

SISTEMA DE VISIÓN ARTIFICIAL PARA LA VIGILANCIA DE PROPIEDADES MEDIANTE
PROCESAMIENTO DIGITAL DE IMÁGENES DE VIDEO Y RECONOCIMIENTO DE PATRONES

PAOLA ANDREA CÓRDOBA ESTRADA
CAMILO SÁNCHEZ FERREIRA

UNIVERSIDAD DEL CAUCA
FACULTAD DE CIENCIAS NATURALES EXACTAS Y DE LA EDUCACIÓN
INGENIERÍA FÍSICA
POPAYÁN
2008

SISTEMA DE VISIÓN ARTIFICIAL PARA LA VIGILANCIA DE PROPIEDADES MEDIANTE
PROCESAMIENTO DIGITAL DE IMÁGENES DE VIDEO Y RECONOCIMIENTO DE PATRONES

PAOLA ANDREA CÓRDOBA ESTRADA
CAMILO SÁNCHEZ FERREIRA

Trabajo de grado presentado como requisito parcial para optar al título de
Ingeniero Físico

DIRECTOR

Ing.EDGAR LEONAIRO PENCUE FIERRO

UNIVERSIDAD DEL CAUCA
FACULTAD DE CIENCIAS NATURALES EXACTAS Y DE LA EDUCACIÓN
INGENIERÍA FÍSICA
POPAYÁN
2008

Nota de aceptación:

Ing. Edgar Leonairo Pencue Fierro
Director

Ing. Elena Muñoz España
Jurado

Mg. Jaury León Téllez
Jurado

Fecha de Sustentación:
Popayán, 10 de Abril de 2008.

Dedico este trabajo especialmente a mi Mamá, a quien amo y admiro profundamente, por su inmenso apoyo en todas mis decisiones y por la formación que me ha brindado; la que me ha llevado hasta este punto de mi vida y gracias a la cual soy quien soy.

A la memoria de mi Papá, a quien amo y extraño mucho, por todo lo que me enseñó y por presentarme esta carrera en la que estoy muy contenta.

A mis hermanos Adriana y Francisco, por su apoyo, cariño y por aguantarme todo este tiempo.

A mis Tíos por su apoyo en momentos difíciles y por la confianza depositada en mí.

A mi amigo y compañero de tesis, Camilo, de quien he aprendido mucho, y con el que he trabajado a gusto por su responsabilidad, entrega y trabajo en equipo.

A mis amigos: Nathalie, Leonardo, Carlos Augusto, Carlos Andrés, Claudia, Sandra, Edison, no sólo por brindarme su amistad sino también por su apoyo y por todo lo que he aprendido de cada uno de ellos.

Paola Andrea Córdoba E.

Dedico éste trabajo a:

Mis Padres, quienes me han brindado todo su amor y apoyo incondicional. Mi Mamá, a quien quiero con todo mi corazón y que me ha enseñado mucho con el amor que sólo ella podría hacerlo. Mi Padre, a quien quiero y admiro muchísimo, y que me ha enseñado que debo poner todo mi empeño para lograr lo que quiero y que no debo rendirme hasta lograrlo. Gracias a los dos. LOS QUIERO MUCHÍSIMO.

A Claudia, mi amor, mi amiga y mi compañera de carrera, quien me ha acompañado incondicionalmente en este duro caminar. Si no fuera por Ti no sería la persona que soy hoy, porque tu me enseñaste que la vida, sin esos momentos difíciles, sería muy aburrida. Gracias por haberme dado tanto sin pedir nada a cambio y por amarme como lo haces. TE AMO MUCHÍSIMO.

A mi Hermano, que ha estado conmigo desde que yo tenía un año y medio de edad, y siempre ha querido lo mejor para mí. Gracias, porque mi vida sin Ti hubiera sido muy aburrida. TE QUIERO MUCHO

A mi pequeño sobrino Sebastián, quien ha llenado de alegría nuestras vidas y que, con su inocente sonrisa, nos motiva todos los días a seguir adelante.

A Paola Andrea, mi amiga y compañera en esta última etapa de mi carrera, por creer que esto era posible, por su gran optimismo y porque sin ella esto habría sido muchísimo más difícil.

A mis amigos: Leonardo, Carlos Augusto, Carlos Andrés, Edison, Sandra y Natalie, quienes han hecho agradable mi paso por esta Universidad, con quienes reí muchísimo y aprendí aun más. A Lilly y Vivian, mis amigas de siempre, con quienes he compartido muchos momentos bonitos. A José Luís, Álvaro, Diego e Izjak, mis amigos del colegio, quienes a pesar del tiempo, siguen brindándome todo su apoyo. A Sara, Sulgenis, Johana, Patricia y Diego Alejandro, mis amigos y compañeros de trabajo, porque, en tan poco tiempo de conocerlos, me han apoyado en mi nueva aventura como docente y me han dado una gran amistad.

Muchisímas gracias a todos.

Camilo Sánchez F.

AGRADECIMIENTOS

Los autores queremos agradecer a:

El Ing. Leonairo Pencue Fierro por sus valiosos aportes al trabajo, por su apoyo y por creer en nosotros.

A los jurados evaluadores por sus conocimientos, su dedicación y su valioso tiempo.

Al Grupo de Óptica y Láser y sus integrantes por habernos permitido hacer parte de él y por aportar el espacio para desarrollar este proyecto.

A aquellos profesores que nos enseñaron muchas cosas, y que hicieron de nosotros buenos profesionales.

Al profesor Luis Fernando Echeverri, quien ha logrado que los Ingenieros Físicos tengan cada día más acogida y sean más conocidos.

A la Universidad del Cauca.

CONTENIDO

	pág
INTRODUCCIÓN	14
1. VIGILANCIA Y SISTEMAS DE CCTV	16
1.1. PANORAMA MUNDIAL DE LA VIGILANCIA Y SEGURIDAD PRIVADA	16
1.2. PANORAMA NACIONAL DE LA VIGILANCIA Y SEGURIDAD PRIVADA	18
1.3. SISTEMAS ACTUALES DE VIGILANCIA CON BASE TECNOLÓGICA	20
1.4. SISTEMAS DE CIRCUITO CERRADO DE TELEVISIÓN	21
BIBLIOGRAFÍA	22
2. COMPONENTES DE UN SISTEMA DE VISIÓN ARTIFICIAL	23
2.1. ADQUISICIÓN DE IMÁGENES	24
2.1.1. Imágenes Digitales y Video	24
2.1.2. Iluminación	30
2.1.3. Cámaras Digitales	37
2.1.4. Tarjetas de Adquisición y Procesamiento de Imágenes	40
2.2. PREPROCESAMIENTO DE IMÁGENES	41
2.2.1. Eliminación de Ruido	42
2.3. SEGMENTACIÓN DE IMÁGENES	48
2.3.1. Umbralización	49
2.3.2. Etiquetado	50
2.4. OPERACIONES MORFOLÓGICAS Y DESCRIPCIÓN DE OBJETOS	50
2.4.1. Dilatación	51
2.4.2. Erosión	52
2.4.3. Apertura y Cierre	52
2.4.4. Esqueletización	53
2.4.5. Características de Regiones	54
2.5. RECONOCIMIENTO DE OBJETOS	56
2.5.1. Estructura General de un Clasificador	56
BIBLIOGRAFÍA	57

3. SISTEMAS DE VIGILANCIA BASADOS EN VISIÓN ARTIFICIAL	58
3.1. DETECCIÓN DE MOVIMIENTO	58
3.1.1. Sustracción del Fondo	58
3.1.2. Métodos Estadísticos	61
3.1.3. Diferenciación Temporal	61
3.1.4. Flujo Óptico	63
3.1.5. Detección de Cambios de Luz y Sombras	63
3.2. SEGUIMIENTO DE OBJETOS EN MOVIMIENTO	64
3.2.1. Filtro de Kalman	64
3.2.2. Seguimiento de Regiones	67
3.2.3. Seguimiento por Contornos	67
3.3. CLASIFICACIÓN DE OBJETOS	69
3.3.1. Clasificación Basada en Forma	69
3.3.2. Clasificación Basada en Movimiento	70
3.3.3. Clasificación Basada en Modelos	70
BIBLIOGRAFÍA	71
4. DESCRIPCIÓN DEL SISTEMA DE VIGILANCIA BASADO EN VISIÓN ARTIFICIAL	73
4.1. SISTEMA DE ADQUISICIÓN DE LAS IMÁGENES	73
4.1.1. Cámara de Video-Vigilancia	74
4.1.2. Tarjeta Digitalizadora de Video	74
4.1.3. Video Adquirido por el Sistema	75
4.2. PREPROCESAMIENTO DE LAS IMÁGENES ADQUIRIDAS	76
4.2.1. Conversión de las Imágenes a Color a Imágenes en Escala de Grises	76
4.2.2. Control de Iluminación	77
4.2.3. Filtro de Paso Bajo	78
4.3. DETECCIÓN DE MOVIMIENTO	79
4.3.1. Inicialización de Fondo	81
4.3.2. Sustracción del Fondo	81
4.3.3. Segmentación del Movimiento	81
4.3.4. Eliminación del Ruido	82
4.3.5. Actualización del Fondo	82
4.4. SEGUIMIENTO DE LOS OBJETOS DE INTERÉS	82
4.5. CLASIFICACIÓN DE LOS OBJETOS EN MOVIMIENTO	85
4.6. ALARMAS Y REPORTES	86

5. RESULTADOS Y ANÁLISIS	89
5.1. CONTROL DE ILUMINACIÓN	89
5.2. DETECCIÓN DE MOVIMIENTO	91
5.3. SEGUIMIENTO DE OBJETOS DE INTERÉS	99
5.4. RECONOCIMIENTO DE OBJETOS	103
5.5. RENDIMIENTO DEL SISTEMA	107
CONCLUSIONES	110
RECOMENDACIONES Y TRABAJO FUTURO	112
A. MANUAL DEL USUARIO	114

LISTA DE FIGURAS

	pág
1.1. Descripción del mercado global.	17
1.2. Concentración del mercado global.	17
1.3. Tipos de servicios de clase global.	18
1.4. Comportamiento de los servicios de vigilancia y seguridad privada en Colombia.	19
1.5. Participación de servicios de vigilancia y seguridad privada por sectores.	20
2.1. Componentes de un sistema de visión artificial.	23
2.2. Convenio de ejes utilizado para la representación digital de imágenes.	25
2.3. Transferencia de una señal de video analógica.	27
2.4. Campos entrelazados.	27
2.5. Iluminancia	27
2.6. Tono	28
2.7. Saturación	29
2.8. Video digital	30
2.9. Iluminación direccional.	31
2.10. Iluminación difusa.	32
2.11. Iluminación posterior o a contraluz.	33
2.12. Fuentes de luz incandescentes.	34
2.13. Fuentes de luz Flourescente.	34
2.14. Iluminación con LED.	35
2.15. Fuentes de luz estroboscópica.	36
2.16. Haz de luz láser de longitud de onda de 532 nm.	36
2.17. Fibra Óptica como medio de propagación de la luz.	37
2.18. Cámara digital y dispositivo CCD.	37
2.19. Sensor CCD con arquitectura de transferencia interlínea.	39
2.20. Sensor CCD con arquitectura de transferencia de cuadro.	40
2.21. Filtros espaciales de paso bajo	44
2.22. Filtrado espacial de paso bajo.	45
2.23. Filtro espacial de paso alto.	45
2.24. Filtrado espacial de paso alto.	46
2.25. Sustracción de imágenes.	47

2.26. Operaciones morfológicas.	53
2.27. Estructura general de un clasificador.	56
3.1. Organigrama de algoritmos de detección de movimiento basados en sustracción del fondo.	59
3.2. Detección de movimiento mediante sustracción del fondo.	60
3.3. Ejemplo de diferenciado temporal.	62
3.4. Seguimiento de regiones usando el criterio de mínima distancia.	68
3.5. Seguimiento de regiones usando el criterio de superposición de <i>bounding box</i>	68
3.6. Área del objeto de interés.	69
3.7. Perímetro del objeto de interés.	70
4.1. Cámara de video-vigilancia utilizada en el sistema.	74
4.2. Tarjeta importadora de video marca Zoltrix.	75
4.3. Imágenes adquiridas por la cámara y la tarjeta en diferentes condiciones de iluminación.	76
4.4. Conversión a escala de grises.	77
4.5. Control de iluminación.	79
4.6. Esquema general del algoritmo de detección de movimiento empleado.	80
4.7. Detección de movimiento.	83
4.8. Seguimiento de regiones.	84
4.9. Solución de problemas de oclusión.	85
4.10. Seguimiento de regiones.	86
4.11. Clasificación de los objetos en movimiento.	87
5.1. Efectos del control de iluminación.	90
5.2. Pasos para la detección de movimiento y filtrado de ruido.	91
5.3. Detección de movimiento sin filtro gaussiano.	93
5.4. Detección de movimiento con filtro gaussiano.	94
5.5. Actualización del fondo.	94
5.6. Detección de movimiento en la escena 4.	96
5.7. Detección de movimiento en la escena 5.	97
5.8. Problemas de detección de objetos muy pequeños.	98
5.9. Problemas de la detección de grupos.	101
5.10. Interacción entre objetos.	102
5.11. Agrupamiento y separación de objetos.	103
5.12. Reconocimiento de objetos en movimiento.	104
5.13. Diferencia en reconocimiento de acuerdo a la zona de interés.	106
5.14. Errores en la asignación de las respectivas etiquetas a los objetos de interés.	108

5.15. Error en la clasificación de objetos.	109
A.1. Interfaz gráfica de usuario del sistema.	114
A.2. Opción Abrir Video.	115
A.3. Interfaz gráfica de usuario después de iniciar la captura o abrir el video.	116
A.4. Opción Guardar Video.	117
A.5. Interfaz gráfica cuando la opción Fondo está activada.	118
A.6. Interfaz gráfica cuando la opción Movimiento está activada.	119
A.7. Interfaz gráfica cuando la opción Configurar está activada.	120
A.8. Ventana de configuración y selección de la zona de interés.	121
A.9. Ventana de configuración para fijar el tiempo de restricción.	121
A.10. Ventana de configuración y selección de la zona restringida.	122
A.11. Ventana de configuración para fijar la distancia patrón.	123
A.12. Selección del patrón de medida.	123
A.13. Ventana de entrenamiento para la etapa de reconocimiento.	124
A.14. Interfaz gráfica de usuario después del entrenamiento.	124
A.15. Reporte generado por el sistema.	126

LISTA DE TABLAS

	pág
5.1. Porcentaje de detección de movimiento del sistema.	95
5.2. Resultados de la detección de movimiento en el rango de operación.	99
5.3. Eficiencia del sistema para realizar el seguimiento de objetos.	100
5.4. Matriz de confusión del sistema para los peatones.	105
5.5. Matriz de confusión del sistema para los vehículos.	107
5.6. Tiempos de procesamiento del sistema por etapas.	108

INTRODUCCIÓN

El deseo y la necesidad de encontrar mecanismos que permitan mantener protegidas todo tipo de propiedades ha ido creciendo, no sólo con el aumento de la inseguridad, sino también con la necesidad de automatizar y mejorar cualquier tipo de proceso en el cual se requiera un monitoreo constante del sistema. Se han desarrollado numerosos sistemas que cambian su complejidad y costo de acuerdo a las condiciones del lugar que se desea vigilar, es así que en la actualidad se han utilizado diferentes tecnologías para el desarrollo de estos sistemas, tales como: la interrupción de un láser, el monitoreo con cámaras de video, sensores de temperatura, detectores de infrarrojo, sensores de movimiento, entre otros.

Estos sistemas presentan ciertas ventajas y desventajas. Por ejemplo, el sistema de interrupción láser es de bajo costo, pero presenta dificultades en el momento de cubrir grandes espacios y de reconocer los objetos que pueden generar la alarma. En contraste, el reconocimiento de objetos es una ventaja para los sistemas de circuito cerrado de televisión (CCTV), ya que estos captan la imagen del objeto y cubren extensas áreas; sin embargo, requieren de una persona que identifique y distinga los diferentes tipos de eventos que se puedan presentar, además de ser costosos. Los sensores de infrarrojo y temperatura, por su parte, pueden generar alarmas automáticamente a un costo moderado, pero no pueden distinguir objetos eficientemente.

En otras palabras, los sistemas de seguridad empleados actualmente para la vigilancia de propiedades, presentan algunos problemas cuando se utilizan en lugares cuyas condiciones ambientales y de iluminación no están controladas.

Por otro lado, la técnica de visión artificial ha permitido el desarrollo de sistemas automatizados que son capaces de realizar tareas bastante complejas, por ejemplo: el reconocimiento de objetos, control de calidad, conteo automático, entre otras.

El objetivo principal de este trabajo es implementar un sistema que permita la vigilancia en lugares cuyas condiciones tanto de iluminación como ambientales sean cambiantes. Para cumplir con este objetivo, se propone emplear la técnica de visión artificial, la cual, como se mencionó ya, es una herramienta con muchas ventajas para la automatización de sistemas, y que ha permitido realizar tareas bastante complejas.

Lo más importante de este trabajo es que el sistema funciona correctamente en espacios exteriores, donde las condiciones de iluminación, clima, entre otras, no están controladas, lo que, normalmente, aumentaría la complejidad del sistema y aumentaría la probabilidad de falsas alarmas.

El presente documento se encuentra estructurado así:

Capítulo 1 - Describe brevemente cómo están conformados actualmente los sistemas de seguridad, tanto a nivel internacional como en Colombia. También hace una breve descripción sobre los sistemas de circuito cerrado de televisión, que son los más empleados hoy en día.

Capítulo 2 - Describe uno a uno los componentes básicos de un sistema de visión artificial, como son la adquisición de las imágenes, el preprocesamiento de la imagen, y la segmentación y reconocimiento de los objetos de interés para un sistema basado en esta técnica.

Capítulo 3 - Describe los componentes básicos de un sistema de vigilancia basado en visión artificial, se tratan técnicas como la detección de movimiento, seguimiento y clasificación de objetos en movimiento en una secuencia de imágenes de una escena que se desee vigilar.

Capítulo 4 - Se muestra como está conformado el sistema de vigilancia implementado en este trabajo, así como las técnicas empleadas en la aplicación final.

Capítulo 5 - Se muestran los resultados obtenidos con este sistema, así como un análisis de los mismos. En este capítulo se pueden observar también las ventajas y desventajas de la aplicación.

Posteriormente, se presentan las conclusiones obtenidas con base en los diferentes resultados obtenidos, problemas, ventajas y desventajas del sistema.

Por último, se anexa un manual para el usuario del sistema, en el cual podrá conocer las diferentes herramientas de la aplicación.

Capítulo 1

VIGILANCIA Y SISTEMAS DE CCTV

El término vigilancia se define como la observación sistemática de espacios, personas o cosas mediante fotografía, videos, dispositivos electrónicos, entre otros [1]. Actualmente la vigilancia se ha convertido en una necesidad para garantizar la seguridad de personas o bienes materiales en espacios públicos o privados con el fin de reducir el riesgo de accidentes y prevenir acciones delictivas como robo, vandalismo o terrorismo.

Esta realidad se ve reflejada en el incremento de los servicios de vigilancia y seguridad privada, tanto a nivel mundial como nacional.

1.1. PANORAMA MUNDIAL DE LA VIGILANCIA Y SEGURIDAD PRIVADA

Es de gran importancia conocer cómo se encuentra este sector a nivel mundial, ya que permitirá visualizar su impacto en la economía global, las tendencias e innovaciones en la prestación de los servicios del sector.

El primer acercamiento que se desea brindar al mercado global de la vigilancia y seguridad privada, se da desde el punto de vista de su magnitud en términos de los ingresos operacionales. Es decir, desde la cuantificación del volumen global de la prestación de sus servicios. Se proyecta que para el año 2008 los ingresos operacionales globales del sector ascenderán a 137 billones de dólares.

Como se puede apreciar en la figura 1.1, el 70% del mercado de la vigilancia y seguridad privada se concentra en los mercados de Estados Unidos y Europa Occidental. Igualmente, en estas regiones se presenta un elevado grado de concentración. En el mercado de Estados

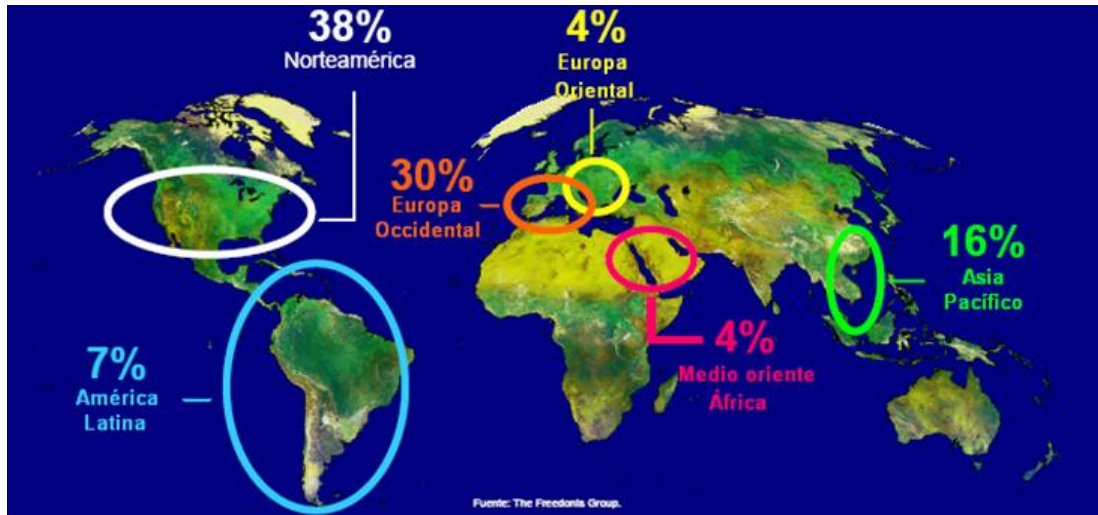


Figura 1.1. Descripción del mercado global. Porcentajes de participación por ingresos operacionales.

Unidos las cinco empresas líderes concentraban el 35 % en 1990; para el 2006, dichas empresas concentraban el 51 %. En el mercado de Europa Occidental, las cuatro empresas líderes concentraban el 15 % en 1990; para el 2006, las cuatro empresas líderes concentraban el 35 % como se puede observar en la figura 1.2.

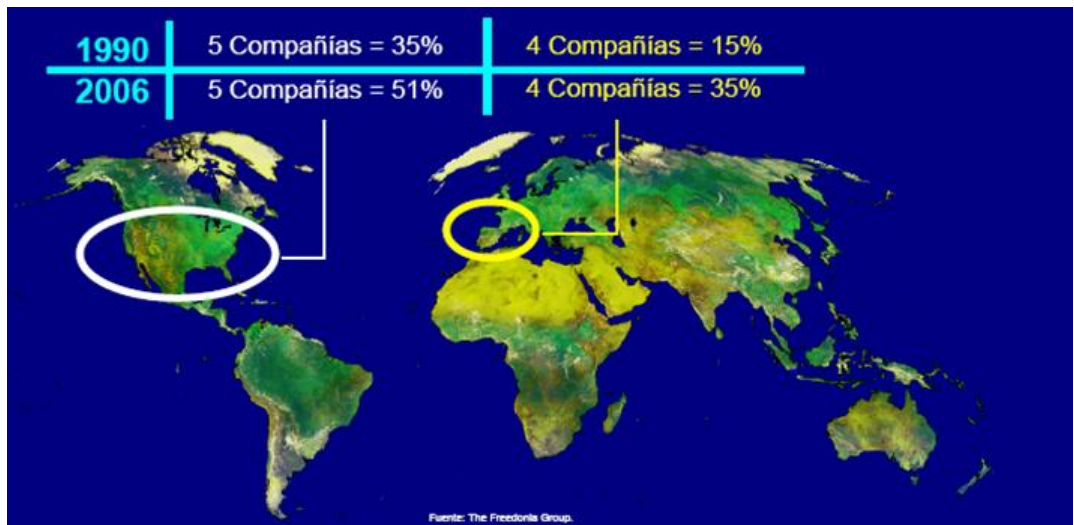


Figura 1.2. Concentración del mercado global. Participación del mercado de las empresas líderes.

La composición total de los ingresos del mercado global se distribuye en los siguientes tipos de servicios: el 45 % lo conforma los servicios de vigilancia y seguridad privada armados, el 35 % los servicios prestados con medios tecnológicos, el 9.6 % lo constituye los servicios de

transporte de valores, el 12.6% lo conforman un paquete de servicios varios en los cuales se incluye las asesorías, consultoría e investigadores, y un 2.3% lo conforma el servicio de vigilancia y seguridad privada en prisiones y correccionales (Figura 1.1) [2].

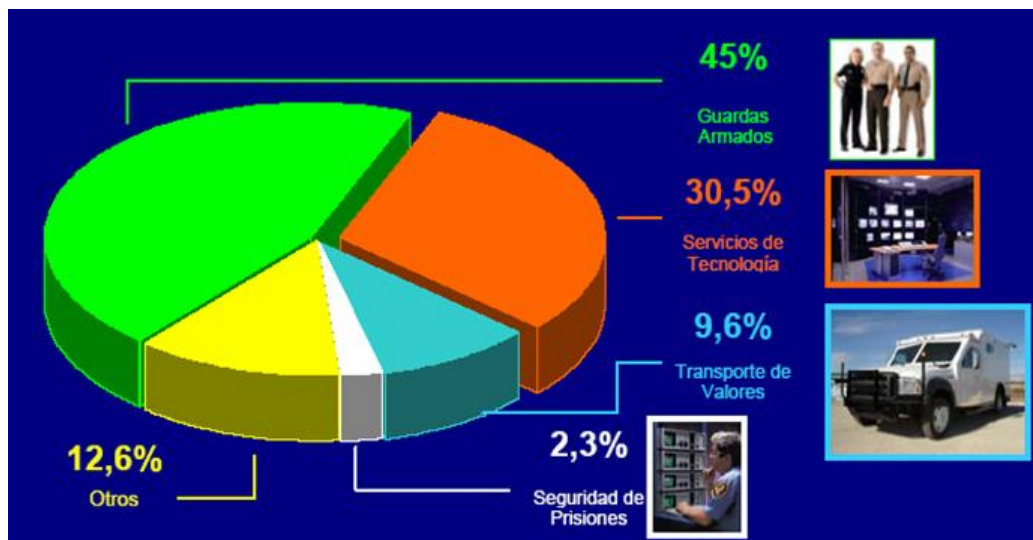


Figura 1.3. Tipos de servicios de clase global. Composición del mercado mundial de la seguridad privada.

1.2. PANORAMA NACIONAL DE LA VIGILANCIA Y SEGURIDAD PRIVADA

El ejercicio del control al sector en Colombia se da con la creación de la Superintendencia de Vigilancia y Seguridad Privada, la cual inicia sus funciones a partir de 1.994. Para esta fecha se registró un total de 763 servicios de vigilancia y seguridad privada, los cuales se encontraban conformados por 395 empresas de vigilancia y seguridad privada con armas; 360 departamentos de seguridad; y 8 empresas transportadoras de valores.

Desde entonces el sector ha presentado un importante crecimiento, en la actualidad los servicios de vigilancia y seguridad privada se encuentran conformados por 3.511 servicios (Figura 1.4), lo que presenta un crecimiento del 360%. Para el periodo comprendido entre enero de 1994 y agosto de 2007, los servicios de vigilancia y seguridad privada han presentado un crecimiento anual del 15%.

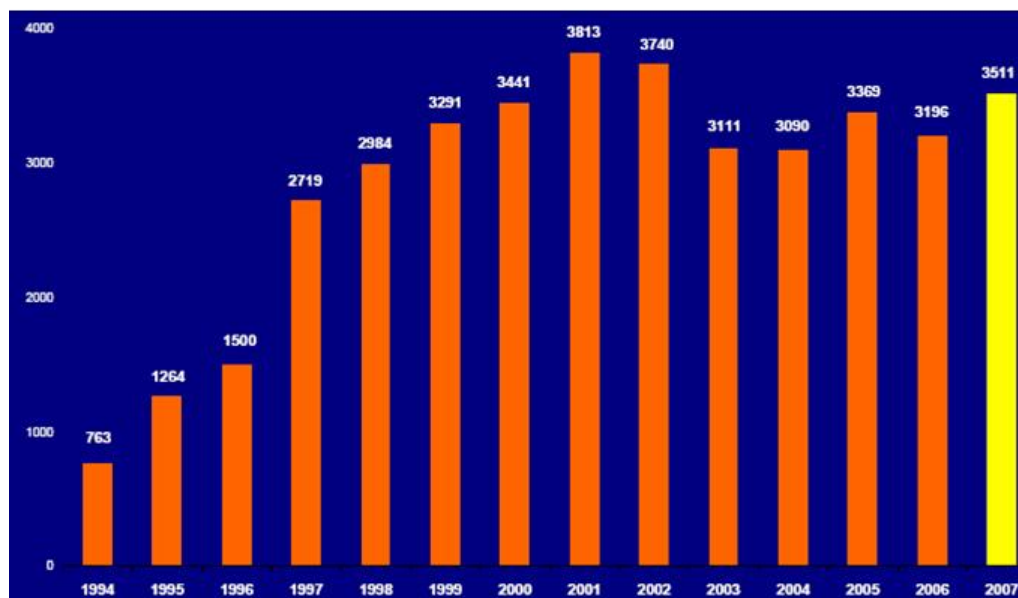


Figura 1.4. Comportamiento de los servicios de vigilancia y seguridad privada en Colombia (enero 1994 - agosto 2007).

Actualmente y en comparación con el mercado internacional, en Colombia, la prestación de los servicios de vigilancia y seguridad privada se concentra en un 76 % en los servicios armados y solamente un 6 % del mercado lo abarca los servicios prestados con medios tecnológicos.

Al analizar los tipos de clientes que demandan estos servicios permite establecer el grado de experiencia del sector en la prestación de sus servicios. Como se aprecia en la figura 1.5, el sector de la vigilancia y seguridad privada concentra la prestación de sus servicios en los sectores residenciales, comerciales, de servicios, e industriales, con una participación del 69 %. Sectores para los cuales se puede afirmar que el sector cuenta con una alta experiencia.

En la prestación de sus servicios a sectores como el bancario y de empresas públicas, el sector concentra el 15 % de sus actividades. Lo cual se puede interpretar, en términos de experiencia, que cuenta con un nivel adecuado. Mientras que refleja una mínima participación y un bajo nivel de experiencia en sectores tales como transporte, comunicación, petrolero, portuario, entre otros [3].

Según lo mencionado anteriormente, en Colombia, el porcentaje de sistemas de vigilancia y seguridad privada, basados en medios tecnológicos, es muy bajo en comparación a los otros países, en los cuales hay una tendencia a aumentar este tipo de servicios por su versatilidad y bajo coste económico.

