



POLITÉCNICA

Universidad Politécnica de Madrid

Escuela Técnica Superior de Ingeniería
y Sistemas de Telecomunicación

**CONTRIBUCIÓN A LA DETECCIÓN DE OBJETIVOS
ESPECÍFICOS PARA APLICACIONES DE VIGILANCIA CON
REALIDAD AUMENTADA**

TESIS DOCTORAL

Henry O. Cruz Carrillo

Master en Redes y Telecomunicaciones

2017

Centro de Investigación en Tecnologías de Software y
Sistemas Multimedia para la Sostenibilidad
Escuela Técnica Superior de Ingeniería y Sistemas de Telecomunicación

**CONTRIBUCIÓN A LA DETECCIÓN DE OBJETIVOS
ESPECÍFICOS PARA APLICACIONES DE VIGILANCIA CON
REALIDAD AUMENTADA**

TESIS DOCTORAL

Ing. Henry O. Cruz Carrillo

Master en Redes y Telecomunicaciones

Directores:

Dr. Juan Manuel Meneses Chaus.

Dr. Ingeniero de Telecomunicación

Dra. Martina Eckert

Dra. Ingeniera de Telecomunicación

Universidad Politécnica de Madrid

2017

**DOCTORADO EN INGENIERÍA DE SISTEMAS Y SERVICIOS
PARA LA SOCIEDAD DE LA INFORMACIÓN**

| Tesis Doctoral | | |
|---------------------------------|---|-------|
| Título | Contribución a la detección de objetivos específicos para aplicaciones de vigilancia con realidad aumentada | |
| Autor | Henry Omar Cruz Carrillo | |
| Director | Juan Manuel Meneses Chaus | VºBº. |
| Co-Directora | Martina Eckert | VºBº. |
| Tribunal | | |
| Presidente | | |
| Secretario | | |
| Vocal | | |
| Vocal | | |
| Vocal | | |
| Suplente | | |
| Suplente | | |
| Lugar y fecha de lectura | E.T.S.I. y Sistemas de Telecomunicación (U.P.M.) | |
| Calificación | | |

El Presidente

El secretario

El vocal 1º

El vocal 2º

El vocal 3º

Tesis Doctoral para la obtención del título de Doctor
por la Universidad Politécnica de Madrid.

2017

“La imaginación es más importante que el conocimiento.”

Albert Einstein (1879-1955)

AGRADECIMIENTOS

Como creyente que soy en primer lugar doy gracias a Dios por su ayuda y los dones que me ha concedido, sin su bendición no hubiera podido realizar nada en mi vida, todo se lo debo a él. De igual forma le doy gracias a la virgen María por los cuidados y protección del cual he sido objeto. Doy gracias a la vida por la oportunidad que me ha dado de experimentar y conocer nuevas formas de existir.

En el plano familiar le doy gracias a mis hijos Henry Sebastián, José David y Alejandro Israel y a mi esposa Daicy, por el amor y comprensión que han tenido conmigo así como por soportar mis altibajos, les debo tiempo y les quiero mucho.

También agradezco a mis padres Israel y Marianita, a mi hermano Ronny y su esposa Miri, a mi hermano menor Juan Pablo, así como a mi familia radicada en España por la ayuda y el soporte de todo tipo, que he recibido durante este tiempo.

Un agradecimiento especial a mis Directores que a la vez han sido tutores y mentores, Juan Meneses y Martina Eckert, sin su ayuda, consejos, supervisión y guía no hubiera sido posible concluir con este trabajo de tesis doctoral, mis más sinceros agradecimientos.

También agradezco a todos los profesores, amigos (as), compañeros (as), estudiantes y colegas tanto del CITSEM como de la ETSIST, que me han aportado con su experiencia, consejos, conocimientos y buena voluntad durante todo este proceso.

Finalmente agradezco a la SENESCYT y a mi país el Ecuador, por concederme una beca de estudios y permitirme de esta forma realizar mis estudios en este prestigioso centro de educación superior como es la “Universidad Politécnica de Madrid”.

TABLA DE CONTENIDO

| | |
|---|------|
| AGRADECIMIENTOS | i |
| INDICE DE FIGURAS | vii |
| INDICE DE TABLAS | xi |
| RESUMEN | xiii |
| ABSTRACT | xiv |
| Capítulo I: INTRODUCCIÓN Y OBJETIVOS | 1 |
| 1.1. Motivación | 3 |
| 1.2. Objetivos | 3 |
| 1.3. Estructura de la tesis..... | 4 |
| Capítulo II: ANTECEDENTES Y ESTADO ACTUAL DE LA TÉCNOLOGÍA | 7 |
| 2.1. Introducción | 9 |
| 2.2. Introducción a los sistemas de videovigilancia | 9 |
| 2.2.1. Evolución de los sistemas de videovigilancia | 9 |
| 2.2.2. Estructura de un sistema de videovigilancia típico | 11 |
| 2.2.3. Videovigilancia inteligente | 12 |
| 2.2.4. La percepción visual y su relación con la detección de objetivos en videovigilancia inteligente..... | 13 |
| 2.3. Detección de objetos específicos en la escena | 15 |
| 2.4. Realidad Aumentada (RA) en los sistemas de vigilancia..... | 27 |
| 2.4.1. Introducción a la RA | 27 |
| 2.4.2. Empleo de la Realidad Aumentada en los sistemas de videovigilancia..... | 30 |
| 2.5. Detección de ROI's a través del color..... | 38 |
| 2.6. Uso de la umbralización en la detección de objetivos específicos en tiempo real | 38 |
| 2.7. Extracción de objetivos relevantes en la escena a través de índices de color | 39 |
| 2.7.1. Los índices de color - inicios y desarrollo..... | 39 |
| 2.7.2. El uso de los índices de color en la detección de ROI's en tiempo real..... | 42 |
| 2.8. Conclusiones | 44 |
| Capítulo III: METODOLOGÍA PARA LA DETECCIÓN DE OBJETIVOS ESPECÍFICOS MEDIANTE EL TRATAMIENTO DEL COLOR | 47 |
| 3.1. Introducción | 49 |
| 3.2. Terminología empleada..... | 49 |
| 3.2.1. El objetivo específico..... | 50 |
| 3.2.2. Detección en tiempo real de objetivos específicos..... | 50 |
| 3.2.3. El tiempo de procesamiento (<i>TP</i>) | 50 |

| | | |
|---|--|-----|
| 3.2.4. | Índice de color específico (<i>ICE</i>) y general (<i>ICG</i>) | 50 |
| 3.3. | Métricas de evaluación..... | 51 |
| 3.3.1. | Métricas geométricas | 51 |
| 3.3.2. | Métricas estadísticas..... | 53 |
| 3.3.3. | Métricas estadísticas de precisión, sensibilidad y rendimiento para evaluar la detección | 54 |
| 3.4. | Metodología de detección de objetivos con formas irregulares por índices de color | 56 |
| 3.4.1. | Etapa 1: Creación de Índices de Color Específicos..... | 58 |
| 3.4.2. | Etapa 2: Segmentación del objetivo específico | 66 |
| 3.4.3. | Etapa 3: Extracción de información y uso para generación de Realidad Aumentada | 72 |
| 3.5. | Proceso general de detección de zonas deforestadas..... | 73 |
| 3.5.1. | Relación entre componentes cromáticas para zonas deforestadas | 74 |
| 3.5.2. | Extracción del Exceso de Marrón (<i>ExBr</i>)..... | 75 |
| 3.5.3. | Extracción del índice de zonas no forestadas (<i>NFDI</i>)..... | 76 |
| 3.5.4. | Umbral global para la detección de zonas no forestadas (<i>TNFDI</i>)..... | 78 |
| 3.5.5. | Evaluación del rendimiento global del <i>NFDI</i> | 80 |
| 3.5.6. | Pruebas de detección con variaciones de iluminación | 89 |
| 3.6. | El Índice de Detección de Fuego Forestal (<i>FFDI</i>)..... | 91 |
| 3.6.1. | Extracción y aplicación del <i>FFDI</i> | 91 |
| 3.6.2. | Relación entre componentes para la detección de fuego..... | 91 |
| 3.6.3. | Obtención del <i>FDI</i> y del <i>FFDI</i> | 93 |
| 3.6.4. | El coeficiente de detección de fuego y humo ρ | 94 |
| 3.6.5. | El proceso de clasificación y detección del fuego forestal..... | 96 |
| 3.6.6. | Evaluación del <i>FFDI</i> | 97 |
| 3.7. | Conclusiones | 100 |
| Capítulo IV: APLICACIONES Y GENERACIÓN DE REALIDAD AUMENTADA..... | | 103 |
| 4.1. | Introducción | 105 |
| 4.2. | Sistema de Detección de Fuego Forestal..... | 105 |
| 4.2.1. | Descripción general del sistema..... | 105 |
| 4.2.2. | El sistema de detección de fuego forestal implementado para consola..... | 106 |
| 4.2.3. | El sistema de detección de fuego forestal implementado para dispositivos móviles | 110 |
| 4.2.4. | Generación de RA en el módulo de correlación de datos | 114 |
| 4.3. | Sistema de asistencia a la conducción a través de dispositivos inteligentes | 114 |
| 4.3.1. | Descripción general del sistema..... | 115 |
| 4.3.2. | Generación de Realidad Aumentada | 117 |
| 4.3.3. | Pruebas de rendimiento | 119 |

| | | |
|---|--|-----|
| 4.4. | Conclusiones | 120 |
| Capítulo V: CONTRIBUCIONES Y TRABAJOS FUTUROS | | 121 |
| 5.1. | Introducción | 123 |
| 5.2. | Contribuciones | 123 |
| 5.2.1. | Contribuciones Metodológicas..... | 123 |
| 5.2.2. | Aportaciones derivadas de las aplicaciones desarrolladas | 124 |
| 5.2.3. | Publicaciones realizadas..... | 125 |
| 5.2.4. | Conferencias y ponencias internacionales..... | 126 |
| 5.2.5. | Proyectos de investigación que han dado soporte al desarrollo de la Tesis Doctoral | 126 |
| 5.3. | Trabajos Futuros..... | 127 |
| 5.3.1. | Aplicación de los <i>ICE</i> en otros contextos | 127 |
| 5.3.2. | Patrón de reconocimiento a través de ecualización de valores del color. | 128 |
| ANEXOS | | 129 |
| A. | Teoría del color | 131 |
| 1. | El espacio de color RGB (Red, Green, Blue)..... | 131 |
| 2. | El Modelo YCbCr | 133 |
| 3. | Modelo de Color HSI..... | 133 |
| 4. | Modelo RGBa | 135 |
| B. | Umbralización propuesta por Otsu..... | 136 |
| C. | Bases de datos de imágenes y video utilizadas | 138 |
| 1. | Base de datos CIDFAE | 138 |
| 2. | Base de datos Europea de Fuego Forestal (EFFIS, 2015)..... | 139 |
| 3. | Bugwood Images Database | 140 |
| 4. | Base de datos del ICMC/USP | 140 |
| 5. | Base de datos Caltech 101..... | 140 |
| 6. | Base de datos del GEDEM..... | 140 |
| D. | Formatos, equipamiento y software empleado..... | 141 |
| 1. | Formatos de imagen y video | 141 |
| 2. | Software y lenguajes de programación | 141 |
| 3. | Hardware empleado | 142 |
| E. | Sistema de evaluación de la segmentación y detección basada en multihilos | 144 |
| 1. | Descripción general del funcionamiento del sistema | 144 |
| 2. | Esquema de paralelización de los procesos de la segmentación y evaluación..... | 146 |
| REFERENCIAS | | 149 |

INDICE DE FIGURAS

| | |
|--|----|
| Figura 2.1. Estructura típica de un sistema profesional de videovigilancia. | 11 |
| Figura 2.2. Descripción general de subprocesos en un sistema de videovigilancia inteligente. | 12 |
| Figura 2.3. Mapa de características prominentes a través del análisis de la luminancia y el color en diversas regiones, método presentado por Achanta (Achanta et al., 2008). | 17 |
| Figura 2.4. Esquema de adquisición de objetos relevantes considerando operaciones algebraicas de adición y sustracción de características de color y luminosidad en diferentes niveles, tomado de (Sandor et al., 2010)..... | 18 |
| Figura 2.5. Obtención del mapa de características prominentes a través de la <i>IDCT</i> propuesta por (Hou et al., 2012)..... | 19 |
| Figura 2.6. Modelo “Sparse and Dense Reconstruction” para la detección de objetos específicos en la escena, tomado de (X. Li et al., 2013). | 19 |
| Figura 2.7. Resultado del empleo del modelo <i>GMM</i> para detectar las diferentes áreas de color en la imagen considerando los diferentes niveles del histograma de cuantización de la imagen original (Cheng et al., 2013). | 20 |
| Figura 2.8. La detección de objetivos de interés en la escena a través de la <i>DCT</i> según el modelo propuesto por (F. Li et al., 2014). | 21 |
| Figura 2.9. Muestra visual comparativa del método “Bipartite Dictionary (<i>BD</i>)” frente a otros 12 métodos propuestos en el estado del arte (Seo et al., 2014). | 22 |
| Figura 2.10. Resultados alcanzados por Kim et al. En (a) se encuentra la imagen original, en (b) la extracción del objeto relevante en la escena y en (c) el resultado de la aplicación de refinamiento de la escena (J. Kim et al., 2014)..... | 23 |
| Figura 2.11. Diagrama general de la detección de objetivos específicos a través del método PISA. | 24 |
| Figura 2.12. Esquema grafico del modelo de fusión de características de detección de objetos relevantes en la escena con modelos predictivos y de aprendizaje propuesto por (J. Wang et al., 2016)..... | 26 |
| Figura 2.13. Detección de objetivos específicos en imágenes con perspectiva aérea captadas por sensores SAR: En (a) Estimación colorimétrica en escala y detección de objetivos predominantes a través de Chan-Vese, (b) mapa de objetivos predominantes en la escena, (c) binarización de la escena (Tu & Su, 2016). | 27 |
| Figura 2.14. Interfaz de RA en el sistema de asistencia a la conducción propuesto por (Abdi et al., 2015). | 29 |

| | |
|--|----|
| Figura 2.15. Hyundai Virtual Guide, una muestra de guía virtual de operación y mantenimiento para el modelo de Hyundai Sonata 2015, gráfico tomado de (Avila & Bailey, 2016). | 29 |
| Figura 2.16. Combinación de mapas de objetos relevantes para producir RA (Sandor et al., 2010). | 31 |
| Figura 2.17. Teledetección, control y manejo teledirigido a través de cámaras de vigilancia e interfaces de RA (Sauer et al., 2010)..... | 31 |
| Figura 2.18. Detección e identificación de secuestros o actitudes hostiles presentado en el interfaz gráfico de RA tomado de (Gwon et al., 2013)..... | 32 |
| Figura 2.19. Interfaz de RA que muestra el mapa de detección de vehículos para estimación de densidad de tráfico desarrollado para sistemas de vigilancia (Filonenko et al., 2014). | 33 |
| Figura 2.20. Información y alarma virtual de detección de humo para sistemas de vigilancia medioambiental (Bugarić et al., 2014)..... | 33 |
| Figura 2.21. Información virtual sobre eventos inusuales en carretera en (a) muestra la detección de una colisiona si como un vehículo estacionado (b) etiquetado de los elemento que conforman la escena (Lan et al., 2016)..... | 34 |
| Figura 2.22. Información virtual de la estimación del número de personas en una escena de videovigilancia (Hashemzadeh & Farajzadeh, 2016)..... | 34 |
| Figura 2.23. Generación de RA considerando información obtenida por cámaras simulando retrovisores en un coche (Kane et al., 2016)..... | 35 |
| Figura 2.24. Información de RA en el reconocimiento de objetos a través de etiquetado semántico (Gori et al., 2016)..... | 35 |
| Figura 2.25. Interfaz gráfico del sistema CASSANDRA para detectar y alertar sobre agresiones. | 36 |
| Figura 2.26. Presentación de información virtual sobre detecciones realizadas luego de alcanzar estabilidad en los sensores (Lee et al., 2015)..... | 36 |
| Figura 2.27. Interfaz gráfica de RA en la detección de fuego forestal..... | 37 |
| Figura 2.28. Sistema de detección y etiquetado virtual de rocas y horizonte visual marciano empleado en la sonda Mars Pathfinder (a) a corta distancia y (b) a larga distancia (Gao et al., 2014b). | 37 |
| Figura 3.1. Evaluación comparativa entre la segmentación automática frente al GT. En (a) imagen original, (b) GT en binario, (c) binario de la segmentación automática, (d) clasificación de regiones por coordenadas de la matriz de confusión. | 54 |
| Figura 3.2. Detección de un objetivo específico con formas irregulares basado en ICE: (a) Etapa 1. Extracción del ICE e ICG en la imagen. (b) Etapa 2. Refinamiento de la detección. (c) Etapa 3. Uso de la información de la ROI para generar Realidad Aumentada. | 57 |

| | |
|---|----|
| Figura 3.3. Extracción de la componente R y r de una imagen: En (a) imagen de R y su respectivo histograma. En (b) imagen de r con su histograma..... | 61 |
| Figura 3.4. Evaluación de regiones de píxeles en imágenes. | 63 |
| Figura 3.5. Histograma de la distribución de color para RGB y en escala de grises..... | 67 |
| Figura 3.6. Histogramas en escala lineal empleados en la umbralización: En (a) Histograma bimodal o ideal. En (b) Histograma multimodal. | 71 |
| Figura 3.7. Histograma encontrado a partir de la aplicación de un ICG en la imagen..... | 71 |
| Figura 3.8. Esquema general de extracción de zonas no forestadas a través del NFDI | 74 |
| Figura 3.9. Aplicación del ExBr en (a) imagen con contrastes bien definidos entre el suelo y la vegetación, en (b) imagen con mayor cantidad de vegetación y contrastes poco definidos. | 76 |
| Figura 3.10. Proceso de obtención del NFDI en (a) mapa de la vegetación a través del ExG en (b) mapa de las zonas deforestadas encontradas a partir del ExBr en (c) se muestra el resultado de la detección a través del NFDI. | 77 |
| Figura 3.11. Análisis gráfico comparativo del rendimiento del NFDI frente al ExGR..... | 78 |
| Figura 3.12. Detección de zonas no forestadas (segmentación) a través del NFDI. En (a) se presenta la binarización y en (b) el contorno en la imagen original. | 79 |
| Figura 3.13. Extracción de NFDI en diversas condiciones, C1 representa la zona no forestada y C2 la vegetación. | 80 |
| Figura 3.14. Segmentación por umbrales a través del NFDI (a) y GT de expertos (b)..... | 82 |
| Figura 3.15. Segmentaciones de zonas no forestadas obtenidas por varios métodos..... | 84 |
| Figura 3.16. Comparación vía multihilos. En (a) se muestra el binario de la imagen patrón (GT), en (b) el binario del NFDI, en (c) el resultado de la comparación y las coordenadas TP, TN, FP y FN en color..... | 85 |
| Figura 3.17. Curvas de precisión de la detección frente al tiempo de procesamiento para dos tamaños de imágenes. | 88 |
| Figura 3.18. Prueba de esfuerzo simulando diferentes condiciones de saturación de color y cambios extremos de iluminación. A la izquierda se muestra la imagen modificada, en el centro la distribución del color, a la derecha clasificación a través del NFDI y etiquetado..... | 90 |
| Figura 3.19. Extracción del fuego y humo y cálculo de la media de RGB a través de sus respectivos histogramas. | 92 |
| Figura 3.20. Diagrama de detección de fuego forestal a través del FFDI, (a) Relación r - b; (b) Relación r - g; (c) Extracción del FDI; (d) Extracción del ExG; (e) Mapa de objetos relevantes vía FFDI;(f) Binarización;(g) Fuego forestal detectado. | 93 |
| Figura 3.21. Influencia del coeficiente ρ en la detección del fuego y el humo circundante. | 95 |
| Figura 3.22. Comportamiento de r, g, b en el FFDI frente a la variación de ρ | 95 |

| | |
|---|-----|
| Figura 3.23. Histograma de una AOI del humo y los valores respectivos para R, G y B. | 96 |
| Figura 3.24. Ejemplos de detecciones de fuego forestal vía FFDI bajo perspectiva aérea y lejana. | 98 |
| Figura 3.25. Relación del de la detección en general (DI) frente al tiempo de procesamiento. .. | 99 |
| Figura 3.26. Detecciones de fuego forestal en imágenes con perspectiva cercana y en inicio del conato de incendio. | 100 |
| Figura 4.1. Esquema general del sistema de detección de fuego vía FFDI. | 108 |
| Figura 4.2. Prototipo de vigilancia de fuego forestal en un puesto fijo (consola). | 109 |
| Figura 4. 3. Plataforma de desarrollo de FFDI en Android. | 111 |
| Figura 4.4. El Lanzador de la aplicación con su menú de opciones. | 111 |
| Figura 4.5. Adquisición de cuadros de video de una escena de fuego forestal (simulación). ... | 112 |
| Figura 4.6. Aplicación del FFDI en la escena a través del dispositivo móvil. | 113 |
| Figura 4.7. Detección de fuego forestal en tiempo real. | 113 |
| Figura 4.8. Presentación de información virtual sobre la detección del fuego forestal. | 114 |
| Figura 4.9. Diagrama general de funcionamiento del sistema de asistencia de conducción nocturna. | 116 |
| Figura 4.10. La aplicación en funcionamiento muestra diferentes datos de velocidad, distancia, posición espacial, coche detectado, alarmas, entre otros. | 118 |
| Figura 4.11. Pruebas realizadas en diferentes condiciones climáticas: (a) detección con escasa lluvia, (b) detección con fuerte lluvia, (c) detección con nieve poco copiosa. | 119 |
| Figura A.1. Representación espacial del modelo RGB. | 132 |
| Figura A.2. Modelo HSI, representación gráfica. | 134 |
| Figura C.1. Ubicación de las zonas de estudio y obtención de imágenes que forman parte de la reserva de producción faunísticas “Cuyabeno”. | 139 |
| Figura E.1. Proceso general de evaluación de la segmentación y detección vi múltiples hilos. | 145 |
| Figura E.2. Procedimiento de comparación entre un método de segmentación automático frente a GT. | 146 |

