“MI TEMA DE TESIS DE MAESTRÍA EN EL CENIDET”

DURANTE EL PRIMER SIMPOSO 2018 DE LA RAMA ESTUDIANTIL IEEE -UACAM
CAMPUS V, SAN FRANCISCO DE CAMPECHE, DEL 20 AL 22 DE MAYO DE 2018

BR. ANGEL ULISES LOREDO VARGAS
PRESIDENTE
RAMA ESTUDIANTIL IEEE UACAM

M. EN C. GUADALUPE MANUEL ESTRADA
DIRECTOR
FACULTAD DE INGENIERÍA UACAM







Otorga la presente

Constancia

A

ING. DAMIAN URIEL ROSADO CASTELLANOS

Por su valiosa participación con la conferencia:

“TECNICAS PARA LA VIDEOVIGILANCIA INTELIGENTE”

Realizada el 25 de Octubre de 2018.

2do Congreso Internacional
 Villa Montemorelos, Dgo., a 25 de Octubre de 2018.



TECNIM
 INSTITUTO TECNOLÓGICO
 Y DE ESTUDIOS SUPERIORES
 DE GUANAJUATO
 DIRECCIÓN



Ing. Agustín Cervantes Gómez
 Director del I.T. del Valle del Guadiana