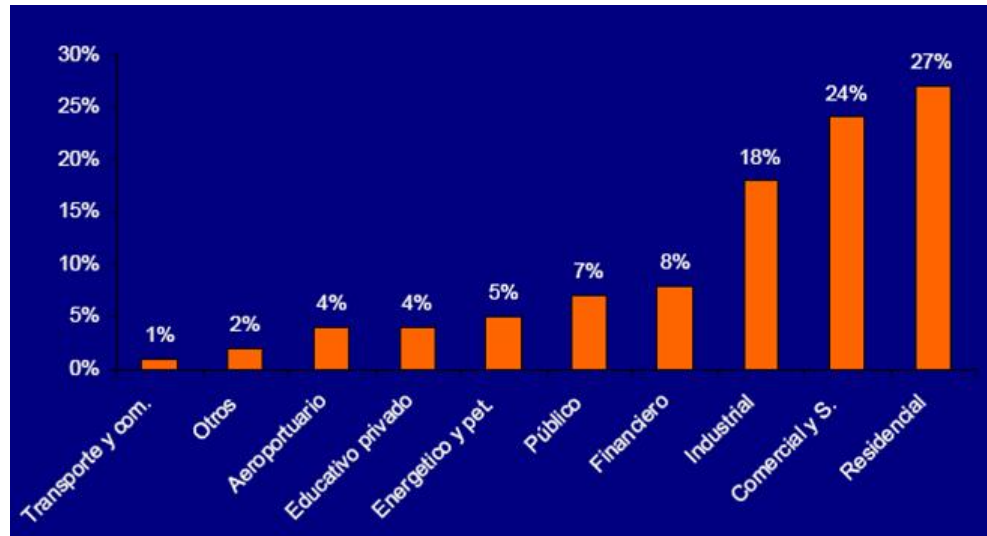


Figura 1.5. Participación de servicios de vigilancia y seguridad privada por sectores.

1.3. SISTEMAS ACTUALES DE VIGILANCIA CON BASE TECNOLÓGICA

En la actualidad existen diferentes tipos de sistemas con base tecnológica empleados para la vigilancia y seguridad de espacios como aeropuertos, bancos, supermercados, establecimientos comerciales, residencias, entre otros. Éstos cambian su complejidad y costo de acuerdo a las condiciones del lugar que se desea vigilar, es así que en la actualidad se han utilizado diferentes tecnologías para el desarrollo de estos sistemas, tales como: la interrupción de un láser, el monitoreo con cámaras de video, sensores de temperatura, detectores de infrarrojo, sensores de movimiento, entre otros.

Estos sistemas presentan ciertas ventajas y desventajas. Por ejemplo, el sistema de interrupción láser es de bajo costo, pero presenta dificultades en el momento de cubrir grandes espacios y de reconocer los objetos que pueden generar la alarma. Los sensores de infrarrojo y temperatura, por su parte, pueden generar alarmas automáticamente a un costo moderado, pero no pueden distinguir objetos eficientemente. Por esta razón es muy común que este tipo de sistemas se empleen simultáneamente en un mismo sitio.

En contraste, el reconocimiento de objetos es una ventaja para los sistemas de monitoreo con cámara de video, ya que estos captan la imagen del objeto y cubren extensas áreas, por lo que

los sistemas de video vigilancia tienen cada vez mayor demanda, especialmente, para garantizar la seguridad de interiores y alrededores de edificios. La presencia de numerosas cámaras de seguridad en cualquier entorno urbano es un hecho. Gracias a la evolución tecnológica se ha logrado que instalar cámaras de captura de video no precise altas inversiones económicas y, por tanto, la mayoría de bancos, estaciones, aeropuertos, tiendas, parqueaderos, etc., incorporan en sus instalaciones algún sistema, más o menos complejo, de seguridad basado en video vigilancia [4].

1.4. SISTEMAS DE CIRCUITO CERRADO DE TELEVISIÓN

El video brinda múltiples ventajas dentro de un sistema de seguridad. Una de ellas es que permite el monitoreo localizado de bienes o propiedades además de la actividad que se presente en los sectores de interés. Es además empleado para detectar la entrada de una persona no autorizada en un determinado lugar y su seguimiento dentro de éste. En la función de protección de un perímetro, el video se usa en complemento con otros dispositivos de detección de intrusos y alarma así como también la detección de movimiento a partir del video para alertar al guardia o al personal encargado, de la intrusión. Si ocurre un suceso de este tipo, las cámaras que componen el circuito cerrado de televisión (CCTV) facilitan la tarea de alertar acerca de este evento.

El monitoreo de video permite al guardia ser más efectivo, pero, además mejora la seguridad al transmitir la señal a otros centros de control o personal. La imagen de video puede ser grabada por medio de VCR o DVR y copiada como evidencia.

La mayoría de este tipo de aplicaciones de video se pueden clasificar en dos clases: las aplicaciones para recintos cerrados (*indoor*) y las aplicaciones para espacios abiertos (*outdoor*). Esta división se explica por el hecho de que estas dos condiciones ambientales requieren de equipos adecuados para tales condiciones. Los dos parámetros principales son los factores ambientales y los factores de iluminación. Los sistemas que se emplean dentro de recintos cerrados requieren de luz artificial que puede o no ser aumentada por la luz natural. Estos sistemas además están sujetos a leves variaciones de temperatura, humedad, polvo y humo. En contraste, los sistemas que operan en espacios abiertos deben resistir temperaturas extremas, precipitación, neblina, suciedad, polvo, humo, etc. Este tipo de sistemas emplean la luz natural en el día y luz artificial en la noche. Algunas cámaras pueden cambiar au-

tomáticamente de modo a color durante el día a monocromático en la noche o cuando la iluminación cae por debajo de algún nivel determinado [4].

Por otro lado, el campo de la visión artificial ofrece un panorama ideal para complementar este tipo de sistemas mediante el análisis de las imágenes de video. El video digital, por su arquitectura, es ideal para ser evaluado por un sistema de visión por computador; de él se puede extraer la información necesaria para generar alarmas en el momento justo y analizar eventos eficientemente. Ésta es una de las razones que ha motivado el desarrollo de este trabajo que será descrito detalladamente en los capítulos siguientes.

BIBLIOGRAFÍA

- [1] <http://www.answers.com/topic/surveillance?cat=biz-fin>.
- [2] Superintendencia de vigilancia y seguridad privada. Ministerio de defensa de la República de Colombia. *Panorama Mundial de la Vigilancia y Seguridad Privada*.
- [3] Superintendencia de vigilancia y seguridad privada. Ministerio de defensa de la República de Colombia. *Dimensionamiento Nacional del Sector de la Vigilancia y Seguridad Privada e Impacto Económico y Social*.
- [4] H. Kruegle, *CCTV Surveillance: Video Practices and Technology*. Segunda Edición. Elsevier. USA. 2007. ISBN 0-7506-7768-6.

Capítulo 2

COMPONENTES DE UN SISTEMA DE VISIÓN ARTIFICIAL

La percepción es la relación entre una entrada visual y ciertos modelos previamente existentes en el mundo. Hay un gran espacio entre la imagen y dichos modelos (ideas, conceptos, etc.) los cuales explican, describen o resumen la información en la imagen. Para cubrir este espacio, los sistemas de visión artificial tienen una gran cantidad de elementos capaces de “conectar” la información visual con una salida, que puede ser una descripción final, una interpretación o la toma de una decisión [1]. La visión artificial se encarga entonces del diseño de herramientas e implementación de algoritmos capaces de adquirir, digitalizar y procesar la información de una imagen con el fin de extraer de ella determinadas propiedades de forma equivalente a la inspección visual humana.



Figura 2.1. Componentes de un sistema de visión artificial.

Para desarrollar un sistema que requiera el análisis de imágenes por medio de algún tipo de CPU (computadora, microcontrolador, DSP, etc.) se deben tener a disposición un amplio conjunto de elementos que van desde sistemas para la adquisición de la imagen hasta algoritmos para procesar la información obtenida de ella (Figura 2.1). Un sistema de visión artificial consta básicamente de los siguientes elementos:

2.1. ADQUISICIÓN DE IMÁGENES

La primera etapa del procesamiento de imágenes es la *adquisición de la imagen*, es decir, la adquisición de una imagen digital. Para ello hay que tener en cuenta ciertos parámetros que influyen en la calidad de la imagen adquirida. Las características de elementos como el sistema de iluminación, la cámara o sensor de imágenes utilizado y la tarjeta para la digitalización de la imagen son factores importantes para la obtención de una imagen de buena calidad. En esta sección se tratará cada una de estos elementos.

2.1.1. Imágenes Digitales y Video

Todo sistema de visión artificial tiene que procesar cierta información proveniente de una imagen. Para esto es necesario que esa información pueda ser comprendida y analizada por el computador. Por esta razón es importante comprender qué es una imagen o video y cómo está conformada.

2.1.1.1. Imágenes Digitales

El término *imagen* se refiere a una función bidimensional de intensidad de luz $f(x, y)$, donde x e y representan las coordenadas espaciales y el valor de f en un punto cualquiera (x, y) es proporcional al brillo (o *nivel de gris*) de la imagen en ese punto [2]. El nivel de gris de la imagen en cierto punto depende de la cantidad de luz que incide sobre la escena, así como también de la parte que sea reflejada por los objetos que componen dicha escena. Estos

componentes son llamados *iluminación* y *reflexión*, siendo descritos por $i(x, y)$ y $r(x, y)$ respectivamente. El producto de ambas funciones proporciona la función $f(x, y)$.

$$f(x, y) = i(x, y)r(x, y) \quad (2.1)$$

La naturaleza de la iluminación viene determinada por la fuente de luz, mientras que la reflexión depende de las características del objeto en la escena [3].

Una *imagen digital* es una imagen $f(x, y)$ que se ha discretizado tanto en las coordenadas espaciales como en el brillo o valor de nivel de gris. Una imagen digital puede considerarse como una matriz cuyos índices de fila y columna identifican un punto de la imagen y el valor del correspondiente elemento de la matriz indica el nivel de gris en ese punto (Figura 2.2). Los elementos de una distribución digital de este tipo se denominan *elementos de la imagen*, o más comúnmente *pixeles* [2].

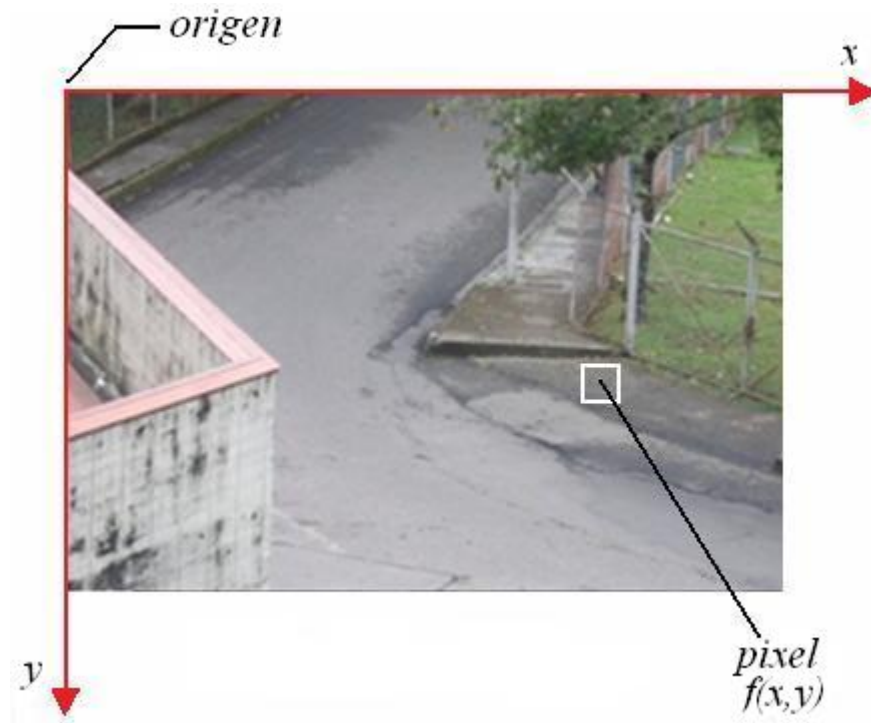


Figura 2.2. Convenio de ejes utilizado para la representación digital de imágenes.

2.1.1.2. Video

Es importante revisar algunos conceptos claves relacionados con el video con el fin de entender como éste juega un papel importante en lo que conocemos como video digital. La palabra video viene del verbo latín *videre*, “ver”, y comúnmente es usado para hacer referencia a dispositivos como monitores o grabadoras de video [4].

El video es una tecnología de captura electrónica, grabación, tratamiento, almacenamiento, transmisión, y reconstrucción de una secuencia de imágenes que representan escenas en movimiento. Cada una de las imágenes de la secuencia recibe el nombre de *frame*. El número de imágenes o *frames* por unidad de tiempo se conoce como *frame rate*, que puede variar entre seis a ocho imágenes por segundo (fps) para viejas cámaras mecánicas y 120 imágenes por segundo o más para las nuevas cámaras profesionales. Los estándares PAL (Europa, Asia, Australia, etc) y SECAM (Francia, Rusia, partes de África, etc) especifican 25 fps, mientras que NTSC (EE.UU., Canadá, Japón, etc) especifica 29,97 fps. El cine es más lento con una velocidad de 24 fps, lo que complica un poco el proceso de transferir una película de cine a video. Para lograr la ilusión de una imagen en movimiento, la velocidad mínima de carga de las imágenes debe ser de al menos quince imágenes por segundo [5].

“Nosotros vivimos en un mundo analógico, y la visión es una función analógica”¹. Una cámara analógica “observa” un evento que posteriormente se transforma en una señal electrónica, esta entonces es transmitida y desplegada en un monitor (Figura 2.3). En países como Estados Unidos, Canada, Japón, entre otros, la señal de video esta formada por 525 líneas horizontales, que se crean alternadamente, de acuerdo con el estándar NTSC. Un *frame* de video se compone de barridos de 262.5 líneas horizontales cada uno. La cámara barre, inicialmente, 262.5 líneas. Posteriormente la imagen es escaneada en las 262.5 líneas restantes. Cada mitad de *frame* o 262.5 se conoce como *campo de video*. Para el estándar NTSC hay 60 campos cada segundo en una secuencia de video. La Figura 2.4. muestra como el campo 1 barre primero las líneas impares (1, 3 , 5, etc.) y después son barridas las líneas pares (2, 4, 6, etc.). Los campos están *entrelazados*. El mismo proceso toma lugar en las cámaras con formato PAL, excepto que en éstas hay 50 campos por segundo, es decir, 25 fps [4].

Una señal eléctrica de video analógico representa *iluminancia*, *tono (hue)* y *saturación* [4].

¹Tomado de: *Digital CCTV. A Security Professional's Guide*. [4]



Figura 2.3. Transferencia de una señal de video analógica.

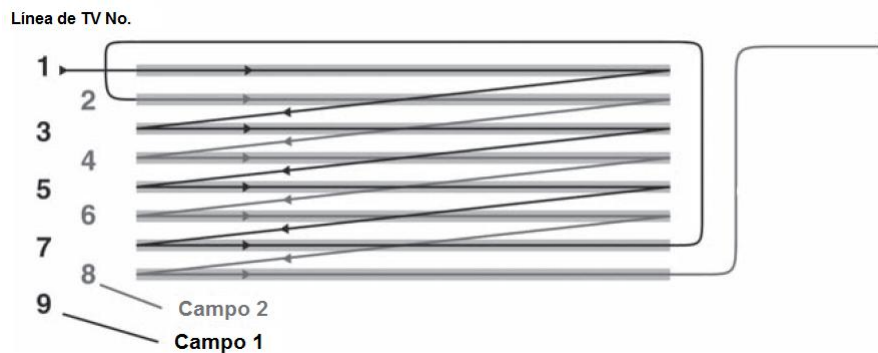


Figura 2.4. Campos entrelazados.

2.1.1.2.1. Iluminancia

La iluminancia es un índice representativo de la densidad del flujo luminoso sobre una superficie. Se define como la relación entre el flujo luminoso que incide sobre una superficie y el tamaño de esta superficie. A su vez la iluminancia no se encuentra vinculada a una superficie real, puede ser determinada en cualquier lugar del espacio. La iluminancia se puede deducir de la intensidad luminosa. Al mismo tiempo disminuye la iluminancia con el cuadrado de la distancia de la fuente de luz (Figura 2.5).

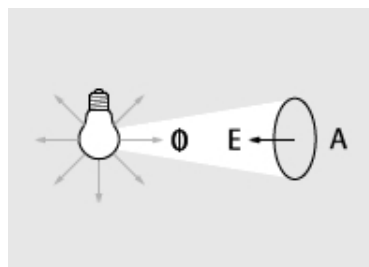


Figura 2.5. Iluminancia E como índice representativo del flujo luminoso que incide por cada unidad de superficie A .

Es el término que describe los valores de oscuridad y luz en una imagen. En otras palabras la iluminancia es la porción de la señal de video que lleva información para el brillo, la oscuridad y el contraste.

2.1.1.2.2. Tono (Hue)

Definimos tono como la propia cualidad que tiene un color. Tonos son todos los colores del círculo cromático, primarios, secundarios e intermedios. Podemos decir que cuando se va a la izquierda o a la derecha en el círculo cromático se produce un cambio de tono (Figura 2.6).

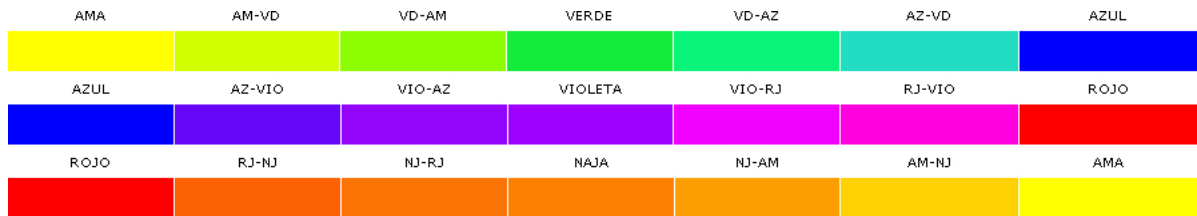


Figura 2.6. Los diferentes tonos que se ven no aparecen en círculo. Están colocados en tiras disponiendo los secundarios e intermedios correspondientes entre cada par de primarios.

Es el término que describe los valores de color en una imagen a color. A estos valores se les han asignado nombres, como por ejemplo, verde, amarillo, azul, etc. Los tonos son generalmente valores de frecuencia que recibe el ojo y el cerebro se encarga de percibir los colores a los que corresponden. Estos rangos de valores representan colores desde el azul, que es el valor mas bajo, hasta el rojo, que es el mas alto. En este espectro se incluyen también los colores verde, violeta, naranja y amarillo. El tono es el término utilizado para determinar de qué color es un objeto.

2.1.1.2.3. Saturación

Esta propiedad describe que tan intenso es el color o el valor que obtenemos del tono. Para un color muy intenso podemos decir que, este, está bastante saturado. Una imagen a color adhiere los términos *tono* y *saturación* para complementar la información de la iluminancia.

Cuando un color pertenece al círculo cromático se dice que está saturado, que tiene el máximo poder de pigmentación, de coloración. Pero no siempre nos encontramos los colores puros,

sino que se suelen ver compuestos por mezclas complejas, con cantidades desiguales de colores primarios. Para cambiar la saturación de un color hay que mezclarlo con su complementario y, así, se obtiene la escala de saturación o de grises (Figura 2.7).



Figura 2.7. Escala de colores con diferente saturación.

2.1.1.3. Video Digital

La información de la señal analógica ahora es “traducida” al lenguaje digital. En efecto, uno de los mejores caminos para entender las tecnologías analógica y digital, es considerarlas como dos lenguajes diferentes.

Hay un gran número de formas en las cuales el video se puede representar digitalmente. Una forma es utilizando Modulación por Codificación de Pulsos (*Pulse Code Modulation PCM*), en la cual una forma de onda analógica es muestreada en intervalos regulares de tiempo. En la tecnología digital, la onda analógica es muestreada y convertida en una serie de números que son almacenados en un dispositivo digital. Este proceso es llamado *muestreo*. La frecuencia a la cual se toman los datos de la señal se conoce como *frecuencia de muestreo*, la cual, según el teorema de Nyquist, debe ser al menos el doble de la frecuencia de la señal muestreada.

El resultado de muestrear una señal de video es el *video digital*. Hay muchas formas de lograr el muestreo de una señal de video analógica. El estándar que ha surgido para la toma de muestras para video digital es el ITU-R BT.601, mas comúnmente conocido como CCIR 601. ITU, establecido por la Unión Internacional de Comunicaciones.

CCIR 601 es basada en múltiplos de una frecuencia fundamental de muestreo de 3,375 MHz. Esta frecuencia fue cuidadosamente escogida debido a que está fijada tanto para NTSC como para PAL. La señal de iluminancia es muestreada a 13,5 MHz, cuatro veces la frecuencia

fundamental de muestreo. Cada una de las diferentes señales de color es muestreada a 6,75 MHz, dos veces la frecuencia fundamental de muestreo. La conversión completa, cuando cada número está representado por un número discreto, se presenta en un proceso llamado *cuantización*.

El proceso de *cuantización* convierte los números muestreados a *bits*. Este proceso ocurre generalmente después del muestreo, aproximando el grupo valores continuos de la imagen a un grupo finito de valores preferiblemente pequeños. La entrada de un cuantizador es la imagen original, y la salida es uno entre un finito grupo de valores. En otras palabras un buen cuantizador representa la señal original con una mínima pérdida o distorsión. Hay dos tipos de cuantización: Cuantización escalar y cuantización vectorial. En la cuantización escalar cada símbolo de entrada es tratado individualmente cuando se produce la salida. En la cuantización vectorial, los símbolos de entrada son organizados en vectores y procesados para producir una salida [4].

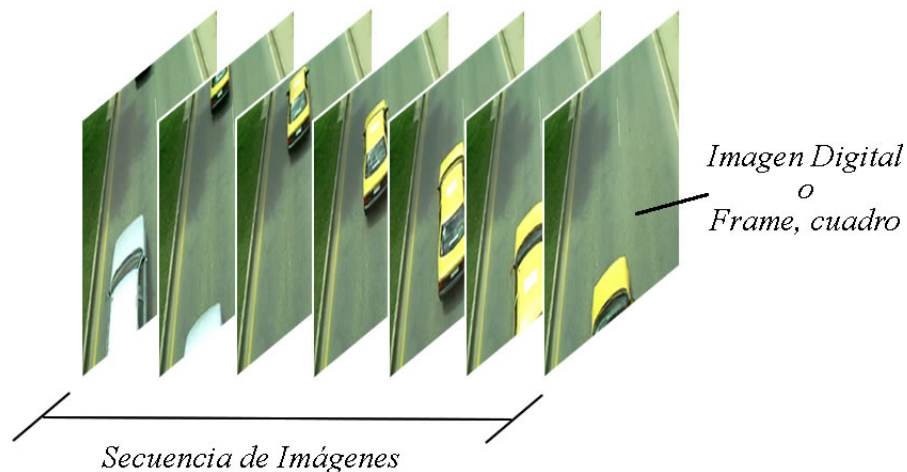


Figura 2.8. Video digital.

2.1.2. Iluminación

La iluminación juega un papel vital en la visión artificial, pues de un buen sistema de iluminación depende también la calidad de la imagen captada. Cuando se realiza una iluminación adecuada sobre el objeto o la escena de interés las etapas posteriores a la adquisición de la imagen resultan más cómodas, pues se simplifica de manera considerable el análisis y

posterior interpretación de la escena. La iluminación suele afectar a la complejidad de los algoritmos de visión; por otro lado siempre será más rápido solucionar un problema cambiando la iluminación que modificando el algoritmo [3].

El objetivo principal de la iluminación es resaltar las características de interés en el objeto que se quiere estudiar. Por esta razón el sistema de iluminación debe ser escogido según la aplicación en particular.

2.1.2.1. Tipos de Iluminación

Los diversos tipos de iluminación son [2]:

2.1.2.1.1. Iluminación Direccional

Este tipo de iluminación se utiliza con mayor frecuencia en aplicaciones como localización y reconocimiento de piezas, o inspección de la superficie de objetos. Consiste en aplicar una iluminación orientada al objeto usando un haz altamente direccional (Figura 2.9). Este tipo de iluminación es de fácil uso. La forma del haz luminoso depende de la aplicación y del objeto a analizar.

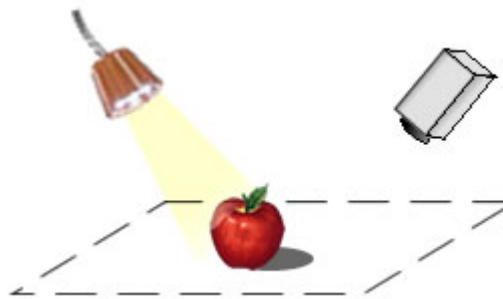


Figura 2.9. Iluminación direccional.

2.1.2.1.2. Iluminación Difusa

Con este tipo de iluminación se busca que los haces de luz incidan sobre el objeto desde todas las direcciones (Figura 2.10). Generalmente este tipo de iluminación es utilizada cuando se necesitan imágenes desde varios puntos de vista, por lo que no se pueden favorecer unas zonas mas que otras.

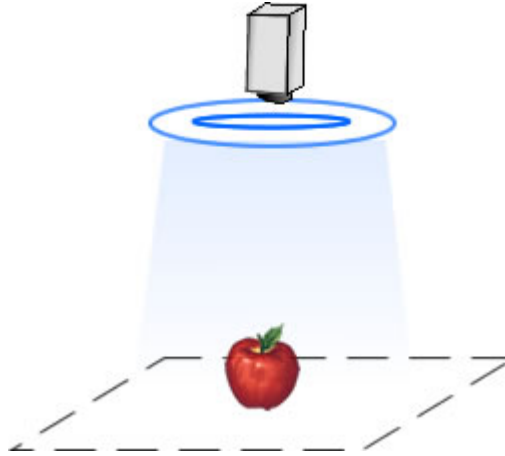


Figura 2.10. Iluminación difusa.

2.1.2.1.3. Iluminación a Contraluz

Este tipo de iluminación produce imágenes prácticamente con sólo dos niveles de gris. Este sistema de iluminación consiste en iluminar el objeto por detrás de forma que la fuente luminosa, el objeto y la cámara estén alineados (Figura 2.11). Esta técnica es ampliamente utilizada en aplicaciones en las cuales la silueta del objeto es suficiente para su reconocimiento.

2.1.2.2. Fuentes de Luz

Para obtener una imagen de calidad que permita realizar un procesamiento más fácil, además de la forma en la cual se va a iluminar, hay que escoger una fuente de luz idónea para lograr dicho fin.

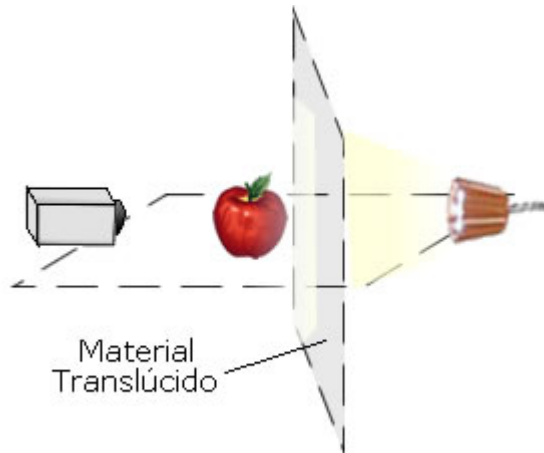


Figura 2.11. Iluminación posterior o a contraluz.

A continuación se presentan las fuentes de luz más comúnmente utilizadas:

2.1.2.2.1. Fuentes Incandescentes

Las luces incandescentes (Figura 2.12) presentan un coste bajo, son fáciles de utilizar y además pueden funcionar durante periodos largos de tiempo. Una ventaja importante de este tipo de fuentes es que se puede variar su intensidad lumínica simplemente variando la intensidad de la corriente aplicada a la lámpara.

El principal inconveniente que presentan las fuentes incandescentes es que emiten una gran cantidad de calor, lo cual limita su campo de aplicación. Otro inconveniente es el paso por cero de la corriente alterna (60 o 50 Hz dependiendo de la ubicación geográfica) aplicada a la lámpara, ya que se requiere una perfecta sincronización con la cámara, la cual funciona también a la misma frecuencia [3].

2.1.2.2.2. Fuentes Fluorescentes

Estas son fuentes muy eficientes que no generan calor. También se pueden encontrar en diversas formas y tamaños, así como en diversos colores. Generalmente se emplean para iluminación difusa. El principal inconveniente es la disminución del nivel de intensidad con el paso del tiempo. Una de las aplicaciones más comunes de las fuentes fluorescentes es la iluminación de objetos con superficies altamente reflectoras (Figura 2.13) [3].



Figura 2.12. Fuentes de luz incandescentes.



Figura 2.13. Fuentes de luz Fluorescente.

2.1.2.2.3. Diodos LED

Los diodos LED (Figura 2.14) proporcionan una luz monocromática disponible en una gran variedad de colores. Este tipo de iluminación puede durar años funcionando y su coste es el mas bajo de todas las fuentes de luz. Uno de los principales inconvenientes es que alcanza una intensidad de iluminación bastante pequeña, por lo que generalmente hay que utilizar varios. Este tipo de iluminación es ampliamente utilizado en iluminación a contraluz, e iluminación difusa [3].



Figura 2.14. Iluminación con LED.

2.1.2.2.4. Luz Estroboscópica

Son ideales cuando hay que analizar objetos en movimiento. Brindan una iluminación intensa, por lo que la iluminación del ambiente influye poco. Estas fuentes presentan algunos inconvenientes: necesitan una fuente de tensión especial, hay que sincronizarlos con las cámaras y su intensidad va decreciendo con el tiempo [3](Figura 2.15).

2.1.2.2.5. Láser

El láser es una fuente de luz utilizada en iluminación estructurada (Figura 2.16), que permite entre muchas otras aplicaciones, tomar medidas tridimensionales. El uso del láser concede la posibilidad de generar diversos patrones de luz como puntos, líneas muy finas o rejillas [3].

2.1.2.2.6. Fibra Óptica

El uso de la fibra óptica como guía de luz está asociado a la no transmisión del calor en el proceso de iluminación. Es por esto que su utilización va asociada casi siempre a las fuentes



Figura 2.15. Fuentes de luz estroboscópica.

incandescentes. Su uso se centra principalmente en la iluminación de pequeñas áreas y cavidades y en aplicaciones que presentan un espacio constante para la adquisición de imágenes (Figura 2.17) [3].