INDICE DE TABLAS

| | |
|---|-----|
| Tabla 3. 1. Muestra del análisis cromático sobre 7 regiones..... | 64 |
| Tabla 3. 2. Relación entre coordenadas cromáticas y sus correspondientes normalizadas en 7 regiones..... | 65 |
| Tabla 3.3. Resultados de la medición de GCE, OCE, RI, VOI en la segmentación para NFDI y otros métodos..... | 83 |
| Tabla 3.4. Promedio de evaluación de precisión de la detección sobre 30 Imágenes | 86 |
| Tabla 3.5. Promedio de Tp sobre 30 imágenes | 87 |
| Tabla 3.6. Promedio de R sobre 30 imágenes | 87 |
| Tabla 3.7. Promedio de evaluaciones del NFDI en imágenes HD (1296×894 píxeles)..... | 89 |
| Tabla 3.8. Valores promedio de las métricas obtenidas sobre 50 imágenes de prueba redimensionadas | 98 |
| Tabla 3.9. Análisis de rendimiento sobre 50 imágenes | 99 |
| Tabla 3.10. Valores promedios de las métricas obtenidas sobre 30 imágenes a corta distancia. | 100 |
| Tabla 4.1. Resultados de las pruebas de rendimiento efectuadas sobre diferentes condiciones meteorológicas..... | 120 |
| Tabla C.1. Especificaciones Técnicas del UAV “Gavilán”..... | 138 |
| Tabla D.1. Características técnicas de la cámara y especificaciones de las imágenes. | 142 |

RESUMEN

En la presente tesis doctoral, se ha identificado que existen deficiencias a la hora de detectar objetos y regiones con formas irregulares de manera eficaz y eficiente en los sistemas de videovigilancia por imagen y video. La eficacia está relacionada con la capacidad de detectar objetivos específicos con precisión, y la eficiencia es concordante con la oportunidad con la que se presenta la información de detección, es decir que pueda ser mostrada en tiempo real.

Para solventar estas deficiencias, se presentan una serie de propuestas que decantan en una estrategia metodológica basada en el tratamiento de la información del color. Esta metodología contribuye a mejorar la detección de objetivos específicos de este tipo de formas y en donde la detección ha sido enfocada a resolver problemas de vigilancia ambiental. De esta forma, se plantean tres etapas que son la extracción de los *ICE*, el cálculo y la aplicación del umbral T_{ICG} para obtener la imagen binaria de la detección y finalmente el uso de información de interés para generar Realidad Aumentada. Dentro de la primera etapa, se proponen dos índices de color, el *Forest Fire Detection Index (FFDI)* y el *Non Forest Detection Index (NFDI)*, que representan umbrales óptimos de detección tanto de fuego forestal como de zonas deforestadas respectivamente. La precisión de detección alcanzada a través de estos índices bordea un máximo de 96% de eficacia, que es muy alta en comparación a otros métodos, permitiendo además su uso en aplicaciones en tiempo real.

Las capacidades y las potencialidades encontradas en los algoritmos implementados a partir del método propuesto, permiten el uso de estos algoritmos tanto en sistemas tradicionales fijos (consola) como en sistemas de vigilancia móviles como pueden ser los drones, esto permite cambiar el enfoque tradicional de videovigilancia a un nuevo escenario que garantiza movilidad, flexibilidad así como ampliación de la cobertura de vigilancia.

Por otra parte se ha desarrollado dos prototipos, que han permitido verificar la eficacia y efectividad de las propuestas planteadas así como han permitido disponer de información virtual a partir de la información intrínseca de las regiones de interés detectadas.

ABSTRACT

In this Doctoral Thesis, deficiencies have been identified in video surveillance systems regarding the effective and efficient detection of objects and regions with irregular shapes. The effectiveness is related to the capacity to detect specific objectives precisely, while efficiency is related to the possibility to present this information in real time to the user.

Different proposals based on a methodological strategy of the treatment of color information are formulated in this work. This methodology contributes to the improvement of detections of specific targets with the mentioned irregular shapes, focused on resolving problems of environmental surveillance. In this context, the following three stages are presented: the extraction of the *ICE*, the calculation and application of the threshold T_{ICG} , applied to obtain the binary image of the detection and finally, the use of the relevant information to generate Augmented Reality. In the first stage, two color indices are proposed, the *Forest Fire Detection Index (FFDI)* and the *Non Forest Detection Index (NFDI)*, which represent the optimal thresholds of for detecting forest fires and non-forested areas respectively. The level of precision reached through these indices is almost 96%, which is very high in comparison with other methods. They also have been proved to be applicable to real time applications.

The capabilities and potentials found in the algorithms implemented throughout the proposed method allow the use of these algorithms as well as in traditional permanently installed systems as in mobile surveillance systems like drones. In this way, the change of the traditional video surveillance approach to a new scenario that guarantees mobility, flexibility and the extension of the coverage of surveillance is possible.

In addition, two prototypes have been developed in this Thesis. Results verified the effectiveness and efficiency of the proposals and showed the possibilities to obtain and present virtual information from intrinsic information of the detected regions of interest.

Capítulo I: INTRODUCCIÓN Y OBJETIVOS

1.1. Motivación

Los sistemas de videovigilancia en la actualidad forman parte intrínseca del vivir y la cotidianidad de los seres humanos, permitiendo el control y la supervisión del ambiente que los rodea. Estos sistemas demandan cada vez mayores funcionalidades que les permiten ser más efectivos y eficientes a la hora de monitorear áreas o las respectivas zonas de cobertura.

Sabiendo las deficiencias que presentan los sistemas de vigilancia en cuanto a la detección de formas irregulares, la principal motivación de este trabajo fue encontrar un método alternativo que permita la robusta detección de este tipo de formas, considerando al color como fuente de información. Aunque se conoce de antemano que el color como tal no asegura detecciones efectivas en escenas con condiciones de iluminación inestables, se busca estudiar, indagar y plantear una estrategia que combine diferentes técnicas para alcanzar detecciones precisas.

Por otra parte, otra de las motivaciones ha sido generar los algoritmos de procesamiento para que funcionen en los drones y dispositivos inteligentes de tal forma que suministren información en tiempo real y de esta manera estimular a que germinen nuevas formas de videovigilancia diferentes a las tradicionales. Precisamente en la presente tesis doctoral se ha buscado proporcionar soluciones alternativas para que este tipo de dispositivos permitan un mejor control de desastres y una efectiva supervisión y vigilancia medioambiental (deforestaciones, incendios forestales, inundaciones), sobre todo considerando los nuevos retos a los que se enfrenta la humanidad por causas de la elevación paulatina de las temperaturas.

Una motivación final ha sido el conseguir generar realidad aumentada de manera intuitiva para los sistemas de vigilancia, considerando la información que presentan las regiones de interés detectadas, así como la información proporcionada por diversos sensores de los dispositivos móviles usados por los humanos.

1.2. Objetivos

Por lo tanto, el objetivo general de la tesis doctoral es contribuir a la detección precisa de objetivos específicos con formas irregulares, enfocando la detección a actividades de vigilancia, proporcionando además información asociada a este tipo de sistemas en tiempo real.

Los objetivos específicos son:

- Conseguir detecciones precisas de objetivos con formas irregulares que se encuentran en fondos ambiguos o no uniformes a través del análisis de la información de color.
- Desarrollar y proponer nuevos índices de color.
- Desarrollar e implantar la algoritmia basada en los métodos propuestos de tal manera que permita la detección automática de objetivos específicos tanto para sistemas de vigilancia con cámaras fijas como para móviles (drones, dispositivos inteligentes).
- Mostrar aplicaciones de los métodos y estrategias planteadas para solucionar problemas reales de detección en sistemas de vigilancia medioambiental, como son la detección de zonas deforestadas y de incendios forestales.
- Garantizar detecciones con altas tasas de rendimiento de tal forma que puedan ser empleados en los sistemas de videovigilancia y que sean útiles para generar realidad aumentada de un tipo de evento en concreto.

1.3. Estructura de la tesis

Esta tesis doctoral se ha estructurado en cinco capítulos y seis anexos, una breve descripción de cada capítulo y de los anexos se muestra a continuación.

En el Capítulo II se presenta una revisión del estado actual de los sistemas de vigilancia, su evolución, configuración y variedad. Además se realiza un compendio sobre la videovigilancia inteligente y se expone el progreso de los sistemas de visión artificial para emular el sistema visual humano por medio de sensores ópticos.

Por otra parte se realiza una revisión pormenorizada de los últimos trabajos presentados sobre la detección de objetos relevantes en una escena. Este estudio se ha enfocado en dos aspectos, primero en que la fuente de información esté vinculada con el color para conseguir la detección del objetivo específico. Por otra parte se realiza un análisis tanto de la precisión como del tiempo de procesamiento que se emplea para ejecutar las detecciones, todo esto con el enfoque en su uso en tiempo real. También se realiza un estudio de los principales trabajos que emplean la realidad aumentada en los sistemas de vigilancia.

Otro aspecto importante que se muestra es el estudio de la evolución de los índices de color y su enfoque de empleo, ya que en posteriores capítulos se realiza la propuesta de nuevos índices así como de su uso en aplicaciones en tiempo real.

El Capítulo III constituye el núcleo central de la Tesis Doctoral y es donde se muestran las respectivas propuestas metodológicas a fin de alcanzar la detección de objetos específicos con formas irregulares y en fondos no uniformes. Aquí se plantea una estrategia general de adquisición de índices de color basado en la extracción de áreas de interés y evaluación estadística de tal forma que se consigue establecer las relaciones cromáticas entre las diferentes regiones que conforman la escena. A partir de las relaciones cromáticas se proponen dos nuevos índices de color: el *NFDI* (*Non Forest Detection Index*) empleado en la detección de zonas no forestales y el *FFDI* (*Forest Fire Detection Index*) de utilidad en la detección de fuego forestal.

Se realizan sendas pruebas de calidad de segmentación, detección y toma de tiempos de procesamiento con la finalidad de establecer el rendimiento de los índices así como analizar su aplicabilidad tanto en sistemas de vigilancia fijos como móviles.

En el Capítulo IV se presentan dos prototipos que han permitido poner a prueba la funcionalidad de las diferentes propuestas metodológicas desarrolladas.

El primer prototipo presenta un sistema de detección de fuego forestal tanto para dispositivos fijos como móviles, que tiene como base el *FFDI*. De esta manera se confirma la utilidad de este índice en aplicaciones prácticas con enfoque en la vigilancia de tipo ambiental.

El segundo prototipo incorpora el diseño y la implementación de un sistema de asistencia a la conducción en condiciones nocturnas y de visibilidad reducida. Este sistema tiene como base la estrategia de extracción de áreas de interés, la normalización y la toma de un umbral óptimo de detección.

Un aspecto importante en este capítulo es que en todas las aplicaciones se muestran los procedimientos ejecutados para generar realidad aumentada, considerando no solo la información que presenta una región de píxeles detectada y de su entorno, sino también la información que se puede obtener de los diferentes sensores de los dispositivos inteligentes.

Por último, en el Capítulo V se realiza un compendio de las principales conclusiones que se ha obtenido del presente trabajo de tesis doctoral. En forma adicional se muestran las aportaciones tanto de tipo metodológico como de aplicación fruto del trabajo de investigación y desarrollo. Se presentan además un resumen de las capacidades y mejoras que se han obtenido al aplicar la estrategia metodológica propuesta. Posteriormente se exponen los resultados alcanzados a través de la difusión de la información así como actividades y proyectos que giran en torno al desarrollo de la Tesis Doctoral.

Finalmente se presenta una sección de trabajos futuros, en donde se exponen posibles líneas de investigación en relación a las propuestas presentadas. La tesis doctoral también cuenta con anexos que se han considerado importantes añadir y en donde se incluye la siguiente información:

En el Anexo A se explican los espacios de color usados en el entorno de videovigilancia para presentar la base teórica en la que se sostiene el procesado de imagen presentado.

En el Anexo B se presenta la información teórica y la formulación matemática sobre la umbralización a través del modelo planteado por N. Otsu.

En el Anexo C se muestra un compendio de las bases de datos de imágenes y vídeo utilizadas y que ha servido para ejecutar las diferentes pruebas experimentales.

En el Anexo D se muestra información sobre los formatos de imagen y video empleados así como las principales características del hardware y software usado durante el proceso de investigación.

El Anexo E incluye una breve descripción del funcionamiento del sistema multihilos basado en computación paralela, que ha sido desarrollado para garantizar eficiencia y exactitud en la evaluación de los resultados obtenidos.

Capítulo II: ANTECEDENTES Y
ESTADO ACTUAL DE LA
TECNOLOGÍA

2.1. Introducción

Se orienta el estudio presentado en conocer la evolución de los sistemas de videovigilancia, su estructura y los tipos de vigilancia existentes. Además, se ha realizado un análisis sobre la percepción visual en sistemas de videovigilancia y su relación con métodos empleados en visión artificial.

Un aspecto importante es la revisión de los trabajos más destacados sobre la detección de objetivos específicos mediante técnicas de extracción de regiones en la escena. En todos los casos se hace un escrutinio de aspectos como, precisión de la detección, tiempo de procesamiento y rendimiento del sistema.

Por otro lado se ha hecho una revisión de la evolución del uso de la realidad aumentada (RA) y su empleo en sistemas de videovigilancia con distintas finalidades como son proporcionar alarmas de colisiones vehiculares, detección de congestiones vehiculares y aglutinamiento de personas, alertas sobre señales e infracciones tránsito, alertas de fuego forestal, etc. Finalmente se presentan las conclusiones sobre los estudios realizados.

2.2. Introducción a los sistemas de videovigilancia

Los sistemas de videovigilancia tienen como misión principal el monitorizar áreas específicas de un campo visual determinado. Los sistemas de vigilancia a través de los sistemas de visión artificial buscan procesar la información visual en forma análoga a la percepción visual natural. Los sensores envían la información para que una unidad central de procesamiento (CPU) ejecute algoritmos matemáticos de tal manera que se pueda interpretar el entorno de vigilancia.

Los sistemas de videovigilancia se componen de un conjunto de sensores ópticos que realizan una cobertura visual de áreas definidas y que adquieren la información de diversos entornos. Esta capacidad se ha visto potenciada por su automatización y los sistemas han pasado de ser dependientes de la actuación por humanos a ser capaces de generar alarmas de detección de forma automática, así como a interpretar escenas y tomar decisiones autónomas. Es decir, la vigilancia ha pasado de ser meramente pasiva a ser autónoma y activa.

2.2.1. Evolución de los sistemas de videovigilancia

Tradicionalmente, los sistemas de videovigilancia disponen de una topología punto-multipunto, conocida como circuito cerrado de televisión (CCTV). La primera versión de CCTV para vigilancia del tráfico vehicular se presentó en 1969 en la ciudad de Olean, estado de Nueva York. La arquitectura de los CCTV en sus inicios fue analógica y conforme aparecieron los ordenadores y la comunicación alámbrica entre ellos, todo el conjunto del CCTV se fue modificando hacia la arquitectura digital. El empleo de la tecnología digital en la fabricación de procesadores, sensores ópticos, elementos transmisores y receptores así como el desarrollo de protocolos eficientes de comunicación y compresión de la información ha garantizado la transmisión de información cada vez con mayor calidad de servicio QoS (*Quality of Service*) y en menor tiempo. Precisamente la implementación de las primeras redes de CCTV digitales se presentaron en 1995 con la aparición de las cámaras IP (*Internet Protocol*), las redes de datos, repositorios y terminales digitales. Las cámaras IP a través de sensores CCD (*Charge Coupled Device*) o los CMOS (*Complementary Metal-Oxide-Semiconductor*) captan y transmiten una señal de video digital a través de la red IP

estándar a otros dispositivos de red interconectados como pueden ser un PC de escritorio, una consola o un dispositivo inteligente (teléfono, tableta).

En la actualidad, la miniaturización de componentes permite desarrollar arquitecturas descentralizadas en donde los sensores ópticos pueden llevar a cabo por si solos tareas de interpretación y reconocimiento de forma automática.

Mediante una dirección IP privada, un servidor web y protocolos de retransmisión de video (*video streaming*), los usuarios autorizados pueden visualizar, almacenar y gestionar video de forma local o remota y en tiempo real. Cada usuario autorizado es capaz de controlar y gestionar varias cámaras al mismo tiempo desde cualquier lugar donde haya conexión de red. Estas redes pueden ser de área local LAN (*Local Area Network*) o de cobertura extendida WAN (*Wide Area Network*). Por otra parte, el uso de las redes de datos y telecomunicaciones permiten la integración de las redes de CCTV con sistemas biométricos y de control de accesos, lo que ha posibilitado la videovigilancia, el monitoreo y el control en forma remota.

La estimación del número de cámaras profesionales instaladas para vigilancia en 2014 es de más de 245 millones alrededor del mundo y con un crecimiento del 7% en 2015, lo cual da una idea del potencial y la importancia de su uso en la actualidad y en el futuro (Jon Cropley, 2016). Tradicionalmente las cámaras de vigilancia externa han sido instaladas en postes, torres, o habitáculos, etc., lo que ha limitado la cobertura visual al ser infraestructuras fijas, este problema se ha solucionado instalando cada vez un mayor número de cámaras. Este concepto de vigilancia a través de estructuras fijas ha venido cambiando en los últimos años, según el informe 2016 de la “*Industrial High Surveillance*” (IHS) la vigilancia muestra cambios debido principalmente a la presencia de nuevas tecnologías que le permiten ser más dinámica, interactiva e inteligente (Cropley & Industrial High Surveillance, 2016). Precisamente estas características de inteligencia, dinamismo e interactividad tienen relación al uso de algoritmos de procesamiento digital de imágenes, visión artificial y aprendizaje automático con los sistemas de videovigilancia (Hampapur et al., 2005; Jon Cropley, 2016; Tian, Feris, Liu, Hampapur, & Sun, 2011). El estudio citado concluye que en un futuro inmediato el desarrollo de soluciones y las tendencias relacionadas con videovigilancia deberán incluir:

- Vigilancia móvil a través de drones.- La movilidad, flexibilidad y robustez que proporcionan tanto los vehículos aéreos no tripulados (*Unmanned Aerial Vehicles - UAV's*), como drones promete una nueva etapa para la vigilancia. Se considera a la vigilancia a través de drones como la versión 2.0, por el cambio en la concepción tradicional de la vigilancia (*mobile video surveillance*).
- Videovigilancia bajo el concepto 4K.- Que proporcione alta definición, estabilización y que permita observar en coberturas de 180°. La implementación de tecnología de “ojo de pez” (*fish eye*) permite observar ya en 180° sin distorsiones. Este tipo de tecnología se ha incorporado en las cámaras de vigilancia de drones como en las versiones Parrot®.
- Mayor almacenamiento.- Mejorar las opciones y capacidades del almacenamiento de la información es otra tendencia actual, sobre todo por el incremento de cobertura visual y creciente demanda de sensores.
- Integración en el internet de las cosas (IoT).- Se buscará desarrollar tecnologías mixtas que permitan la interacción de los sensores ópticos a través del internet de las cosas (*Internet of Things - IoT*).