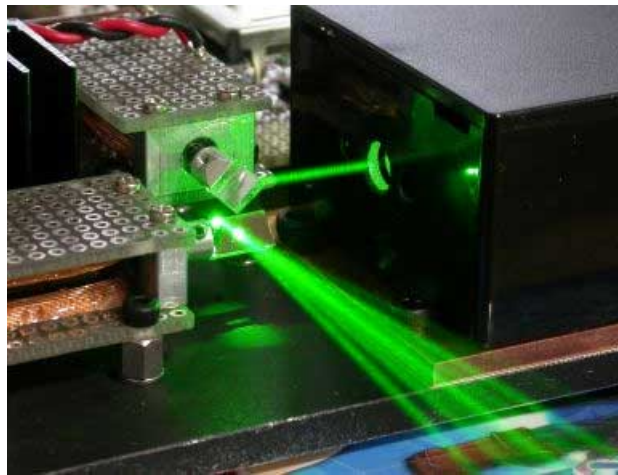


Figura 2.16. Haz de luz láser de longitud de onda de 532 nm.

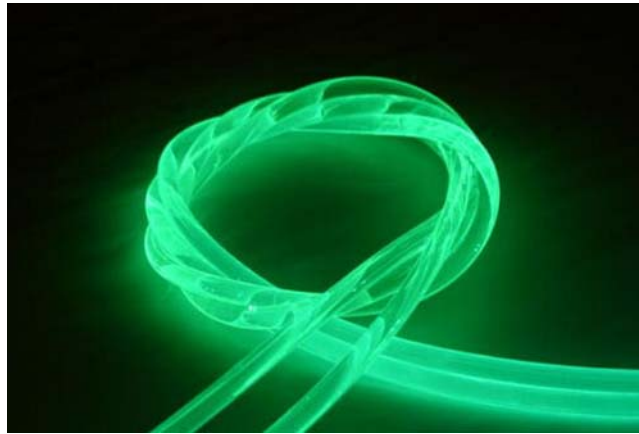


Figura 2.17. Fibra Óptica como medio de propagación de la luz.

2.1.3. Cámaras Digitales

Las cámaras o sensores de imágenes (Figura 2.18) son los elementos encargados de captar la información luminosa de la escena, y transmitirla al computador como una señal analógica o digital. Aunque inicialmente se utilizaron las cámaras de tubo (el *vidicom* por ejemplo), en la actualidad son las denominadas *cámaras de estado sólido* las que se utilizan en la inmensa mayoría de las aplicaciones [3].



Figura 2.18. Cámara digital y dispositivo CCD.

Hay dos tipos de cámaras digitales: las cámaras CCD (*Charged Coupled Device*), y las más recientes, las cámaras CMOS (*Complementary Metal Oxide Silicon*). A continuación se analizan cada una de ellas.

2.1.3.1. Dispositivos CCD (*dispositivos de carga acoplada*)

La lente de la cámara enfoca la luz desde un punto de la escena hacia cada pixel, la luz incidente sobre cada uno de estos genera un “paquete” de carga electrónica, cuya intensidad es proporcional a la luz incidente. Cada fila de pixeles representa una línea horizontal de la información del video. Si el patrón de radiación incidente es la luz de la imagen llevada por el sistema óptico, entonces los paquetes de carga creados en el arreglo de pixeles son una fiel reproducción de esa imagen [6].

La cámara CCD (*charge-coupled-device*) es un sensor de imagen que utiliza elementos semiconductores fotosensibles en forma de arreglos matriciales. Los receptores activos de este arreglo son distribuidos en pixeles individuales. En un sensor CCD se transforma la energía de la luz incidente en energía eléctrica. La carga eléctrica almacenada en la celda es posteriormente transportada utilizando un registro de corrimiento (*shift register*) para conformar una señal de video. Cabe señalar que en las cámaras CCD se discretiza la imagen en pixeles, sin embargo el valor de la carga eléctrica almacenado en cada celda no se digitaliza en el arreglo CCD sino en una conversión posterior realizada por un conversor análogo-digital.

Las cámaras CCD son consideradas como dispositivos que poseen una muy baja deformación geométrica de la imagen, una muy buena sensibilidad a la luz y una muy buena resolución llegando típicamente a los 400.000 pixeles. El tiempo de captura de la imagen está típicamente en el rango de 1/60 s y 1/10000 s [7].

Existen diferentes arquitecturas para los sensores CCD usadas por diferentes fabricantes. Las dos más comunes son la de *transferencia interlínea* (*interline transfer*) y la de *transferencia de cuadro* (*frame transfer*).

2.1.3.1.1. Dispositivos CCD de transferencia interlínea

La organización de los pixeles tiene alineados los fotosensores con los registros de corrimiento verticales, y un registro horizontal enlazado con los registros verticales como se muestra en la figura 2.19. El fotosensor responde a las variaciones de luz generando cargas eléctricas proporcionales a la intensidad de la luz. Las cargas son transferidas a los registros de corrimiento verticales y después a los registros horizontales sucesivamente hasta que llegan al amplificador de salida del sensor. La electrónica de la cámara realiza la amplificación y el

procesamiento de la señal. Cada pixel y línea en los dispositivos de transferencia interlínea son transferidos fuera del arreglo de sensores línea por línea hasta transferir las 525 líneas, es decir, la imagen completa [6].

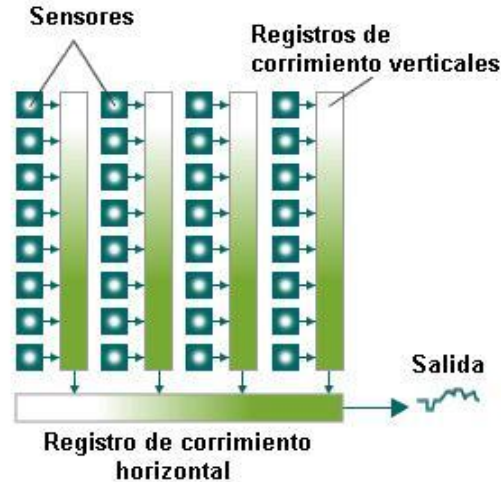


Figura 2.19. Sensor CCD con arquitectura de transferencia interlínea.

2.1.3.1.2. Dispositivos CCD de transferencia de cuadro

En esta arquitectura las 525 líneas son transferidas completamente desde el arreglo de sensores hacia un *buffer* de almacenamiento adyacente (Figura 2.20).

La estructura básica de esta arquitectura consta de dos elementos: el arreglo de sensores y una sección de memoria. Primero el arreglo de sensores es expuesto a la luz. Después la carga producida es rápidamente transferida a la sección de memoria. Mientras la memoria está siendo leída, el arreglo de sensores es expuesto para la siguiente imagen [6].

2.1.3.2. Dispositivos CMOS

En el pasado los sensores de imagen CMOS estaban relegados a aplicaciones de baja resolución, pero ahora tienen una gran cantidad de píxeles, lo que ha permitido utilizarlos en

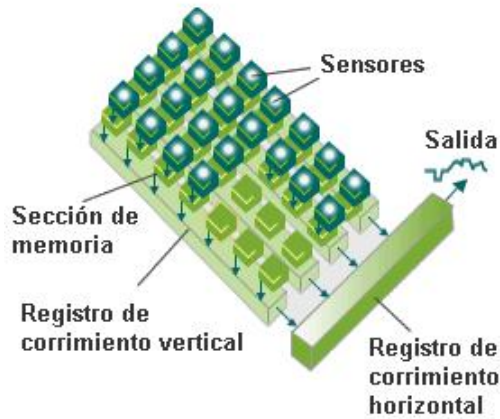


Figura 2.20. Sensor CCD con arquitectura de transferencia de cuadro.

aplicaciones más complejas.

Los sensores del tipo CMOS brinda una imagen de alta calidad, pero tienen una sensibilidad más baja que los sensores CCD. En los dispositivos CMOS, las señales eléctricas son transferidas directamente a través de un arreglo de transistores.

Los sensores CMOS han tenido gran acogida y tienen la ventaja de incorporar el convertidor análogo-digital, circuitos de tiempo, relojes, y circuitos de sincronización, en el mismo chip. Las señales digitales de los sensores CMOS son siempre transmitidas, sin ser almacenadas como en el sensor CCD, por la que no necesitan un DSP [6].

2.1.4. Tarjetas de Adquisición y Procesamiento de Imágenes

Una vez que la imagen ha sido recogida por la cámara de video, la tarjeta de adquisición y procesamiento de imágenes recibe la señal analógica enviada por ella, para convertirla en una señal digital así como para realizar el procesamiento de la imagen. La cantidad de información que proporcionan las cámaras es muy alta, y para poder procesarla correctamente existen básicamente dos posibilidades: conseguir un microprocesador más rápido o bien, poner varios procesadores más lentos en paralelo. Ninguno de los dos enfoques es claramente superior al otro, ya que depende de la tecnología existente en cada instante el que una opción sea más rápida que otra.

Existen tres tipos de tarjetas:

- * Digitalizadores de imágenes o *frame grabbers*.
- * Digitalizadores de imágenes con procesador específico.
- * Tarjetas de procesamiento paralelo.

Los digitalizadores de imágenes muestrean la señal de la cámara y guardan la imagen digital en una zona de memoria, que pueda ser accedida por el microprocesador del computador para su tratamiento. Básicamente están constituidos por los puertos de entrada de la señal de video, y el conversor analógico-digital para obtener la imagen. Como puede verse es la elección más simple. Para aplicaciones en las que se necesitaba una gran rapidez no eran la opción adecuada por la poca velocidad de transferencia entre la memoria de la tarjeta y el poco rendimiento de los microprocesadores. Esto ha cambiado últimamente con el bus PCI y el microprocesador Pentium o el bus VME y las estaciones de trabajo.

Si se necesita mayor potencia de cálculo hay que trabajar al menos con dos microprocesadores. La manera más sencilla de perfeccionar el modelo anterior es dotar a la digitalizadora de imágenes con su propio microprocesador, para que realice una de las partes más complicadas y libere de trabajo al computador. Suele ser generalmente un DSP (*Digital Signal Processor*).

Por último, si el algoritmo exigiese más poder de cálculo, habrá que acudir a las tarjetas de procesamiento paralelo. Aquí existen dos opciones: repartir el algoritmo en varias partes independientes, cada una de las cuales se realizará en un módulo distinto, o dividir las imágenes en varios trozos, cada uno de los cuales será tratado por un módulo [3].

2.2. PREPROCESAMIENTO DE IMÁGENES

El preprocesamiento de imágenes busca modificar y preparar los píxeles de una imagen digital para darle una forma mas adecuada para posteriores operaciones. Hay dos grandes ramas del preprocesamiento de imágenes: el mejoramiento de imágenes y la restauración de imágenes.

El mejoramiento de imágenes trata de mejorar la calidad de la imagen o enfatizar aspectos particulares dentro de la imagen. Este objetivo implica por lo general un grado de juicio subjetivo acerca de la calidad y el resultado dependerá de la operación y de la aplicación en cuestión. Los resultados pueden producir una imagen bastante diferente de la original y algunos aspectos pueden ser deliberadamente sacrificados con el fin de mejorar otros.

El objetivo de la restauración de imágenes es recuperar la imagen original después de que ésta ha sido degradada por factores conocidos como una distorsión geométrica dentro del sistema de una cámara.

Ambos tipos de operación toman la imagen adquirida como una entrada y produce una imagen modificada como salida, y así estas operaciones representan procesamiento de imagen puro. Muchas de las operaciones más comunes están relacionadas con la aplicación en filtros lineales sobre la imagen original. Un ejemplo obvio es la eliminación del ruido de la imagen [8].

2.2.1. Eliminación de Ruido

Todas las imágenes tienen una cierta cantidad de ruido, valores distorsionados, bien debidos al sensor CCD de la cámara o al medio de transmisión de la señal. El ruido se manifestará generalmente en pixeles aislados que toman un nivel de gris diferente al de sus vecinos. El ruido puede clasificarse en cuatro tipos [3]:

2.2.1.0.1. Gaussiano

Produce pequeñas variaciones en la imagen. Es debido, principalmente, a las diferentes ganancias en el sensor, ruido en los digitalizadores y perturbaciones en la transmisión de la señal. Se considera siempre que el valor final del pixel es el ideal mas una cantidad correspondiente al error, que puede describirse como una variable gaussiana.

2.2.1.0.2. Impulsional (*sal y pimienta*)

El valor que toma el pixel no tiene relación con el valor ideal sino con el valor del ruido que toma valores muy altos o bajos. Se caracteriza entonces porque el pixel toma un valor máximo, causado por la saturación del sensor, o mínimo, si se ha perdido su señal.

2.2.1.0.3. Frecuencial

La imagen obtenida es la suma entre la imagen ideal y otra señal, la interferencia, caracterizada por ser una función sinusoidal con frecuencia determinada.

2.2.1.0.4. Multiplicativo

La imagen obtenida es el fruto de la multiplicación de dos señales, por ejemplo, en una escena donde la iluminación no es constante, algunas zonas pueden quedar más oscuras que otras.

2.2.1.1. Filtrado Espacial

El empleo de máscaras espaciales para el procesamiento de las imágenes se denomina frecuentemente *filtrado espacial*, y las propias máscaras se denominan *filtros espaciales*.

Los filtros denominados de *paso bajo* atenúan o eliminan las componentes de alta frecuencia a la vez que dejan inalteradas las bajas frecuencias. Los componentes de alta frecuencia caracterizan los bordes y los restantes detalles muy marcados de la imagen, y por ello el efecto de introducir un filtrado de paso bajo es el de hacer mas borrosa la imagen. De forma similar, los filtros de *paso alto* atenúan o eliminan las componentes de baja frecuencia. Como estas componentes son responsables de las características lentamente variables de la imagen, como el contraste global o la intensidad media, el resultado neto de un filtrado paso alto es la reducción de estas características y, en correspondencia, una aparente agudización de los bordes y de los detalles finos restantes [2].

2.2.1.1.1. Filtrado espacial de paso bajo

La forma de la respuesta de un impulso necesaria para implementar un filtro espacial de paso bajo indica que el filtro ha de tener todos sus coeficientes positivos. Aunque la forma del filtro espacial de la figura 2.21 (a) pueda ser descrita, por ejemplo, por una función gaussiana predeterminada, el requisito clave es que todos los coeficientes sean positivos. Para un filtro espacial de 3x3, la construcción más simple consistiría en una máscara en la que todos los coeficientes fuesen iguales a 1 [2]. Si al resultado se le multiplica por 1/9 se tiene la media de todos los píxeles, por lo que el ruido disminuye.

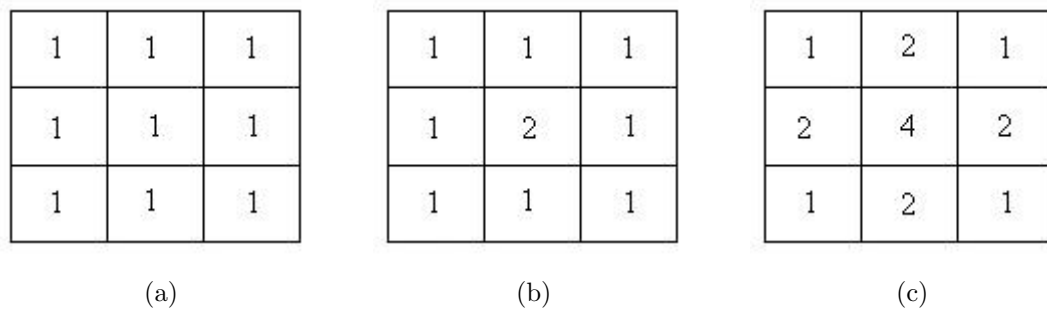


Figura 2.21. Filtros espaciales de paso bajo

Sin embargo este filtro presupone que la influencia de todos los píxeles es igual. Otra consideración es que, cuanto más alejado esté el píxel del central, su valor será menor, se tiene entonces la máscara mostrada en la figura 2.21 (b). Si se le vuelve a querer dar más importancia al píxel central que a sus vecinos se puede utilizar la máscara de la figura 2.21 (c) [3].

Los resultados de esta operación de filtrado se muestran en la figura 2.22.

2.2.1.1.2. Filtrado espacial de paso alto

El perfil de la respuesta a un impulso necesaria para implementar un filtro espacial de paso alto indica que el filtro debe tener coeficientes positivos cerca de su centro y coeficientes negativos en la periferia. Para una máscara de 3x3 esta condición se cumple escogiendo un valor positivo en el centro y tomando coeficientes negativos en el resto.

La figura 2.23 muestra la implementación más clásica de un filtro de realce 3x3. Obsérvese que la suma de los coeficientes es igual a 0, así cuando la máscara esta sobre un área de nivel



Figura 2.22. Filtrado espacial de paso bajo: (a) Imagen original. (b) Resultado de la aplicación del filtro de paso bajo mostrado en la figura 2.21 (a).

-1	-1	-1
-1	8	-1
-1	-1	-1

Figura 2.23. Filtro espacial de paso alto.

de gris constante o lentamente variable, la salida proporcionada por la máscara es cero o un valor muy pequeño (figura 2.24) [2].

2.2.1.1.3. Suma de imágenes

Si se supone que el ruido es una señal que se añade a la original, el nivel de gris de un pixel puede definirse como la suma entre el nivel de gris ideal y el ruido:

$$f(x, y) = f_i(x, y) + r(x, y) \quad (2.2)$$



Figura 2.24. Filtrado espacial de paso alto: (a) Imagen original. (b) Resultado de la aplicación del filtro de paso alto mostrado en la figura 2.23.

Aunque el ruido esté siempre presente, el que afecte más o menos a un pixel determinado es aleatorio. Si se tratase de ruido gaussiano vendría definido por una distribución normal de media cero y desviación típica σ . Por ello, en el dominio temporal, la manera más sencilla es simplemente la media de varias imágenes (*filtro paso bajo temporal*); pues al suponer que el ruido tiene una distribución espacial aleatoria la influencia es menor cuantas más imágenes se tengan, ya que el valor de un pixel será:

$$f(x, y) = f_i(x, y) + \frac{r(x, y)}{n} \quad (2.3)$$

Siendo n el número de imágenes tomadas. Ahora el ruido de la imagen sigue una distribución normal de la forma [2]:

$$N\left(0, \frac{\sigma}{\sqrt{n}}\right) \quad (2.4)$$

2.2.1.1.4. Sustracción de imágenes

La diferencia entre dos imágenes $f(x, y)$ y $g(x, y)$, expresada de la forma

$$g(x, y) = f(x, y) - h(x, y) \quad (2.5)$$

se obtiene calculando la diferencia entre todos los pares de píxeles correspondientes de f y h . Esta técnica es frecuentemente empleada para detectar objeto que no pertenecen a la imagen original. La sustracción de imágenes también es ideal para la detección de movimiento en una secuencia de video (Figura 2.25).



(c)

Figura 2.25. Sustracción de imágenes: (a) Imagen original f . (b) Imagen h . (c) Diferencia entre las imágenes f y h

2.3. SEGMENTACIÓN DE IMÁGENES

La segmentación se encarga de dividir o separar la imagen en regiones con atributos similares. El atributo más básico para la segmentación es el nivel de gris para una imagen monocroma o las componentes de color para una imagen a color. Los bordes y las texturas también son atributos útiles para la segmentación [9].

La segmentación se basa en tres propiedades:

- * Similitud. Cada uno de los píxeles de un elemento tiene valores parecidos para alguna propiedad.
- * Discontinuidad. Los objetos destacan del entorno y tienen por tanto unos bordes definidos.
- * Conectividad. Los píxeles pertenecientes al mismo objeto tienen que ser contiguos, es decir, deben estar agrupados.

Estas suposiciones no son fáciles de cumplir y sin embargo son fundamentales para la segmentación. Respecto a la primera condición, la similitud, los objetos deberían presentar una apariencia uniforme. Esto no va a ser posible debido a que la iluminación no es constante existiendo diferentes variaciones en el material, la diferencia entre las ganancias para cada píxel de la cámara, la existencia de brillos o la presencia de ruido en la imagen. Respecto a la segunda condición los bordes no van a estar siempre bien definidos. Por último se puede dar el caso de ocultamientos parciales de un objeto por parte de otro lo que dificultará el cumplimiento de la conectividad de todos los puntos de los objetos.

Las técnicas de segmentación se basan por tanto en la búsqueda de las partes uniformes de la imagen o justo lo contrario, aquellas partes en donde se produce un cambio. Dependerá del caso concreto el que se siga un método u otro o la unión de ambos. Después de haber detectado los puntos que presentan esta discontinuidad (bordes) habrá que encontrar un camino entre el píxel P_1 y el píxel P_N . Un *camino* es la secuencia de puntos P_2, P_3, \dots, P_{N-1} , donde el píxel P_{i+1} es vecino del píxel P_i . Se dirá que es un camino sencillo si ningún punto aparece más de una vez y ningún píxel tiene más de dos vecinos. Un camino es un camino sencillo en el que el primer punto es vecino del último. Una *región* es un conjunto de píxeles en los que existe un camino entre cualquier pareja de sus puntos, y todos los píxeles del camino pertenecen a la región [3].

2.3.1. Umbralización

Con la umbralización se convierte una imagen con varios niveles de gris a una nueva con sólo dos, de manera que los objetos quedan separados del fondo. La Umbralización se basa en que los pixeles de un determinado objeto tienen el mismo nivel de gris. Como esto no será del todo cierto se buscan intervalos de gris, cada uno de ellos pertenecientes a un objeto. Esta técnica será útil si los objetos tienen una superficie parecida y el fondo es uniforme. Se aplicarán por lo tanto las condiciones siguientes:

Sea T el umbral definido para la segmentación, y $f(x, y)$ el nivel de gris del objeto en un punto determinado, la imagen umbralizada será:

$$g(x, y) = \begin{cases} 1 & \text{si } T \leq f(x, y) \\ 0 & \text{en cualquier otro caso} \end{cases} \quad (2.6)$$

cuando el valor buscado sea mayor que uno dado, o de la forma:

$$g(x, y) = \begin{cases} 1 & \text{si } T \geq f(x, y) \\ 0 & \text{en cualquier otro caso} \end{cases} \quad (2.7)$$

si es menor que el umbral, y de la siguiente manera

$$g(x, y) = \begin{cases} 1 & \text{si } T_a \leq f(x, y) \leq T_b \\ 0 & \text{en cualquier otro caso} \end{cases} \quad (2.8)$$

si el nivel de gris de los objetos se encuentra en un intervalo.

2.3.2. Etiquetado

Con la umbralización los objetos han sido separados del fondo. Sin embargo no han sido completamente distinguidos entre sí (todos tienen el mismo nivel de gris alto); es por esto que la siguiente etapa es el *etiquetado* (*labelling*), o lo que es lo mismo, a cada objeto se le asignará un nivel de gris distinto. Para ello, y empezando por el pixel superior izquierdo, se recorre la imagen hacia la derecha. Cuando se encuentra el primer pixel que esté a nivel alto se le asocia la *etiqueta* (*label*) 1, para identificarlo como el primer objeto. Se examinan a continuación sus vecinos para ver si también están a nivel alto, si lo están recibirán la misma etiqueta. Cuando un pixel no sea vecino de uno etiquetado pero esté a nivel alto se le asocia la siguiente etiqueta, 2 y así sucesivamente. Con un solo recorrido de la imagen se pueden producir indeterminaciones. La solución está en llevar una tabla con etiquetas que correspondan al mismo objeto. Una vez terminada esta segunda pasada sobre la imagen se deshacen las ambigüedades asignando a todos los puntos la misma etiqueta [3].

2.4. OPERACIONES MORFOLÓGICAS Y DESCRIPCIÓN DE OBJETOS

La palabra *morfología* indica normalmente una rama de la biología que trata de la forma y la estructura de animales y plantas. Cuando se habla de imágenes, se emplea esta palabra en el contexto de *morfología matemática* como una herramienta para extraer componentes de una imagen que sean útiles en la representación y descripción de una región, tales como contornos, esqueletos, etc. También son de interés las técnicas morfológicas para el pre y postprocesado, tales como el filtrado morfológico, la reducción y el recortado.

Las transformaciones morfológicas son aquellas que modifican la estructura o forma de los objetos presentes en una imagen. Estas herramientas, además de ser útiles para la extracción de características, permiten la eliminación de ruido que se produce en todo proceso de segmentación [3].

La morfología matemática se basa en la teoría de conjuntos y en la topología. A continuación se mostrarán algunas definiciones básicas de este aspecto.

- * *Elemento estructural*: Los elementos estructurales son un conjunto de puntos que servirán para determinar la estructura de un conjunto X . Uno de ellos constituirá el centro del elemento.
- * Sean A y B conjuntos de Z^2 , con componentes $a = (a_1, a_2)$ y $b = (b_1, b_2)$ respectivamente. La traslación de A por $x = (x_1, x_2)$, representada $(A)_x$, se define como:

$$(A)_x = \{c \mid c = a + x, \text{ para } a \in A\} \quad (2.9)$$

- * La reflexión de B , representada por \hat{B} , se define como:

$$\hat{B} = \{x \mid c = -b, \text{ para } b \in B\} \quad (2.10)$$

2.4.1. Dilatación

El proceso de dilatación consiste en obtener la reflexión de B sobre su origen y después cambiar esta reflexión por x . La dilatación de A por B es entonces el conjunto de todos los desplazamientos x tales que \hat{B} y A se solapen en al menos un elemento distinto de cero [2]. Esta operación se representa mediante:

$$A \oplus B = \{x \mid [(\hat{B})_x \cap A] \subseteq A\} \quad (2.11)$$

2.4.2. Erosión

El efecto de la erosión sobre la imagen es contraerla. Técnicamente la erosión de A por B es el conjunto de todos los puntos x tales que B , trasladado por x , está contenido en A . Para los conjuntos A y B de Z^2 , la erosión se define como [2]:

$$A \ominus B = \{x \mid [(B)_x \cap A] \subseteq A\} \quad (2.12)$$

2.4.3. Apertura y Cierre

La *apertura* generalmente suaviza el contorno de una imagen, rompe istmos estrechos y elimina protuberancias delgadas. El *cierre* también tiende a suavizar secciones de contornos pero, en oposición a la apertura, generalmente fusiona separaciones estrechas y entrantes delgados y profundos, elimina pequeños huecos y rellena agujeros del contorno.

2.4.3.1. Apertura

Es simplemente la erosión de A por B , seguida de una dilatación del resultado por B . Está definida como [2]:

$$A \circ B = (A \ominus B) \oplus B \quad (2.13)$$

2.4.3.2. Cierre

El cierre del conjunto A por el elemento de estructura B , representado por $A \bullet B$, se define como [2]:

$$A \bullet B = (A \oplus B) \ominus B \quad (2.14)$$

En la Figura 2.26 se muestra un ejemplo de las operaciones morfológicas mencionadas.

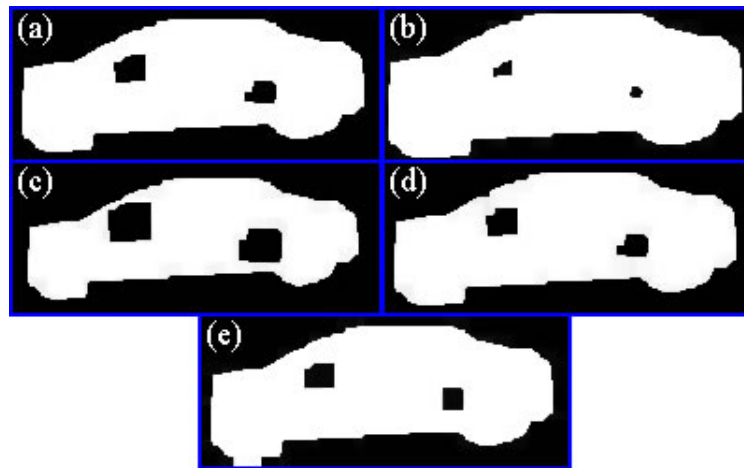


Figura 2.26. Operaciones morfológicas. (a) Imagen binaria original. (b) Dilatación. (c) Erosión. (d) Apertura. (e) Cierre

2.4.4. Esqueletización

Una importante aproximación para representar la forma estructural de una región plana es reducirla a un grafo. En esta reducción se puede conseguir el esqueleto de la región mediante un algoritmo de reducción, denominado también *esqueletización*. El esqueleto de un conjunto A , se puede expresar en función de erosiones y aperturas. Esto es, si $S(A)$ representa el esqueleto de A , se puede mostrar que:

$$S(A) = \bigcup_{k=0}^K S_k(A) \quad (2.15)$$

con

$$S_k(A) = \bigcup_{k=0}^K \{(A \ominus kB) - [(A \ominus kB) \circ B]\} \quad (2.16)$$

donde B es un elemento de estructura, $(A \ominus kB)$ indica k erosiones sucesivas de A ; esto es:

$$(A \ominus kB) = ((\dots(A \ominus B) \ominus B) \ominus \dots) \ominus B \quad (2.17)$$

k veces, y K es el último paso iterativo antes de que A se erosiones hasta un conjunto vacío [2].

2.4.5. Características de Regiones

Una vez que los objetos están separados y se han corregido los errores cometidos en la segmentación, se está en disposición de obtener una serie de valores numéricos que describan los objetos. Este grupo de características ayudará en la última etapa, el reconocimiento de los objetos. En ella se clasificará el objeto detectado en la imagen entre los posibles casos que pueden aparecer en la misma [3].

2.4.5.1. Tamaño, perímetro, compactibilidad

El *tamaño* de un objeto estará dado por la suma de sus píxeles:

$$A = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} g(i, j) \quad (2.18)$$

siendo $g(x, y)$ la imagen binaria.

El *perímetro* P lo constituirá el número de píxeles exteriores al objeto y la compactibilidad viene dada por la fórmula matemática:

$$C = \frac{P^2}{A} \quad (2.19)$$

Fórmula cuyo valor mínimo corresponde al círculo con un valor de 4π .

Una variación es la *rectangularidad*, que se define como el cociente entre el área del objeto y la del rectángulo más pequeño que lo engloba (*bounding box*) [3].

2.4.5.2. Posición

El conocimiento de la posición de un objeto en una imagen es importante ya que en la mayoría de las aplicaciones industriales los objetos se encuentran situados sobre una superficie plana y se conocen los parámetros de la cámara: tamaño del CCD, distancia focal, y la distancia y orientación de la cámara respecto a la superficie sobre la que se encuentran los objetos, por lo que a partir de la posición de la imagen pueden obtenerse las coordenadas del objeto que pueden ser utilizadas por otros elementos del sistema automático como un robot. Usualmente se obtiene el centro de gravedad del objeto, definido por:

$$\bar{x} = \frac{1}{A} \left(\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} jg(i, j) \right) \quad (2.20)$$

$$\bar{y} = \frac{1}{A} \left(\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} ig(i, j) \right) \quad (2.21)$$

2.5. RECONOCIMIENTO DE OBJETOS

El reconocimiento de objetos es la última etapa dentro de un sistema de visión artificial. A partir de las características encontradas y de los posibles objetos que el conocimiento a priori del problema se espera que puedan aparecer, el sistema debe determinar qué objetos están presentes en la imagen. Esta etapa es la de mayor grado de abstracción de todas las que forman un sistema de visión y a menudo se realizan fuertes simplificaciones para que funcione con éxito [3].

2.5.1. Estructura General de un Clasificador

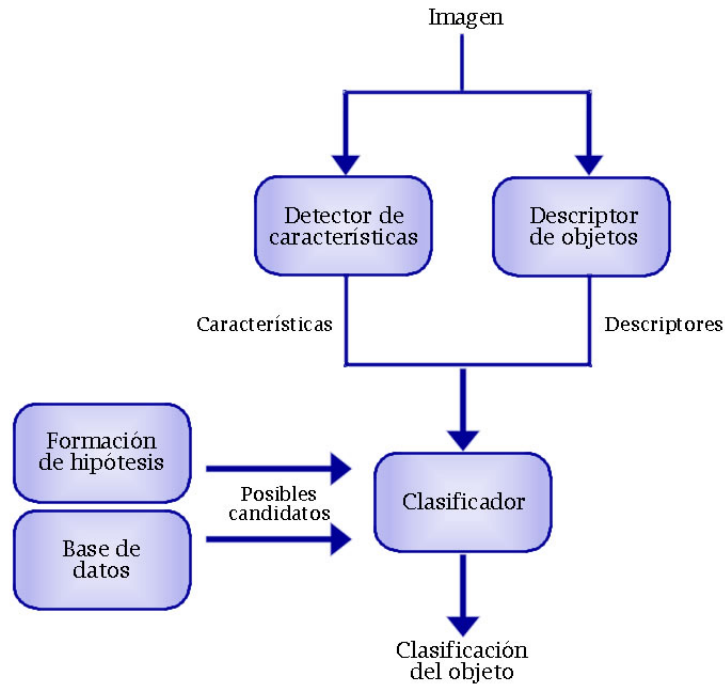


Figura 2.27. Estructura general de un clasificador.

La estructura general puede observarse en la Figura 2.27. A partir de la imagen original o procesada, se obtienen una serie de características que definen cada objeto. Con estas se ha pretendido reducir el volumen de información hasta hacerla manejable pero sin perder ninguna que sea importante. La representación de los objetos por medio de estas características se

habrá realizado mediante el análisis previo de otras imágenes en donde se habrá comprobado que realmente definen a los posibles objetos y los hacen distinguibles unos de otros.

Por otro lado, el análisis por parte de un operador humano de las imágenes de prueba se habrá obtenido una base de datos que contiene los modelos de los objetos que puedan aparecer en la imagen. El contenido de esta base de datos irá muy ligado al tipo de características que se van a obtener de la imagen. También con estas imágenes de prueba se realizarán una serie de hipótesis sobre los posibles candidatos que pueden aparecer en la imagen [3].

BIBLIOGRAFÍA

- [1] D. H. Ballard, ; C. M. Brown, *Computer Vision*. New York, 1982, 523 p. University of Rochester. Department of Computer Science. ISBN 0-13-165316-4.
- [2] R. González, R. Woods, *Tratamiento Digital de Imágenes*. Wilmington, USA: Addison-Wesley, 1996. ISBN 0-201-62576-8.
- [3] A. Escalera, *Visión por computador: Fundamentos y métodos*. Madrid: Prentice Hall, 2001. 304 p. ISBN 84-205-3098-0.
- [4] E. Harwood, *Digital CCTV. A Security Professional's Guide*. Estados Unidos de América: Elsevier, 2007. ISBN 0-7506-7745-7.
- [5] <http://es.wikipedia.org/wiki/Video>
- [6] H. Kruegle, *CCTV Surveillance: Video Practices and Technology*. Segunda Edición. Elsevier. USA. 2007. ISBN 0-7506-7768-6.
- [7] D. Mery, *Visión Artificial*. Santiago de Chile, 2002. Universidad de Santiago de Chile. Departamento de Ingeniería Informática.
- [8] G.J. Awcock, R. Thomas, *Applied Image Processing*. McGraw Hill: Londres, 1996. University of Brighton. Department of Electrical and Electronic Engineering. ISBN 0-07-001470-1.
- [9] W. K. Pratt, *Digital Image Processing*. Los Altos, California, 2001. ISBN 0-471-37407-5.

Capítulo 3

SISTEMAS DE VIGILANCIA BASADOS EN VISIÓN ARTIFICIAL

En la actualidad se han desarrollado numerosos trabajos en el campo de la vigilancia mediante visión artificial, los cuales se referencian en este capítulo. En general, en cada uno de ellos se realizan tareas tales como: detección del movimiento, seguimiento y/o clasificación de los objetos de interés. A continuación se muestran las técnicas más conocidas que se emplean para llevar a cabo dichas tareas.

3.1. DETECCIÓN DE MOVIMIENTO

Los algoritmos de detección de movimiento ayudan a identificar las regiones de la imagen donde hay movimiento y que corresponden a los objetos de interés tales como vehículos y peatones. Este es el primer paso que se realiza en casi todo sistema de visión artificial pues facilita el análisis en las etapas posteriores [1]. Para el desarrollo e implementación de estos algoritmos existen diferentes técnicas, entre las cuales se pueden destacar:

3.1.1. Sustracción del Fondo

La sustracción del fondo es una técnica comúnmente utilizada para la segmentación de movimiento en escenas estáticas. Este método detecta regiones en donde hay movimiento sustrayendo la imagen actual pixel por pixel de una imagen de referencia denominada fondo que es creada promediando imágenes en un periodo de inicialización [1].

Los algoritmos de detección de movimiento basados en la sustracción del fondo se pueden describir mediante el organigrama mostrado en la figura 3.1. Cada nueva imagen de la secuencia de video o *frame* (Figura 3.2 (a)) se resta de la imagen del fondo (Figura 3.2 (b)). Al mismo tiempo, el nuevo *frame* se utiliza para actualizar la imagen del fondo existente. Se aplica un umbral [2,3] a la imagen resultante de la diferencia entre el *frame* actual y la imagen de fondo para producir una imagen binaria (Figura 3.2 (c)) que indica los sectores en los cuales se presenta actividad o movimiento. Finalmente, se aplican métodos de segmentación a esta imagen para extraer y almacenar los objetos en movimiento (Figura 3.2 (d)).

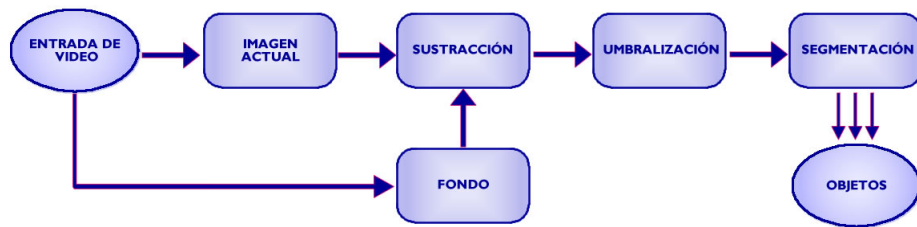


Figura 3.1. Organigrama de algoritmos de detección de movimiento basados en sustracción del fondo [4].

En términos matemáticos, un pixel en la posición (x, y) de la imagen actual I_t (*frame* en un tiempo t) hace parte del primer plano (más comúnmente conocido como *foreground*) o en otras palabras es un objeto en movimiento si satisface la condición:

$$|I_t(x, y) - B_t(x, y)| > \tau \quad (3.1)$$

donde τ es un umbral predefinido y B_t es la imagen del fondo [1].

La imagen del fondo B_t se calcula mediante la relación:

$$B_t(x, y) = \frac{\sum_{i=0}^N b_i(x, y)}{N} \quad (3.2)$$

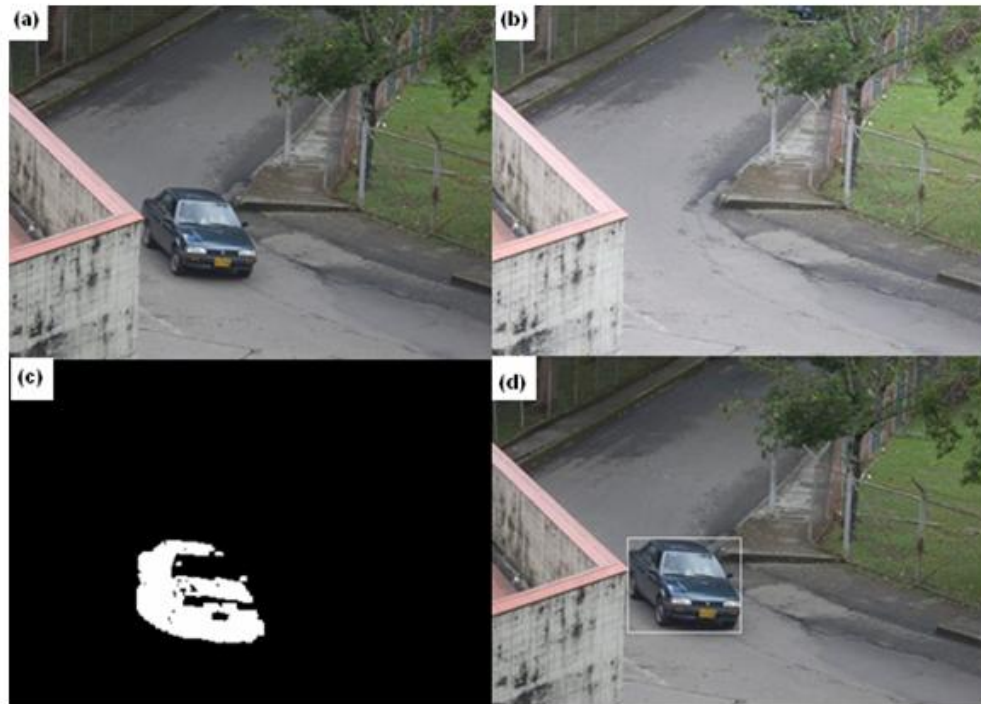


Figura 3.2. Detección de movimiento mediante sustracción del fondo. (a) Imagen Actual, (b) Imagen de fondo, (c) Imagen segmentada después de la sustracción, (d) Objetos detectados.

donde $b_i(x, y)$ es la imagen del fondo en el *frame* i y N es el número de *frames* que constituyen el periodo de inicialización.

Por supuesto, los cambios capturados de esta manera pueden o no corresponder a objetos en movimiento. Las sombras, el ruido de la imagen, la variación en las condiciones de la iluminación, condiciones climáticas, y los cambios pequeños en los parámetros intrínsecos de la cámara de video son algunos ejemplos de los factores que podrían presentarse como movimiento en la imagen diferencia pero, que realmente, no corresponden a objetos en movimiento. También es posible que algunos objetos en movimiento no introduzcan ningún cambio a la imagen diferencia, tal como un objeto en movimiento que sea muy similar en color al fondo de la imagen [5].

3.1.2. Métodos Estadísticos

Son métodos más avanzados que emplean las características estadísticas de los píxeles individuales, se han desarrollado para superar las limitaciones de los métodos básicos de sustracción del fondo. Están basados, principalmente, en métodos de sustracción del fondo, al actualizar dinámicamente la estadística de los píxeles que pertenecen a la imagen de fondo $((x, y) \in BG)$. Los píxeles del primer plano $((x, y) \in FG)$ son identificados comparando la estadística de cada píxel con la del fondo modelo. Esta aproximación se está convirtiendo en la más popular, debido a su confiabilidad en escenas que contienen ruido, cambios de iluminación y sombras. En términos matemáticos, el fondo se actualiza de la siguiente manera:

$$B_{t+1}(x, y) = \begin{cases} \alpha B_t(x, y) + (1 - \alpha) I_t(x, y), & \text{si } x, y \in BG \\ \beta B_t(x, y) + (1 - \beta) I_t(x, y), & \text{si } x, y \in FG \end{cases} \quad (3.3)$$

Donde α y β toman valores entre 0 y 1 y se denominan constantes de aprendizaje que especifican qué tanta información de la imagen actual I_t pasa a ser parte del fondo.

Después de la umbralización, se aplica una iteración de erosión morfológica a los píxeles detectados del primer plano, para remover el ruido que se presenta. Para regresar las regiones detectadas a su tamaño original, se ejecuta una secuencia de erosiones y dilataciones sobre los píxeles pertenecientes al primer plano (píxeles activos). Además, las regiones clasificadas como pequeñas son eliminadas después de etiquetadas para encontrar las regiones que corresponden realmente a los objetos de interés que se encuentran en movimiento. La estadística de los píxeles del fondo, que pertenecen a regiones que no están en movimiento en la imagen actual, se actualiza con los nuevos datos de la imagen [1].

3.1.3. Diferenciación Temporal

Consiste en detectar regiones en movimiento mediante la diferencia temporal entre píxeles de dos o tres *frames* consecutivos en una secuencia de video. Este método se adapta muy bien a cambios de escena dinámicos, sin embargo, falla generalmente en la detección de conjuntos de píxeles relevantes de ciertos tipos de objetos en movimiento. En la figura 3.3 se

puede observar un ejemplo de cómo esta técnica puede tener algunos problemas con aquellas regiones de píxeles que no presentan un cambio significativo entre *frames*.



Figura 3.3. Ejemplo de diferenciado temporal. (a) Un ejemplo de la escena con objetos en movimiento. (b) La diferenciación temporal falla en la detección de todos los píxeles en movimiento del sujeto de camisa amarilla debido a su uniformidad de color. Las regiones en movimiento detectadas se marcan con color azul.

Se requiere de métodos adicionales para detectar los objetos estáticos para lograr un procesamiento de alto nivel.

Lipton, presenta un esquema de diferenciación entre dos *frames* donde los píxeles que satisfacen la siguiente ecuación pertenecen al primer plano de la escena [1,6].

$$|I_t(x, y) - I_{t-1}(x, y)| > \tau \quad (3.4)$$

Con el fin de superar la deficiencia de la diferenciación de dos *frames* en algunos casos, se aplica este procedimiento para tres *frames* [7]. Para obtener un sistema de detección de movimiento más robusto, con frecuencia se emplean algoritmos híbridos, que combinan la diferenciación de tres *frames* con un modelo adaptativo de sustracción del fondo. Este tipo de algoritmos realiza, con gran precisión, la segmentación de regiones en movimiento sin los problemas de diferenciación temporal y sustracción del fondo [1].

3.1.4. Flujo Óptico

Los métodos basados en flujo óptico emplean vectores de flujo de objetos moviéndose en el tiempo, para detectar regiones en movimiento en una imagen. Pueden detectar movimiento en secuencias de video aún si se utilizan cámaras en movimiento. Sin embargo, la mayoría de estos métodos son muy costosos desde el punto de vista computacional, y no pueden ser utilizados en sistemas de tiempo real sin un hardware especializado [1].

3.1.5. Detección de Cambios de Luz y Sombras

Los métodos anteriormente descritos para la detección de movimiento trabajan, bien sea en recintos cerrados, como en zonas que se encuentran al aire libre, y se han utilizado en sistemas de vigilancia de tiempo real por años. Sin embargo, la mayoría de estos algoritmos son susceptibles a cambios locales (presencia de sombras y brillos) y globales de iluminación (días soleados o nublados). Las sombras causan problemas al momento de realizar la segmentación y detección de movimiento, y por tanto afectan sustancialmente los procesos de clasificación y reconocimiento de objetos. Existen en la literatura diversos métodos para superar este tipo de obstáculos. Horprasert y otros autores, por ejemplo, presentaron un novedoso método [8]. En éste, cada pixel es representado por un modelo de color que separa el brillo del componente cromático. Un pixel dado es clasificado dentro de cuatro categorías diferentes (fondo, fondo sombreado o sombra, fondo brillante y objeto en movimiento del primer plano) calculando la distorsión de brillo y cromaticidad entre los pixeles del fondo y la imagen actual. Otros métodos aprovechan el hecho de que un área sombreada presenta cambios significativos en intensidad, más no tanto en cromaticidad, de este modo superan el problema de las sombras mediante medidas de cromaticidad y gradiente de intensidad. Además, se utiliza la información del gradiente en regiones en movimiento para asegurar la confiabilidad del sistema en casos ambiguos.

En algunos sistemas, los cambios globales de iluminación son detectados por medio del conteo del número de pixeles del primer plano y si el número total excede algún umbral (porcentaje del tamaño total de la imagen), el sistema se reinicia para adoptar los nuevos cambios de iluminación [1, 9, 10].

3.2. SEGUIMIENTO DE OBJETOS EN MOVIMIENTO

La finalidad del seguimiento de objetos en movimiento es determinar si un objeto en el *frame* actual es el mismo objeto en el *frame* anterior, esta correspondencia entre objetos, o parte de objetos, en *frames* consecutivos permite extraer información temporal como su trayectoria, postura, velocidad y dirección. Esta tarea constituye una de las más difíciles dentro de un sistema de vigilancia mediante visión artificial y es además crucial para el análisis del comportamiento del objeto de interés [1].

Existen diversos métodos de seguimiento de objetos en movimiento, que se pueden clasificar en tres categorías: los métodos predictivos, los seguidores de regiones y los que siguen contornos.

Dentro de los métodos estadísticos, que predicen la posición de un objeto en el siguiente *frame* a partir de un modelo matemático que se reajusta mediante medidas, la técnica más representativa es el filtro de Kalman.

3.2.1. Filtro de Kalman

En el ámbito de la Visión Artificial el filtro de Kalman es un algoritmo recursivo que se utiliza para estimar la posición de un punto o característica en movimiento y la incertidumbre de la medida, en la siguiente imagen. Se trata de buscar la característica (punto, borde, esquina, región, etc.) en un área determinada de la siguiente imagen alrededor de la posición predicha, en la que estamos seguros de encontrar la característica dentro de un cierto grado de confianza [11].

El objetivo del filtro es la obtención de un estimador óptimo de las variables de estado de un sistema dinámico, basado en observaciones ruidosas y en un modelo de incertidumbre de la dinámica del sistema.

El método estima el estado perteneciente de un proceso controlado en el tiempo discreto que es gobernado por la ecuación diferencial del tipo:

$$x_{k+1} = A_k x_k + B \mu_k + w_k \quad (3.5)$$

Con una medida z correspondiente a la observación y perteneciente a \mathfrak{R}^m que es:

$$z_k = H_k x_k + v_k \quad (3.6)$$

Las variables aleatorias w_k y v_k representan el ruido del proceso y de la medida respectivamente y se asume que son independientes y blancos.

Cuando el filtro de Kalman se aplica a la Visión Artificial, el estado x se corresponde con el vector posición del objeto en la imagen determinado por las coordenadas de posición x_x y x_y , y las coordenadas de velocidad v_x y v_y . La observación z en cambio, es únicamente un vector de dos componentes z_x y z_y , correspondiente a las coordenadas de la posición observada del objeto de interés.

La matriz A_{NxM} relaciona el estado en tiempo k con el estado en tiempo $k + 1$. Esta relación se manifiesta en las siguientes ecuaciones, dando como resultado la matriz A .

$$\begin{cases} x_{x_{k+1}} = x_{x_k} + v_x t \\ x_{y_{k+1}} = x_{y_k} + v_y t \\ v_{x_{k+1}} = v_{x_k} \\ v_{y_{k+1}} = v_{y_k} \end{cases} \quad (3.7)$$

$$A = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (3.8)$$

La matriz B_{Nx1} relaciona la entrada control μ perteneciente a \mathfrak{R}^1 con el estado x . Y la matriz

$H_{N \times N}$ relaciona el estado con la medida z_k .

$$M = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \quad (3.9)$$

El filtro de Kalman proporciona una ecuación que computa un estimador del estado a posteriori \hat{x}_k como combinación lineal del estimador a priori \hat{x}_k^- y la diferencia ponderada entre la observación actual z_k y una predicción de medida $H_k \hat{x}_k^-$:

$$\hat{x}_k = \hat{x}_k^- + K(z_k - H_k \hat{x}_k^-) \quad (3.10)$$

La diferencia $(z_k - H_k \hat{x}_k^-)$ se llama comúnmente innovación de la medida o simplemente residuo y refleja la discrepancia entre la predicción de la medida $H_k \hat{x}_k^-$ y la observación actual z_k . La matriz $K_{N \times M}$ llamada ganancia de Kalman o factor de mezcla establece la cantidad de influencia del error entre nuestra estimación y la medida:

$$K_k = P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1} \quad (3.11)$$

Siendo P_k^- el estimador de la covarianza del error a priori y R_k la covarianza del error medido. Vemos que si R_k se aproxima a 0, la ganancia ponderará el residuo con mayor peso. Por el contrario, cuando P_k^- se aproxime a 0, la ganancia ponderará menos el residuo.

El filtro de Kalman estima variables de estado de un proceso con realimentación. Calcula el estado del proceso en algún instante y entonces obtiene información (se realimenta) de la medida. Por tanto, las ecuaciones del filtro se pueden clasificar en dos tipos: actualización del tiempo y actualización de las medidas. Las primeras son responsables de proyectar hacia el futuro los estimadores del estado actual y de la covarianza del error, para obtener los estimadores a priori del siguiente estado. Las ecuaciones de actualización de las medidas son responsables de la realimentación, incorporando una nueva medida a los estimadores a priori para obtener unos estimadores a posteriori mejorados. Las ecuaciones de actualización del tiempo pueden ser interpretadas como ecuaciones de predicción, mientras que las

de actualización de la medida pueden pensarse como ecuaciones de corrección [11].

3.2.2. Seguimiento de Regiones

Un seguidor de región identifica una región de la imagen, para la cual se usa una medida de similitud con el fin de decidir si la región identificada corresponde a la misma región en la siguiente imagen de la secuencia. La región se toma de tal modo que contenga alguno de los objetos de interés, generalmente los límites de estas regiones son polígonos o rectángulos que encierran estos objetos [12]. Estos límites también son conocidos como *bounding box*.

La medida de similitud anteriormente mencionada es el factor clave para la correspondencia entre objetos en *frames* consecutivos. La escogencia de esta medida puede variar de acuerdo a la aplicación. En la figura 3.4 se muestra un criterio de mínima distancia entre objetos en movimiento, en donde la nueva posición del objeto de interés, en el *frame* siguiente ($t + 1$), será la más cercana a éste en el *frame* anterior (t). De este modo se asignan las etiquetas correspondientes de acuerdo a la condición planteada.

Obsérvese que las distancias se miden a partir de un punto dentro del objeto, que por lo general es el centro de masa aunque en ocasiones se emplea el centro del *bounding box*.

El criterio de medida para el seguimiento de regiones puede variar, en muchos casos no se utiliza una medida de distancia sino que se calcula el área de solapamiento entre *bounding box* como se puede observar en la figura 3.5. Cuando se emplea este criterio la correspondencia entre objetos en *frames* subsecuentes se da siempre y cuando exista una región de solapamiento entre regiones de interés en imágenes consecutivas.

3.2.3. Seguimiento por Contornos

Finalmente los seguidores de bordes constituyen la tercera categoría. Éstos buscan los límites de los objetos de interés, o posiciones de alta luminosidad o también, cambios de color, dentro de una imagen. Este tipo de algoritmos se basan en la teoría de que los límites de un objeto tendrán fuertes variaciones en color, bordes o iluminación, cuando éstos se encuentran

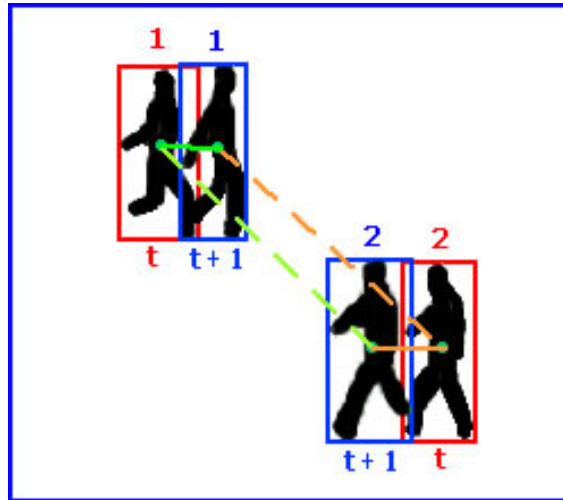


Figura 3.4. Seguimiento de regiones usando el criterio de mínima distancia.

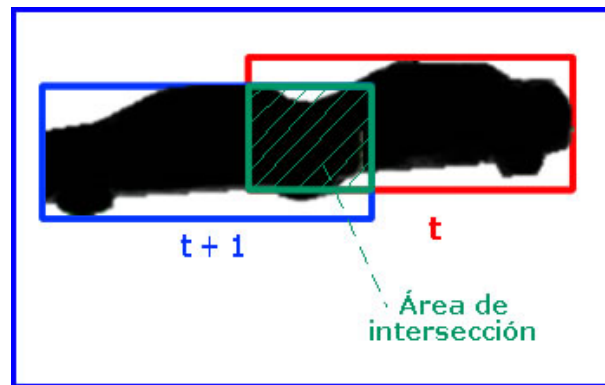


Figura 3.5. Seguimiento de regiones usando el criterio de superposición de *bounding box*.

en movimiento. Los algoritmos de seguimiento de bordes tienen problemas cuando la iluminación de la escena es baja debido a que, en estos casos, los cambios de color e iluminación son muy pequeños y difíciles de detectar. También cuando el fondo presenta una fuerte textura es difícil determinar qué borde seguir. El mayor problema de este tipo de algoritmos es cómo decidir cuando un borde hace parte del objeto y cuando descartarlo por hacer parte del fondo [13]. Dentro de esta categoría se pueden clasificar todos los métodos de contornos activos (*Snakes*). Una *snake* es un contorno elástico que se ajusta a las características detectadas en una imagen. La naturaleza de esta elasticidad depende de cierta configuración que representa la información a priori de la forma que se desea seguir y las medidas reales que se obtienen de la imagen [14].

3.3. CLASIFICACIÓN DE OBJETOS

Las regiones en movimiento detectadas en un video pueden corresponder a diferentes objetos como, por ejemplo, peatones, vehículos, etc. Es de gran importancia reconocer el tipo de objeto detectado con el fin de seguirlo eficientemente y analizar sus actividades correctamente [1]. La etapa de *reconocimiento* o clasificación de objetos es el proceso que asigna una etiqueta a un objeto basándose en la información proporcionada por sus descriptores [15]. Actualmente, existen tres formas de clasificar objetos en movimiento: según su forma, según su movimiento y por correlación. Veamos a continuación de qué se trata cada una de ellas.

3.3.1. Clasificación Basada en Forma

Las características comúnmente utilizadas para este tipo de esquemas de clasificación son: el rectángulo mínimo que encierra el objeto en movimiento (*bounding box*), el área (Figura 3.6), la longitud de la silueta (perímetro) (Figura 3.7) y el gradiente de las regiones del objeto detectado [1].

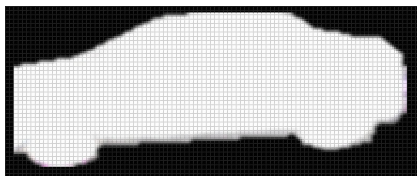


Figura 3.6. Área del objeto de interés. El número de pixeles activos (blanco) constituye el área del objeto.

El modelo presentado en [6] utiliza el perímetro y el área del objeto como información para clasificar objetos detectados dentro de tres grupos: humanos, vehículos y otros. Este método parte de la suposición de que los humanos son, en general, más pequeños que los vehículos y tienen formas más complejas. Este método utiliza la *compactibilidad* como operador métrico de clasificación [1,6]. La *compactibilidad* es una simple relación entre parámetros de forma simples, y esta dado por la ecuación 2.19



Figura 3.7. Perímetro del objeto de interés. El número de píxeles activos (blanco) constituye el perímetro del objeto.

Claramente un humano, con formas más complejas, tendrá mayor *compactibilidad* que un vehículo [6]. La clasificación es realizada en cada *frame* para lograr una clasificación consistente temporalmente [1,6].

3.3.2. Clasificación Basada en Movimiento

Este tipo de clasificación, por lo general, se realiza después del seguimiento, debido a que emplea descriptores como la velocidad y la trayectoria, que esta última etapa entrega. La velocidad, por ejemplo, constituye una buena medida de clasificación cuando se desea distinguir peatones de vehículos, ya que estos últimos se mueven a una velocidad mucho mayor que con la que se desplazan los peatones. La trayectoria trazada por los objetos de interés, es otro parámetro que se tiene en cuenta, pues existe una marcada diferencia entre los trazos dejados por los vehículos y los trazos dejados por los peatones. Los primeros, generalmente, dejan líneas rectas mientras que los segundos, generan trayectorias no lineales y además con cierta periodicidad.

3.3.3. Clasificación Basada en Modelos

El propósito de este tipo de clasificación es determinar cuál de los patrones conocidos o plantillas (*templates*), que se tienen disponibles inicialmente, corresponde o se semeja más al patrón desconocido. En el campo de la visión por computador, estas plantillas no son más que ciertos objetos en una escena. El primer paso para realizar esta tarea es definir una medida o costo de medida de la “distancia” o “similaridad” entre el patrón de referencia

(plantilla u objeto conocido) y el patrón desconocido con el propósito de encontrar la clase correspondiente a dicho objeto. Esta técnica de clasificación se conoce con el nombre de *template matching* por su traducción al inglés [16].

BIBLIOGRAFÍA

- [1] Y. Dedeoglu, *Moving Object Detection, Tracking And Classification For Smart Video Surveillance*, Bilkent University, Department of Computer Engineering and the Institute of Engineering and Science, Agosto, 2004.
- [2] Paul L. Rosin, Tim Ellis, *Image difference threshold strategies and shadow detection*, 6th British Machine Vision Conference, Birmingham, UK, 1995.
- [3] P. Rosin, *Thresholding for Change Detection*, 8th British Machine Vision Conference, Cochester, UK, 1997.
- [4] D. Makris, *Visual Learning in Surveillance Systems*, City University, Department of Electrical, Electronic and Information Engineering Machine Vision Group, London, Mayo 2001.
- [5] O. Masoud, N. Papanikolopoulos, *A Robust Real-time Multi-level Model-based Pedestrian Tracking System*, Artificial Intelligence, Robotics, and Vision Laboratory, Department of Computer Science, University of Minnesota, 1997
- [6] A. J. Lipton, H. Fujiyoshi, y R.S. Patil, *Moving Target Classification and Tracking from Real-Time Video*. In Proc. of Workshop Applications of Computer Vision, pp 129-136, 1998.
- [7] L. Wang, W. Hu, y T. Tan, *Recent developments in human motion analysis*. Pattern Recognition, 36(3):585-601, Marzo 2003. Referenciado por [2].
- [8] T. Horprasert, D. Harwood, y L.S. Davis, *A Statistical Approach for Real-Time Robust Background Subtraction and Shadow Detection*. In Proc. of IEEE, Frame Rate Workshop, pp 1-19, Kerkyra, Grecia, 1999. Referenciado por [2].
- [9] J. Owens y A. Hunter. *A Fast Model-Free Morphology-Based Object Tracking Algorithm*. British Machine Vision Conference, pp 767-776, Cardiff, UK, Septiembre 2002. Referenciado por [2].