- Arquitecturas descentralizadas.- Este tipo de arquitecturas deberá garantizar el análisis y la comprensión de video para ejecutar tareas de vigilancia inteligente en tiempo real.
- Sensores ópticos portables.- Una tendencia en crecimiento representa los denominados sensores ópticos o cámaras portables (*body worn*) que son empleados como herramientas de vigilancia y transmisión de información visual en tiempo real. Este tipo de sensores están integrados en los dispositivos inteligentes o embebidos en elementos de uso cotidiano como son las prendas de vestir, accesorios de moda, etc.
- Desarrollos dirigidos a la vigilancia medioambiental.- La actividad de control medioambiental a través de sistemas de videovigilancia cada vez cobra una mayor importancia debido al incremento de desastres naturales debido al cambio climático y a los factores asociados a este.

2.2.2. Estructura de un sistema de videovigilancia típico

Un sistema de vigilancia típico está conformado principalmente por tres elementos: una red de sensores ópticos, una infraestructura de transporte de la información y finalmente, un centro de control en donde se incluye el hardware y software para ejecutar tareas de control, análisis e interpretación de la información de video. La Figura 2.1 muestra un esquema típico.

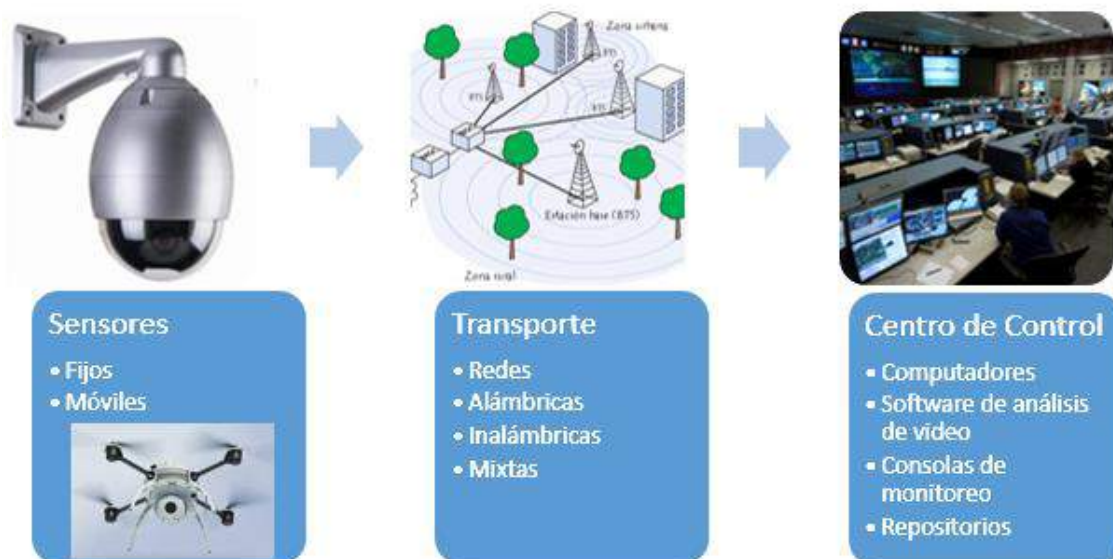


Figura 2.1. Estructura típica de un sistema profesional de videovigilancia.

Las cámaras alojadas en plataformas fijas y en móviles (drones) son instrumentos para adquirir información (secuencias de video), que se comprime y codifica para ser transportada a través de redes de comunicación tanto alámbricas (Ethernet, fibra óptica) como inalámbricas (Wifi, Wimax, 5G). Esta información se transporta a centros de control en donde se realiza el análisis y procesamiento de la misma.

El diseño y puesta a punto de este tipo de infraestructuras de videovigilancia puede llegar a ser una tarea que demanda muchos recursos y que está supeditada al perfecto funcionamiento de los subsistemas que la conforman. En especial las redes de vigilancia de área amplia, son utilizadas en tareas para la seguridad de las personas, infraestructuras, recursos, áreas estratégicas así como en el control del tránsito y transporte terrestre.

Las cámaras IP de videovigilancia utilizan en forma general los formatos de compresión MJPEG, MPEG-4 y H.264, especialmente el formato MPEG-4 es una norma de aceptación universal (Mata, 2010). En este tipo de sistemas, el formato MPEG-4 garantiza flexibilidad y calidad ya que permite transmitir vídeo e imágenes a través de un ancho de banda estrecho, permitiendo además mezclar vídeo con texto, gráficos y capas de animación 2D y 3D. El MPEG-4 garantiza la codificación y transmisión de datos de video/audio hasta de 1,5 Mbit/s.

2.2.3. Videovigilancia inteligente

Mediante el término de videovigilancia inteligente, se refiere a la capacidad de los sistemas para detectar e interpretar una escena de forma automática proporcionando alarmas o información de una escena, situación o evento en concreto. En el centro de control o centro de procesamiento de datos de video es donde los diferentes algoritmos de visión artificial interactúan con la información obtenida por los sensores. La interacción permite detectar los objetivos que son de interés, ejecutar una clasificación de los objetivos en base a sus características, realizar el seguimiento del objetivo de interés, establecer un reconocimiento de patrones y finalmente proporcionar información o interpretación del evento. En el esquema de la Figura 2.2, se muestra los subprocesos que presenta un sistema de vigilancia inteligente.

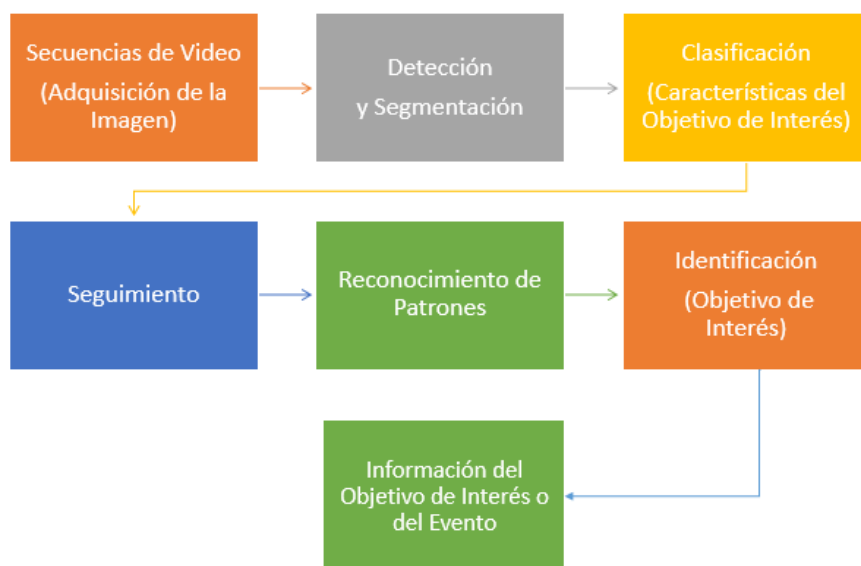


Figura 2.2. Descripción general de subprocesos en un sistema de videovigilancia inteligente.

Como se puede observar, el módulo de detección y segmentación es genérico, es decir todos los objetos en la escena son detectados sin distinguir unos de otros y la distinción y extracción de un objetivo específico, se realiza en el módulo de clasificación. Una vez extraída la región de interés (*Region of Interest - ROI*), se realiza el seguimiento de esta ROI considerando diversos patrones espaciales, temporales, predictivos, estadísticos, probabilísticos, etc. La correlación de características y patrones del objetivo permite su identificación en la secuencia de video. Además, el análisis de las características del objetivo así como de su entorno ayuda a identificar patrones de comportamiento. Para los sistemas de videovigilancia inteligente, lo ideal es que todo este conjunto de módulos actúe en forma interdependiente y en forma eficiente.

Los sistemas inteligentes de videovigilancia deben proporcionar alertas en tiempo real, de tal manera que permitan una reacción oportuna. Estas alertas son clasificadas de dos maneras: alertas definidas por el usuario y alertas automáticas ante comportamientos inusuales (Hampapur et al., 2003a; Mata, 2010). Las alertas definidas por el usuario, se refieren a aquellas que se presentan en el seguimiento de un evento o de un área específica como por ejemplo los incendios forestales. Los sistemas de videovigilancia en este tipo de alertas deben proveer información con oportunidad para que la reacción sea en igual medida, es decir la información, las alertas o la notificación de un suceso (siniestro, emergencia, anomalía) debe ser al instante que se produce el mismo, para que las medidas de mitigación o control rindan efecto. Existen los siguientes tipos de alertas definidas por el usuario:

- Alarmas de tipo genérico.- que detectan el movimiento del objetivo (personas u objetos), y que en base al análisis del movimiento proporcionan información sobre las características del mismo, como la velocidad, la posición espacial y la presencia en la escena. Este tipo de información permite conocer por ejemplo cuando un objeto ha sido abandonado, cuando una persona ha entrado en un área restringido o cuando un vehículo ha sobrepasado un límite de velocidad.
- Alarmas de tipo específico.- se muestran cuando un objetivo o evento definido se presenta en un tipo de escena determinada poco común. Este tipo de alertas es muy útil para evitar o mitigar catástrofes o accidentes; por ejemplo en aplicaciones de vigilancia de zonas forestales, cauces de ríos, líneas costeras, pistas de aterrizaje, etc.
- Alarmas de tipo estadístico.- se presentan cuando se sobrepasa o no se alcanza un número definido de objetivos en la escena. El análisis estadístico permite por ejemplo establecer cuando el aforo de un local ha excedido su capacidad o cuando existen aglomeraciones en áreas determinadas (*crowed scenes*). El uso de umbrales es muy frecuente en la interpretación de datos en este tipo de alarmas.

Las alertas automáticas ante comportamientos inusuales se presentan cuando existe una interrupción de los patrones normales, por ejemplo al detectar anomalías en la vigilancia de zonas peatonales por la presencia de bicicletas, vehículos motorizados, etc., así mismo en la vigilancia de estaciones de trenes cuando se produce la caída de una persona u objeto a los rieles del tren, o cuando se salta los torniquetes de pago del tren, etc.

Este tipo de alertas se vinculan también al proceso de recuperación automática de información de video. Esto es posible a través de la extracción y comparación de ciertas características de un objeto como puede ser su forma, tamaño, apariencia, trayectoria espacial y temporal, etc. Existen dos formas de análisis de objetivos de interés, una basada en la recuperación de espacio temporal del objetivo y otra basada en la minería de datos. Otro aspecto que destaca Hampapur es la importancia del módulo de detección del objetivo, pues de su eficacia dependerá el rendimiento del resto de módulos.

2.2.4. La percepción visual y su relación con la detección de objetivos en videovigilancia inteligente

La detección automática de objetivos es una tarea clave en la videovigilancia inteligente, posiblemente entre las más importantes, de ahí que esta tarea ha sido objeto de investigación y desarrollo permanente. Es por ello que materias como la visión artificial tratan continuamente de emular la percepción visual biológica de los seres humanos. Para conseguir este objetivo, ha sido

necesario estudiar cómo funciona el sistema visual humano y cómo afectan las condiciones externas como la iluminación en la adquisición de la información.

Los sistemas artificiales de visión intentan reproducir en los procesadores la percepción visual humana y su mecanismo de abstracción, es decir que un organismo autónomo de vigilancia disponga de la capacidad de interpretar una acción o suceso, por ejemplo distinguir cuando se ha movido un objeto, interpretar que es un incendio o simplemente dar una alarma cuando se ha producido este evento, identificar cuando una zona ha sido deforestada, etc. Para conseguir este tipo de interpretaciones se requiere primeramente detectar los elementos que conforman una escena o separar un objeto en específico de la escena, para lo cual se analiza la información que proporciona tanto de la ROI como el entorno de la misma así como la respuesta que pueden existir a estímulos externos (variaciones de iluminación, respuestas de reflectancia de los objetos).

En los sistemas de videovigilancia inteligente, tradicionalmente el módulo de detección y segmentación utiliza el movimiento como una fuente de adquisición genérica de información. Especialmente en la década de los noventa se plantea la extracción de características basada en las diferencias que presentan los píxeles, lo que permite encontrar patrones de movimiento y su correspondiente representación en dos dimensiones. Trabajos como el propuestos por Cedras et al., da cuenta del desarrollo de diversas técnicas que fundamentalmente se basan en el manejo de patrones espacio temporales, para conseguir detecciones (Cedras & Shah, 1995). A partir del año 2000, se plantean diversos métodos que permiten no solo la detección sino que permite la interpretación de escenas de video, la revisión realizada por H. Weiming et al., presenta un extenso estudio de todos estos métodos. De este trabajo se desprende que las principales aplicaciones de los sistemas de videovigilancia están dirigidas a desarrollar controles de acceso, identificación de personas, detección de flujo de tráfico vehicular, aglomeraciones y congestiones de personas, detección de anomalías y alarmas en áreas comunes, además plantea el uso de múltiples cámaras para una vigilancia interactiva. H. Weiming et al., dan cuenta también que la vigilancia puede desarrollarse bajo dos esquemas uno a través de escenas dinámicas y otro a través de escenas no dinámicas (estáticas) (Hu, Tan, Wang, & Maybank, 2004a). Otros autores conciben a las escenas no dinámicas como un primer plano estacionario (*stationary foreground*) (Cuevas, Martínez, & García, 2016). Las escenas dinámicas están relacionadas al movimiento de los objetos en donde se percibe un cambio constante de características tanto del fondo (*background*) como del primer plano (*foreground*), por ejemplo en la vigilancia de calles, carreteras, plazas, autopistas, interiores, etc. Entre los principales métodos que basan la vigilancia en base al esquema de escenas dinámicas está la sustracción del fondo, los métodos basados en la diferencia espacio temporal, flujo óptico así como métodos basados en reconocimiento de patrones y clasificación automática (K. Kim & Davis, 2011; Ojha & Sakhare, 2015).

En general estos métodos funcionan sin mayores inconvenientes en escenarios controlados como es el caso de videovigilancia de calles, plazas, personas, vehículos, etc., sin embargo presentan fallos cuando la vigilancia se ejecuta sobre fondos muy complejos o cuando los objetivos a detectar presentan formas irregulares, no representan objetos sólidos, son poco predecibles o no siguen un patrón determinado de tal manera que la clasificación es muy compleja. Este tipo de escenarios se presentan en la detección y clasificación de especies vegetales, bosques, líneas costeras, erosión, derrames de combustibles, detección de humo, etc.).

La vigilancia en escenas no dinámicas corresponde a aquellas en las que los cambios en los píxeles se perciben en períodos más largos de tiempo, por ejemplo; en la vigilancia de zonas erosionadas

(d'Oleire-Oltmanns, Marzloff, Peter, & Ries, 2012a), en la agricultura automatizada (Herwitz et al., 2004; Primicerio et al., 2012), en el control de crecimiento de la vegetación (Tang & Shao, 2015a), en trabajos de identificación de suelos (Gao, Spiteri, Pham, & Al-Milli, 2014a), análisis e identificación de video (Hung, Hsieh, Kuo, & Pan, 2011). Una de las formas de detectar los cambios que se producen en este tipo de escenas es a través del análisis de diferencia o similitud de las segmentaciones. La segmentación es una tarea asociada a la detección de un objeto, esta representa la separación de las regiones de interés (*Regions of Interest- ROI's*) del resto de la escena. En los sistemas de videovigilancia, la detección a través de escenas no dinámicas requiere de un pre-procesamiento de la imagen o cuadro de video. La separación de las ROI's en una secuencia de cuadros de video es conocida como la segmentación semántica de los objetos en video. Para llevar a cabo la segmentación semántica de objetos, se han presentado propuestas basadas en el crecimiento de regiones, energía difusa, *superpixels*, contornos activos, bordes, extracción de características prominentes, etc. Para ello se han extraído características de color, contornos, texturas y apariencias (Ojha & Sakhare, 2015). El uso de estas técnicas en aplicaciones de videovigilancia no ha producido los resultados esperados pues no siempre guardan un compromiso entre la calidad de la detección y el tiempo de procesamiento empleado para llevarla a cabo (Minetto et al., 2012).

Otras clasificaciones muestran la detección de objetivos específicos dentro de dos categorías, una de detección objetos con formas invariantes y otra con variabilidad de formas. Según Hsiao y Herbert, alcanzar precisión con eficiencia computacional en la detección de objetivos con formas invariantes es un reto por alcanzar (Hsiao & Hebert, 2013).

La extracción de las características de una imagen o cuadro de video es un elemento clave para ejecutar la segmentación, detección y clasificación del objetivo de interés. Las principales características que sirven para la detección de un objetivo específico o de interés están relacionadas con la extracción de información del color, forma, textura y características geométricas del objeto.

2.3. Detección de objetos específicos en la escena

La detección de un objetivo específico está ligada a la extracción de características de los elementos más relevantes que se pueden obtener de la imagen, para ello se han considerado características de color, intensidad y orientación del objeto a detectar. Un objeto relevante en la escena es un elemento representativo de la misma, que es importante y por sus características intrínsecas es fácilmente distinguible del resto de objetos. Durante décadas, diversos métodos se han presentado con la finalidad de detectar estos elementos relevantes y que han servido para desarrollar un sin número de aplicaciones referentes a detección de objetos, seguimiento, seguridad, recuperación de imágenes, reconocimiento de patrones, análisis de contenido, etc., (Cheng et al., 2013).

Según W. Li et al., en el diseño de un algoritmo de extracción automática de objetivos específicos, es necesario considerar tres condiciones básicas. La primera tiene que ver con categorizar los objetivos, es decir el algoritmo debe resaltar el objetivo específico del resto de una imagen o un cuadro de video. Segundo, el algoritmo debe ser capaz de detectar el objeto aunque su movimiento no sea uniforme. Finalmente, el algoritmo debe ser capaz de reconocer el objetivo específico en la escena aunque exista ambigüedad debido a otros objetos o fondo con colores similares, bajos

contrastes, poca iluminación, escenarios irregulares o texturas no uniformes etc. (W. Li, Chang, Lien, Chang, & Wang, 2013).

La percepción visual humana es capaz de identificar cuando un objeto destaca del resto y realiza una abstracción del suceso. Los sistemas de visión artificial buscan simular este tipo de capacidades, para ello un paso inicial consiste en ejecutar un tratamiento de la imagen considerando las características de la imagen (color, texturas, relaciones espacio-temporales, etc.).

La apreciación de los objetos relevantes o de los que destacan en la escena es conocida como *visual saliency*. La percepción de un objeto relevante puede ser confundida con atención visual, A. Borji define claramente que la percepción visual de un objeto relevante se refiere al conjunto de procesos que permiten que ciertas áreas o regiones de una imagen sean más visibles que el resto, por ejemplo a través del contraste que existe entre tonalidades (Borji, Sihite, & Itti, 2013).

Por otro lado, Y. Li et al., define que la detección de objetivos específicos está vinculada a la extracción de regiones de interés. Las ROI's se agrupan en relación a características comunes de color, intensidad, forma, etc., entre el objeto y la escena. Múltiples técnicas han utilizado estas características, analizando únicamente una o la combinación de ellas para alcanzar precisión en la detección. Sin embargo, la precisión de detección debe estar acompañada de una eficiencia computacional, esta premisa pasa a ser de cumplimiento obligatorio para sistemas que funcionan en tiempo real, pues si la tasa de detecciones es alta pero se requiere un mayor tiempo para ejecutarla (debido a la complejidad de los métodos planteados, redundancias de procesos, hardware o software deficiente, etc.), entonces la información que se proporciona no es oportuna y por lo tanto no es de utilidad (Y. Li, Wang, Tian, & Ding, 2015).