- [10] X. Zhou, R. T. Collins, T. Kanade, y P. Metes. *A Master-Slave System to Acquire Biometric Imagery of Humans at Distance*. First ACM SIGMM International Workshop on Video Surveillance, pp 113 - 120. ACM Press, 2003. Referenciado por [2].
- [11] P. Rodríguez M., A. Sanz M. y J. J. Pantrigo, *Aplicación del Filtro de Kalman al seguimiento de objetos en secuencias de imágenes*, Universidad Rey Juan Carlos, Madrid, España, 2003.
- [12] L. R. Rabiner, B-H. Juang, *An Introduction to Hidden Markov Models*, IEEE ASSP Magazine, pp 4-16, 1986. Referenciado por [14].
- [13] K. Shearer, K. D. Wong, S. Venkatesh *Combining multiple tracking algorithms for improved general performance*. Pattern Recognition, 2003.
- [14] A. Blake, M. Isard, *Active Contours*. Visual Dynamics Group, University of Oxford. April 1998, from, Springer-Verlag.
- [15] R. González, R. Woods, *Tratamiento Digital de Imágenes*. Wilmington, USA: Addison-Wesley, 1996. ISBN 0-201-62576-8.
- [16] S. Theodoridis, K. Koutroubas, *Pattern Recognition*, Segunda edición, Elsevier, USA, 2003.

Capítulo 4

DESCRIPCIÓN DEL SISTEMA DE VIGILANCIA BASADO EN VISIÓN ARTIFICIAL

El sistema de vigilancia desarrollado consta de seis partes básicas, estas son:

- * Adquisición de las imágenes
- * Preprocesamiento
- * Detección de movimiento
- * Seguimiento
- * Clasificación
- * Entrenamiento

Cada una de estas partes es un elemento esencial para el buen funcionamiento del sistema, por lo que es de gran importancia hablar de cómo se desarrolló cada una de ellas.

4.1. SISTEMA DE ADQUISICIÓN DE LAS IMÁGENES

Como se mencionó en la sección 2.1, un sistema de adquisición de imágenes está formado por elementos importantes, tales como la iluminación, la cámara y la tarjeta importadora de video. El sistema no cuenta con algún tipo de iluminación específico, ya que está diseñado para trabajar en zonas abiertas, donde las condiciones de iluminación no se mantienen constantes. Para suplir esta situación se diseñó un sistema de control de iluminación, que se

menciona con mayor detalle en la sección 4.2. En la presente sección se muestran la cámara y la tarjeta utilizadas en el sistema y las imágenes adquiridas.

4.1.1. Cámara de Video-Vigilancia

La cámara utilizada en el sistema (Figura 4.1) es una cámara de video-vigilancia genérica a color que cuenta con un sensor de tecnología CMOS (sección 2.1). Es una cámara, de bajo costo, proporciona imágenes de buena calidad. Cuenta con un sistema de ganancia automática, que cambia la iluminancia de la imagen cuando se presentan cambios de iluminación en la escena, lo que puede no ser conveniente en algunas situaciones.



Figura 4.1. Cámara de video-vigilancia utilizada en el sistema.

4.1.2. Tarjeta Digitalizadora de Video

La tarjeta utilizada por el sistema para la digitalización de las imágenes es una tarjeta marca Zoltrix, mostrada en la Figura 4.2. Esta tarjeta cuenta con las siguientes características:

- * Captura en tiempo real.
- * Captura de video de movimiento completo.



Figura 4.2. Tarjeta importadora de video marca Zoltrix.

- * Captura de imágenes a 30 fps o mas.
- * Colores desde 256 hasta 16 millones.

4.1.3. Video Adquirido por el Sistema

El sistema es capaz de adquirir video, bien sea directamente desde una cámara conectada a la tarjeta de video del PC, o bien desde un archivo de video previamente guardado en el equipo. Las imágenes que conforman el video tienen dimensiones de 320 pixeles de ancho por 240 de alto. El número de cuadros por segundo, cuando se adquiere video de forma directa desde la cámara de vigilancia, es de 30. Sin embargo, el programa desarrollado en este trabajo carga videos desde archivo sin importar el *frame rate*.

Por razones prácticas se grabaron videos de distintas escenas que contenían flujo tanto de vehículos como de peatones, pues son los objetos de estudio en este proyecto, así como también distintas interacciones entre ellos, que fueron analizadas como se verá en las secciones posteriores.

Como se mencionó en capítulos anteriores, un sistema de vigilancia para zonas abiertas está expuesto a condiciones cambiantes tanto de iluminación como ambientales y debe funcionar a pesar de esta variabilidad. Por esta razón se seleccionaron escenarios que presentan diferencias de este tipo. En la figura 4.3 se observan cada uno de ellos. En esta figura se puede observar que, por ejemplo, los escenario numero 1 y 5 son los mismos, sin embargo presentan condiciones climáticas diferentes. Los escenarios 3 y 4 también son los mismos y sin embargo sus nivel de iluminación es diferente.



Figura 4.3. Imágenes adquiridas por la cámara y la tarjeta en diferentes condiciones de iluminación.

4.2. PREPROCESAMIENTO DE LAS IMÁGENES ADQUIRIDAS

La etapa de preprocesamiento de las imágenes, en este sistema, consta de tres partes: La conversión de las imágenes a niveles de gris, el control de la iluminación en la imagen y la aplicación de un filtro gaussiano de paso bajo. Este último es opcional y se utiliza sólo en escenas en donde se presenta movimiento que no corresponde a objetos de interés, como por ejemplo las hojas de los árboles, el césped, etc.

4.2.1. Conversión de las Imágenes a Color a Imágenes en Escala de Grises

Las imágenes adquiridas mediante la cámara y la tarjeta importadora son imágenes a color, esto implica que cada imagen tiene una gran cantidad de información, y si a esto se le suma que se está analizando una secuencia de imágenes, es decir, más de una imagen, entonces el sistema deberá manejar mucha información en muy poco tiempo, ya que, por lo general, se tendrán 30 imágenes cada segundo.

Por esta razón se decidió trabajar con imágenes en escala de grises para reducir el coste com-

putacional y hacer más fácil la tarea del procesador reduciendo el tiempo de procesamiento. En la figura 4.4 (a) se muestra una imagen adquirida mediante la cámara y la tarjeta de video. Así mismo se puede observar su transformación a escala de grises en la figura 4.4 (b).



Figura 4.4. Conversión a escala de grises: (a) Imagen original. (b) Imagen en escala de grises.

4.2.2. Control de Iluminación

Debido a que el sistema se utiliza para el monitoreo de zonas abiertas, normalmente, hay cambios bruscos en la iluminación, los cuales pueden generar falsas detecciones en la etapa de detección de movimiento. Para evitar estas falsas detecciones se creó e implementó un algoritmo de control de iluminación que consiste en promediar la iluminancia de la imagen de referencia o fondo (sección 4.3.1) de acuerdo a la relación que se describe a continuación (4.1), y calcular la diferencia entre este valor y la iluminancia promedio de la imagen actual (Figura 4.5 (a)).

$$I_{ref} = \frac{1}{M \cdot N} \left(\sum_{i=0}^M \sum_{j=0}^N F(i, j) \right) \quad (4.1)$$

donde I_{ref} es la iluminancia promedio de la imagen del fondo, la cual se denota como $F(i, j)$. M y N son respectivamente el ancho y el alto, en píxeles, de la imagen. En este sistema estos valores son 320 y 240 por ser el tamaño con el que se adquieren las imágenes de video.

$$s = |I_{ref} - I_{act}| \quad (4.2)$$

$$I_c(x, y) = \begin{cases} I(x, y) - s, & \text{si } I_{ref} < I_{act} \\ I(x, y) + s, & \text{si } I_{ref} > I_{act} \end{cases} \quad (4.3)$$

Si la iluminancia promedio de la imagen actual, denotada por I_{act} , es menor que la iluminancia de referencia, se suma la diferencia (s), dada por 4.2, a la imagen actual denotada por I (Figura 4.5 (b) y (c)). Esta imagen resultante, que es una imagen compensada, se denota I_c . Así mismo, si la iluminancia promedio de la imagen actual es mayor que la iluminancia de referencia, se resta la diferencia a la imagen actual (Figura 4.5 (d) y (e)), obteniendo así una imagen con un nivel de iluminancia parecido al de referencia y evitando las falsas detecciones. Esta compensación se puede expresar en términos matemáticos mediante la ecuación 4.3.

4.2.3. Filtro de Paso Bajo

Existen escenas en donde hay un leve movimiento de objetos que no son de interés, como por ejemplo, el de las ramas de los árboles gracias al viento, el césped, etc. Por esto se realiza, como procedimiento opcional del sistema, un filtro gaussiano a cada una de las imágenes de la secuencia de video. El usuario del sistema define si es conveniente activar esta opción. Los resultados de esta etapa se muestran con mayor claridad en la sección 5.2, y la forma en la cual se activa esta opción se puede conocer en el manual del usuario descrito en el Anexo A.



(a)



(b)



(c)



(d)



(e)

Figura 4.5. Control de iluminación: (a) Imagen de referencia o fondo. (b) Imagen actual de iluminancia más baja. (c) Imagen con iluminación controlada mediante el algoritmo de control. (d) Imagen actual de iluminancia más alta. (e) Imagen con iluminación controlada mediante el algoritmo de control.

4.3. DETECCIÓN DE MOVIMIENTO

Una vez que la imagen está en mejores condiciones para ser analizada, se desarrolla la siguiente etapa, que consiste en la detección de las regiones en movimiento presentes en la

escena. Esta etapa se realizó utilizando la técnica de sustracción del fondo descrita en la sección 3.1.1. Debido a que el ambiente es cambiante, se implementó un algoritmo de actualización del fondo de referencia, de manera que las variaciones de iluminación a las que está sometida la escena no afecten la detección de los objetos. El esquema general del algoritmo de detección de movimiento se presenta en la figura 4.6. El algoritmo está formado por cinco partes fundamentales: La inicialización, la sustracción y la actualización del fondo, la segmentación de las regiones en movimiento y la eliminación del ruido generado en la segmentación.

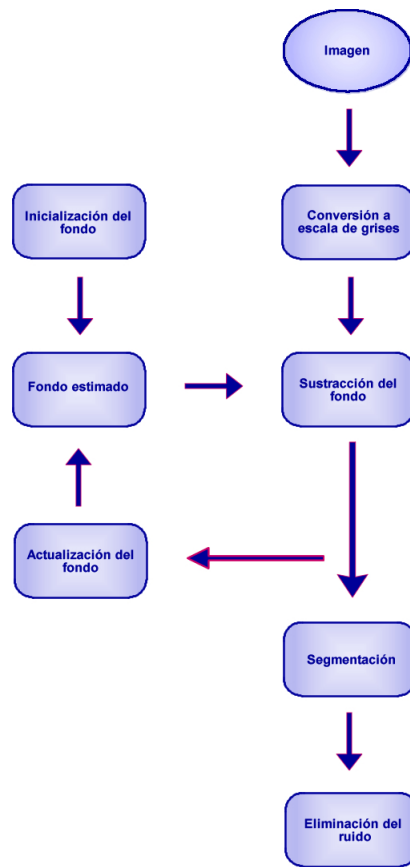


Figura 4.6. Esquema general del algoritmo de detección de movimiento empleado.

4.3.1. Inicialización de Fondo

La inicialización del fondo se realiza mediante una orden dada por el usuario del sistema, el proceso consiste en promediar cien imágenes inmediatamente después de haberse dado la orden. Como se mencionó en la sección 2.2.1.1, el hecho de promediar imágenes de una misma escena genera como resultado un filtrado de paso bajo que elimina ruidos de alta frecuencia. Si algún objeto ha pasado por la escena, éste no estará incluido en el fondo, debido a la cantidad de imágenes tomadas para la inicialización del fondo. Un ejemplo para el fondo estimado inicialmente en una escena determinada se muestra en la figura 4.7 (a).

4.3.2. Sustracción del Fondo

El siguiente paso es la sustracción del fondo. Aquí se toma la imagen de entrada y se resta pixel a pixel del fondo de referencia. El resultado de esta operación es una imagen en escala de grises, como se puede observar en la figura 4.7 (c).

4.3.3. Segmentación del Movimiento

Para la segmentación del movimiento se empleó el método de umbralización mencionado en la sección 2.3.1. El umbral se escoge empíricamente. Esta operación es realizada por el usuario, ya que cada escena puede tener un umbral de segmentación diferente, debido a que las condiciones de iluminación no están controladas y a que no se conoce previamente el nivel de gris de los objetos que aparecen en la escena. En el ejemplo mostrado en la figura 4.7 (d) se utilizó un umbral de 40.

4.3.4. Eliminación del Ruido

Después de la segmentación suelen aparecer objetos que no son de interés para el sistema. Debido a esto es necesaria la aplicación de filtros morfológicos que permitan obtener una imagen mas limpia. El filtrado se realizó de dos formas; la primera consiste en aplicar el filtro morfológico conocido como cierre, el cual se mencionó en la sección 2.4.3.2, para eliminar los “huecos” presentes en los objetos de interés (Figura 4.7 (e)). Debido a que el cierre no elimina objetos pequeños, y a que no se quiere cambiar demasiado la forma de los objetos, el siguiente método de filtrado consiste en etiquetar (sección 2.3.2) los objetos presentes en la imagen, y aquellos que sean muy pequeños, en comparación con las dimensiones normales de los objetos, son eliminados. De esta manera se obtiene una imagen que segmenta perfectamente los objetos presentes en la escena como se puede observar en la figura 4.7 (f).

4.3.5. Actualización del Fondo

Como se mencionó anteriormente, el objetivo principal de esta etapa es hacer que los cambios en la iluminación, debido a la variación en las condiciones ambientales del lugar, sean imperceptibles para el sistema. La técnica empleada para la actualización del fondo es la mencionada en la sección 3.1.2. Los valores de α y β fueron escogidos empíricamente. Se tuvo en cuenta que si los valores de α y β son muy grandes, el fondo actualizará de manera muy rápida, y si son muy bajos, la actualización será muy lenta. Debido a que los cambios bruscos de iluminación se controlaron como se explicó en la sección 4.2.2, se escogió un valor de α bajo para que el fondo se adaptara lentamente a los cambios de iluminación habituales del ambiente, los cuales son lentos. El valor de α escogido fue 0,02, y el de β igual a uno.

4.4. SEGUIMIENTO DE LOS OBJETOS DE INTERÉS

Una vez detectados los objetos, estos tienen que ser seguidos en cada uno de los *frames* posteriores. Para el seguimiento de estos objetos se utilizó la técnica de seguimiento de regiones, mencionada en la sección 3.2.2. El parámetro de similitud utilizado fue el valor del

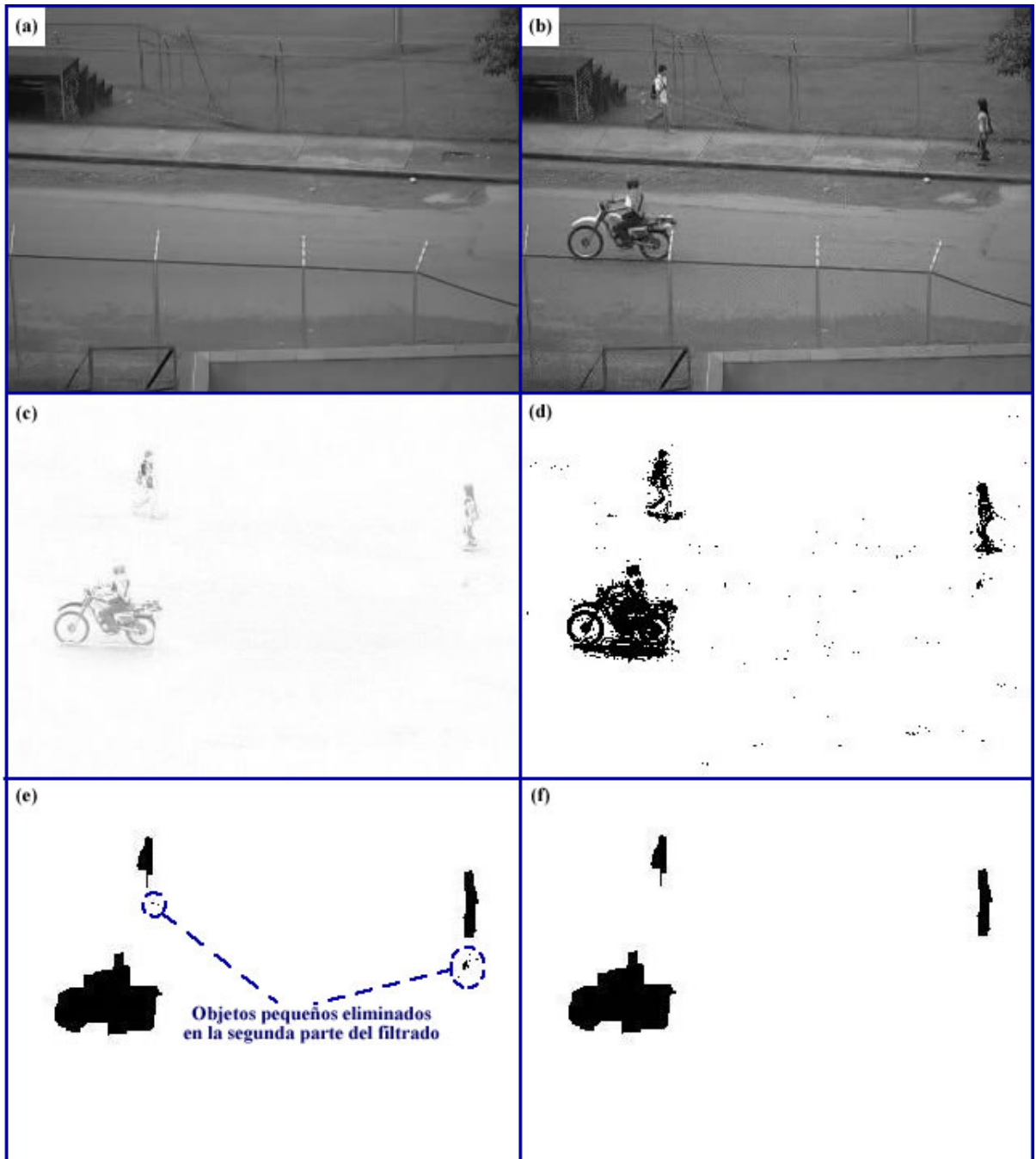


Figura 4.7. Detección de movimiento: (a) Fondo estimado. (b) Imagen de entrada. (c) Sustracción del fondo. (d) Segmentación de las regiones en movimiento. (e) Cierre. (f) Eliminación de los objetos pequeños.

área intersectada entre los *bounding boxes* de un mismo objeto en dos *frames* consecutivos.

Si entre los *bounding boxes* de un mismo objeto, en dos *frames* t y $t + 1$, existe un solapamiento o intersección, se puede afirmar que esa región es la misma del *frame* anterior, de lo contrario, corresponde a otra diferente (Figura 4.8). Para llevar un control de cada región

u objeto detectado, se almacenan ciertos parámetros en una base de datos, lo cual permite monitorear los valores de posición, área, dirección de movimiento, velocidad, etiqueta, entre otros parámetros que describen características importantes del objeto.

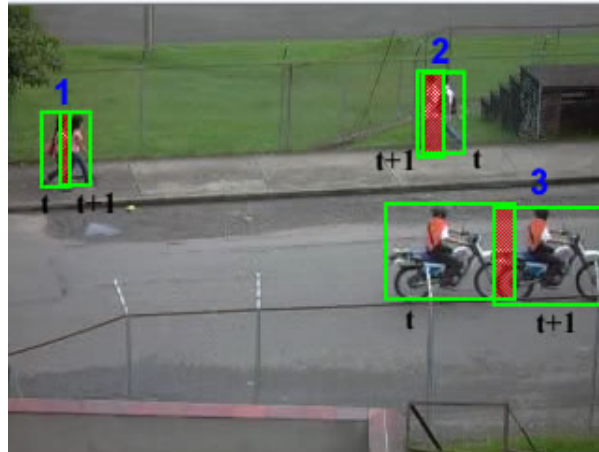


Figura 4.8. Seguimiento de regiones.

En esta etapa se presentan varios casos que pueden dificultar el correcto funcionamiento del sistema, como por ejemplo, dos objetos que se ocluyen, total o parcialmente, mutuamente en cierto punto de la escena y después se separan. En estos casos, el sistema detecta, inicialmente, la conformación de un grupo. Esta detección se realiza teniendo en cuenta que, si se solapan dos o más objetos en una misma región, estos son parte de un grupo (Figura 4.9 (b)). Una vez detectado el grupo, se almacena, en una base de datos, información sobre dicho grupo, como por ejemplo, el número de objetos que forman el grupo, etiquetas de los objetos del grupo, etc., además, el grupo es seguido utilizando el mismo criterio que para los objetos individuales (Figura 4.9 (c)). Posteriormente, el sistema detecta la separación del grupo, teniendo en cuenta que, si los *bounding boxes* de dos o más objetos se intersectan con el *bounding box* del grupo en el tiempo t , se ha separado un grupo (Figura 4.9 (d)).

Después de que el grupo se ha separado, el sistema conoce, gracias a la base de datos, por cuales objetos estaba conformado el grupo. Inicialmente, el sistema asigna, de manera aleatoria, las etiquetas de la base de datos a los objetos que se separaron del grupo. Como es de vital importancia que estas etiquetas no se asignen de esta manera, ya que es de interés seguir siempre el mismo objeto, el sistema mide la magnitud y la dirección de la velocidad de cada objeto y, al compararla con la información almacenada en la base de datos, se reasignan a los objetos las etiquetas que, anteriormente, les estaban asignadas (Figura 4.10).



Figura 4.9. Solución de problemas de oclusión: (a) Objetos 1 y 2 moviéndose en direcciones contrarias. (b) Formación de un grupo. (c) Seguimiento, como objeto individual, del grupo conformado. (d) Separación del grupo.

4.5. CLASIFICACIÓN DE LOS OBJETOS EN MOVIMIENTO

Es de gran interés para el sistema identificar cada uno de los objetos que están presentes en la escena. Para la identificación de estos, el sistema realiza una clasificación basada en la forma (sección 3.3.1). El único descriptor utilizado es el *área* de las regiones en movimiento, ya que los objetos de interés (peatones y vehículos) presentan una diferencia significativa en el valor de este parámetro, siendo el área de un vehículo mucho mayor que la de un peatón. El reconocimiento de cada objeto se realiza sólo una vez (en el momento en que ingresa en la escena). El resultado de la clasificación es almacenado en la base de datos de cada objeto, y gracias a la etapa de seguimiento (sección 4.4), no es necesario hacer el reconocimiento del objeto en cada *frame*.

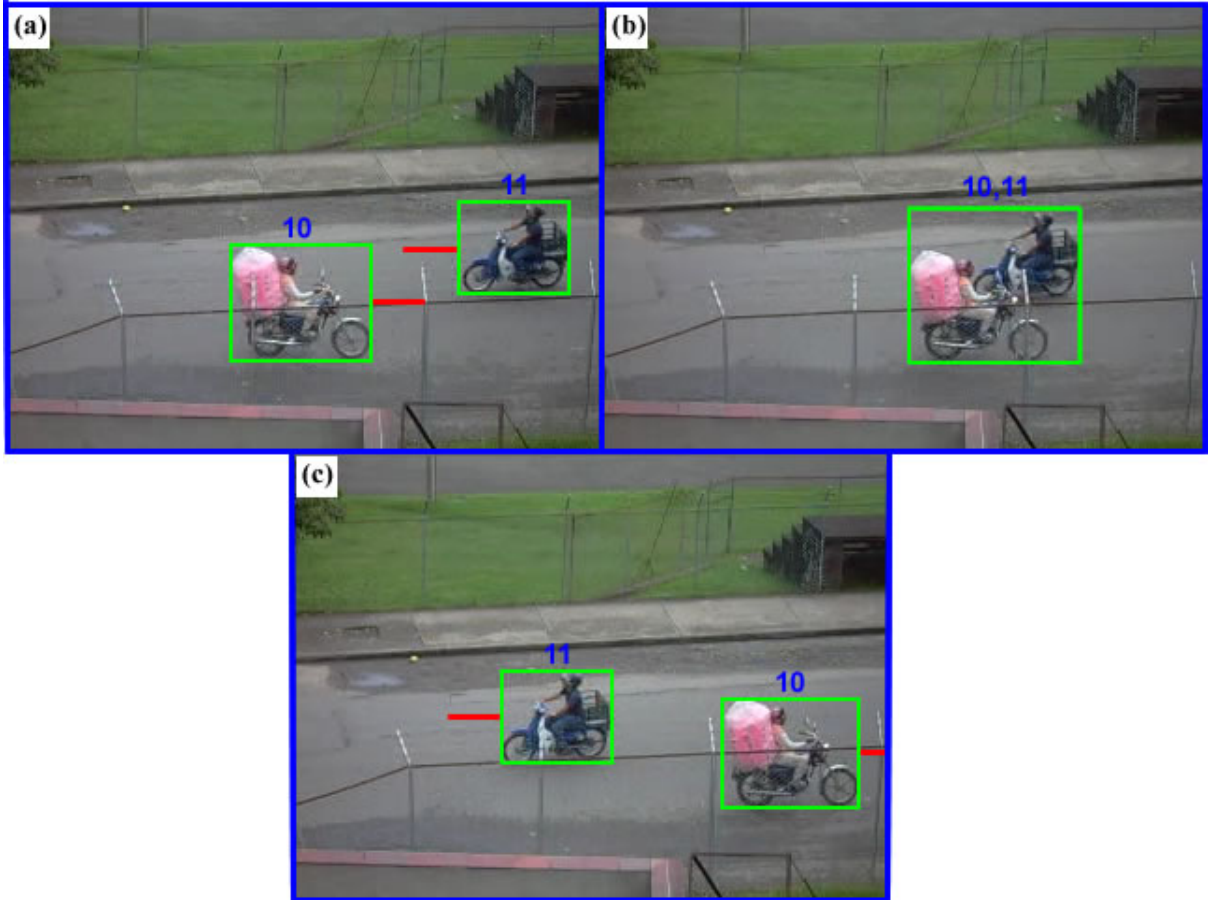


Figura 4.10. Seguimiento de regiones: (a) Objetos 10 y 11 moviéndose en direcciones contrarias. (b) Grupo formado por las etiquetas 10 y 11. (c) Asignación de las etiquetas correctas después de la separación.

Debido a que el tamaño o área de los objetos depende de la ubicación física de la cámara, es decir, que tan lejos o que tan cerca esté de la escena, fue necesario incluir al sistema una etapa de entrenamiento, la cual se realiza de forma manual por el usuario tal como se indica en el Manual del Usuario (Anexo A). Una vez que ya se conoce el tipo de objeto al cual corresponde cada región, se asigna un símbolo a cada una de ellas como muestra la figura 4.11.

4.6. ALARMAS Y REPORTE

El sistema, además de detectar, seguir y reconocer objetos en movimiento, también es capaz de generar alarmas y reportes según las situaciones que se presenten en la escena.

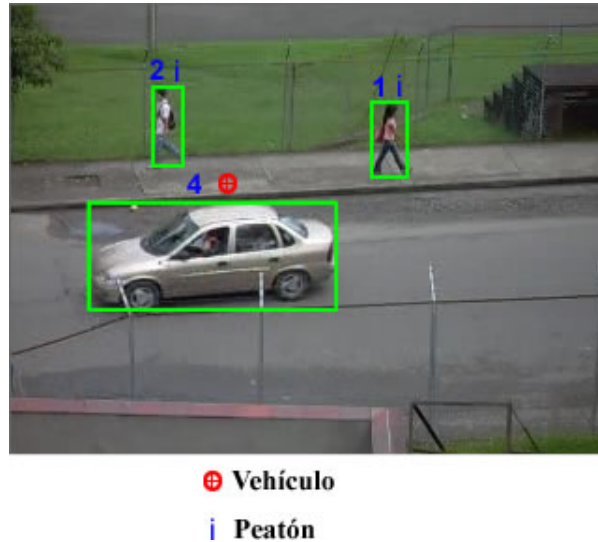


Figura 4.11. Clasificación de los objetos en movimiento.

El sistema genera una alarma cuando un objeto entra en una zona restringida, la cual es definida por el usuario del sistema (Anexo A). El usuario puede activar una de dos opciones para la generación de la alarma:

- * *Primera Opción:* Si el objeto entra en zona restringida, el sistema inmediatamente genera una alarma.
- * *Segunda Opción:* El usuario puede definir un tiempo límite durante el cual, el objeto, puede estar dentro de la zona restringida. Si el objeto permanece dentro, más del tiempo fijado por el usuario, el sistema da la alarma.

Además, es de gran utilidad para el usuario poder generar un reporte que contenga información relevante sobre lo que ha sucedido en la escena. Este reporte se puede generar en cualquier momento, según lo necesite el usuario (Anexo A). Este se almacena en el disco duro del computador en donde se esté ejecutando el programa como un archivo de Excel con extensión *.xls*. En el reporte se incluye la siguiente información:

- * *Número de peatones:* Indica el número total de peatones que aparecieron en la escena mientras se estaba generando el reporte.
- * *Número de vehículos:* Indica el número total de vehículos que aparecieron en la escena mientras se estaba generando el reporte.

- * *Fecha*: Muestra el día, el mes y el año en el que se genera el reporte.
- * *Etiqueta*: Muestra el número de la etiqueta asignada al objeto.
- * *Tipo de objeto*: Indica si el objeto en la escena es un peatón o un vehículo.
- * *Observaciones*: Contiene la dirección en la que se mueve el objeto.
- * *Velocidad*: Muestra la velocidad media del objeto en el intervalo de tiempo que estuvo en escena.
- * *Hora*: Contiene la hora exacta en la que apareció el objeto.
- * *Infracción*: Indica si el objeto entró o no en zona restringida

Capítulo 5

RESULTADOS Y ANÁLISIS

En este capítulo se mostrarán los resultados obtenidos con el sistema que se describió en el capítulo 4. Además, se dará a conocer la eficiencia del sistema en términos cuantitativos y los resultados parciales por cada etapa que se siguió.

5.1. CONTROL DE ILUMINACIÓN

Un sistema de visión artificial que opera en zonas abiertas requiere de un control de iluminación para adecuar la imagen actual a sus variaciones. En el capítulo anterior (sección 4.2.2), se describió el proceso que se llevó a cabo para efectuar este control mediante compensación de los niveles de gris.

En la figura 5.1, se observa el efecto que tiene dicha compensación en la detección de movimiento. La imagen superior izquierda, de esta misma figura, representa el fondo de la escena; a la derecha, se muestra la imagen actual de la secuencia de video. La figura 5.1 (c) y 5.1 (d) son imágenes binarias, en las que los píxeles activos (en negro) representan los objetos en movimiento y los píxeles inactivos (en blanco) representan el fondo. Estas imágenes constituyen la segmentación de movimiento y de su análisis podemos obtener el resultado mostrado en las figuras 5.1 (e) y 5.1 (f).

A simple vista podemos observar el cambio de iluminación entre el fondo y la imagen actual, y ésta a su vez, se ve reflejada en la segmentación. Cuando ocurre una variación de la iluminación, ésta se detecta como movimiento, generando regiones amplias como las que se observan en la figura 5.1 (c). Esto constituye uno de los mayores obstáculos para la detección de movimiento y puede considerarse también como ruido pues opaca los objetos de interés.

Mediante el algoritmo de control de iluminación se logra resolver este problema por completo. En la figura 5.1 (d) se observa la segmentación obtenida al realizar este control y la eliminación total del ruido debido a cambios de iluminación bruscos como el que se muestra en esta escena.

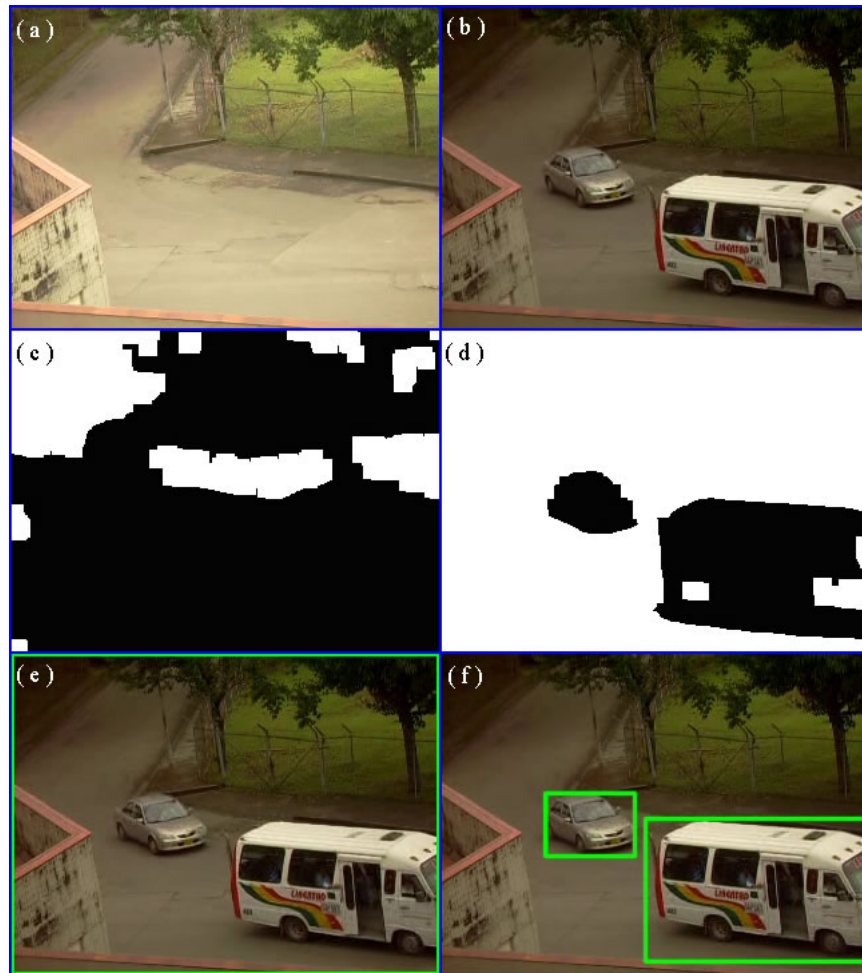


Figura 5.1. Efectos del control de iluminación: (a) Fondo. (b) Imagen actual con cambio en la iluminación. (c) y (e) Detección de movimiento sin compensación. (d) y (f) Detección de movimiento aplicando el algoritmo de control.

5.2. DETECCIÓN DE MOVIMIENTO

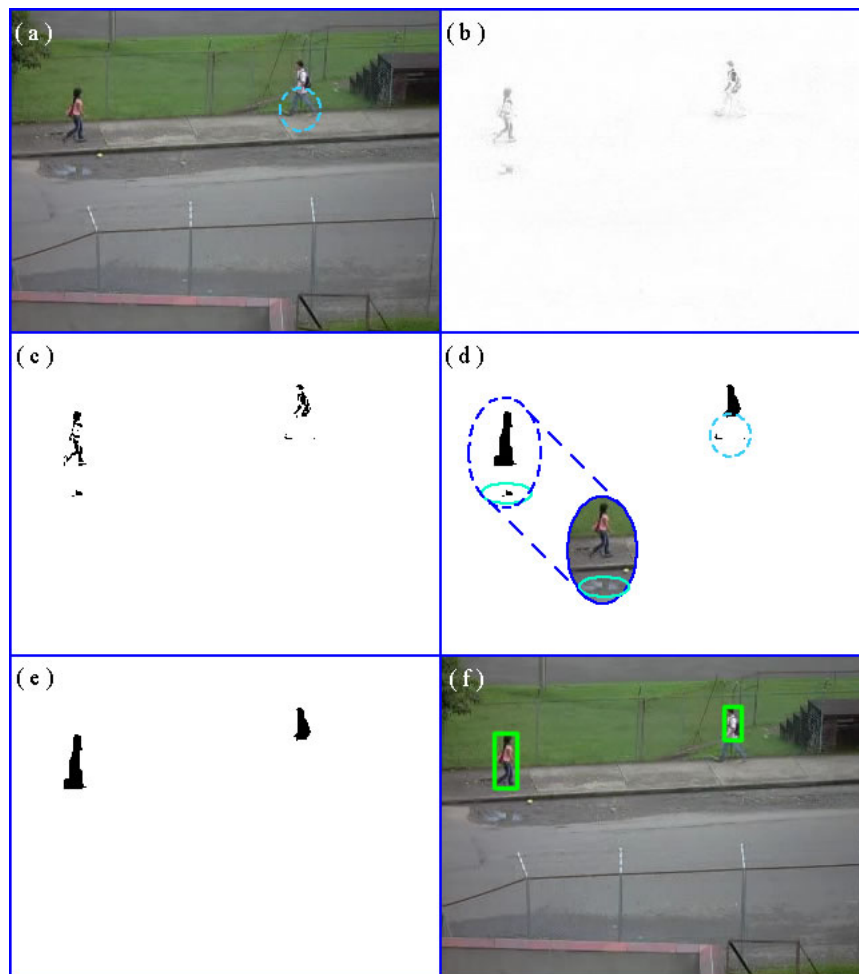


Figura 5.2. Pasos para la detección de movimiento y filtrado de ruido: (a) Frame actual. (b) Diferencia entre el fondo y el frame actual. (c) Umbralización. (d) Filtrado morfológico. (e) Filtrado por tamaño. (f) Resultado de la detección de movimiento.

La figura 5.2 muestra el resultado de cada uno de los pasos seguidos para llevar a cabo la tarea de la detección de movimiento hasta obtener finalmente la respuesta del sistema en esta etapa de segmentación.

La primera imagen de la figura 5.2, representa una de las escenas de prueba, en la que se tienen dos objetos en movimiento, en este caso se trata de dos peatones que avanzan en direcciones opuestas.

El paso inicial es la diferenciación con respecto al fondo (Figura 5.2 (b)). Lo siguiente consiste en la umbralización de esta imagen resultante, para ello se empleó una umbralización manual, en la que el usuario determina el valor del umbral en la etapa de calibración inicial que es expuesta con más detalle en el anexo A. Para esta escena y en general para todas las escenas de prueba se encontró de forma experimental que, el umbral que mejor segmentó los objetos fue de 40 dentro de un rango de niveles de gris entre 0 y 255. De este modo se obtiene la imagen de la figura 5.2 (c).

En esta última se puede ver que los objetos de interés no conforman regiones conexas definidas que los representen como una unidad (regiones en negro), por esta razón se utilizaron operaciones morfológicas de dilatación y erosión. Al aplicar este filtro morfológico se obtiene la imagen de la figura 5.2 (d), en la que se observa la presencia de ruido u objetos en movimiento que no corresponden a vehículos o peatones, y que en este caso específico se debe a la reflexión del peatón de la izquierda sobre el cúmulo de agua que hay en el pavimento.

El ruido también se debe, como se puede apreciar en esta misma figura, a pequeñas regiones que hacen parte del objeto pero que no están conectadas a él. La razón de ello, es que cuando los niveles de gris de los píxeles que conforman la región en movimiento son similares a los del fondo, la diferencia tiende a ser nula y en la etapa de umbralización estos píxeles desaparecen por completo. Esto conlleva a una incompleta segmentación de los objetos de interés, el cual es el caso del peatón de la derecha (Figura 5.2 (d)) cuya región, que desaparece tras la etapa de umbralización, está enmarcada en azul.

Luego del filtrado morfológico, la imagen pasa por otra etapa de filtrado en la que objetos pequeños son eliminados de la imagen (Figura 5.2 (e)). El criterio de selectividad de tamaño se determinó experimentalmente. Los objetos cuya suma de su ancho y su alto sea menor a 30 píxeles se consideran ruido y aquellos que sobrepasan este valor se consideran objetos de interés.

Es importante aclarar que el tamaño de los objetos depende de la distancia a la cual se ubique la cámara de video. En este trabajo se grabaron videos de escenas en las que la cámara se ubicó a diferentes distancias con el propósito de evaluar el alcance efectivo del sistema, sin embargo se encontraron ciertos inconvenientes cuando la distancia es grande, como se verá más adelante.

En las figuras 5.3 y 5.4, se puede observar ruido por motivos diferentes como el movimiento

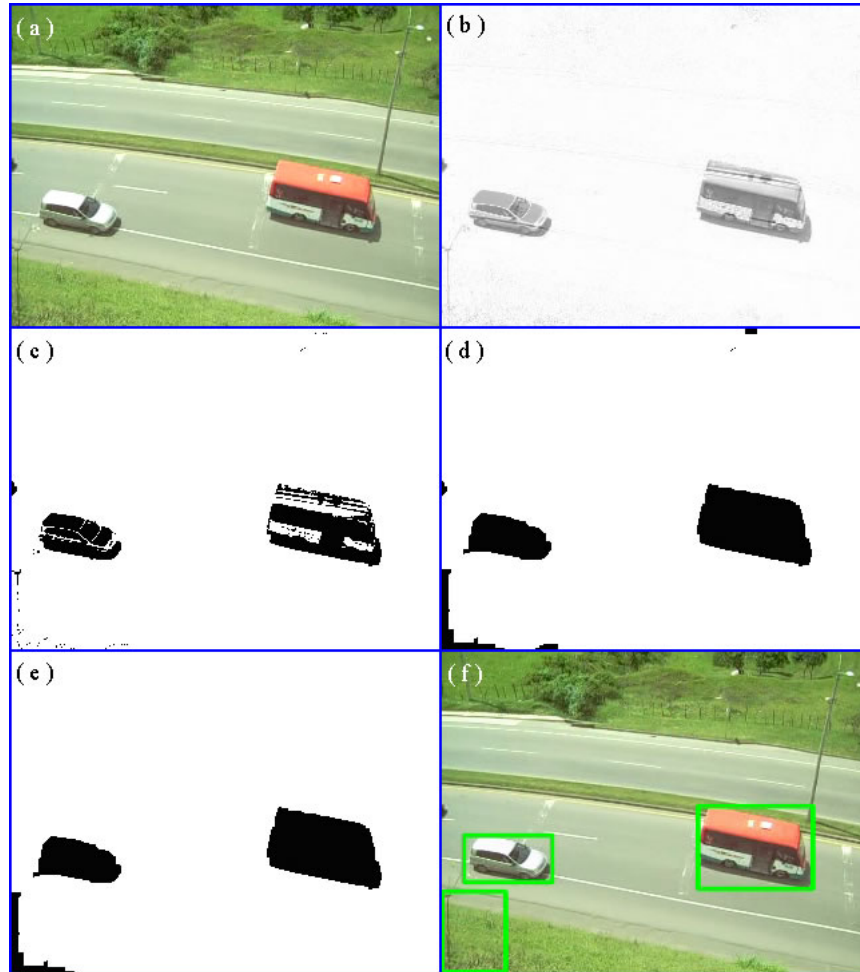


Figura 5.3. Detección de movimiento sin filtro gaussiano: (a) Frame actual. (b) Diferencia entre el fondo y el frame actual. (c) Umbralización. (d) Filtrado morfológico. (e) Filtrado por tamaño. (f) Resultado de la detección de movimiento.

de árboles y arbustos, a causa del viento y que hacen parte del fondo. Éste también es eliminado mediante el criterio de tamaño que se describió anteriormente. Sin embargo, existen casos en los cuales se presentan objetos en movimiento que hacen parte del fondo y que no alcanzan a ser eliminados en las etapas de filtrado por su tamaño significativo tal caso se puede ver en la figura 5.3, donde se produce una detección incorrecta.

Este problema se soluciona en la etapa de preprocesamiento. Como se comentó en el capítulo anterior, existe la opción de aplicar un filtro gaussiano a la imagen del frame actual, de este modo se atenúa el movimiento de los objetos que pertenecen al fondo como se puede ver en la figura 5.4. Aquí se puede apreciar con claridad la minimización del ruido de este tipo y su total eliminación tras pasar por el filtrado basado en el tamaño de los objetos.

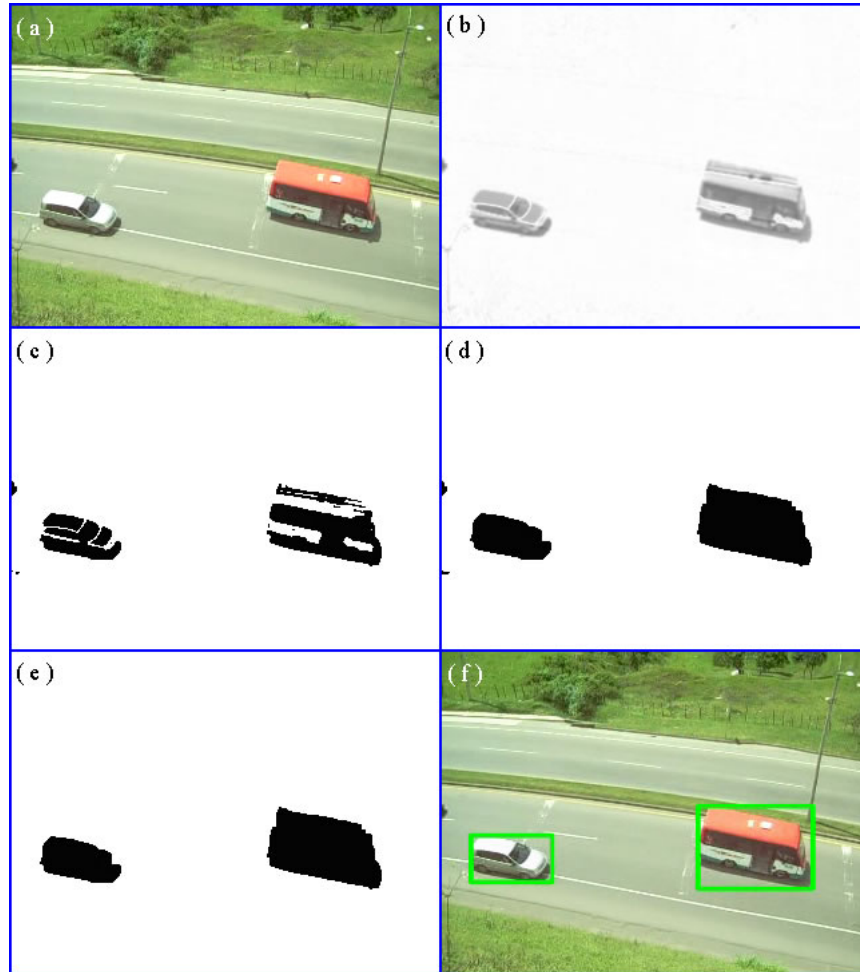


Figura 5.4. Detección de movimiento con filtro gaussiano. (a) Frame actual. (b) Diferencia entre el fondo y el frame actual. (c) Umbralización. (d) Filtrado morfológico. (e) Filtrado por tamaño. (f) Resultado de la detección de movimiento.

En general, el ruido debido a objetos en movimiento que hacen parte del fondo no presenta mayores problemas y es eliminado con relativa facilidad.

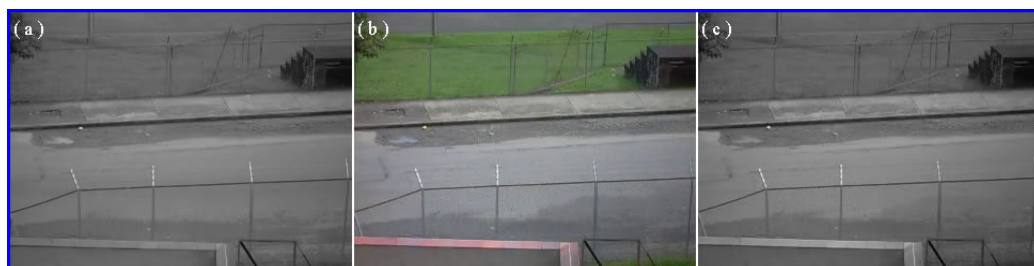


Figura 5.5. Actualización del fondo. (a) Inicialización del fondo o fondo promedio. (b) Variación leve de la iluminación en la imagen actual. (c) Actualización del fondo debida a esta variación.

Como se vio en la sección anterior (Sección 5.1), el control de iluminación reacciona cuando se presentan cambios bruscos de iluminación. Adicionalmente el sistema desarrollado en este trabajo implementa un algoritmo de actualización del fondo con el fin de eliminar ruido causado por leves variaciones de este tipo. En la figura 5.5 se muestra la imagen de fondo promedio que se obtiene al inicio de la aplicación (Figura 5.5 (a)), cuando la iluminación de la escena varía de forma natural (Figura 5.5 (b)), el algoritmo de actualización del fondo reajusta el fondo promedio para obtener, en este caso, la imagen de la figura 5.5 (c). La constante de actualización (α) se determinó de forma experimental y es de 0,02.

La tabla (tabla 5.1) muestra el porcentaje de detección de movimiento que se obtuvo para los diferentes escenarios de prueba que se presentaron y enumeraron en la sección 4.1.3 (Figura 4.3).

Escena (Video)	Número real de objetos	Número de objetos detectados	Porcentaje de detección de movimiento (%)
1	81	85	95,06
2	15	15	100,0
3	32	33	96,88
4	78	146	12,82
5	32	49	46,88
6	115	142	76,52
Total	353	470	71,36

Tabla 5.1. Porcentaje de detección de movimiento del sistema. Se muestran los resultados obtenidos por cada escena de prueba así como también el total de los objetos detectados y el porcentaje general de detección de movimiento.

A partir de esta última, se puede calcular el porcentaje de error y de detección del sistema desarrollado en el presente trabajo, teniendo en cuenta la totalidad de las escenas empleadas para probarlo. Obteniendo así que el porcentaje de error es del 28.64% y el porcentaje de detección, de 71.36%. Obsérvese que para las escenas 4, 5 y 6 se tienen los porcentajes más bajos de detección, incluso en la primera de ellas se llega a un porcentaje del 12.82%. Esto supondría que la detección de movimiento que se implementó no es muy eficiente, sin embargo, estas tres escenas (escena 4, 5 y 6), presentan características especiales.

El primero de estos escenarios, el cual se ilustra en la figura 5.6, está compuesto por zonas en las que la perspectiva varía de un punto a otro (Figura 5.6 (a)). El sistema presenta

dificultades en este tipo de escenarios debido a que no se ha tenido en cuenta el plano de la escena (muchas veces conocido como *ground plane* por su traducción al inglés) y sobre el cual se mueven los objetos.

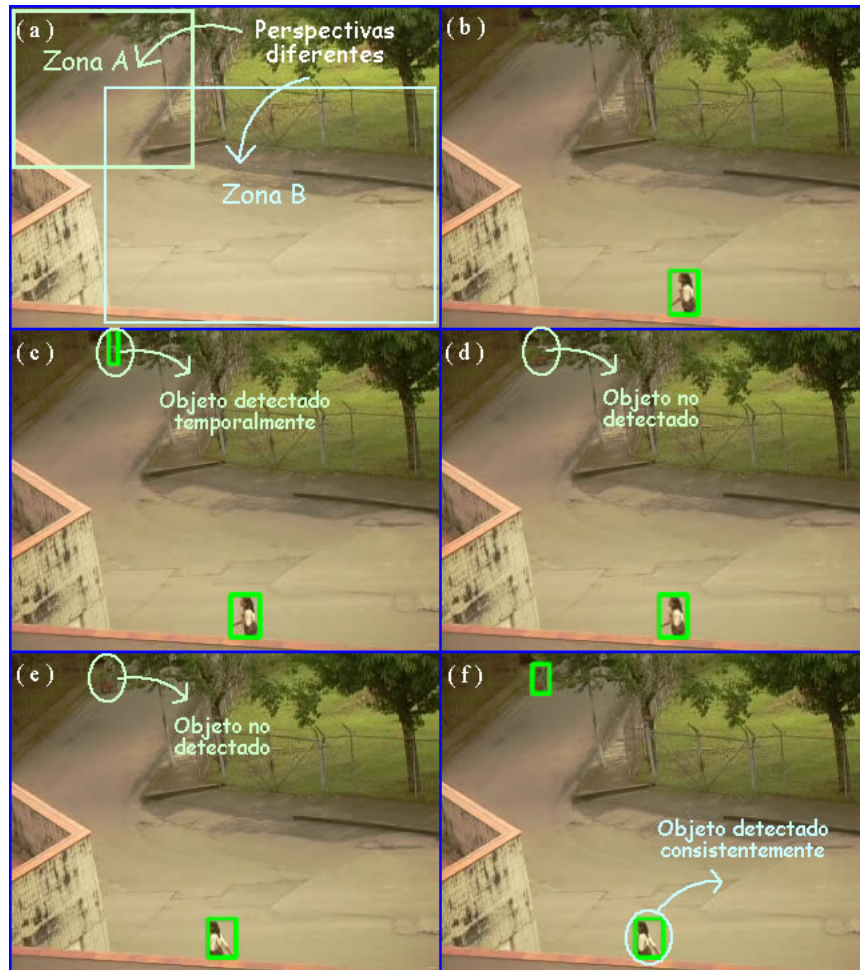


Figura 5.6. Detección de movimiento en la escena 4. Aquí se muestran algunos problemas de la detección de movimiento por el cambio de perspectiva dentro de una misma escena: (a) Zonas con perspectivas diferentes dentro del escenario. (b) Imagen actual. (c) Aparición de un objeto en la escena después de transcurridos unos cuantos cuadros (frames). (d) y (e) Imágenes en la secuencia de video en la que no se detecta el movimiento del objeto señalado. (f) detección de movimiento en el instante siguiente.

Debido a que el tamaño de los objetos en la imagen depende en gran medida del punto de vista de la cámara o del plano de la escena, los objetos que por ejemplo ingresan en la zona A, son muy pequeños y tienden a ser eliminados por el filtro de tamaño del que se habló anteriormente. Por esta razón, la detección de movimiento se muestra inconsistente ante este tipo de variaciones en el plano de la escena. Por el contrario, los objetos que ingresan en la

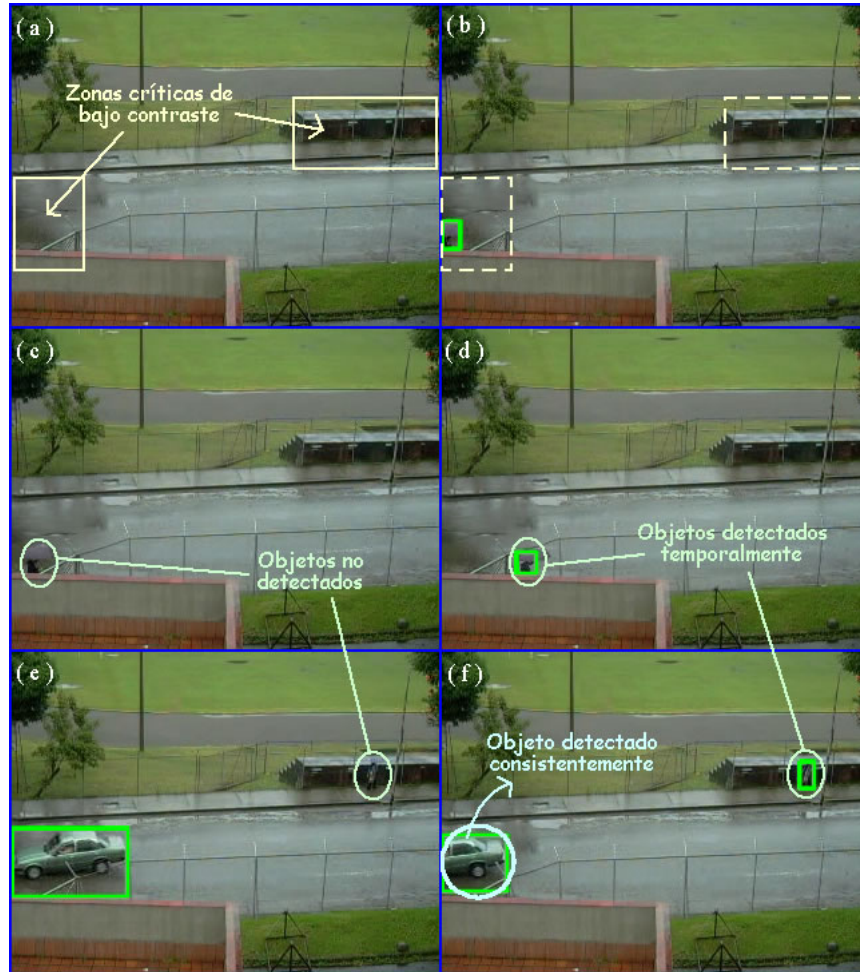


Figura 5.7. Detección de movimiento en la escena 5. Se muestran problemas de la detección de movimiento por la posición de la cámara y el bajo contraste del fondo con los objetos de interés. (a) Zonas críticas de bajo contraste. (b) Imagen actual con peatón que ingresa a la escena. (c), (d), (e) y (f) Muestra la consistencia e inconsistencia de la detección de algunos objetos dentro del escenario en instantes diferentes.

zona B son detectados eficientemente pues el tamaño de los objetos permanece casi constante a lo largo de este sector.

Las escenas 5 y 6, representan un caso en el que el video se graba desde un punto más alejado de los objetos de interés en comparación con los otros escenarios de prueba (Figuras 5.7 y 5.8).

En la primera (Figura 5.7), se tienen dos problemas, en primer lugar, existen dos regiones que presentan problemas de contraste con algunos objetos con píxeles cuyos valores de niveles de gris son muy parecidos al fondo. Otro aspecto importante es que los objetos como peatones, se ven muy pequeños desde el ángulo con el cual se toma el video y conforman regiones con un número reducido de píxeles, lo que se traduce en una menor resolución

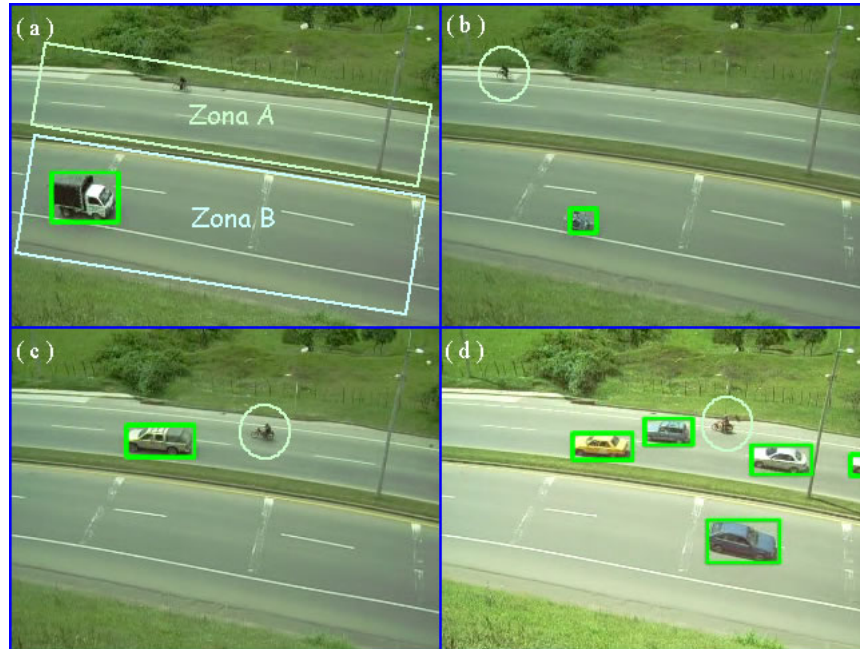


Figura 5.8. Problemas de detección de objetos muy pequeños. (a) Identificación de zonas en las que los objetos varían de tamaño. (b), (c) y (d) Casos en los cuales falla la detección de movimiento según su ubicación dentro de la escena.

para identificar los objetos de estudio. La disminución en la resolución influye de manera importante pues con ello se disminuye el contraste entre los objetos y el fondo. La figura 5.7 muestra las inconsistencias en la detección de movimiento por las razones aquí expuestas.

En la escena 6 (Figura 5.8), se presenta un caso parecido al de la escena 5, se tienen dos regiones de la imagen en las que los objetos se ven más o menos pequeños dependiendo de su ubicación en estas zonas, de este modo regiones en movimiento pequeñas como las correspondientes a peatones y vehículos pequeños tales como motos y ciclistas, que se encuentran o se desplazan dentro de la zona A, no son detectados consistentemente debido al filtro de tamaño que se aplica para eliminar ruido.

Debido a este problema fue necesario implementar una opción para seleccionar una región de interés en la que el usuario pueda delimitar una zona de la imagen en la cual se haga todo el análisis de movimiento de los objetos de estudio.

La tabla 5.2 muestra los resultados de la detección de movimiento en la que se seleccionaron regiones de interés en las escenas 4, 5 y 6, con el fin de eliminar las inconsistencias producidas por los problemas mencionados anteriormente y limitar el sistema a su rango de operación.

Escena (Video)	Número de objetos	Número de objetos detectados	Porcentaje de detección de movimiento (%)
1	81	85	95,06
2	15	15	100,0
3	32	33	96,88
4	78	80	97,44
5	32	34	93,75
6	54	53	98,15
Total	292	300	96,88

Tabla 5.2. Resultados de la detección de movimiento limitando el problema a escenas que no presentan cambios radicales del plano de la escena.

Con estos nuevos resultados tenemos que el porcentaje de error del sistema es de 3.12% y el de detección, 96.88%. Esta es la eficiencia del sistema adquiriendo video de escenas que se encuentran a una distancia similar a las de las escenas 1, 2 y 3.