Para que un método de detección de objetivos específicos pueda ser utilizado en sistemas de vigilancia, es necesario que cumpla con dos condiciones primordiales, la primera es que disponga de una alta tasa de precisión en la detección y por otra parte lo realice con un bajo tiempo de procesamiento de tal manera que alcance su ejecución en tiempo real. Tiempo real para estos casos significaría que el algoritmo de detección interactúa en forma continua con la información de entrada de tal manera que el observador pueda identificar una ROI sin retardos. A continuación se presentan los métodos más representativos que consideran estos aspectos.

Uno de los trabajos que son referentes en la literatura es el presentado por R. Achanta et al., en 2008. Achanta extrae características tanto de la luminancia como del color en la imagen para realizar comparaciones y calcula las diferencias de valores medios entre regiones de píxeles, lo cual le permite generar segmentaciones consiguiendo de esta manera extraer las ROI's del resto. A partir de las ROI's, Achanta extrae un mapa de características sobresalientes, conocido como *saliency map*. Este mapa es concebido como una forma de representar las zonas resaltadas del resto de la imagen. En este trabajo se destaca el bajo tiempo de procesamiento con una media de 0,12 s para imágenes de 320x240 píxeles, el tiempo de procesamiento alcanzado permite inferir que a través de la extracción de características del color es posible alcanzar detecciones en tiempo real. En cuanto a valores de precisión de la detección, los autores no presentan datos concretos y más bien las conclusiones presentadas motivan a establecer bajas tasas de detección (Achanta, Estrada, Wils, & Süssstrunk, 2008). En la Figura 2.3, el resultado de la extracción del objeto relevante a partir del método planteado por Achanta.

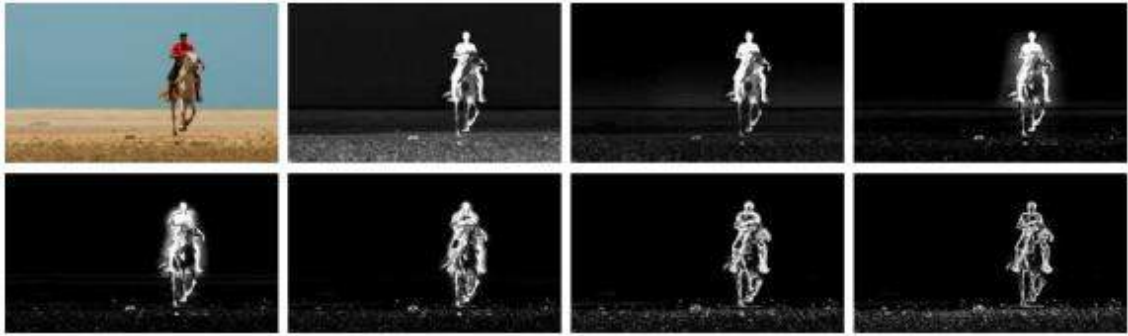


Figura 2.3. Mapa de características prominentes a través del análisis de la luminancia y el color en diversas regiones, método presentado por Achanta (Achanta et al., 2008).

Por otra parte, C. Sandor et. al., realizan la detección de objetos sobresalientes considerando la ejecución de operaciones algebraicas de adición y diferencia de valores de luminosidad y las componentes normalizadas r , g , b (Anexo A). Las operaciones algebraicas son efectuadas en diferentes niveles considerando el redimensionamiento de la imagen. El desarrollo de este método está dirigido a aplicaciones de realidad aumentada en base al solapamiento de imágenes relevantes (Sandor, Cunningham, Dey, & Mattila, 2010). El rendimiento del sistema permite ejecutar la detección en tiempo real considerando una resolución de 640x480 píxeles y únicamente objetivos de hasta 16 píxeles, objetivos de mayor tamaño reducen las tasas de precisión significativamente. Un esquema del proceso de adquisición del objetivo relevante se muestra en la Figura 2.4.

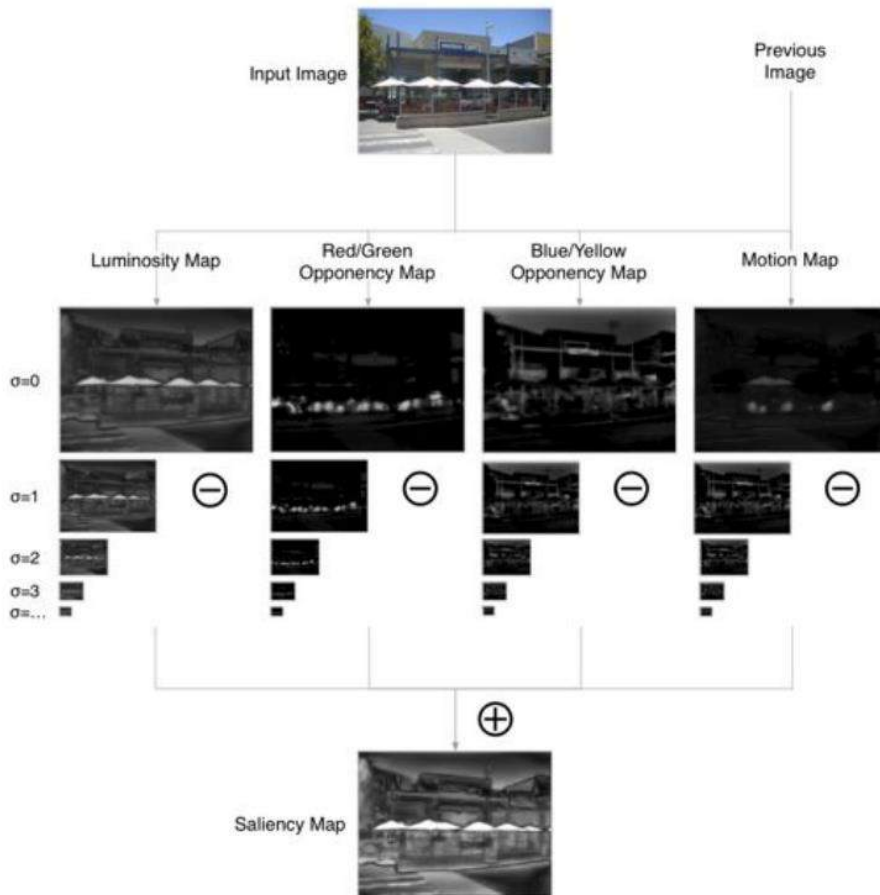


Figura 2.4. Esquema de adquisición de objetos relevantes considerando operaciones algebraicas de adición y sustracción de características de color y luminosidad en diferentes niveles, tomado de (Sandor et al., 2010).

Por otra parte, (Hung et al., 2011) utiliza la diferencia entre las componentes Cb y del Cr del YCbCr (luminancia, crominancia de rojo, crominancia de azul) y la función de densidad de probabilidad (*probability density function - pdf*) para encontrar la distribución de una misma tonalidad de color en la escena. Este método es utilizado en la detección de diferentes regiones de acuerdo a la tonalidad de color en escenas deportivas. El método presentado alcanza una precisión de detección del 94,49% y un tiempo de procesamiento de 2,53 s. Este último valor no asegura una detección en tiempo real. Cabe destacar además, que el tiempo de procesamiento muestra mejoras en relación a otros como *mean shift based* y los basados en Modelos Gaussianos Mixtos (*Gaussian Mixture Models - GMM*).

En 2011 W. Kim et al., plantea el uso de la detección de objetos relevantes tanto en escenas estáticas como dinámicas a través de la obtención de histogramas de color y bordes. En escenas dinámicas se establece una correlación de los valores de posición de los píxeles de interés en los diversos cuadros de video. Kim parte de la identificación de regiones comunes en los diferentes cuadros de video a través de histogramas de color y detección de bordes. Cuando existe una correspondencia de estos dos parámetros, se establecen regiones locales (W. Kim, Jung, & Kim, 2011), para obtener los histogramas de color se utiliza el espacio HSV (*Hue, Saturation, Value*). Por otra parte se obtiene los vectores de los píxeles de los bordes a través de sus histogramas correspondientes. Finalmente, una evaluación de correspondencia de valores permite identificar el objetivo específico o relevante en la escena. La evaluación del rendimiento del algoritmo muestra una tasa de precisión de detección del 51% y un tiempo de procesamiento de 23 ms en cuadros de video de baja resolución (320x240 píxeles).

En el trabajo presentado por F. Kan et al., las características de color se combinan con la extracción de características de textura para conseguir la detección del objetivo. Específicamente se analiza el grado de intensidad que presentan los colores en la imagen por un lado y por otro se analiza el cambio de intensidad en base a histogramas orientados a gradientes (*Histograms of Oriented Gradients - HOG*), el análisis y correlación de resultados permite obtener el objetivo de interés. El método propuesto alcanza precisiones hasta de un 61% y da cuenta de una reducción de coste computacional aunque no se especifica el valor (Khan et al., 2012).

Por otra parte, X. Hou et al., utiliza la transformada discreta del coseno inversa (*Inverse Discrete Cosine Transform - IDCT*) para realzar regiones específicas de color que forman parte del primer plano separándolo del fondo, obteniendo así mapas de características prominentes (Hou, Harel, & Koch, 2012). Este método tiene un tiempo de procesamiento de 0,024 s para imágenes con un tamaño reducido de 64 x 48 píxeles. La precisión de la detección bordea el 69,49% debido a la pérdida de información por baja resolución. En este trabajo no existe un compromiso entre tiempo de procesamiento y precisión de la detección. En la Figura 2.5, se describe como obtener los objetos relevantes en la escena a través del método planteado por X. Hou. Se muestra cómo funciona el método, en donde se aplica la IDCT en cada canal de color para obtener mapas parciales y la posterior adquisición de un mapa final.

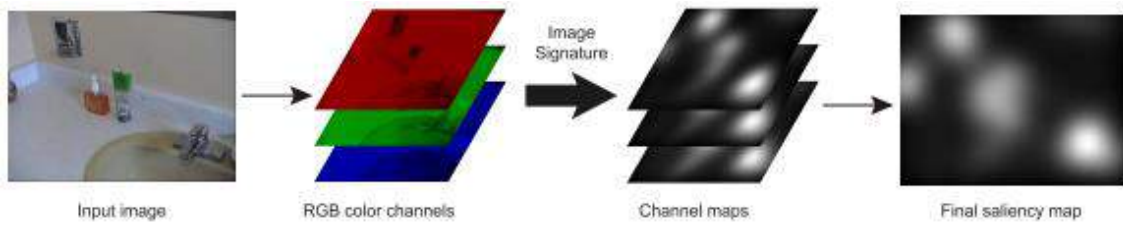


Figura 2.5. Obtención del mapa de características prominentes a través de la IDCT propuesta por (Hou et al., 2012).

S. Goferman et al., en cambio, utiliza el contexto que rodea al objetivo específico para realizar su detección. Para ello se analiza la concentración de color en la imagen, destacando una región de interés donde existe la mencionada concentración. Luego para asegurar precisión en la detección, se calcula la distancia Euclídea entre los píxeles que conforman la ROI con los píxeles de su entorno, de esta manera se establece su similitud o diferencia. Un píxel es considerado de interés y pertenece a una ROI, cuando el valor de la distancia euclídea es más alto en la comparativa. A través de este procedimiento se establece límites entre la ROI y el resto. Las pruebas son realizadas en el espacio de color CIE L^*a^*b , el cual es normalizado considerando escala binaria de [0,1] (Goferman, Zelnik-Manor, & Tal, 2012). La tasa de detección de verdaderos positivos (*TPR*) alcanza un 87% con un tiempo de procesamiento medio de 4,8 s por imagen, este tiempo de procesamiento impediría su aplicación en tiempo real.

Por otra parte, X. Li et al., utilizan el algoritmo “*Simple Linear Iterative Clustering*” (*SLIC*), que permite la extracción de *superpíxeles* para obtener una serie de modelos (*templates*) del fondo. Estos modelos son comparados para determinar errores de correlación y para reconstruir la imagen. En la detección final de las zonas prominentes se utiliza refinamientos basados en modelos Gaussianos y Bayesianos. La tasa de rendimiento global de detección (F_m) bordea el 85% de eficacia presentando un alto coste computacional debido al sin número de tareas a ejecutar en la separación del objetivo específico del resto. En la Figura 2.6 se observa el esquema general de este método (X. Li, Lu, Zhang, Ruan, & Yang, 2013).

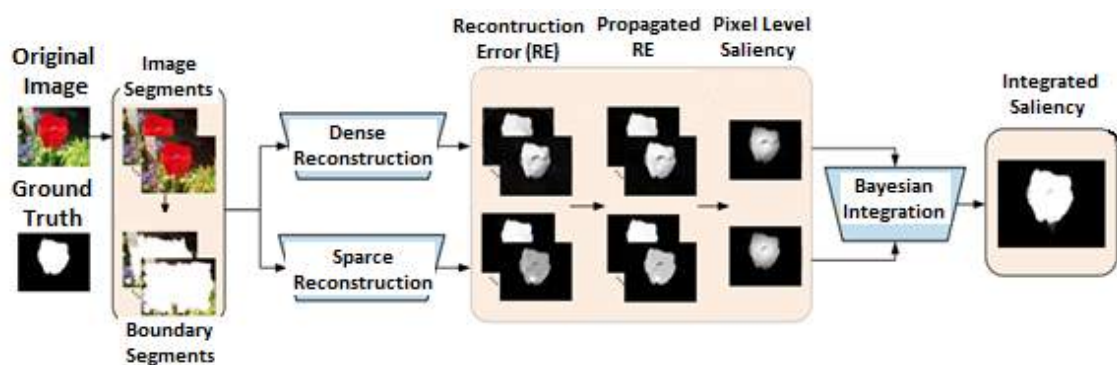


Figura 2.6. Modelo “*Sparse and Dense Reconstruction*” para la detección de objetos específicos en la escena, tomado de (X. Li et al., 2013).

J. Shuai et al., realizan una división de la imagen en regiones y la detección de formas a través del algoritmo SLIC. A diferencia del método planteado por X. Li, J. Shuai ejecuta un pre-procesamiento de la imagen para eliminar la textura de la imagen para ello se emplea el algoritmo “*Structure Extraction*”. La división en *superpíxeles* que se ejecuta es más bien homogénea, y

permite encontrar los valores que corresponden al fondo específicamente. J. Shuia infiere que la homogeneidad varía cuando se evalúan píxeles cercanos a los bordes o contornos de los objetos relevantes. La extracción de los píxeles de interés se realiza considerando las mínimas distancias de color entre los *superpixels* del fondo y de los contornos (Shuai, Qing, Miao, Ma, & Chen, 2013). El resultado de precisión de la detección bordea 91%, mientras el F_m presenta un 89% de eficacia. No se presentan valores de tiempos de procesamiento (T_p), sin embargo se infiere un alto T_p debido a las diversas tareas a cumplir para alcanzar la detección del objetivo específico.

W. Li et al., en cambio propone el uso de modelos de extracción de objetivos de interés en escenas de video utilizando la información de los píxeles que rodean al objetivo. Esta información es obtenida a través del análisis visual de la imagen considerando *turbopixels* (versión modificada de *superpixels*) y la detección de objetos en movimiento a través de flujo óptico. La combinación de esta información permite resaltar e identificar el objetivo de interés en la escena de video. Los resultados de evaluación muestran un $F_m = 86.17\%$, el autor destaca que la principal contribución del trabajo es la de poder prescindir de un entrenamiento previo y aun así garantizar detecciones (W. Li et al., 2013).

En el mismo año, M. Cheng et al., propone el uso de modelos estadísticos de similitud a través del GMM en la imagen (Cheng et al., 2013). El objetivo es establecer zonas homogéneas de color, para ello se considera el valor del píxel, la media del color y la matriz de covarianza. La aplicación de estas variables conlleva a establecer la probabilidad de pertenencia de un píxel a una región determinada de la imagen (*cluster*). El autor destaca que este procedimiento evita la complicada adquisición de contornos del objeto relevante en la escena, como se presenta en el caso de la utilización de *superpixels*. La correspondencia entre el modelo gaussiano y la distribución espacial de los colores permite identificar la ROI en la escena. La precisión de la detección a través de este algoritmo alcanza un 87%. Un aspecto a destacar es que el tiempo de procesamiento registra una media de 0,09 s para imágenes con resolución de 300x400 píxeles, dando de esta manera otra pista para establecer que es posible alcanzar el tiempo real con precisión en la detección a través del manejo de las características de color. En la Figura 2.7, se muestra como de una misma imagen se ha adquirido diversas capturas de regiones homogéneas entre ellas y que se han obtenido a través del análisis diferentes niveles que presenta el histograma de cuantificación fruto de la aplicación del GMM en la imagen original.

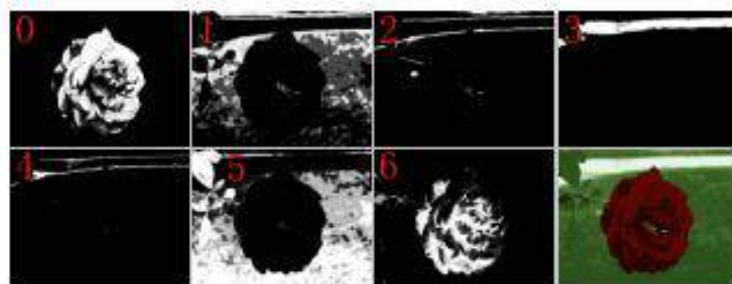


Figura 2.7. Resultado del empleo del modelo GMM para detectar las diferentes áreas de color en la imagen considerando los diferentes niveles del histograma de cuantificación de la imagen original (Cheng et al., 2013).

Por otra parte, A. Borji et al., presenta una completa revisión sobre diversos trabajos enfocados en la detección de objetivos específicos a través de mapas de objetos relevantes en la escena

(*saliency maps*). Este estudio se centra sobre todo en métodos probabilísticos y predictivos que permiten clasificación automática para usos en modelos de entrenamiento. Entre las actividades desarrolladas muestra un análisis comparativo de rendimiento, considerando una muestra de 32 métodos. Borji resume su estudio en dos conclusiones, la primera expone que los modelos predictivos presentan resultados inadecuados, en especial para escenas complejas y que no siguen patrones definidos como son los relieves del terreno. Por otra parte, este tipo de algoritmos presentan un buen resultado en escenas con objetos más bien definidos como son personas, animales, caras y textos (Borji, Tavakoli, Sihite, & Itti, 2013). Por otra parte, W. Li también expone la imposibilidad de crear modelos de aprendizaje para escenarios complejos debido a la ingente cantidad de información que debería ser entrenada tanto de los objetos prominentes como para los diversos fondos en las secuencias de video (W. Li et al., 2013). Estas conclusiones son básicas para determinar la línea de investigación en torno a conseguir contribuciones en la detección de objetivos de interés de formas no definidas o simétricas.

Siguiendo la línea de X. Hou, en 2014 F. Li et al., se utiliza la transformada discreta del coseno (*Discrete Cousine Transform - DCT*) para analizar la disimilitud de concentración de la energía en las diferentes regiones de una imagen. En primer lugar, Hou obtiene sub imágenes que no es otra cosa que la aplicación de GMM en cada canal de color, luego realiza un cálculo de la distancia entre los píxeles de las sub imágenes, cuando la distancia es menor entre píxeles, se establece que existe un mayor número de píxeles por lo tanto Hou establece que hay mayor concentración de energía. Finalmente se correlaciona valores entre las sub imágenes y la diferencia entre los valores del fondo y los píxeles de interés permitirá resaltar el objetivo relevante del resto (F. Li, Li, Zou, & Chen, 2014). La evaluación se realiza considerando modelos de predicción de acierto, para ello utiliza los valores del área bajo la curva de características operativas del receptor (*Receiver Operating Characteristic - ROC*), alcanzando de esta manera una probabilidad de acierto del 78% en la detección. No se registran valores de tiempos de procesamiento. En la Figura 2.8, se muestra el esquema general de desarrollo del método propuesto F. Li.