5.3. SEGUIMIENTO DE OBJETOS DE INTERÉS

El seguimiento es una tarea del procesamiento de imágenes de video de muy alta complejidad por las múltiples interacciones que pueden ocurrir dentro de un escenario en el que existen objetos en movimiento. Para un seguimiento eficiente se requiere tener en cuenta los tipos de interacciones que se puedan dar, entre ellas tenemos: el agrupamiento de objetos, separación de objetos y oclusión, bien sea entre objetos en movimiento o entre un objeto en movimiento y objetos estáticos que pertenecen al fondo.

En este trabajo se tuvo en cuenta sólo dos de ellos: agrupamiento y separación de objetos en movimiento por ser los casos que más se presentaron en los videos de prueba del sistema. Cabe agregar que en el presente trabajo, la oclusión entre objetos en movimiento se trata como un caso de agrupamiento.

Al hacer el seguimiento de los objetos en movimiento, se hace posible contar los objetos presentes en la escena, por tal razón, el número de objetos que el sistema es capaz de contar en relación con el número real de objetos, constituye una medida de la eficiencia para realizar

la tarea de seguimiento.

La tabla 5.3 muestra esta relación, en cada escenario y la eficiencia total del sistema en cuanto al seguimiento. Cabe resaltar que para estos resultados se tuvieron en cuenta los rangos de operación efectivos del sistema, en otras palabras, en las mismas condiciones que para la detección de movimiento cuyos resultados se mostraron en la tabla 5.2, por las desventajas que presenta este sistema para realizar análisis de imágenes de video cuando se adquieren secuencias de imágenes a distancias relativamente lejanas de los objetos de estudio.

Escena (Video)	Número real de objetos	Número de objetos contados	Porcentaje de eficiencia (%)
1	81	83	97,53
2	15	14	93,33
3	32	37	84,38
4	78	82	94,87
5	32	38	81,25
6	54	53	98,15
Total	292	307	91,58

Tabla 5.3. Eficiencia del sistema para realizar el seguimiento de objetos. Se muestran los valores por cada escena y el total.

La tabla anterior (Tabla 5.3) muestra el número de objetos contados por el sistema independientemente de cómo se desarrollen las interacciones entre objetos durante toda la secuencia de video. No obstante, las interacciones entre objetos influyen en gran medida en el correcto seguimiento de los objetos en movimiento. En la figura 5.9 se ilustran algunos casos relacionados con el agrupamiento y separación de objetos, y que muestran cómo éstas influyen en el correcto conteo de objetos.

La figura 5.9 (a), muestra un caso muy frecuente en el que dos o más objetos ingresan juntos a la escena conformando una sola región en movimiento. El sistema al detectar una sola región conexas, etiqueta al objeto de acuerdo al orden de entrada en la escena y no como un grupo. Este hecho constituye un error en el conteo, pues está sumando un solo objeto a la cuenta y no dos, como correspondería en este caso en particular.

El caso b de esta misma figura, es cuando ocurre totalmente lo contrario, un objeto que está constituido por dos regiones que no están conexas y que el sistema asume como dos

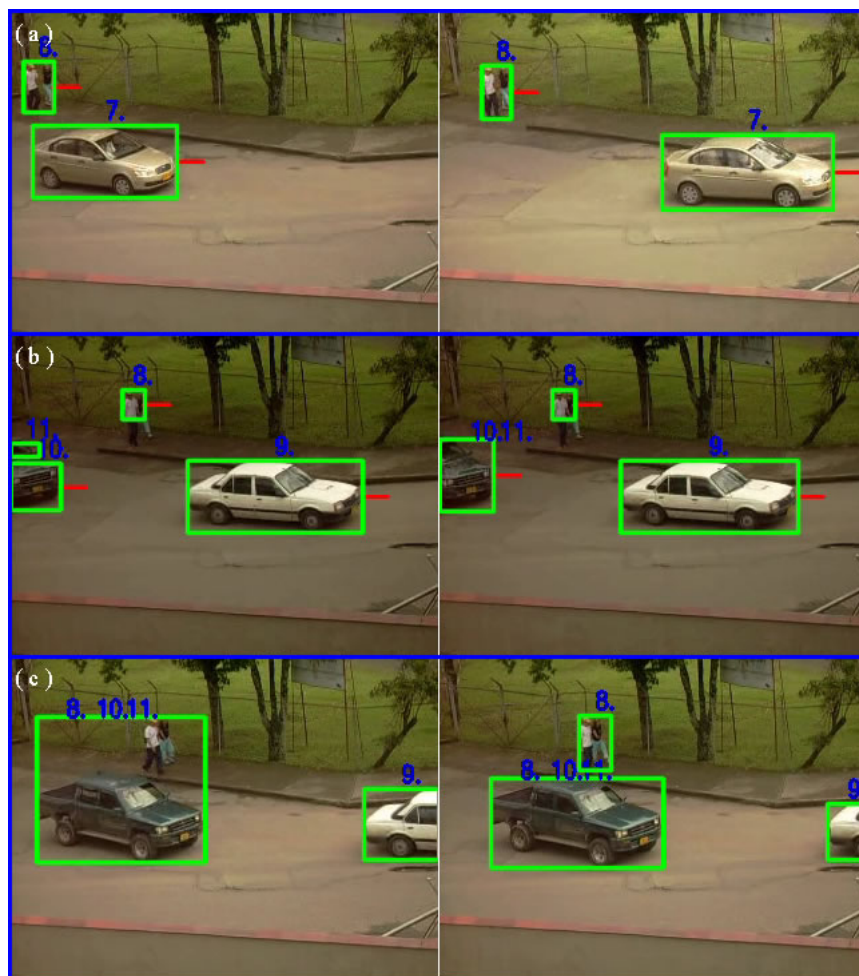


Figura 5.9. Problemas de la detección de grupos. (a) Caso de un grupo que entra en la escena 3 y que es detectado como un solo objeto. (b) Caso de un objeto que se detecta como grupo por la segmentación. (c) Separación con asignación incorrecta de etiquetas.

objetos independientes, aquí suceden dos problemas, el primero, el sistema está contando un objeto de más y el segundo, en el siguiente frame, se detecta un grupo que no existe realmente. Esto genera otro problema que se puede apreciar en la figura 5.9 (c), en donde este objeto, que el sistema asume como grupo conforma otro grupo con otro objeto y al momento de separarse, el sistema asigna de forma incorrecta las etiquetas debido a que hay menos objetos que etiquetas.

La figura 5.10 muestra una ampliación del primer caso mostrado en la figura 5.9 (a), donde un grupo, que es detectado como un solo objeto en un principio, se separa unos segundos después. Éste no constituye un error muy crítico ya que el sistema tiene en cuenta el objeto que aparece y lo suma como uno nuevo. Por el contrario, el evento b (Figura 5.9(b)), si representa un problema porque cuenta un solo objeto siendo en realidad tres y éste se

mantiene de este modo durante todo el tiempo que permanece en la escena.

En la figura 5.11 podemos ver diferentes interacciones entre objetos en movimiento que son identificadas correctamente.

Cabe resaltar que para realizar el seguimiento se implementaron dos técnicas, una de ellas y la que finalmente fue la escogida para realizar esta tarea, es la de seguimiento de regiones. La otra opción que se ensayó fue el filtro de kalman, que no fue utilizada en el sistema porque no superó a la anterior en simplicidad y costo computacional, además, los resultados que arrojó fueron muy similares a los obtenidos por la técnica de seguimiento de regiones.

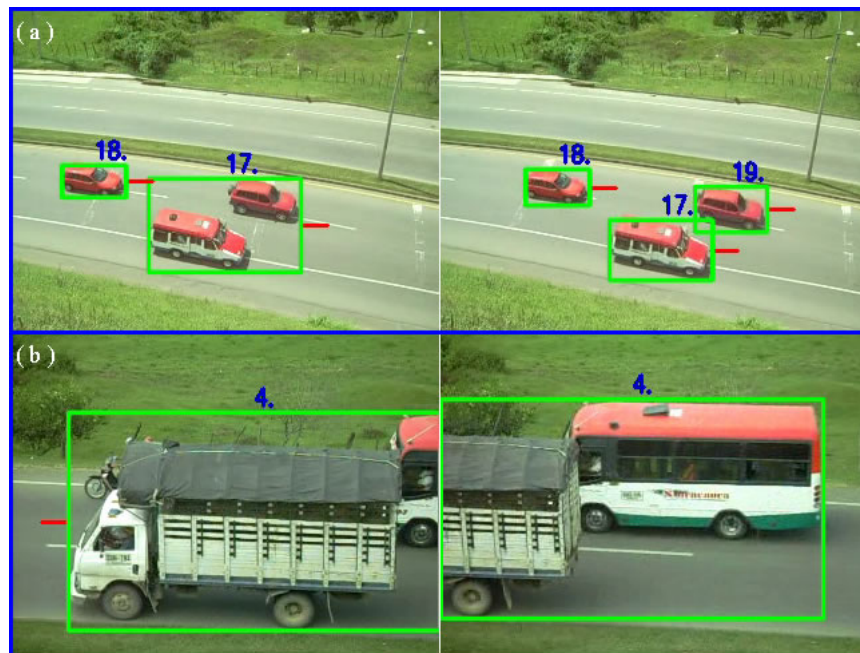


Figura 5.10. Interacción entre objetos. (a) Caso de un grupo que ingresa a la escena 6 y se separa después de un tiempo. (b) Grupo que es detectado como un solo objeto de principio a fin.

La eficiencia del sistema, como se puede observar en la tabla de resultados 5.3, es ligeramente superior al 90%, la cual es muy buena considerando que no se tienen condiciones ambientales y de iluminación controladas, y a pesar de algunos inconvenientes con el manejo de agrupamiento y separación de objetos.

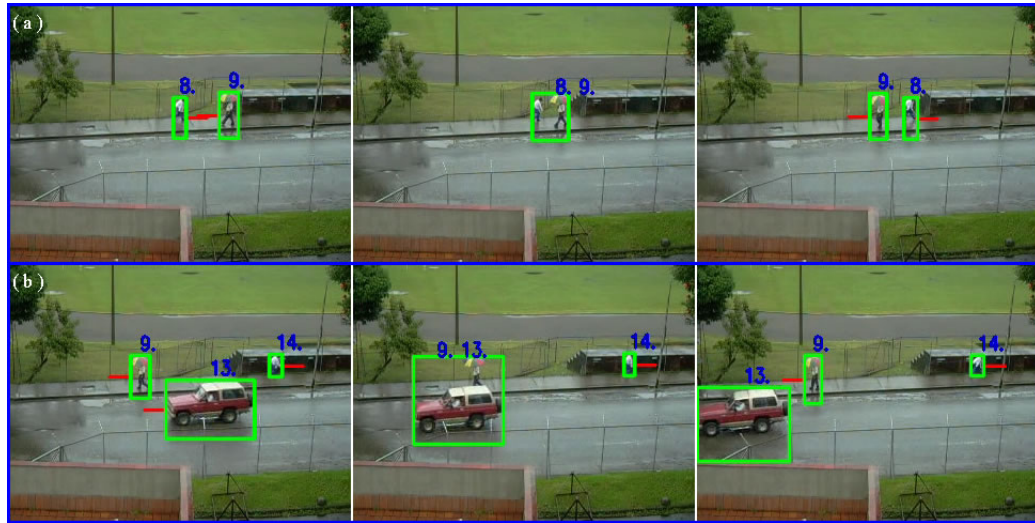


Figura 5.11. Agrupamiento y separación de objetos. Muestra casos de interacción entre objetos que son detectadas correctamente. (a) y (b) Agrupamiento y separación de objetos que se mueven en direcciones opuestas y en la misma dirección, respectivamente.

5.4. RECONOCIMIENTO DE OBJETOS

En un principio se tuvo en cuenta varios descriptores como el factor de forma, la compactibilidad, el color, el esqueleto del objeto, de las cuales se ha comentado en las secciones 2.4.4 y 2.4.5 con más detenimiento, sin embargo, el descriptor que se seleccionó para realizar la etapa de diferenciación entre clases de objetos fue el área, la razón de ello es que arrojó los mejores resultados de reconocimiento.

Las otras características presentaron inconvenientes por diferentes motivos. La compactibilidad, por ejemplo, no diferenció adecuadamente los objetos segmentados porque tanto vehículos como peatones, presentan formas complejas y esta característica se usa partiendo del hecho de que los peatones poseen mayor complejidad en su forma que los vehículos. Esto es cierto cuando se controlan las condiciones de iluminación, el cual no es el caso para este sistema.

Para utilizar el color como característica que distingue un objeto, se hizo el promediado de color de todos los píxeles que lo conforman. Este descriptor sería una opción si permaneciera constante a lo largo del tránsito del objeto por la escena, sin embargo, no ocurre así por el tipo de segmentación empleada en el que en algunos casos se obtiene sólo una parte de objeto.

Por otro lado, la esqueletización, otra característica tenida en cuenta, presentó problemas de implementación por su alto costo computacional, recordemos que la esqueletización es un proceso de iteración y en un sistema en el que se procesa tanta información en tan poco tiempo, esto es, cada frame, añadir un proceso como el de esta característica representa disminuir la eficiencia del sistema en términos de velocidad de procesamiento.

Finalmente el área del objeto constituye el descriptor más adecuado en este sistema, debido a la gran diferencia de tamaño que existe entre vehículos y peatones y también porque va acompañado de un proceso de entrenamiento en el que se tienen en cuenta la posición de la cámara con respecto a los objetos de interés.

Este entrenamiento se realiza en la etapa de calibración inicial (en el Anexo A se muestran los pasos necesarios para hacerlo), éste es un paso definitivo ya que mediante él se establecen los rangos de tamaño dentro de los cuales se clasifican los objetos en movimiento, en las dos categorías principales: vehículos y peatones.

La figura 5.12 muestra diferentes escenas en las que se hizo reconocimiento y en donde se puede apreciar la asignación de etiquetas según el tipo de objeto.

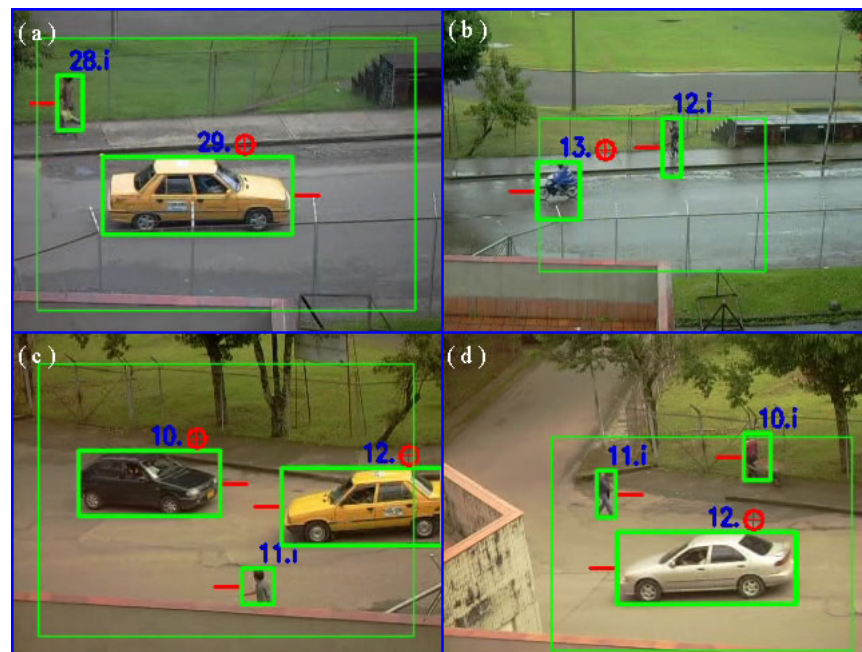


Figura 5.12. Reconocimiento de objetos en movimiento. (a), (b), (c) y (d) Etiquetado de los objetos de interés según la clase a la que pertenecen en las escenas 1, 5, 3 y 4 respectivamente.

Un aspecto importante a la hora de entrenar el sistema es la región de interés. La figura 5.13 muestra el reconocimiento realizado sin zona de interés, es decir, haciendo el reconocimiento en toda la imagen y el realizado limitando la imagen a una región rectangular dentro de la misma imagen.

Cuando un objeto entra en la escena (Figura 5.13 (a)), en este caso un vehículo, se mide el tamaño del objeto con el fin de determinar si pertenece a una determinada clase. Por lo general, los objetos que entran en escena generan regiones de movimiento inicialmente pequeñas que son fáciles de confundir con las de los peatones. Por esta razón es necesario dar al menos unos diez píxeles del borde de la imagen para dar lugar a la formación completa del objeto y por tanto al adecuado reconocimiento de los objetos. La figura 5.13 (b), muestra el caso en el que se emplea una región de interés y la diferencia en cuanto a reconocimiento entre estos dos casos.

Las tablas 5.4 y 5.5 muestran las matrices de confusión para peatones y vehículos respectivamente. En ellas se observan cuántos objetos se reconocen como peatones y cuántos como vehículos de un total determinado de peatones. Así mismo para el caso de los vehículos.

Escena (Video)	Número total de peatones	Peatones bien clasificados	Peatones clasificados como vehículos	Porcentaje de eficiencia (%)
1	9	7	2	77,78
2	0	0	0	----
3	7	5	2	71,43
4	20	16	4	80,00
5	9	7	2	77,78
6	0	0	0	----
Total	45	35	10	76,75

Tabla 5.4. Matriz de confusión del sistema para los peatones.

Vemos la gran diferencia entre la eficiencia del sistema en cuanto al reconocimiento de vehículos y de peatones. Con un porcentaje de reconocimiento de 76.75% para el caso de los peatones, más bajo que el de los vehículos, con 99.04%, aunque hay que destacar que es un valor relativamente alto considerando los problemas que generan estos errores y que veremos a continuación.



Figura 5.13. Diferencia en reconocimiento de acuerdo a la zona de interés. (a) La región de interés es toda la imagen, el sistema hace la clasificación en el momento en el que el objeto entra en escena. (b) Zona de interés enmarcada en verde, se muestra la clasificación del objeto una vez éste se forme adecuadamente.

Las figuras 5.14 y 5.15, son ejemplos de cómo el sistema confunde algunos objetos y les asigna etiquetas que no corresponden en realidad a la clase a la que pertenecen. Vemos en la figura 5.14, el caso en el que inicialmente se asignan de forma correcta las etiquetas a los objetos en movimiento (Figura 5.14 (a) y 5.14 (d)), sin embargo, luego del agrupamiento (Figura 5.14 (b) y 5.14 (e)), se genera el error al etiquetar al peatón como si fuera vehículo y a este último como peatón. Claramente es un error en el manejo de grupos y no del precisamente del reconocimiento.

Otro evento que se da y que se puede ver en la figura 5.15, es aquel en el que un grupo de objetos avanzan juntos y el sistema los detecta como uno sólo (Figura 5.15 (a)) y luego de cierto tiempo se separan (Figura 5.15 (b)), al separarse, uno de ellos continúa con la misma

Escena (Video)	Número total de vehículos	Vehículos clasificados como peatones	Vehículos bien clasificados	Porcentaje de eficiencia (%)
1	72	0	72	100,00
2	15	0	15	100,00
3	25	1	24	96,00
4	58	1	57	98,27
5	23	0	23	100,00
6	54	0	54	100,00
Total	247	2	245	99,04

Tabla 5.5. Matriz de confusión del sistema para los vehículos.

etiqueta y el nuevo objeto se enumera y se le asigna la etiqueta correspondiente a la clase. En este caso el peatón 19 permanece como vehículo y el 20 es el nuevo objeto. Esta es una de las desventajas de clasificar los objetos por el tamaño, ya que los grupos forman regiones conexas de gran tamaño que se confunden con los vehículos.

5.5. RENDIMIENTO DEL SISTEMA

Es importante conocer el coste computacional de cualquier sistema de visión artificial para saber las características básicas que debe tener el equipo en el cual se implementará la aplicación, por lo que para este sistema, se han medido los tiempos que éste tarda en realizar cada una de las etapas que lo componen, así como se muestra en la tabla 5.6.

Estos resultados fueron obtenidos, para una captura de video de 30 fps, en un equipo con las siguientes características:

- * Sistema Operativo Windows XP Profesional.
- * Procesador Intel Pentium 4 de 2.40 GHz
- * Memoria RAM de 256 MB



Figura 5.14. Errores en la asignación de las respectivas etiquetas a los objetos de interés. (a) y (d) Etiquetado, inicialmente correcto. (b) y (e) Agrupamiento de objetos de etiquetas diferentes. (c) y (f) Error en la asignación de las etiquetas correspondientes.

La aplicación fue desarrollada en la plataforma Visual C++ 6.0, haciendo uso de las librerías para procesamiento de imágenes OpenCV de Intel.

Etapa	Tiempo de procesamiento (ms)	Frames procesados en un segundo
Captura	33	30,00
Detección de movimiento	40	24,67
Seguimiento	41	24,42
Reconocimiento	42	23,61

Tabla 5.6. Tiempos de procesamiento del sistema por etapas.

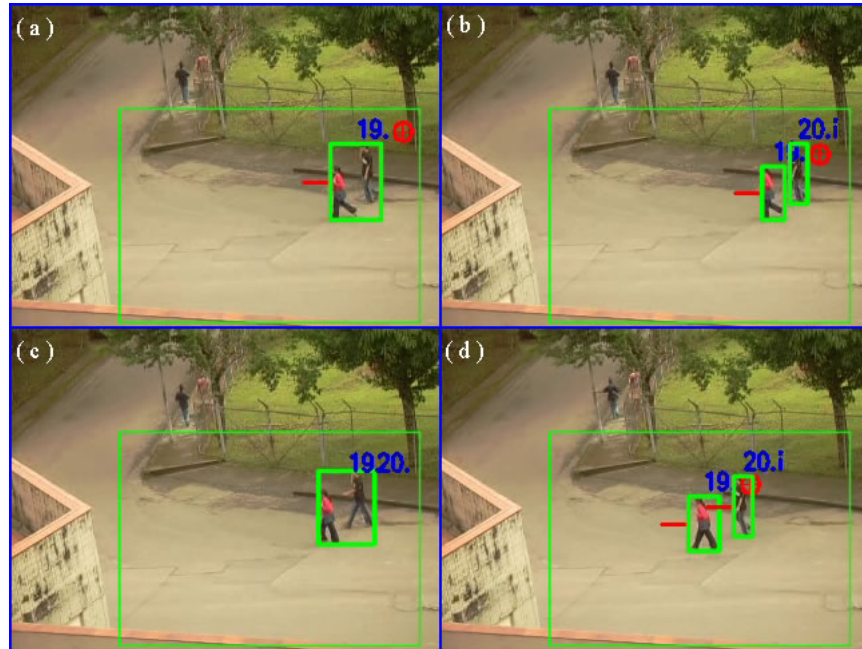


Figura 5.15. Error en la clasificación de objetos. (a) Un grupo de peatones es detectado como un solo objeto y además etiquetado como vehículo. (b) Separación del grupo, uno de los objetos permanece con la misma etiqueta mientras que aquel que se detecta como nuevo es reconocido correctamente como peatón. (c) Agrupamiento de estos dos objetos. (d) Separación, con incorrecta asignación de etiquetas.

En la tabla anterior se puede observar que los tiempos de procesamiento aumentan a medida que se implementa una nueva etapa. Sin embargo este incremento no es significativo, ya que éstos son bajos y el número de frames procesados cada segundo está por encima de los 20 fps y no muy lejano de los 30 fps. Esto significa que el retardo producido en el sistema es muy pequeño y el coste computacional es bajo.

CONCLUSIONES

Se implementó un sistema de visión artificial automatizado para la vigilancia de propiedades, públicas o privadas, mediante procesamiento digital de imágenes de video y reconocimiento de patrones. Este sistema fue probado con escenas de exteriores que presentaban diferentes condiciones de iluminación y ambientales, con un buen grado de funcionalidad y bajo coste computacional.

El algoritmo de control de iluminación que se diseñó en este trabajo elimina por completo el ruido que se genera debido a los cambios bruscos en la iluminación. Esto es importante, ya que permite que el sistema responda correctamente en ambientes en donde es difícil controlar este tipo de variables.

La técnica de sustracción del fondo para la detección de movimiento es idónea para la segmentación de objetos en movimiento en escenas con condiciones de iluminación variables, si se hace uso de los métodos estadísticos para la actualización del fondo. Estos permiten que el sistema se acople a los cambios lentos de iluminación, debidos a la variación normal de la intensidad de la luz en el ambiente.

El seguimiento por regiones es una técnica que permite realizar eficientemente el seguimiento de objetos en movimiento y que además es de bajo coste computacional, lo que permite que el sistema funcione bien sin hacer gran uso de los recursos del PC.

La etapa de reconocimiento de los objetos en movimiento depende directamente de la rigurosidad con la que se realiza el entrenamiento del sistema y los parámetros de configuración escogidos por el usuario.

Una función adicional del sistema es la opción de guardar el video adquirido por las cámaras de seguridad, lo que puede facilitar la labor de un guarda de seguridad en algún momento dado, como por ejemplo, para recolectar evidencia de algún crimen.

Se diseñó un sistema (cámara de video - tarjeta digitalizadora - PC) capaz de realizar un conteo del número peatones y/o vehículos que entraron en la escena. Esto es de gran utilidad en los sistemas de control de tráfico.

La interfaz gráfica implementada en el sistema facilita la realización de todas las tareas necesarias para la vigilancia de una zona, así como la visualización de la información más relevante mediante reportes generados por la aplicación.

RECOMENDACIONES Y TRABAJO FUTURO

La vigilancia de propiedades mediante visión artificial es una tarea altamente compleja, que como se puede ver en este trabajo, es difícil de realizar con eficiencia y exactitud sin limitar el problema a ciertas condiciones de operación. Sin embargo, los resultados del presente trabajo muestran que la visión artificial es el mecanismo ideal por el cual se puede lograr un sistema de vigilancia automatizado y robusto si se tienen en cuenta las siguientes recomendaciones:

En cuanto a la detección de movimiento, tarea indispensable para proseguir con los análisis posteriores, se requiere que el método de sustracción del fondo vaya de la mano con otra técnica de detección de movimiento, como la de diferenciado temporal, por ejemplo, para mejorar el proceso de segmentación y reducir los problemas que se tienen cuando los objetos en movimiento tienen niveles de gris muy similares a los del fondo.

Es necesario, además, implementar un algoritmo de detección de luz y sombras con el fin de minimizar el número de grupos innecesarios y reducirlo sólo a los objetos que en realidad se muevan juntos. Existen muchos casos en los cuales objetos que se mueven dentro de la escena generan sombras, que son detectadas como movimiento, y que además conectan objetos cercanos formando regiones únicas que el sistema tiende a ver como grupos, pero que en realidad no lo son. Esta técnica también evitaría que las regiones que pertenecen a los objetos aumenten o disminuyan su tamaño por las sombras y produzca una deficiente clasificación.

El número y el tipo de interacciones que se pueden presentar entre los objetos en movimiento con los del fondo es amplio y cada una de ellas requiere de un análisis específico. En este trabajo sólo se consideraron dos de ellas, pero el sistema puede ampliarse a los otros casos como la oclusión entre objetos y de objetos con el fondo. La consideración de nuevas formas de interacción de objetos de interés, es necesaria para que el sistema comprenda los diferentes eventos que se presentan en un determinado escenario y evitar errores en el seguimiento de los objetos de estudio. Una técnica ideal para tratar los casos de oclusión es el filtro de kalman por su carácter predictivo y del cual se habló en la sección 3.2.1.

Otro aspecto a tener en cuenta es la consideración del plano de la escena en las medidas que se requieran de la imagen, como por ejemplo el tamaño y la dirección de los objetos, esto,

con el fin de obtener una escala de tamaño que permita una mejor clasificación bien sea por el área o por otra característica de forma que se desee emplear como descriptor.

Finalmente, un sistema de vigilancia ideal, es aquel cuyo funcionamiento efectivo sea de 24 horas, por esta razón un trabajo futuro interesante, sería el de expandir el funcionamiento de este sistema en horas de la noche, empleando cámaras de video infrarrojas o un control de la iluminación externa en la escena.

Anexo A

MANUAL DEL USUARIO

El algoritmo es la parte más importante del sistema, sin embargo se debe tener en cuenta que la interfaz gráfica del usuario también es importante para que el programa se pueda ser utilizado fácilmente.

Una vez ejecutado el programa, aparecerá una ventana como la que se muestra en la Figura A.1.

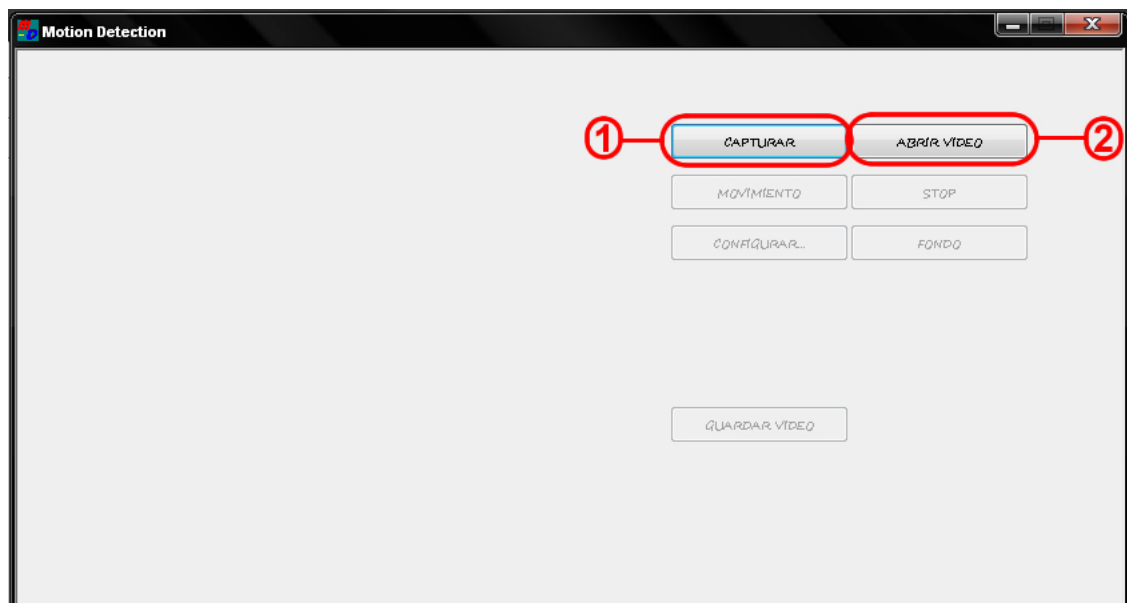


Figura A.1. Interfaz gráfica de usuario del sistema.