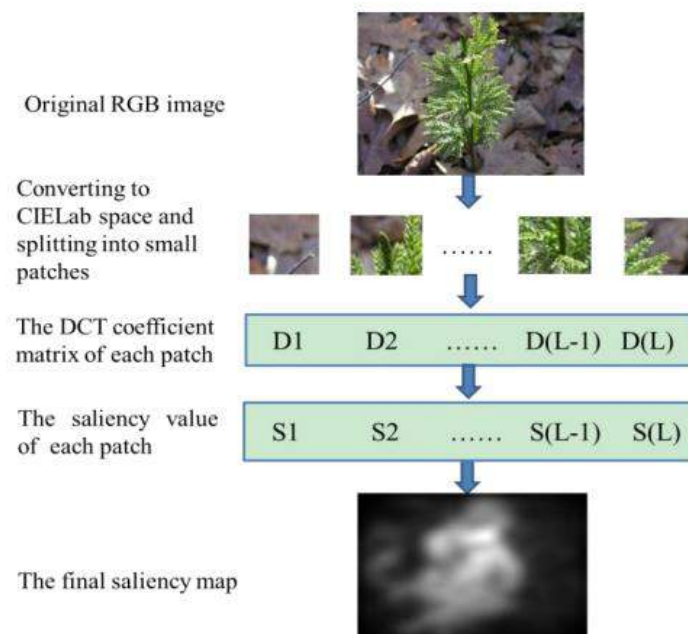


Figura 2.8. La detección de objetivos de interés en la escena a través de la DCT según el modelo propuesto por (F. Li et al., 2014).

Y. Seo et al., con base en el modelo planteado por X. Li en 2013, muestra la extracción de las características prominentes de una imagen a través de la actualización de la información de etiquetado obtenida por medio del algoritmo SLIC. Para ello utiliza la media aritmética de los valores de las componentes de color RGB y la luminancia de los píxeles en el espacio CIELAB¹, de esta manera se crean regiones que corresponden a los *superpixels*. Los autores denominan a la actualización de la información de etiquetado como diccionario bipartito (*Bipartite Dictionary - BD*). Este trabajo presenta una tasa global de detección (F_m) del 87% y mejores resultados frente a 12 algoritmos, por otra parte no registra datos de tiempos de procesamiento (Seo, Lee, & Yoo, 2014), en la Figura 2.9, se muestra los resultados comparativos entre el BD frente a 12 algoritmos, que se refieren a la detección de objetos relevantes en la escena y son los siguientes: *Visual Attention for Rapid Analysis (IT)*, *Graph-Based (GB)*, *Spectral Residual Approach (SR)*, *Visual Attention by Spatio Temporal Cues (LC)*, *Salient Region Detection and Segmentation (AC)*, *Frequency Tuned Saliency Region Detection (FT)*, *Context-Aware (CA)*, *Global Contrast Based Salient Region Detection (HC-RC)*, *Automatic Salient object Detection Context and Shape Prior (CB)*, *Saliency Detection via Dense and Sparse Reconstruction (DSR)*, *Hierarchical Saliency Detection (HS)*.

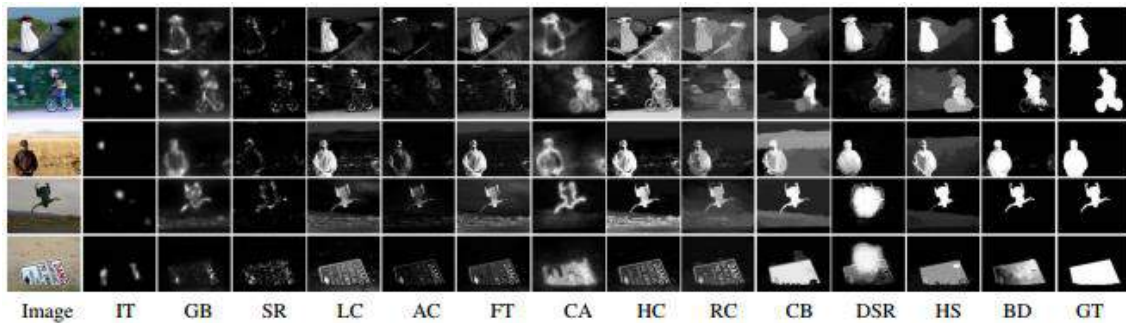


Figura 2.9. Muestra visual comparativa del método “Bipartite Dictionary (BD)” frente a otros 12 métodos propuestos en el estado del arte (Seo et al., 2014).

En ese mismo año otros autores proponen modificaciones del algoritmo SLIC en la detección de objetos relevantes en la escena, por ejemplo P. Chang et al., divide la imagen en regiones de pequeño y mediano tamaño, lo que es conocido como *superpixels* de medio y bajo nivel. Luego agrupa los píxeles que corresponden al objeto relevante a través del modelo probabilístico de Markov, también conocido como *Markov Random Field (MRF)*. Este método presenta una tasa F_m de 73% (Pei, Chang, & Shen, 2014). Como se puede observar, la combinación del algoritmo SLIC con el modelo probabilístico de Markov no genera los mejores resultados.

Siguiendo la estrategia presentada, W. Kim (2011) en 2014 J. Kim et al., retoman la extracción de características de color y luminosidad para alcanzar detecciones de objetos relevantes en la imagen. Precisamente, J. Kim presenta un método que permite encontrar los contornos del objetivo de interés a través de la evaluación de la diferencia entre estas características. Para ello se definen bloques y regiones conexas dentro de la imagen considerando las disimilitudes de características de forma local y luego se utiliza la distribución estacionaria de Markov para agrupar las ROI's. Un aspecto a destacar es la propuesta de extracción de mapas de intensidad, estos mapas permiten escoger los píxeles que presentan mayor diferencia en valor, para ello se

¹ CIE 1976 L*a*b*

formulan gradientes. Los gradientes se obtienen de la diferencia de las componentes de color y luminancia en el modelo YCbCr ($\Delta RG = R - G$; $\Delta BY = B - Y$) (J. Kim, Sim, & Kim, 2014). Este método muestra una precisión hasta del 71% y un F_m de 36,4%, el tiempo de procesamiento refleja un promedio de 7,75 s para imágenes de 400x300 píxeles. En la Figura 2.10 se muestra en forma visual los resultados alcanzados por J. Kim, en la extracción de objetos relevantes en la escena.

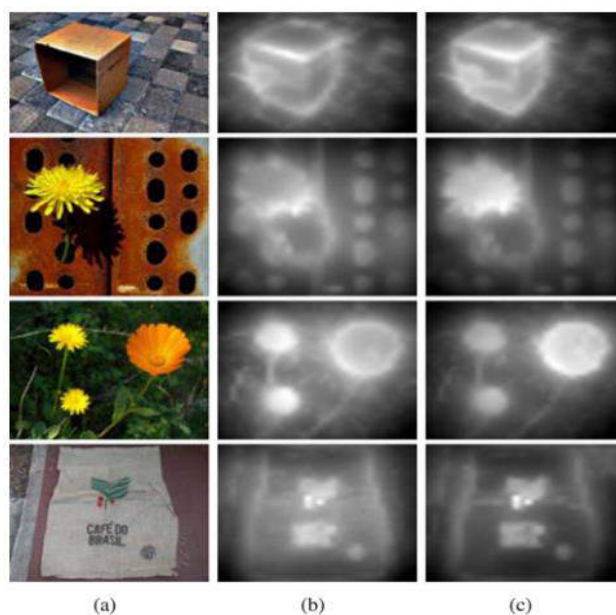


Figura 2.10. Resultados alcanzados por Kim et al. En (a) se encuentra la imagen original, en (b) la extracción del objeto relevante en la escena y en (c) el resultado de la aplicación de refinamiento de la escena (J. Kim et al., 2014).

L. Fan y A. Loui proponen el uso del modelamientos gráfico (*graph model*) para realizar la segmentación de objetos relevantes en secuencias continuas de video. Para ello, se emplea primeramente GMM para la detección del objetivo y luego se analizan las características del fondo y del objeto de interés en diferentes niveles de la imagen. Esto significa que la extracción de características de color, intensidad y contraste del objetivo de interés está asociada a un patrón de localización espacio temporal. De esta manera pueden ser identificados tanto los objetos en movimiento (personas, vehículos, olas del mar, etc.) así como el movimiento en escenarios no dinámicos (movimiento de hojas en los árboles, flujo de agua en el suelo, etc.). Los patrones y características son extraídos del modelo YUV y la localización espacial del objetivo es bidimensional (x, y) (Fan & Loui, 2015). Las estimaciones de precisión de la detección y tiempos de ejecución son más cualitativas que cuantitativas, reportando eficiencia general en estos parámetros.

K. Wang et al., propone la detección de objetos relevantes en la escena considerando un entorno común de características de los píxeles en la imagen, este entorno es conocido como “*Pixelwise Image Saliency Aggregating*” (*PISA*). Para la identificación del objetivo específico, Wang extrae características de la estructura a través de gradientes en la imagen, por otra parte extrae las

características de color del objeto predominante en relación al fondo de la imagen. Para ello considera el centro geométrico del objeto relevante para hacer comparaciones estadísticas sobre la similitud entre los valores del píxel de interés del centro con los el resto de píxeles, definiendo de esta manera los que corresponden al objeto analizado (K. Wang, Lin, Lu, Li, & Shi, 2015). Esta evaluación puede ser hecha también en diversas regiones de la imagen para conocer la pertenencia del objeto en cuestión, el proceso en forma general se lo puede observar en la Figura 2.11. Finalmente, se realiza una valoración global, considerando los resultados obtenidos de la extracción de características vía su estructura y el color. El método muestra un F_m del 93% y un tiempo de procesamiento de 0,650 s en imágenes de 300x400 píxeles.

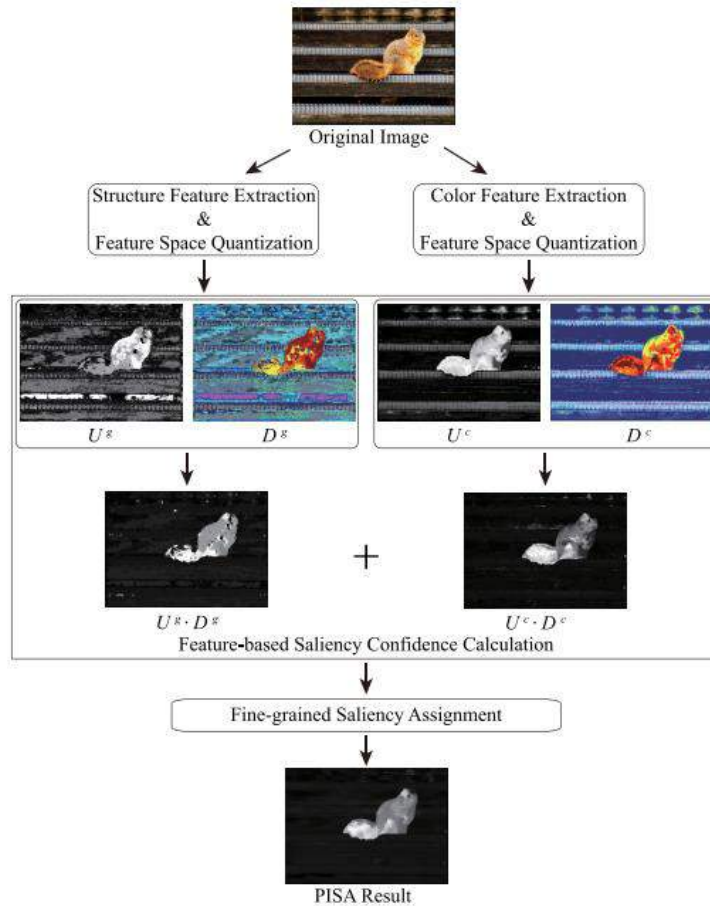


Figura 2.11. Diagrama general de la detección de objetivos específicos a través del método PISA.

Siguiendo la línea de Y. Seo (2014), L. Zang et al., propone el uso de la detección de objetivos específicos vía mapas de extracción de características prominentes considerando realzar contornos del objetivo a través del algoritmo SLIC. Para ello la imagen se divide en regiones de *superpíxeles* considerando dos espacios de color por un lado RGB y por otro CIE L*a*b. Los resultados obtenidos en cada espacio de color son comparados, ratificando las concordancias y rechazando las divergencias. La tasa de la precisión alcanzada por este método es de alrededor del 89% y su tiempo de procesamiento medio en alrededor de 2,15 s para imágenes de 400x300 píxeles (L. Zhang, Zhao, Liu, & Lu, 2015).

En el trabajo presentado por X. Peng, la detección de los objetivos de interés se realiza gracias a la combinación de la extracción de las características de color y la forma de la ROI, un aspecto importante es que consigue realizar la detección en tiempo real a una tasa de 8,8 cuadros por segundo o *frames* por segundo (fps) en un tiempo aproximado de 0,113 s. La precisión de la detección varía desde 86% al 99%, esto se debe a que el método funciona bien para objetos con estructuras definidas y que no presenten contornos irregulares (Peng, 2015). Este método presentaría altas tasas de falsos positivos en la vigilancia de escenas que demandan la detección de formas irregulares.

Recientes trabajos se implican en la detección de objetivos específicos en escenarios complejos y con fondos no definidos como son las imágenes obtenidas en la naturaleza. Específicamente H. Li et al., realiza una propuesta para la detección de objetivos relevantes, para la cual considera la construcción de una matriz de afinidad. Para ello se evalúa la similitud entre los valores de regiones de *superpixels* a través del algoritmo SLIC en un nodo determinado. Un nodo representa una región en donde se concentran valores similares de regiones de *superpixels*. Luego se realiza una evaluación de similitud de los valores de la matriz con el resto de regiones de *superpixels* que conforman la imagen en el espacio CIE L*a*b. Posteriormente se etiquetan las subregiones en la imagen considerando la similitud encontrada, y finalmente se aplica el algoritmo de correlación denominado de “co-transducción”, para agrupar las subregiones que permiten definir el objeto relevante en la escena (H. Li, Lu, Lin, Shen, & Price, 2015). El resultado final muestra una tasa media de rendimiento global de detección $F_m = 90\%$ con un tiempo de procesamiento de 0,87 s.

El uso de unidades *hardware* de procesamiento gráfico (*Graphics Processing Unit - GPU*), ayudan a realizar tareas de procesamiento gráfico que demandan una alta carga computacional. Precisamente la arquitectura de cálculo paralelo (*Compute Unified Device Architecture - CUDA*) aprovechan la redistribución de tareas que se asigna a la GPU para proporcionar un incremento extraordinario del rendimiento del sistema y de esta forma alcanzar el tiempo real. B. Hell et al., proponen el uso de este tipo de arquitecturas para realizar la detección de objetivos específicos en la escena, pero como es lógico suponer este tipo de implementaciones son muy costosas. Hell concretamente propone una optimización del algoritmo de Mumford-Shah a través de la extracción de gradientes, considerando los cambios de intensidad de los píxeles, como forma de encontrar los límites entre los objetos en la escena. El procesamiento de datos lo realiza a través de multihilos (Hell, Kassubeck, Bauszat, Eisemann, & Magnor, 2015). Los resultados muestran tiempos de procesamiento menores a 0,095 s para imágenes de 0,2 megapíxeles (450x450 píxeles aprox.) y alcanzando hasta los 0,08 ms en imágenes de 0,014 megapíxeles (120x120 píxeles) con apenas dos interacciones. Al tratarse de un algoritmo que utiliza interacciones, para asegurar eficacia en la detección se requieren al menos 900 interacciones, este número de interacciones limita su tiempo de ejecución a 1,4 s en imágenes de 640x491 píxeles.

En 2016 se presenta un trabajo que combina modelos predictivos y de entrenamiento con la extracción de características de color y posición espacial para alcanzar la detección de objetos relevantes en escena complejas y ambiguas. Específicamente en el trabajo presentado por J. Wang et al., se propone fusionar la detección de objetos relevantes en la escena considerando el contraste, la distribución espacial del color y la información espacial del centro y contorno de los objetos con modelos de predictivos y de entrenamiento basados en máquinas de soporte vectorial (*Support Vector Machine - SVM*) y a través de la estimación de la densidad de probabilidad. Los resultados alcanzados presentan alrededor de un 90% de precisión en la detección lo cual significa

que todavía el uso de este tipo de clasificadores no consigue un avance significativo en relación a los modelos tradicionales. Wang destaca sin embargo que la estrategia adoptada ha permitido mejoras en relación a anteriores trabajos que han utilizado modelos predictivos y de entrenamiento en la detección de objetivos relevantes en escenas complejas (J. Wang, Borji, Kuo, & Itti, 2016). No se presentan resultados de los tiempos procesamiento considerando tanto en la extracción de características como el requerido para el entrenamiento para alcanzar la detección del objetivo específico. El proceso general del método planteado por J. Wang se presenta en la Figura 2.12.

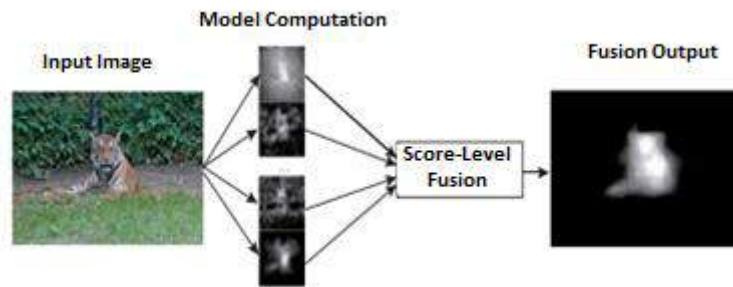


Figura 2.12. Esquema gráfico del modelo de fusión de características de detección de objetos relevantes en la escena con modelos predictivos y de aprendizaje propuesto por (J. Wang et al., 2016).

Siguiendo el modelo “*Sparse and Dense Reconstruction*” propuesto por X. Li en 2013 y H. Lu et al., en 2016 formula la minimización del error basado en la generación de un descriptor de objeto relevante en la escena. H. Lu sigue exactamente el criterio de extracción de regiones a través del algoritmo SLIC para encontrar los contornos del fondo, como se ha podido apreciar en anteriores propuestas. La diferencia radica al proponer un análisis de coincidencias entre plantillas o *templates* generados. La comparación de plantillas permite extraer los errores de coincidencia para posteriormente agruparlos utilizando el algoritmo “*K-means*” y finalmente obtener la integración del error multinivel. Esto al final permite conocer las zonas donde la presencia de errores es más densa y donde es escasa. A través del criterio de compactación de un objeto y considerando la escasas de errores se consigue detectar las ROI’s en la escena. La reconstrucción del mapa de objetos relevantes en la escena se lo realiza considerando una integración de tipo probabilística (Lu, Li, Zhang, Ruan, & Yang, 2016). Los resultados muestran mejoras en relación al método propuesto por X. Li y a otros métodos basados en algoritmos de clasificación automática con un $F_m = 86\%$, una precisión de detección $p = 95\%$ y con un tiempo de 5,010 s por imagen.

El trabajo presentado por X. Liu et al., en 2016 da cuenta de la segmentación de objetos relevantes vinculada a la representación de objetos en tres dimensiones. Expone que la segmentación en malla permite una extracción eficiente de objetos con formas más bien definidas que irregulares, en donde la extracción adecuada de las características del color juega un papel preponderante. Además este trabajo destaca que, en un futuro próximo, la detección en dos dimensiones de objetos específicos podrá ser modelada en 3D gracias al permanente mejoramiento computacional y servirá sobre todo en dispositivos de realidad virtual. De esta forma los algoritmos desarrollados en 2D serán compatibles en el uso de herramientas de realidad virtual. Destaca también que por ahora las segmentaciones en 3D vía malla no son posibles de efectuarlas en tiempo real (Liu, Liu,

Song, Liu, & Ma, 2016). La información virtual que se obtenga de los objetos en 2D, en base a la extracción de características de color, forma, textura, posición, contexto, etc., será de mucha utilidad a la hora de implementar componentes de realidad aumentada a los objetivos específicos modelados en 3D.