En esta ventana se presentan dos opciones (Según la numeración de la Figura A.1):

1. CAPTURAR

Esta opción permite capturar video desde una cámara externa. Es indispensable que los controladores de la cámara se encuentren instalados y que sean compatibles con el sistema

operativo sobre el cual se está ejecutando el programa. En caso de que se esté empleando una tarjeta digitalizadora, los controladores de esta deben estar también instalados. El programa mostrará el video en una resolución de 320x240 píxeles.

2. ABRIR VIDEO

Esta opción permite cargar un archivo de video con extensión (*.avi). Después de dar click sobre esta opción se desplegará en pantalla una ventana como la que se muestra en la Figura A.2, en donde se podrá escoger la ruta y el archivo de video deseado. Es indispensable que el video cargado se encuentre sin compresión. El programa mostrará el video en una resolución de 320x240 píxeles.

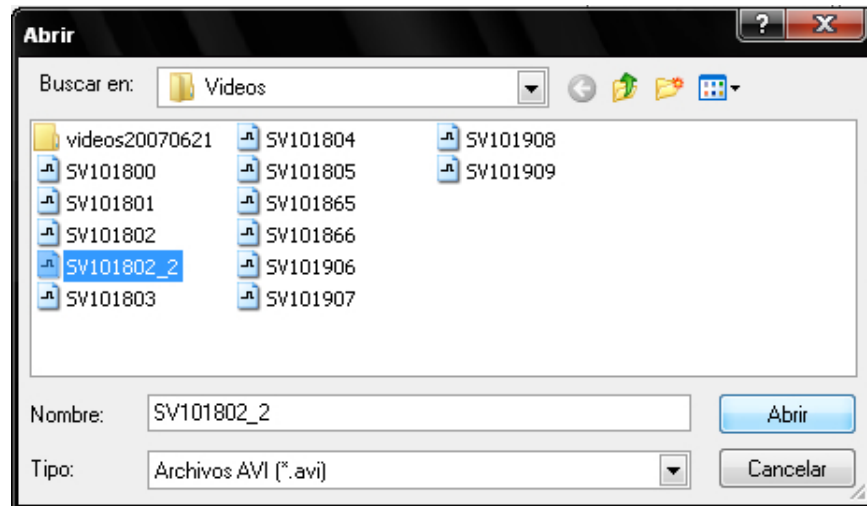


Figura A.2. Opción Abrir Video.

Después de iniciar la captura o cargara el archivo de video, este es desplegado en la ventana del programa como se muestra en la Figura A.3.

La nueva ventana presenta los siguientes elementos (Según la numeración de la Figura A.3):

3. PROGRESO DEL VIDEO Y NÚMERO DE *frames*

Este es un elemento de información del sistema que se presenta solamente cuando se ha cargado un video desde archivo. En él se muestra información sobre el progreso del video y el número de *frames*.

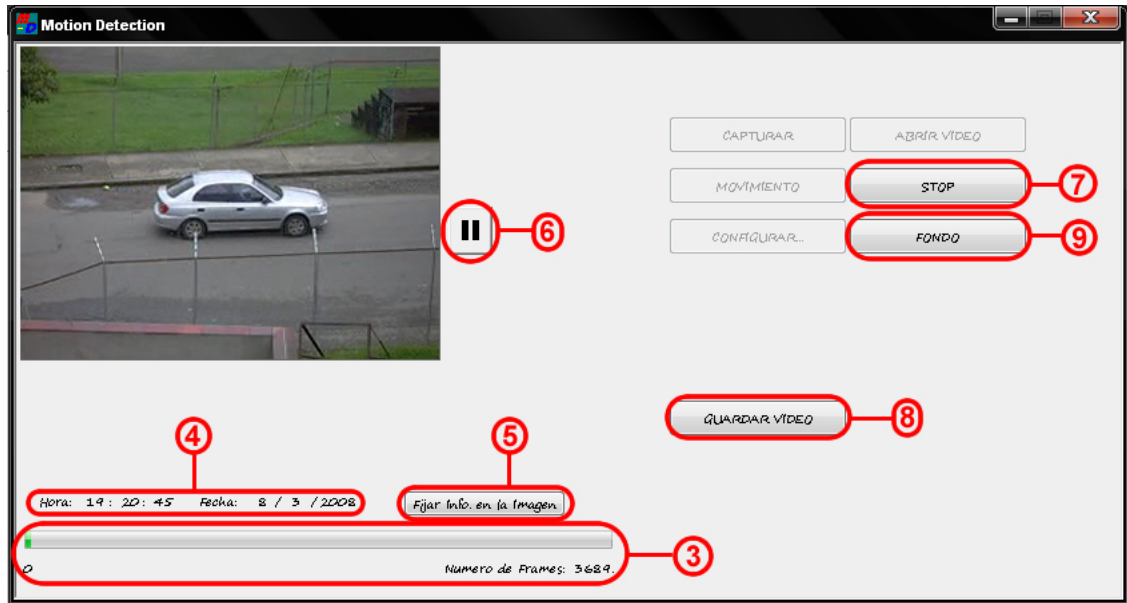


Figura A.3. Interfaz gráfica de usuario después de iniciar la captura o abrir el video.

4. INFORMACIÓN DE HORA Y FECHA

Muestra la hora y la fecha actual. La primera parte es la información de hora con el formato: hora : minuto : segundo. La segunda parte es la fecha con el formato: día : mes : año.

5. FIJAR INFORMACIÓN EN LA IMAGEN

Esta opción fija en el video la información de hora y fecha que se muestra en el elemento número 4.

6. PAUSA

Este botón está activado únicamente cuando el video que se muestra en pantalla es cargado desde archivo. Al hacer click sobre este, el video será detenido temporalmente hasta volver a presionar el botón.

7. STOP

Al presionar este botón el video o la captura se detendrán completamente la captura o el video.

8. GUARDAR VIDEO

Esta opción permite guardar en un archivo el video desplegado por el programa. Al activar esta opción, aparece una ventana como la que se muestra en la A.4.

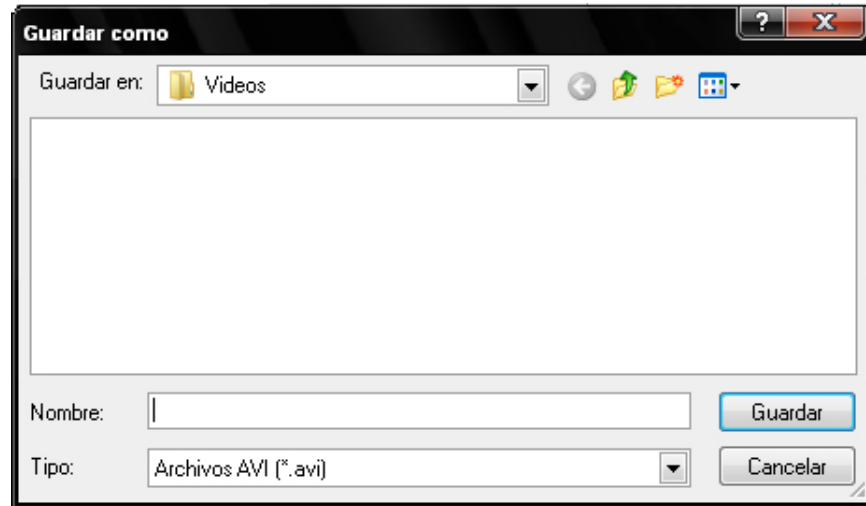


Figura A.4. Opción Guardar Video.

9. FONDO

Esta opción permite inicializar el fondo de referencia para la detección de movimiento. Al presionar este botón, se crea una nueva ventana en donde se muestra el fondo obtenido después de promediar 100 imágenes (Figura A.5).

Una vez inicializado el fondo, se activa una nueva opción como se observa en la Figura A.5:

10. MOVIMIENTO

Este botón inicia el algoritmo de detección de movimiento mediante la técnica de sustracción del fondo. La detección de las regiones en movimiento es mostrada en una nueva ventana (Figura A.6).

Una vez activada la detección de movimiento, se presentaran los siguientes elementos (según la numeración mostrada en la Figura A.6):

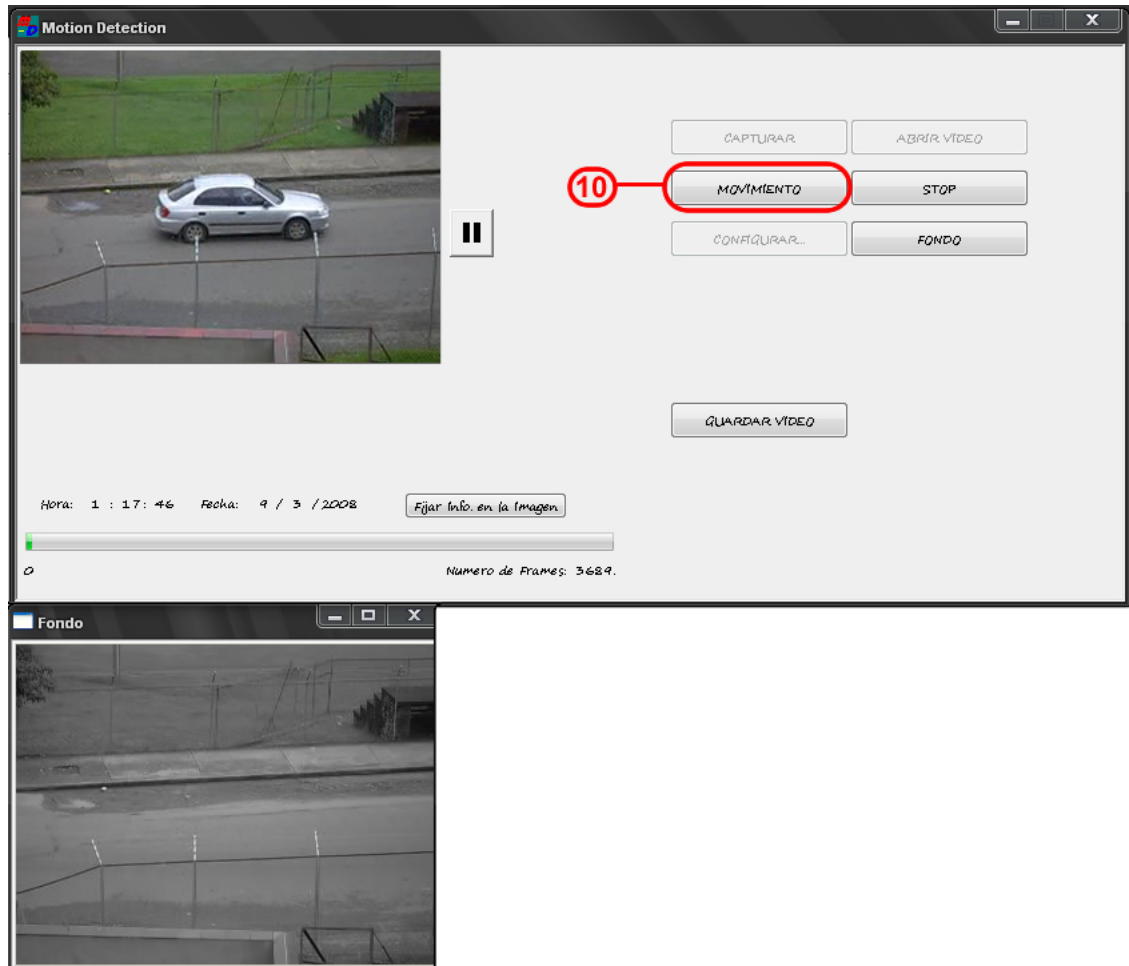


Figura A.5. Interfaz gráfica cuando la opción Fondo está activada.

11. SENSOR DE MOVIMIENTO

Muestra, en color verde, la cantidad de píxeles en movimiento sobre la barra mostrada en la Figura A.6. Esta medida se encuentra en kilopíxeles.

12. ALERTA

Este indicador muestra cuando se produce una intrusión en una zona restringida. Cuando este indicador se encuentra de color verde, quiere decir que no existen intrusos, de lo contrario se encontrará de color rojo. La definición de las zonas restringidas se tratará mas adelante.

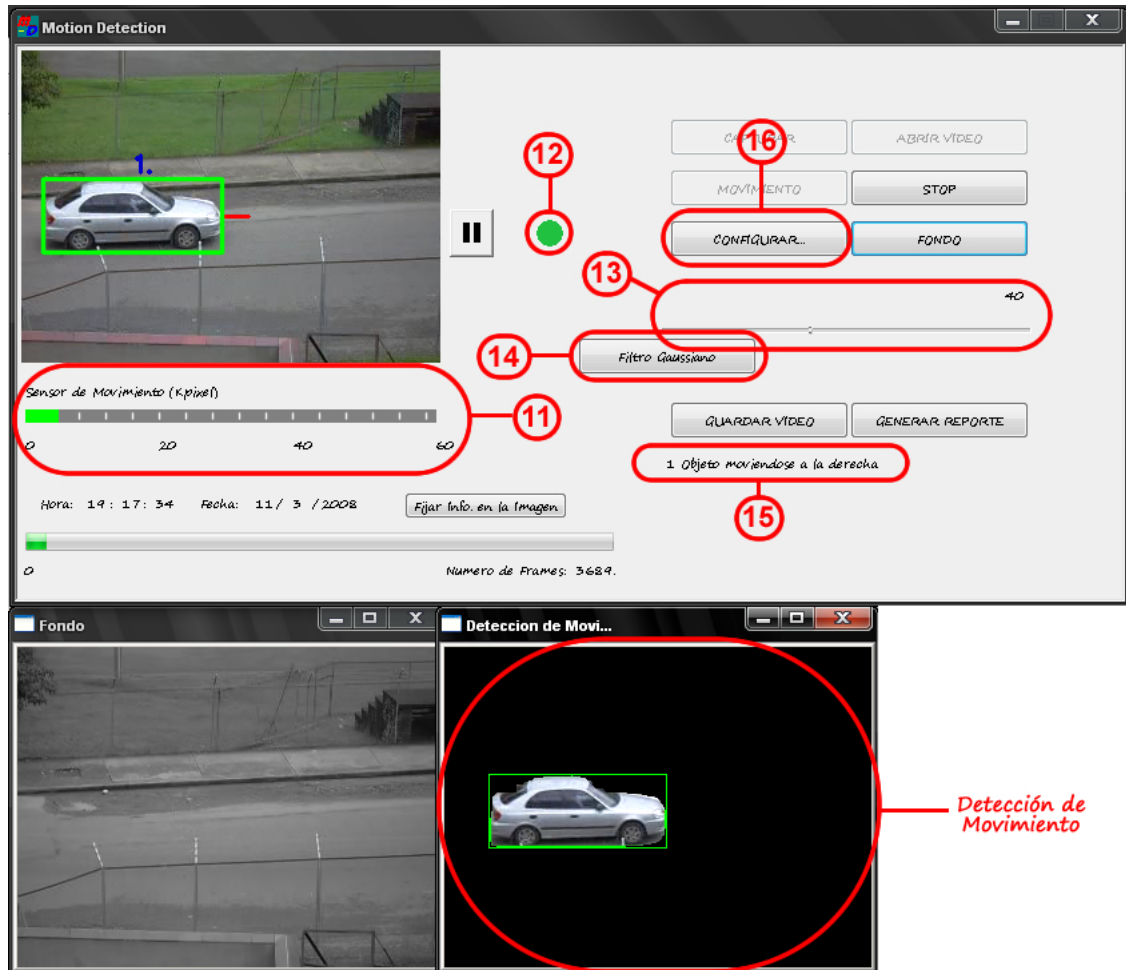


Figura A.6. Interfaz gráfica cuando la opción Movimiento está activada.

13. UMBRAL PARA LA SEGMENTACIÓN DEL MOVIMIENTO

Este control permite variar el umbral para la segmentación del movimiento tras la sustracción del fondo. Este puede tomar valores entre 0 y 255, sin embargo, por defecto, inicia con un valor de 40.

14. FILTRO GAUSIANO

Esta botón permite aplicar un filtro de paso bajo a cada *frame* de la secuencia de video. Aplicar esta opción puede ser muy útil cuando se está trabajando en escenas que tienen objetos del fondo en movimiento, como por ejemplo, las ramas de los árboles, las hojas, el césped, etc.

15. REPORTE INSTANTÁNEO

En esta región se muestra un reporte instantáneo con información sobre los objetos que se encuentran en la escena. Este reporte presenta información como la etiqueta, el tipo de objeto (si está activada la opción de reconocimiento), la velocidad media y la dirección de movimiento.

16. CONFIGURAR

Esta opción permite realizar ciertas configuraciones para el sistema como, por ejemplo, definir las zonas restringidas, entrenar el sistema para el reconocimiento, entre otras. Al activar esta opción, la interfaz gráfica se muestra como en la Figura A.7.

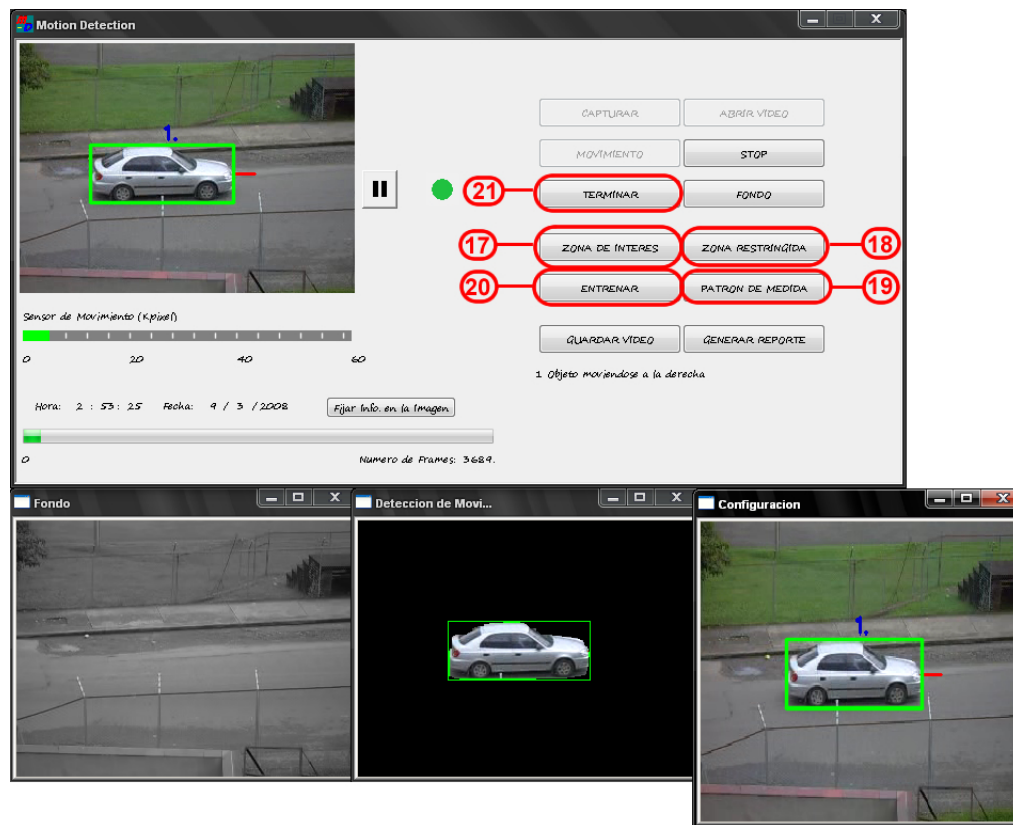


Figura A.7. Interfaz gráfica cuando la opción Configurar está activada.

17. ZONA DE INTERÉS

Este botón sirve para definir la zona en donde se va a aplicar el algoritmo de seguimiento y reconocimiento. Para esto se hace uso de la ventana de configuración que aparece en la

parte inferior derecha de la pantalla (ver Figura A.7).

Al presionar este botón, se podrá seleccionar el área de interés en la ventana de configuración arrastrando el ratón mientras se click izquierdo. La zona de interés se presentará en color verde como se puede observar en la Figura A.8.

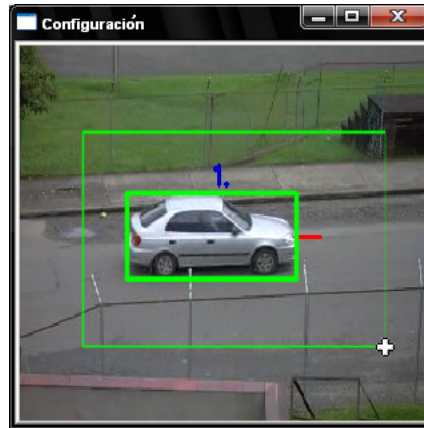


Figura A.8. Ventana de configuración y selección de la zona de interés.

18. ZONA RESTRINGIDA

Este botón permite restringir una zona determinada. Cuando se activa esta opción aparece en pantalla una nueva ventana como la que se muestra en la Figura A.9.

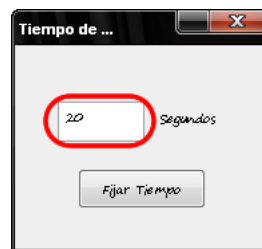


Figura A.9. Ventana de configuración para fijar el tiempo de restricción.

En esta ventana se definirá cuanto tiempo puede permanecer el objeto en la zona restringida antes de activar la alarma. En el ejemplo de la Figura A.9, el objeto sólo podrá permanecer 20 segundos dentro de esta zona. Una vez se ha escrito el valor deseado de tiempo, presione el botón *Fijar Tiempo*. Si el valor del tiempo fijado es cero, la alarma se activará inmediata-

mente un objeto entre en la zona restringida.

Después de haber fijado el tiempo, la zona restringida se selecciona de la misma manera que se seleccionó la zona de interés. Esta vez, la zona seleccionada se mostrará de color rojo como muestra la Figura A.10.



Figura A.10. Ventana de configuración y selección de la zona restringida.

19. PATRÓN DE MEDIDA

Esta opción permite definir un patrón de medida en la imagen para las medidas de distancia y velocidad. Es necesario definir este patrón, ya que en la imagen se miden las distancias en número de píxeles, lo que depende directamente de la distancia a la cual está ubicada la cámara. Para definir este patrón, la distancia debe ser conocida previamente por el usuario del sistema.

Una vez presionado este botón, aparecerá una nueva ventana como la que se muestra en la Figura A.11. En esta ventana se fija la distancia que va a ser utilizada como patrón (en el ejemplo de la Figura A.11, se define una distancia de 2 metros). Cuando se haya presionado el botón *aceptar*, se podrá definir el patrón en la ventana de configuración, arrastrando el ratón mientras se mantiene presionado el botón izquierdo como se muestra en la Figura A.12 (a). El patrón definido se mostrará en el video del sistema y en la ventana de configuración como se muestra en la Figura A.12 (b). En caso de no ser definido un patrón, el sistema tomará por defecto 10 píxeles como 1 metro.

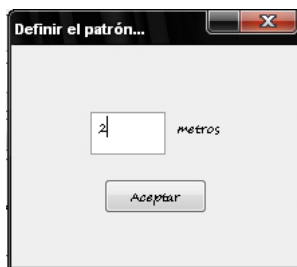


Figura A.11. Ventana de configuración para fijar la distancia patrón.



Figura A.12. Selección del patrón de medida: (a) Definición del patrón de medida en la imagen; (b) Visualización del patrón de medida en la imagen.

20. ENTRENAR

Esta opción permite entrenar el sistema para el reconocimiento de los objetos en movimiento. Al presionar este botón se puede seleccionar en la ventana de configuración los objetos con los que se quiere entrenar el sistema. Cuando se ha dado click sobre el objeto, aparecerá una ventana como la que se muestra en la Figura A.13. Se recomienda que se realice este proceso con la mayor cantidad de objetos posible.

21. TERMINAR

Este botón termina el proceso de configuración y fija los valores definidos en esta etapa. Una vez se presione este botón, el video, o la captura, parará y será necesario iniciarlos de nuevo.

Después de reiniciar la captura o el video, se mostrará una interfaz como en la Figura A.14.

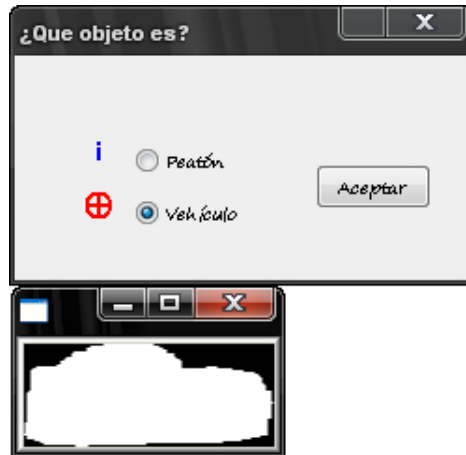


Figura A.13. Ventana de entrenamiento para la etapa de reconocimiento.

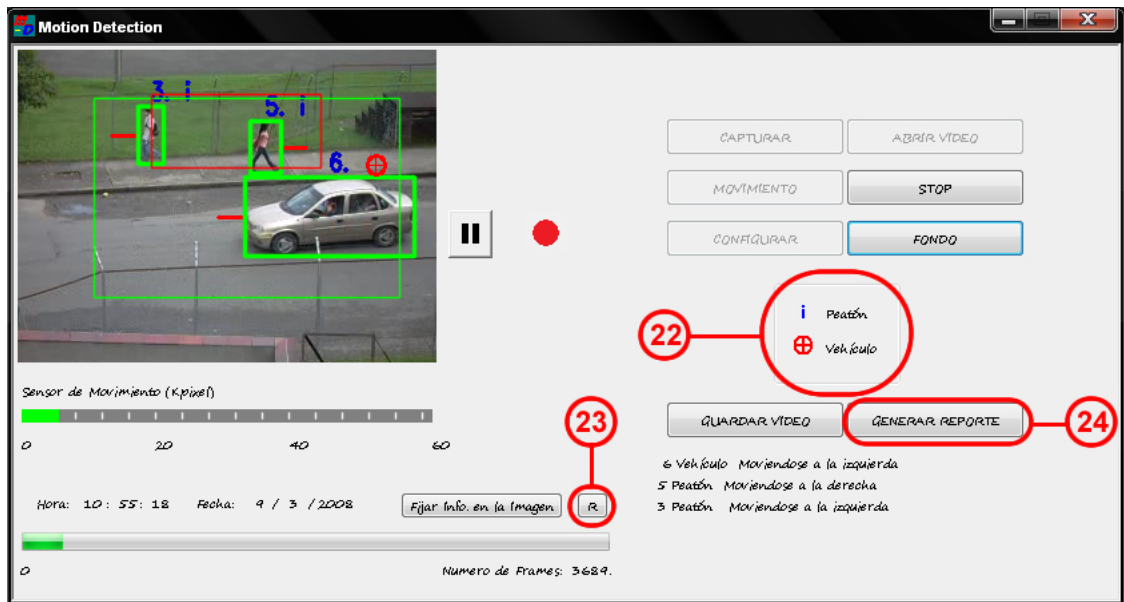


Figura A.14. Interfaz gráfica de usuario después del entrenamiento.

22. CUADRO DE INFORMACIÓN ACERCA DE LAS ETIQUETAS

Este cuadro informativo muestra cuales son las etiquetas utilizadas para los vehículos y los peatones y permanece fijo después de haber entrenado el sistema.

23. RESET

Esta opción es para volver todas las variables del entrenamiento a su estado inicial. Al presionar este botón se detendrá la visualización del video o de la captura. Después de activar el botón de *Reset* será necesario volver a entrenar el sistema.

24. GENERAR REPORTE

Esta opción permite generar un reporte completo sobre lo que ha pasado en la escena. El reporte empezará en el momento que se presione el botón *Generar Reporte*. En este momento el botón cambiará su texto al de *Finalizar Reporte* (Para terminar el reporte presione este botón).

El reporte se guardará en un archivo con extensión (*.xls), y puede ser visualizado en el programa Excel de Microsoft. Este reporte contiene la siguiente información:

- * *Número de peatones*: Indica el número total de peatones que aparecieron en la escena mientras se estaba generando el reporte.
- * *Número de vehículos*: Indica el número total de vehículos que aparecieron en la escena mientras se estaba generando el reporte.
- * *Fecha*: Muestra el día, el mes y el año en el que se genera el reporte.
- * *Etiqueta*: Muestra el número de la etiqueta asignada al objeto.
- * *Tipo de objeto*: Indica si el objeto en la escena es un peatón o un vehículo.
- * *Observaciones*: Contiene la dirección en la que se mueve el objeto.
- * *Velocidad*: Muestra la velocidad media del objeto en el intervalo de tiempo que estuvo en escena.
- * *Hora*: Contiene la hora exacta en la que apareció el objeto.
- * *Infracción*: Indica si el objeto entró o no en zona restringida

Un ejemplo del reporte generado por el sistema se muestra en la Figura A.15.

Microsoft Excel - Mdetect_BS4_5_v_0.1

Archivo Edición Ver Insertar Formato Herramientas Datos Ventana ?

M32 fx Arial 10

	A	B	C	D	E	F	G	H
1	UNIVERSIDAD DEL CUAUCA							
2	FACULTAD DE CIENCIAS NATURALES, EXACTAS Y DE LA EDUCACIÓN							
3	PROGRAMA DE INGENIERÍA FÍSICA							
4	SISTEMA DE VIGILANCIA							
5								
6	Reporte No.	1.	Peatones: 9	Vehículos: 73	04/02/2008			
7	=====							
	ETIQUETA	TIPO DE OBJETO	OBSERVACIONES	VELOCIDAD (m/s)	HORA	INFRACCION		
9	1	Vehículo	Moviendose a la derecha	11.1	16:23:23	Ninguna		
10	3	Vehículo	Moviendose a la derecha	13.2	16:23:29	Ninguna		
11	5	Vehículo	Moviendose a la izquierda	-10.	16:23:34	Ninguna		
12	2	Peatón	Moviendose a la izquierda	-1.7	16:23:28	Entró en zona restringida		
13	4	Peatón	Moviendose a la derecha	2.03	16:23:30	Entró en zona restringida		
14	6	Vehículo	Moviendose a la izquierda	-14.	16:23:41	Ninguna		
15	7	Vehículo	Moviendose a la izquierda	-10.	16:23:43	Ninguna		
16	8	Vehículo	Moviendose a la izquierda	-14.	16:23:44	Ninguna		
17	9	Vehículo	Moviendose a la izquierda	-10.	16:23:46	Ninguna		
18	10	Vehículo	Moviendose a la derecha	6.54	16:23:47	Ninguna		
19	11	Vehículo	Moviendose a la izquierda	-9.6	16:23:48	Ninguna		
20	12	Vehículo	Moviendose a la derecha	9.48	16:23:55	Ninguna		
21	13	Vehículo	Moviendose a la izquierda	-11.	16:23:56	Ninguna		
22	15	Vehículo	Moviendose a la derecha	7.28	16:24:07	Ninguna		
23	16	Vehículo	Moviendose a la derecha	12.1	16:24:09	Ninguna		
24	14	Peatón	Moviendose a la derecha	1.90	16:24:02	Entró en zona restringida		

Figura A.15. Reporte generado por el sistema.