Por otra parte S. Tu y Y. Su, para la detección de objetivos relevantes proponen el uso de contornos activos (*Active Contour Model - ACM*) en imágenes captadas por radares de apertura sintética (*Synthetic Aperture Radar - SAR*). Es conocido que los contornos activos se ajustan a los bordes de los objetivos de interés lo cual garantiza precisión de detección sobre todo de formas irregulares, sin embargo los costes computacionales pueden ser muy altos debido a las interacciones que se requieren para su ejecución. Específicamente los autores proponen una combinación del método Chan-Vese con la obtención de objetos relevantes considerando la concentración de energía en la escena. La ejecución de Chan-Vese se limita únicamente a las regiones de píxeles que muestran la concentración de energía, lo cual permite eficiencia computacional, este proceso representa una modificación del modelo “*Local and Global Intensity Fitting*” (*LGIF*). Los resultados muestran que en imágenes de 258x195 y 242x231 píxeles, se puede alcanzar una precisión de detección considerando el coeficiente *Dice* que va desde 91,97% a 98,01% con tiempos de ejecución de hasta de 0,37s. Sin duda la precisión de la detección a través de este método se garantiza sin embargo el tiempo de procesamiento no permite su uso en tiempo real (Tu & Su, 2016) . En la Figura 2.13, se presenta un ejemplo gráfico de la detección bajo el modelo LGIF modificado.

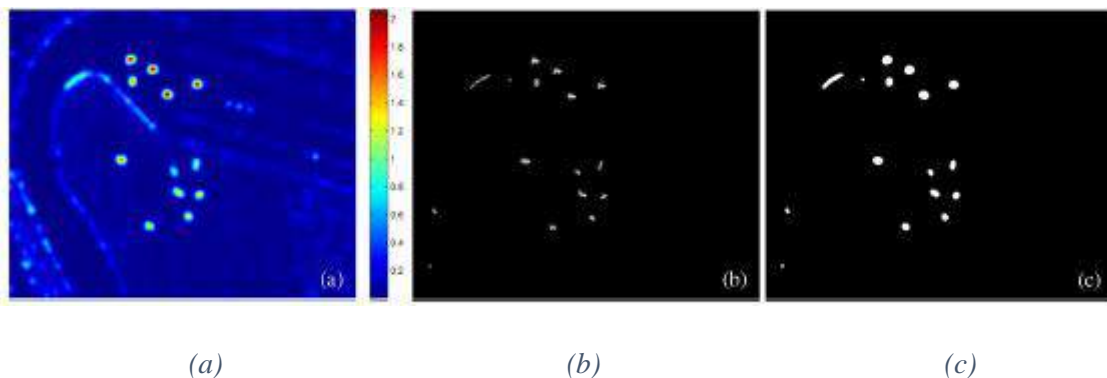


Figura 2.13. Detección de objetivos específicos en imágenes con perspectiva aérea captadas por sensores SAR: En (a) Estimación colorimétrica en escala y detección de objetivos predominantes a través de Chan-Vese, (b) mapa de objetivos predominantes en la escena, (c) binarización de la escena (Tu & Su, 2016).

2.4. Realidad Aumentada (RA) en los sistemas de vigilancia

2.4.1. Introducción a la RA

La Realidad Aumentada (RA) es la posibilidad de añadir información en un ambiente real, esta información puede ser visual en 2D o 3D, audible o sensitiva y debe ser proporcionada en tiempo real. La RA debe ser interactiva, permanente, adaptativa, que considere el contexto de aplicación, que considere diversas fuentes de información y de la misma manera proporcione diversa

información a parte de la visual (Azuma, 1997; Grubert, Langlotz, Zollmann, & Regenbrecht, 2016; Langlotz, Nguyen, Schmalstieg, & Grasset, 2014).

Sus inicios datan de la década de los años 60, cuando M. Heilig desarrolló un aparato llamado *sensorama* que buscaba dar una experiencia completa a la proyección visual introduciendo sonido, vibraciones y sensaciones olfativas. En 1968 surge la primera idea de realidad aumentada computacional cuando I. Sutherland presenta su “*Head Mounted Display*” que permitía observar gráficos virtuales. En 1994, P. Milgram y F. Kishino establecen la línea y composición de la realidad mixta, la cual va desde la realidad virtual hasta la virtualidad total, en donde los pasos intermedios en esta escala son la realidad aumentada y la virtualidad aumentada. En 1999, H. Kato y M. Billinghurst desarrollan el ARToolKit que representa la primera versión de marcadores para la observación virtual de gráficos en 3D. A partir de este trabajo, múltiples propuestas se han venido desarrollado con la finalidad de hacer una RA sobre todo interactiva. En 2009, el ARToolKit fue homologado a la versión de Adobe Flash para llevar la RA a plataformas Web. En ese mismo año se oficializa la realidad aumentada como una tecnología de soporte e información general (Azuma, 1997; Billinghurst, Clark, & Lee, 2015; Van Krevelen & Poelman, 2010a). Recientes trabajos también muestran una modificación substancial al modelo presentado por H. Kato, a través de la creación de marcadores basados en color y considerando los cambios de iluminación (Uranishi, Imura, & Kuroda, 2016).

En la actualidad, tanto los grupos de investigación como la industria de desarrollo de software alrededor del mundo se interesan por presentar aplicaciones de la RA considerando sobre todo las prestaciones que ofrecen los dispositivos inteligentes. Precisamente la versatilidad, movilidad, flexibilidad, entorno amigable de uso, alta tecnología, incremento de la potencia de procesamiento, etc., permite que en los dispositivos inteligentes se alojen aplicaciones de RA dirigidas a la educación, entretenimiento, asistencia a la conducción, vigilancia, seguridad, turismo, comercio, marketing, arquitectura, diseño, entrenamiento y ensamblaje, medicina, rehabilitación física y neurológica, tratamiento de fobias, robótica, etc. (Billinghurst et al., 2015; Papagiannakis, Singh, & Magnenat-Thalmann, 2008; Van Krevelen & Poelman, 2010b).

La RA no solo se limita a proporcionar información visual sin no que también la información puede ser audible y sensitiva. En la curva del hiperciclo de Gartner para tecnologías emergentes del 2016, la RA muestra una próxima transición del posicionamiento como una tecnología que ha disminuido sus expectativas a pasar a un posicionamiento como tecnología madura y de allí al plató de productividad en un período de 5 a 10 años (Billinghurst et al., 2015; Garther I., 2016; Gartner, 2016). Especialmente en los últimos años la industria produce nuevo hardware que permite la interacción directa con RA, por ejemplo en 2013 Google presenta el “*Google Glass*” para RA, en 2015 Hyundai presenta una guía del usuario para la conducción, proporcionando información virtual de componentes, accesorios del vehículo así como procedimientos guiados de mantenimiento, todo esto en tiempo real. En 2015 Microsoft anuncia el desarrollo del prototipo de HoloLens, el cual presenta información visual a través de hologramas en el mundo real, este se puso a disposición de los desarrolladores el primer semestre de 2016.

Otras soluciones no comerciales son puestas a disposición para la asistencia a la conducción. Por ejemplo en 2015, L. Abdi et al., proporcionan un sistema de interacción entre el conductor y el vehículo a través de RA, específicamente se insertan gráficos virtuales sobre detección de coches, alertas por posible colisión, señalización, sentido de conducción, zonas de cruce peatonal, etc. En este caso se utilizan computadores abordo con sensores ópticos externos (Abdi, Abdallah, &

Meddeb, 2015). También M. Rush presenta soluciones de asistencia a la conducción a través de RA para personas mayores, especialmente en control de curvas y eventos de rápida reacción (Rusch, Schall Jr, Lee, Dawson, & Rizzo, 2014). En la Figura 2.14, se puede observar la información que proporciona este sistema en tiempo real.



Figura 2.14. Interfaz de RA en el sistema de asistencia a la conducción propuesto por (Abdi et al., 2015).

En 2016, Qualcomm y la compañía Osterhout Design Group lanzan un prototipo de asistencia a la conducción a través de dispositivos inteligentes que provee información al conductor sin afectar el entorno de conducción y la presencia de vehículos y transeúntes, tal como se muestra en la Figura 2.15. En este mismo año, la empresa Nimax muestra su producto “Universe2Go” para crear experiencias de RA en la observación nocturna del cielo. Este producto proporciona información y localización de constelaciones, estrella, planetas, satélites, la aplicación está desarrollada para dispositivos inteligentes (Avila & Bailey, 2016).



Figura 2.15. Hyundai Virtual Guide, una muestra de guía virtual de operación y mantenimiento para el modelo de Hyundai Sonata 2015, gráfico tomado de (Avila & Bailey, 2016).

En el año 2016, Facebook lanza una campaña para crear tanto experiencias individuales como colectivas de realidad virtual y aumentada a través de dispositivos inteligentes. En ese mismo año se presenta un masivo uso de la aplicación “*Pokemon Go*” desarrollada por la compañía Nintendo para plataformas Android y IOS (*iPhone Operating System*), siendo una clara muestra del uso de realidad aumentada en juegos de entretenimiento e inmersión en la realidad mixta. También se están presentando propuestas para vincular la RA en conjunción con el IoT, de tal manera que proporcione información de los dispositivos, accesorios, electrodomésticos, etc., y permita interactuar con los mismos (Jo & Kim, 2016). Otras propuestas presentan modelos de RA interactiva y colaborativa, es decir la manipulación y el modelamiento de objetos virtuales con más de un participante en un escenario real (Cabral et al., 2016).

Finalmente, a partir del estudio presentado se puede afirmar que la generación de RA es totalmente dependiente de la aplicación que se desarrolla y su producción obedece al uso de información previamente obtenida de procesos como, la detección automática o la información que proporcionan los sensores de los dispositivos.

2.4.2. Empleo de la Realidad Aumentada en los sistemas de videovigilancia

En los sistemas de vigilancia, la información básica que se presenta como RA, está relacionada con la detección, el etiquetado y la identificación virtual del objetivo de interés. De esta manera es posible obtener y manejar información virtual sobre control de accesos a áreas restringidas, identificación de comportamientos individuales y grupales en ciertos escenarios, flujo de tránsito de personas y control de aglomeraciones, flujo y control tráfico vehicular, detección y alarmas por anomalías, información relevante de múltiples cámaras, etc. (Hu, Tan, Wang, & Maybank, 2004b).

En escenarios no dinámicos es posible desarrollar interfaces de RA para detección y estimación de zonas deforestadas o forestadas, alarmas por inundaciones, identificación de especies vegetales, detección y alarmas por incendios, información de gases contaminantes y humo en la atmosfera, detección de fondo e identificación de especies marinas, etc.

La información visual necesaria para producir RA, puede ser recogida a través de diversos sensores ópticos alojados en torres fijas en exteriores, compartimentos en interiores, automotores, vehículos autónomos no tripulados, satélites, dispositivos inteligentes portables, etc. La secuencia de cuadros de video permite producir RA, sin embargo una mayor gama de aplicaciones se obtiene al utilizar la información que proporcionan los sensores de posicionamiento global (*Global Positioning System - GPS*), acelerómetros, giroscopios, sensores de luz ambiental, sensores de proximidad, sensores magnéticos, barómetros, termómetros, etc. Según H. Uchiyama y E. Marchand. La detección y localización del objetivo de interés es clave para el desarrollo de interfaces gráficas de RA, de esta depende el seguimiento en los cuadros de video y la correcta integración de los elementos virtuales en el entorno real (Uchiyama & Marchand, 2012). D. T. Nguyen logró que los sistemas de vigilancia conseguían detectar los objetivos (personas, vehículos, etc.) de forma automática, sin embargo, lo que se pretende en la actualidad es proporcionar información virtual de esas detecciones (Nguyen, Li, & Ogunbona, 2016). Por esto se presentan a continuación algunos trabajos en los que se produce RA para sistemas de vigilancia.

C. Sandor et al., emplea la detección de objetos relevantes en la escena para producir RA, específicamente superpone elementos gráficos virtuales en escenas reales para presentar zonas

interiores ocultas, carteles publicitarios, etc., el sistema es llamado sistema de rayos X para RA. La adquisición de la información es llevada a cabo a través de cámaras externas de alta resolución (Sandor et al., 2010). En la Figura 2.16, se puede observar en forma gráfica el proceso llevado a cabo y los resultados obtenidos.



Figura 2.16. Combinación de mapas de objetos relevantes para producir RA (Sandor et al., 2010).

M. Sauer et al utilizan interfaces de RA para el control remoto de procesos industriales. Específicamente, la información adquirida por las cámaras detecta el movimiento de los brazos robóticos y los diferentes grados de desplazamiento. Los datos de ubicación espacial proporcionan información virtual del funcionamiento y permite ajustar movimientos de los brazos en forma teledirigida (Sauer, Leutert, & Schilling, 2010), una muestra del sistema en ejecución y la presentación de la información a través de consolas fijas se puede observar en la Figura 2.17.



Figura 2.17. Teledetección, control y manejo teledirigido a través de cámaras de vigilancia e interfaces de RA (Sauer et al., 2010)

En 2010, J. Kuo et al. utiliza patrones de color para detectar coches en sistemas de vigilancia. La correspondencia de valores permite identificar un coche en específico una vez concluida esta tarea el sistema proporciona etiquetado virtual de la detección y datos sobre la localización, la fecha y hora del evento. (Kuo, Lai, Huang, & Liu, 2010).

En el campo de la vigilancia y control de desastres ambientales, se ha utilizado la RA para proporcionar alarmas por ejemplo L. Millan et al., desarrollan un sistema para cámaras IP, que detecta e identifica el humo y además genera alertas sobre la presencia del mismo en exteriores y descampados, en este caso se emplea análisis de color para fijar un umbral de detección así como la DCT para reducir coste computacionales e identificar estados normales y estados no normales (Millan-Garcia et al., 2012). En 2013, S. Yoon et al., utilizan modelos GMM para identificar humo en ambientes exteriores. La detección del evento se realiza a través del análisis de diferencias existentes entre cuadros de video considerando cambios de intensidad en los píxeles. Los autores desarrollan una interfaz gráfica proveyendo información visual de la detección y clasificación de humo, niebla y nubes (Yoon & Min, 2013).

Con otro enfoque R.K. Gwon et al., en 2013, presenta alarmas de información virtual sobre secuestros o lesiones de personas, para ello realiza detecciones de anomalías considerando un clasificador automático de muestras denominado “*INHA-VAT*” en combinación de un modelo probabilístico de eventos basado en redes Bayesianas, la correlación efectiva de datos de las muestras dan como resultado la interpretación de la anomalía detectada como un posible secuestro. La interfaz de RA consiste en detecciones, etiquetado gráfico y alarmas cuando ocurre el evento, como se puede observar en la Figura 2.18 (Gwon, Kim, Park, Kim, & Kim, 2013).



Figura 2.18. Detección e identificación de secuestros o actitudes hostiles presentado en el interfaz gráfico de RA tomado de (Gwon et al., 2013).

En 2014, D. Frejlichowski et al., desarrolla un monitor inteligente para detección de personas y presentación de alarmas virtuales de actividades sospechosas en tiempo real, para ello se utiliza HOG y *Haar-Like features* (Frejlichowski, Gościewska, Forczmański, & Hofman, 2014). Por otra parte L. Cheng genera información virtual sobre detección y seguimiento de vehículos sospechosos, estableciendo alarmas por robos, así como establece el itinerario seguido por el vehículo sospechoso a través de un sistema de vigilancia vehicular cooperativo (*Vehicle to Vehicle - V2V*) (Chen, Tseng, & Syue, 2014). También, A. Filonenko et al., presenta un sistema para evaluar la densidad de tráfico vehicular en tiempo real. El sistema provee información virtual sobre ubicación del evento a través de coordenadas GPS, un mapa virtual de detecciones y alarma por congestión vehicular (Filonenko, Vavilin, Kim, & Jo, 2014), una muestra de las detecciones realizadas se presentan en la Figura 2.19.



Figura 2.19. Interfaz de RA que muestra el mapa de detección de vehículos para estimación de densidad de tráfico desarrollado para sistemas de vigilancia (Filonenko et al., 2014).

También en 2014 se presentan, sistemas de vigilancia ambiental especialmente para la detección de humo. En el trabajo presentado por M. Bugarić et al., la detección del humo se realiza en base a modelamientos e información en 3D, presentando alarmas en tiempo real cuando la detección del evento se presenta (Bugarić, Jakovčević, & Stipaničev, 2014). La detección e información visual en color se puede observar en la Figura 2.20.

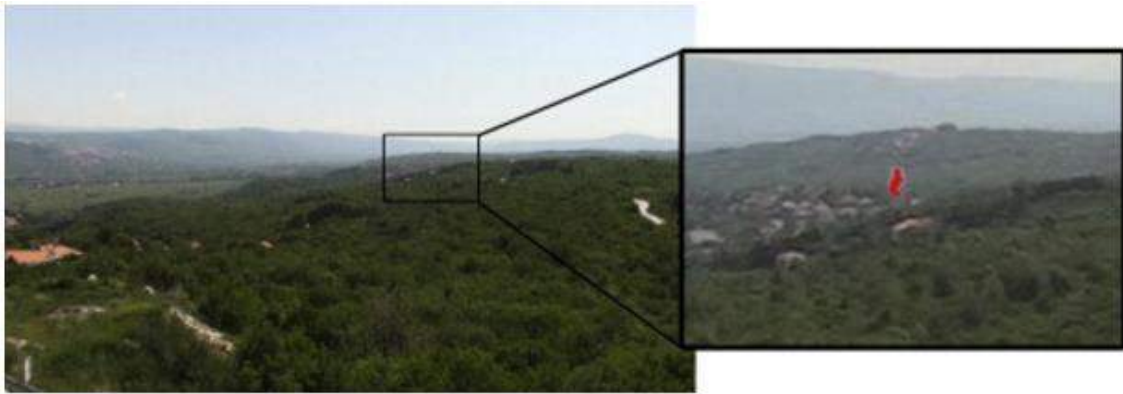


Figura 2.20. Información y alarma virtual de detección de humo para sistemas de vigilancia medioambiental (Bugarić et al., 2014).

Y. Nam, proporciona información virtual sobre la detección de objetos abandonados y alarmas virtuales por desaparición de un objeto en escenarios densos de gente. El interfaz gráfico de RA está desarrollado para sistemas de videovigilancia en tiempo real (Nam, 2015). Siguiendo la misma línea, J. Lan et al., presenta información virtual para sistemas de vigilancia sobre posibles congestionamientos vehiculares por la presencia de obstáculos en la carretera, colisiones, objetos abandonados y aparcamiento ilegal de vehículos. Para ello realiza la detección de obstáculos en el tráfico utilizando el GMM y “*Flushed Region of Interest*” (*FROI*), la clasificación automática la lleva a cabo mediante HOG y SVM. El interfaz de RA muestra el objeto detectado e identifica el tipo de obstáculo, la información proporcionada se puede observar en forma visual, como se presenta en la Figura 2.21 (Lan, Jiang, Fan, Yu, & Zhang, 2016).

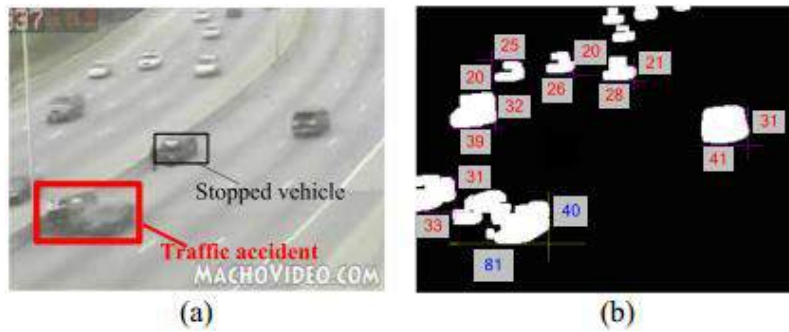


Figura 2.21. Información virtual sobre eventos inusuales en carretera en (a) muestra la detección de una colisión si como un vehículo estacionado (b) etiquetado de los elemento que conforman la escena (Lan et al., 2016).

M Hashemzadeh y N. Farajzadeh combinan la extracción de puntos clave (*keypoints*) y las características de textura, áreas o bordes de los objetivos específicos para proveer información virtual de la cantidad de personas en una escena, como lo muestra en la Figura 2.22 (Hashemzadeh & Farajzadeh, 2016).



Figura 2.22. Información virtual de la estimación del número de personas en una escena de videovigilancia (Hashemzadeh & Farajzadeh, 2016).

También en 2016, V. Kane et al., utiliza la información visual panorámica de los retrovisores adquirida a través de cámaras para proporcionar un enmarcado virtual gráfico sobre la detección de vehículos, estimación de aproximación a través de líneas virtuales, temporizador virtual de aproximación y la adquisición de cuadros de video para monitoreo de la escena (Kane, Smith, Burnett, Gabbard, & Large, 2016). La Figura 2. 23, muestra la diferente información que proporciona los espejos retrovisores.

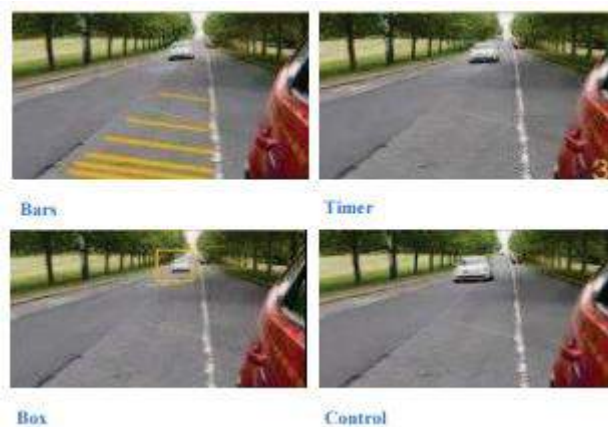


Figura 2. 23. Generación de RA considerando información obtenida por cámaras simulando retrovisores en un coche (Kane et al., 2016)

En la propuesta presentada por M. Jenkins y D. Young se emplea la RA para proporcionar asistencia en la conducción de motocicletas proporcionando información virtual sobre el planeamiento de rutas, velocidad además de proporcionar alarmas visuales, sensitivas (hápticas) y sonoras (Jenkins & Young, 2016).

B. Kaur et al., utiliza la información visual de las cámaras en interiores para establecer mapas virtuales, proporcionando así información de densidad de concurrencia y movilidad en pasillos y corredores (Kaur & Bhattacharya, 2016). Por otra parte M. Gori et al., utiliza el etiquetado semántico para interpretar una escena de videovigilancia. El término “interpretar” en este caso significa que el sistema es capaz de asociar la forma de un objeto determinado a un nombre en específico. Precisamente este autor, utiliza este etiquetado para reconocer los objetos en la escena asignando en forma virtual el nombre a cada elemento que lo conforma, como se puede observar en la Figura 2.24. Este tipo de sistemas es empleados especialmente en ambientes controlados (Gori, Lippi, Maggini, & Melacci, 2016) .



Figura 2.24. Información de RA en el reconocimiento de objetos a través de etiquetado semántico (Gori et al., 2016).

F. Kooji et al., presenta un sistema bautizado con el nombre de “CASSANDRA”, que permite detectar posibles agresiones. La información virtual permite identificar a los posibles agresores y a la víctima así como emitir alarmas del evento como se muestra el interfaz gráfico de presentación de la información y que se puede observar en la Figura 2. 25.

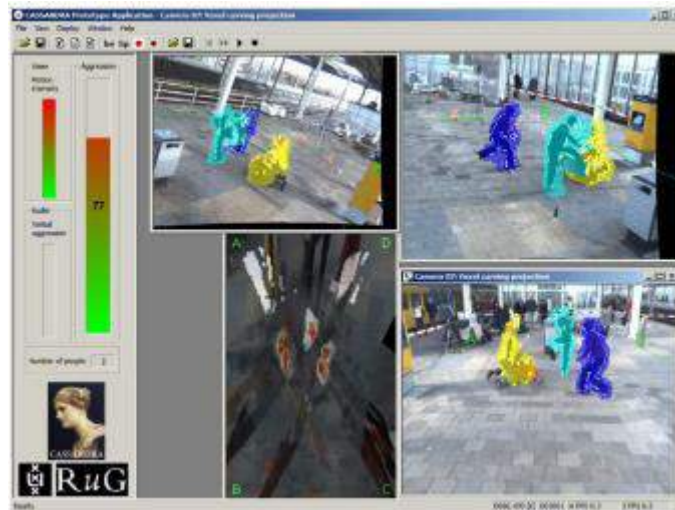


Figura 2. 25. Interfaz grfica del sistema CASSANDRA para detectar y alertar sobre agresiones.

La aplicaci3n de la RA tambi3n est presente en sistemas de vigilancia aut3nomos que operan tanto en aire, mar y tierra. Uno de los aspectos que ms preocupan en la generaci3n de RA en los sistemas no tripulados es la capacidad para alcanzar estabilidad, pues sin ella es difcil localizar el objetivo de inter3s. Esta “estabilidad” se refiere a la capacidad que pueden tener las cmaras para realizar compensaciones electr3nicas internas de tal manera que los cuadros de video permanezcan lo ms fijos posible en un escenario objeto de vigilancia m3vil, afortunadamente en la actualidad se ha conseguido garantizar esa estabilidad tanto para sistemas fijos como m3viles de vigilancia (Lee, Kim, Jeong, Park, & Paik, 2015). En la Figura 2.26, se puede observar las detecciones y remarcado de los objetivos realizadas a trav3s de la informaci3n de video obtenido por drones.



Figura 2.26. Presentaci3n de informaci3n virtual sobre detecciones realizadas luego de alcanzar estabilidad en los sensores (Lee et al., 2015).

En el aire, la informaci3n obtenida por sensores 3pticos alojados en UAV’s y los sistemas a3reos no tripulados (*Unmanned Aerial Systems-UAS*) es utilizada para producir informaci3n sint3tica sobre monitorizaci3n medioambiental (temperaturas, presiones, altitudes, etc.), alarmas sobre catstrofes (detecciones de incendios, inundaciones), flujo de trfico, control de la agricultura y la erosi3n, etc. Por ejemplo, el trabajo presentado por J. Martnez et al., muestra un sistema de alarma temprana de incendios forestales, la informaci3n de video es adquirida a trav3s de drones. En este trabajo se muestra un interfaz de RA que permite mostrar informaci3n de la detecci3n, patrones climticos existentes (presi3n, temperatura, altitud) as como la ubicaci3n del incidente.

Esta información es posible mostrar gracias al uso de varios sensores de monitoreo alojados en el dron y por la interacción con la información visual del objetivo identificado (Martínez-de Dios, Merino, Caballero, & Ollero, 2011). El sistema de detección de fuego forestal y la información que proporciona es presentado en la Figura 2.27.



Figura 2.27. Interfaz gráfica de RA en la detección de fuego forestal.

En otro ámbito aplicativo, las cámaras y otros sensores alojados en vehículos submarinos operados remotamente (*Remote Operated Vehicles - ROV's*), permiten desarrollar interfaces gráficas de RA en asistencia a la vigilancia y exploración submarina (Cardenas & Dutra, 2016).

En vehículos terrestres, la vigilancia está a cargo de robots quienes proporcionan información virtual del escenario, composición y tipos del suelo, configuración del terreno, etc., en tiempo real (Gao, Spiteri, Pham, & Al-Milli, 2014b). En la Figura 2.28, se muestra la detección de rocas, suelo así como el horizonte y su correspondiente etiquetado generado por la sonda espacial *Mars Pathfinder*.

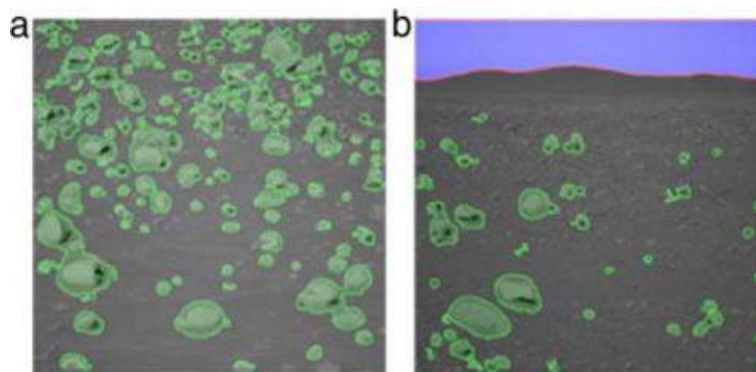


Figura 2.28. Sistema de detección y etiquetado virtual de rocas y horizonte visual marciano empleado en la sonda *Mars Pathfinder* (a) a corta distancia y (b) a larga distancia (Gao et al., 2014b).

2.5. Detección de ROI's a través del color

Las técnicas empleadas en el reconocimiento de regiones permiten la detección de píxeles que son de interés a través de la extracción de características de los mismos. Estas características de los píxeles están asociadas a la detección de bordes (Canny, Sobel, Prewit, Roberts, etc.), en donde se estiman diferencias y similitudes considerando; el incremento o disminución del valor de los píxeles, estimación de gradientes por cambios de intensidad, análisis de componentes de color, etc. (Gonzalez & Woods, 2008; Martín Carabias, 2012). Otras técnicas comparan regiones de píxeles para reconocer formas y manchas (blobs). También la ubicación espacial y la estimación de distancias entre píxeles ayuda a la detección de ROI's, de esta manera es posible la detección de esquinas, aristas, puntos de interés, ubicación de ejes, centros geométricos (De la Escalera & Armingol, 2001; Fukunaga, 2013; Pajares & De la Cruz, 2001). La clasificación de píxeles de acuerdo a diferentes texturas, es otra opción para la detección ROI's. Esta opción se presenta en la identificación de ROI's, especialmente en imágenes sintéticas y que presentan un patrón común así como características invariantes en la imagen, por ejemplo fondos con figuras regulares y proporcionales (Fukunaga, 2013; Maggio & Cavallaro, 2011). Las ROI's se clasifican de acuerdo a sus propiedades topológicas y métricas. Las propiedades topológicas de las ROI's, se definen precisamente considerando grupos de píxeles que contienen un patrón de similitud (color, textura, gradiente, etc.). Por otra parte si se consideran características como centros geométricos, áreas (cantidad de píxeles), distancias, etc. se hará referencia a las propiedades métricas (Pajares & De la Cruz, 2001).

El color es una de las más importantes fuentes de información empleados en la detección de objetos específicos dentro del entorno (Gonzalez & Woods, 2008; Halder, Callaerts, & Gehring, 1995; Qidwai & Chen, 2009). Es por ello que hay múltiples métodos que utilizan esta característica como fuente primaria de información. La umbralización (*thresholding*) es una de las principales técnicas de clasificación de píxeles y detección de ROI's a través del color. Esta detección de regiones permite localizar y separar objetos que disponen de características homogéneas en una escena o secuencia de escenas, esta separación también es conocida como segmentación. La umbralización es una técnica de uso común en la detección de objetivos de interés en materias como el procesamiento digital de imágenes, la visión artificial, la teledetección, el reconocimiento de patrones, etc. A partir de esta técnica se derivan una gran cantidad de aplicaciones como la detección de caras, coches, personas, especias vegetales, suelo, borde de ríos, líneas costeras, fuego, etc. Aunque ciertamente la detección bajo umbrales presenta problemas en aspectos inherentes a confusiones, oclusiones parciales, ruido gaussiano así como fallos por cambios de iluminación, lo cual genera bajas tasas de precisión. El conseguir un umbral ideal u óptimo que permita clasificar correctamente las ROI's, es una tarea de continua investigación y desarrollo (De la Escalera & Armingol, 2001; Gonzalez & Woods, 2008; Plataniotis & Venetsanopoulos, 2013; Qidwai & Chen, 2009).

2.6. Uso de la umbralización en la detección de objetivos específicos en tiempo real

La umbralización como técnica genérica de obtención de ROI's, puede ser utilizada como un método para segmentar y detectar objetivos de interés, separándolos del entorno. En las décadas de los sesentas y setentas, la umbralización estuvo orientada al uso en el procesamiento digital de imágenes en escalas de grises por lo cual métodos como Otsu, Kappur, entropías de Shannon, p-

tile, Kittler y los derivados de estos fueron diseñados para solventar estas necesidades de detección (Sahoo, Soltani, & Wong, 1988), en la actualidad encontrar un umbral óptimo se vuelve una tarea de mayor complejidad debido a que las imágenes que obtienen los sistemas de vigilancia están en color. Según B. Bhanu, la detección de un objetivo de interés se produce cuando las áreas o regiones que pertenecen a este objetivo han sido localizadas en la escena (Bhanu, 1986). Existen múltiples métodos que utilizan la umbralización para alcanzar detecciones en tiempo real, pero la principal dificultad es que no son capaces de mantener un compromiso entre el tiempo de procesamiento y la precisión en la detección. De acuerdo a (Sonka, Hlavac, & Boyle, 2014), para que una umbralización pueda ser empleada en sistemas de vigilancia, debe permitir ejecutar de forma eficiente tareas de detección en tiempo real.

La vigilancia en los sistemas en tiempo real está asociada a la adquisición de información visual de forma continua y ordenada a través de diversos sensores ópticos. Los fotogramas (o cuadros de video) son analizados de tal manera que se pueda obtener sus características. Cuando se obtienen las características de un objetivo de interés, y estas sirven para proporcionar alarmas sobre situaciones y/o comportamientos en tiempo real, esta vigilancia pasa a ser inteligente. Así mismo, la generación de alarmas puede ser causada por estadísticas, detección de movimiento y detección de características específicas del entorno (Hampapur et al., 2003b).

La umbralización permite la detección de un píxel de interés o la agrupación de ellos para obtener ROI's. Las ROI's pueden ser más de una, generando de esta manera este conjunto de regiones homogéneas que representa al objetivo de interés o específico. En el mundo físico, estos objetivos de interés pueden representar personas, vehículos, zonas deforestadas, fuego, agua, etc. Los cuales a través del umbral aplicado en la escena, son separados o identificados del resto de elementos que conforman la escena dada. Aplicado a los sistemas de vigilancia, estas detecciones en tiempo real nos permiten conocer información de un suceso, comportamientos anómalos en personas (no pago de billete en estaciones de trenes), detección de posibles colisiones de vehículos, deforestación, incendios, inundaciones, etc. En el presente trabajo de tesis doctoral, se enfoca en dar utilidad a las detecciones de objetivos específicos de tal forma que se pueda proporcionar información de las detecciones en tiempo real.

2.7. Extracción de objetivos relevantes en la escena a través de índices de color

2.7.1. Los índices de color - inicios y desarrollo

La aplicación de un índice de color en una imagen permite la extracción de objetos relevantes en la escena a partir de la agrupación de los píxeles que tienen una tonalidad de color determinada. Los índices de color tienen su base en el estudio colorimétrico para establecer que cantidad de luz que es reflejada, transmitida y absorbida por las hojas de los árboles, que pueden captar los sensores fotométricos.

Los primeros estudios fotométricos para determinar porcentajes de incidencia de la luz en especies vegetales fue llevado a cabo por Gaussman y Allen en 1973, precisamente estos autores consiguieron identificar 30 especies de plantas, realizando pruebas de reflectancia colorimétrica y conocer de esta manera el nivel de incidencia de luz sobre cada especie de planta en particular (Gausman & Allen, 1973). Siguiendo esta línea de investigación, Gausman, en los años 1975, 1976, 1978 y en 1981, realiza mejoras a su trabajo inicial logrando identificar un mayor número de especies de plantas y con más precisión. En 1981 Peynaldo et al., utiliza la misma metodología

propuesta por Gaussman en la detección de daños en hojas. Todos estos trabajos muestran falencias en la detección por la variabilidad de la iluminación y por requerir una multitud de umbrales locales, p.ej. tantos como especies de plantas estudiadas. No existe hasta ese momento un patrón global de detección (Hatfield, Gitelson, Schepers, & Walthall, 2008). A partir de los umbrales locales obtenidos, varios modelos de sensores y dispositivos fueron desarrollados para efectos de detección de especies agrícolas. Según la revisión efectuada por Meyer et al., los sistemas que se destacan son los implementados por Tucker, 1978, Thomas et al., Hagger, et al (1983-1984), Sorlie et al., en 1988, Tarbell y Reid en 1991 y Franz et al., Tian et al en 1999, (G. E. Meyer, 2011) .

En 1994, Woebbecke et al., realizan un estudio y mediciones sobre los diferentes sensores ópticos (radio espectrómetros y foto transistores ópticos) y térmicos (termómetros infrarrojos) utilizados en la clasificación de plantas y suelo, que han sido desarrollados en base a la toma de umbrales locales. Los resultados son concluyentes y muestran una imprecisión de la medición y por consecuencia una incorrecta clasificación. Siguiendo la metodología de establecer umbrales locales en base al nivel de incidencia de la luz, dependen de muchas variables como la posición del objetivo respecto a la luz, la distancia del sensor, el punto de vista, la cantidad de luz reflejada, etc. Toda esta variabilidad de parámetros genera errores y fallos de clasificación (Woebbecke, Al-Faraj, & Meyer, 1994). A partir de estas conclusiones se ve la necesidad de disponer de una metodología que permita clasificar las plantas del resto, evitando fallos por errores de adquisición de la información. De acuerdo con la evaluación de rendimiento realizada por Hummel y Stoller en 2002, estos sensores presentan problemas de rendimiento debido a la inexactitud de localización de hierbas, malezas etc. Estos problemas están asociados a los expuestos sobre la variabilidad de luz y la excesiva umbralización (Hummel & Stoller, 2002). Además, los materiales electrónicos con los que estaban construidos estos sensores generaban ruido gaussiano, provocando que el error se incremente. La opción de detección vía el procesamiento digital de imágenes que involucre otras variables que mitiguen el efecto de la iluminación, se convertía en una opción prometedora para alcanzar mejores resultados (G. E. Meyer, 2011).

Precisamente, Woebbecke et al., presentan en 1995 una nueva propuesta de detección de las hojas a través del procesamiento digital de imágenes, utilizando relaciones cromáticas entre componentes de color en base a porcentajes de reflectancia de los mismos en las hojas. De esta forma, se plantea un único umbral que se expresa como una fórmula matemática. Un aspecto que se diferencia de los trabajos anteriormente formulados con el mismo propósito es que, utiliza directamente imágenes en color sin pre-procesamiento. Además, el método muestra capacidades de identificación y detección independientemente el tipo de planta, su tamaño, ubicación o perspectiva, la precisión de la clasificación alcanzada fue del 60% al 80%. En las conclusiones, Woebbecke avizora la potencialidad de la metodología descubierta en multiplicidad de aplicaciones diferentes a la propuesta. Woebbecke formaliza su metodología y la presenta en el artículo “*Color indices for weed identification under various soil, residue, and lighting conditions*” (Woebbecke, Meyer, Von Bargen, & Mortensen, 1995), y presenta entre otros al exceso de verdes (*Green Excess - ExG*), un índice de color que es base del trabajo desarrollado.

G. Meyer, en 1998, realiza modificaciones y mejoras al *ExG*, con lo cual existe una mayor precisión en la clasificación, pero sobre todo presenta un promedio de tiempo de procesamiento entre 20 y 30 s, este tiempo es muy rápido considerando los recursos computacionales de la época y demuestra el uso de este tipo de índices en aplicaciones en tiempo real a mediano plazo (G.

Meyer, Mehta, Kocher, Mortensen, & Samal, 1998). Muchos trabajos posteriores (Giltelson et al., 2002; Lamm et al., 2002; and Mao et al., 2003) probaron la eficacia del *ExG* comprobando altas tasas de clasificación y detección (G. E. Meyer, 2011). Siguiendo el trabajo de Woebbecke, en 1998 G. Meyer et al., propone el índice de exceso de rojos (*Red Excess-ExR*) (G. E. Meyer, Hindman, & Laksmi, 1999a), este índice mejora la clasificación de la vegetación, ya que es capaz de resaltar el entorno a las hojas, como es el suelo hojarasca seca. El *ExR* incluye las sombras dentro de las regiones resaltadas, en inclusive a restos vegetales con tonalidades verdosas que no son parte de la planta, esto decanta al final en errores de clasificación. Este índice también ha servido de referencia en el planteamiento de nuevos índices que se presentan en esta tesis doctoral.

Pérez et al., en el año 2000, proponen una mejora al *ExG*, permitiendo que residuos de plantas y hojarasca sean clasificadas como parte del suelo, realzando los verdes de las hojas. De esta manera se presenta el “*Normalized Difference Index*” (*NDI*), un índice que surge de la diferencia entre las componentes rojo y verde en relación a la adición entre estos. Este índice muestra errores de clasificación por confusiones de tonalidades.

Posteriormente en 2008 G. Meyer y Camargo Neto aplican la umbralización basada en Otsu para binarizar las ROI's y así logran separar en forma automática la vegetación del suelo. Esta detección de ROI's lo realiza considerando dos opciones, una basada en la sustracción de los valores del exceso de rojos frente a los verdes (*Excess Green minus Excess Red - ExGR*) propuesto por C. Neto (Camargo Neto, 2004a) y otra a partir de *NDI* (G. E. Meyer & Neto, 2008a), los resultados de clasificación presentan mejores resultados para el *ExGR*. Por otra parte, Gee et al., realizan otras verificaciones de los índices combinados al umbral de Otsu y concluyen que este puede generar pérdidas de píxeles que conforman la vegetación (Gée, Bossu, Jones, & Truchetet, 2008), pues el histograma generado de la clasificación potencia la detección de las zonas que no corresponden a la vegetación.

En 2011, una nueva verificación de precisión de los índices es realizado por M. Guijarro et al., además el autor propone un nuevo índice de enfoque combinado (*Combined Approach - COM*), este índice es el resultado de la aplicación por separado y ponderado del *ExG*, *ExR*, *ExGR*, y el “*Colour Index of Vegetation Extraction*” (*CIVE*) propuesto por Katota et al. (Kataoka, Kaneko, Okamoto, & Hata, 2003) y del “*vegetative index*” (*VEG*) propuesto por Hague et al. (Hague, Tillett, & Wheeler, 2006). Los resultados de la aplicación del *COM* muestran resultados con un bajo porcentaje de error en la clasificación de la vegetación del resto (suelo, matorrales, maleza, mala hierba, residuos sólidos, etc.) (Guijarro et al., 2011a). A partir de este trabajo, más que desarrollar nuevos índices los esfuerzos se han orientado a plantear estrategias que permitan la mejora de la detección así como paralelamente alcanzar eficiencia computacional considerando la reducción de tiempos de procesamiento.

Precisamente en 2012, J. Guerrero utiliza mapas de color obtenidos a través de los índices y clasifica la vegetación del suelo a través de máquinas de soporte vectorial. Se muestra un 93.1% de precisión en la clasificación, lo cual no representa mejoras significativas en relación a anteriores trabajos (Guerrero, Pajares, Montalvo, Romeo, & Guijarro, 2012). Posteriormente el uso de los índices de color se busca combinar con modelamientos espaciales para alcanzar mejor precisión y disminución en los tiempos de procesamiento. Precisamente en 2012, M. Montalvo et al., emplea una variación del modelo de Otsu, la aplicación de los índices de color y un modelo de estimación espacial de surcos en el suelo llamado “*Our Linear Regression*” (*OLR*), para distinguir y clasificar suelo y vegetación. En conjunto esta técnica alcanza una precisión de

detección de 89,3 % con un tiempo de procesamiento de 0,297 s para imágenes de 1392x1044 píxeles, con imágenes de menor resolución la precisión de detección así como el tiempo de procesamiento disminuyen (Montalvo et al., 2012).

En 2013, J. Guerrero et al., busca mejorar la precisión de detección evitando confusiones por pérdida del verde de las plantas, rocas, suelo, restos vegetales, malas hierbas, etc. Justamente se plantea la obtención del mapa binario obtenido de la aplicación del índice *COM* combinado al método de alineamiento de los píxeles que son de interés a través del método de estimación y corrección planteado por Theil-Sen's (Guerrero et al., 2013). En 2015, M. Guijarro et al., combina la extracción del mapa binario en imágenes agrícolas a través de *ExG* combinada con la Transformada Wavelet Discreta (*Discrete Wavelet Transform - DWT*), los resultados alcanzan un 92% de precisión (Guijarro, Riomoros, Pajares, & Zitinski, 2015).

De lo presentado en anteriores párrafos se puede inferir que cada índice permite resaltar una cierta región de píxeles en la imagen y de esta forma obtener objetos relevantes de la escena. Esta adquisición permite identificar en la escena ROI's asociadas a objetivos específicos.

Los índices de color en especial han sido empleados en sistemas autónomos de agricultura, específicamente en la detección automática de vegetación e identificación de plantas, hojas, surcos, suelo, malezas, etc. También se emplean en la producción agrícola automatizada para establecer la calidad de los alimentos y el cumplimiento de estándares de producción (Abdullah, Guan, Lim, & Karim, 2004).

La principal línea de investigación en relación a los índices gira en torno a mejorarlos o en su defecto plantear otros nuevos, todo esto con la finalidad de alcanzar cada vez mayor precisión en la detección. Para ello se han modificado, complementado, combinado métodos y técnicas en busca de alcanzar la precisión ideal. Sin embargo en los últimos años, el interés recae en alcanzar un compromiso entre la precisión de detección frente al tiempo de procesado de tal manera que la ejecución se realice en tiempo real sin afectar al rendimiento general del sistema.

Como se ha presentado en los párrafos anteriores, el *ExR* ha contribuido en la detección del suelo pero con limitaciones considerables, precisamente estas deficiencias impiden que se aplique este índice en la detección de zonas no forestadas. Usualmente en esta tarea de delimitación de zonas no forestadas, se han empleado algoritmos basados en el crecimiento de regiones, contornos activos, *superpixels*, métodos que generan segmentaciones ajustadas al perfil irregular de las zonas deforestadas (Menaka, Kumar, & Bharathi, 2013) .

2.7.2. El uso de los índices de color en la detección de ROI's en tiempo real

Desde el trabajo desarrollado por Woebbecke, el tiempo de procesamiento empleado por los índices para la detección de las ROI's presentó buenas respuestas (Woebbecke et al., 1995). Gracias a la mejora progresiva de la capacidad y la miniaturización electrónica el tiempo de procesamiento de las tareas asignadas se han reducido significativamente, con lo cual el uso en aplicaciones derivadas del uso de los índices de color en tiempo real es posible. Precisamente en anteriores décadas, esta expectativa ha llevado a presentar algunos trabajos que buscan la aplicación de los índices de color en la clasificación de la vegetación del resto en tiempo real pero que no lo han conseguido o han obtenido respuestas cercanas al tiempo real. Son muestra de ello los trabajos presentados por Broddie et al., en 1994, Marchant, Marchant en 1996, Hague et al. en 1997, Gonzalez y Woods en 2003, Astrand y Baerveldt en 2005, Leemans y Destain en 2006,

Tellaeche et al., Gée et al., y Bakker et al., en 2008 (Gée et al., 2008). Otras propuestas han alcanzado mejores respuestas de procesamiento consiguiendo el tiempo real sin embargo dependen de procesos previos de pre procesamiento que no son computados dentro del rendimiento global del sistema o que utilizan arquitecturas computacionales muy complejas y por lo tanto difícilmente pueden de ser replicadas en sistemas convencionales (CUDA, GNU, multi-threads) (Ribeiro, Fernández-Quintanilla, Barroso, García-Alegre, & Stafford, 2005; Søggaard & Olsen, 2003).

Algunos sistemas han empleado los índices en combinación con la transformada de Hough en busca de mejorar tasas de detección (Leemans and Destain, 2006, Tellaeche et al., 2008b, Gée et al., 2008 and Bakker et al) pero su aplicación en tiempo real no es factible debido al alto coste computacional (Burgos-Artizzu, Ribeiro, Guijarro, & Pajares, 2011a).

Burgos et al., presentan en 2011 el “*Robust Crop Row Detection*” (*RCRD*) que es un sistema automático de segmentación y detección de vegetación en zonas agrícolas. Aquí se adquiere la señal en RGB, se procesa la imagen a través del *ExG* y se discrimina la vegetación del resto a través de la umbralización por Otsu. Este sistema consta de una cámara de video empotrada en un tractor que actúa como sensor óptico (sistema de vigilancia), el *RCRD* procesa los cuadros de video y los evalúa cada cierto número. El *RCRD* alcanza una tasa de precisión de detección en tiempo real del 85% para la vegetación y un 69% para el suelo (Burgos-Artizzu, Ribeiro, Guijarro, & Pajares, 2011b). Según C. Zhang y J. Kovacs, el uso de los índices de color presenta inconvenientes para ejecutar la extracción de objetivos relevantes en la escena en especial debido a la baja capacidad de los procesadores en dar respuesta oportuna al control de la iluminación, por lo que es importante conseguir que este factor no afecte en la detección en tiempo real (C. Zhang & Kovacs, 2012). Siguiendo el trabajo de Burgos, 2011, J. Bencochea et al., propone una mejora en la precisión de detección pero esto le genera tasas de rendimiento de 1,5 cuadros por segundo (Bencochea-Guevara, Burgos Artizzu, & Ribeiro Seijas, 2014).

Finalmente en 2016, J. Arroyo et al., evalúa diversos histogramas obtenidos por *ExG* aplicados a las imágenes. Estos histogramas corresponden a diferentes condiciones de detección, considerando la variación de valores de crominancia y luminancia. El autor analiza las similitudes de los histogramas a través de un entrenamiento considerando el algoritmo “*K-Nearest Neighbour*” (*K-NN*). De esta manera se obtiene un histograma de referencia, el cual sirve en la comparación con posteriores cuadros de video y para identificar las ROI's y los objetivos relevantes (Arroyo, Guijarro, & Pajares, 2016). Aunque no presenta valores concretos sobre la mejora de la detección, sí reporta eficiencia computacional que permitirá detecciones en tiempo real.

Otras propuestas innovadoras aprovechan la capacidad de procesamiento de los índices para ser empleados en sistemas de vigilancia como son los drones y los UAV's. Precisamente se busca aprovechar la versatilidad, flexibilidad y capacidad de adquisición de imágenes de alta resolución que proporcionan estos sistemas no tripulados, para alcanzar una mayor efectividad en las tareas de monitorización y detección. Se ha podido observar que la fuente de información primaria constituye imágenes en el espacio RGB, multi-espectrales, imágenes en el infrarrojo cercano (*Near Infrared-NIR*), mosaicos. El tratamiento de estas imágenes sirve como base para establecer la capacidad de los diferentes algoritmos de detección y la eficiencia de procesamiento para su aplicación en tiempo real.

Se ha podido identificar también que para lograr el compromiso de precisión en la detección frente al tiempo de procesamiento, las implementaciones en los sistemas UAV's han requerido de una vasta complejidad. Es decir poseer una serie de componentes como son el subsistema de comunicación, navegación, control, adquisición de señales, transferencia de datos, etc., un ejemplo de esto representan los sistemas Altus I y II, el *Pathfinder-Plus*, GA-ASI ALTAIR, GA-ASI PREDATOR B, desarrollados por la NASA así como el UAV *China Guangxi Forest Inventory* entre otros (Ambrosia, 2002; Salamí, Barrado, & Pastor, 2014; Tang & Shao, 2015b; Yuan, Zhang, & Liu, 2015). Debido a los costes que implica la implementación de la infraestructura tanto tecnológica como física de estos sistemas, la utilización para la obtención de información que permita la detección de ROI's puede ser prohibitiva y exclusiva. Según Watts et al., cuando un sistema de vigilancia a través de UAV's dispone de una excesiva complejidad e infraestructura, las probabilidades de accesibilidad se disminuyen por lo que el empleo de aviones y satélites puede resultar una mejor opción para la obtención de información y para realizar tareas de monitoreo y vigilancia (Watts, Ambrosia, & Hinkley, 2012).

En la actualidad, los drones, debido a la miniaturización de sus componentes, así como el alto rendimiento de sus sensores, surgen como una opción de bajo coste para obtener información de alta calidad. Estos pueden llegar a reemplazar a los UAV's de alta complejidad así como a los sistemas tradicionales como son aviones y satélites, que demandan de altos costes para su puesta a punto, operación y mantenimiento. Uno de los sectores que se ve beneficiado de la vigilancia a través de drones precisamente es la vigilancia medioambiental (Tang & Shao, 2015b; Watts et al., 2012). De hecho, algunos estudios de coste beneficio muestran que resulta más beneficioso el uso de plataformas convencionales de UAVs o drones en tareas de vigilancia en relación con otros sistemas tradicionales como aviones y satélites, empleados para el mismo fin (Christensen, 2015).

Uno de los primeros trabajos desarrollados para conseguir la extracción de la vegetación del resto de la escena a través de UAV's es el presentado por M. Quilter y V. Anderson, aquí se utiliza la normalización y operaciones algebraicas con las componentes de color para conseguir la clasificación y detección de las ROI's (Quilter & Anderson, 2001), los resultados de este trabajo no guardan un compromiso de rendimiento entre la precisión de detección y el tiempo de procesamiento. A partir de ahí se presentan múltiples propuestas para alcanzar este compromiso en la detección de objetivos específicos a través de los índices de color para sistemas UAV's. En anteriores décadas principalmente se ha utilizado índices como; el Índice de vegetación de diferencia normalizada (*Normalized Difference Vegetation Index – NDVI*), el Índice de Reflectancia Fotoquímica (*Photochemical Reflectance Index - PRI*), los resultados no han sido los mejores. Otros recientes trabajos muestran el uso de algunos índices de color como el *ExG*, *ExGR*, el *CIVE*, y la binarización a través de Otsu, en aplicaciones de vigilancia en sistemas UAV's (d'Oleire-Oltmanns, Marzloff, Peter, & Ries, 2012b; Roldán, Joossen, Sanz, del Cerro, & Barrientos, 2015; Salamí et al., 2014; Watts et al., 2012). En especial, el presentado por Torres et al., muestra un buen rendimiento en la detección a través de los índices de color considerando el análisis de imágenes captadas por UAV's (Torres-Sánchez, López-Granados, & Peña, 2015).

2.8. Conclusiones

Los sistemas de videovigilancia están en un proceso de transición de ser sistemas pasivos a sistemas más dinámicos e inteligentes. Estos sistemas en el futuro inmediato utilizarán, otras plataformas de seguimiento móvil como son los dispositivos inteligentes y drones, por lo que se requiere adaptar métodos y algoritmos para su funcionamiento en este tipo de dispositivos.

Se ha identificado que existen deficiencias a la hora de detectar objetos relevantes con formas irregulares, las falencias se incrementan cuando los objetivos específicos están presentes en fondos que no guardan texturas uniformes (fondos ambiguos). Métodos basados en modelos predictivos y probabilísticos así como los basados en el entrenamiento previo, no son los más aconsejables a la hora de detectar este tipo de formas y en escenarios de vigilancia ambiental como son los bosques. Por otra parte, algoritmos de segmentación que utilizan el color como fuente de información como SLIC o algoritmos derivados de métodos gaussianos, no disponen de una alta tasa de rendimiento (precisión de la detección frente al tiempo de procesamiento) como para ser aplicado en sistemas de vigilancia ambiental en tiempo real.

La RA es un complemento importante en la presentación de la información que puede ser empleada en aplicaciones de videovigilancia. Para producir RA en este tipo de sistemas, se requiere disponer de información exacta de la ROI. Una condición común, tanto para conseguir la detección de objetivos específicos en los sistemas de vigilancia como para producir RA, es alcanzar alta precisión en la detección con bajos tiempos de procesamiento, es decir una alta tasa de rendimiento.

El color es una fuente de información que permite la extracción de objetos relevantes de una escena, lo cual posibilita una posterior detección de objetivos específicos. La agrupación de píxeles en torno a una tonalidad de color depende de un umbral determinado, mientras mayor cantidad de tonalidades se agrupen dentro de un mismo umbral este será de tipo global y a la vez óptimo, lo que garantiza eficacia en la detección. La principal ventaja del uso del color en la detección es que la misma se ajusta tanto a formas regulares como a irregulares y en diversas clases de fondos tanto uniformes como ambiguos. Por otra parte, la principal desventaja que se presenta en la detección, es el alto número de falsos positivos que se presenta principalmente debido a los cambios de iluminación.

En el siguiente capítulo se empleará el color como fuente de información, considerando una estrategia metodológica que resalte ciertas tonalidades de color de una escena de tal forma que permita la extracción de objetos relevantes de una escena con formas irregulares y en fondos ambiguos, así como permita la posterior detección de objetivos específicos considerando altas tasas de rendimiento de tal forma que pueda ser utilizado en sistemas de vigilancia.

Capítulo III: METODOLOGÍA PARA
LA DETECCIÓN DE OBJETIVOS
ESPECÍFICOS MEDIANTE EL
TRATAMIENTO DEL COLOR

3.1. Introducción

El estudio del estado tecnológico ha dado lugar a la conclusión que hoy en día todavía no se disponen de algoritmos y métodos que detectan objetos con formas irregulares o poco predecibles encima de fondos ambiguos en tiempo real y con suficiente precisión. Los métodos presentados de segmentación basada en la información del color no producen resultados robustos, sobre todo debido a los cambios de iluminación que impiden una correcta agrupación de píxeles en ROI's, sin embargo la principal ventaja del uso del color en la detección es que la misma se ajusta tanto a formas regulares como a irregulares y en diversas clases de fondos tanto uniformes como ambiguos.

Uno de los objetivos principales de esta tesis es proponer un método nuevo que es capaz de cubrir las deficiencias mencionadas. Por lo tanto, en el presente capítulo se expone un replanteamiento del uso del color como forma de detectar un objetivo específico en la escena, partiendo de la premisa de que si se destaca el valor de los píxeles que conforman una región de interés, puede ser posible que esta sea correctamente identificada. Para ello se calcula índices de color específicamente creados para una determinada relación entre los colores del objetivo y del fondo.

De esta forma se ha logrado contribuir en la detección de dos objetivos de formas irregulares relevantes, que son las zonas deforestadas y los fuegos forestales. Para su detección se propone una novedosa estrategia mediante dos nuevos índices de color, que son el *NFDI* y el *FFDI*. Como se verá en las pruebas presentadas, el rendimiento que presentan estos índices permite detecciones en tiempo real.

A raíz de la creación de los índices mencionados, se plantea un esquema general para la formulación de nuevos índices que permitan solventar otros requerimientos de detección en similares condiciones a la de los índices ya desarrollados.

La propuesta presentada ha sido enfocada al empleo tanto en sistemas de vigilancia fijos como en sistemas móviles (drones). Además se ha buscado que la información que se proporciona pueda ser útil para generar RA.

En este capítulo se presenta todo el proceso de creación de índices de color, su uso en la segmentación y como se llega finalmente a la clasificación de las regiones buscadas. Para esto, se definen primero los términos más usados y se presentan las métricas y procedimientos de evaluación empleados para comprobar la eficacia de los métodos. Después se describe el proceso de detección de formas irregulares mediante índices de color en base a diferentes índices que fueron creados en el pasado y en los que se basa este trabajo. De este proceso se ha derivado un procedimiento genérico para crear cualquier índice deseado, lo cual se presenta en la sección 3.4.1, posteriormente se aplica este procedimiento para la generación de nuevos índices el *NFDI* y el *FFDI*.

3.2. Terminología empleada

En esta sección se definen tanto la terminología utilizada así como su significado dentro del contexto de estudio. Por otra parte se realiza una revisión de las bases teóricas sobre las cuales se fundamentan los métodos para la detección de objetivos específicos considerando el color como fuente de información.

3.2.1. El objetivo específico

El objetivo específico o “*target*”, representa una ROI en donde los píxeles que la conforman han sido clasificados siguiendo un patrón o una regla determinada. En el procesamiento digital de imágenes, una ROI está formada por un grupo de píxeles vecinos que posee similares propiedades o características. En los sistemas de visión artificial, una ROI representa un conjunto de píxeles que definen un objeto. Dentro del entorno de la vigilancia, una ROI describe un objetivo de seguimiento o supervisión como pueden ser personas, vehículos, matrículas, cultivos, bosques, etc. En la vigilancia, la detección de las ROI's es un elemento clave para la interpretación de situaciones de peligro como incendios, inundaciones, accidentes o eventos que se salen de la normalidad (anomalías).

Los objetivos específicos pueden tener formas regulares o irregulares. Las formas regulares se las consideran a aquellas que denotan una similitud con las formas geométricas o predecibles como las de coches, aviones, trenes o personas. En su defecto, las formas irregulares son las que no se pueden definir mediante un patrón reconocible, es decir, son representados por una región de puntos con forma aleatoria. Este tipo de figuras se presentan por ejemplo en llamaradas de fuego, zonas deforestadas, relieve forestal o de vegetación, ríos etc.

En la presente tesis se realiza la detección de objetivos específicos con formas irregulares, siendo esta una tarea complicada de resolver. El enfoque principal está en el reconocimiento de regiones irregulares para aplicaciones de vigilancia como detección de fuegos, inundaciones o deforestación.

3.2.2. Detección en tiempo real de objetivos específicos

El requerimiento principal de sistemas de vigilancia es su capacidad de avisar un acontecimiento en tiempo real, es decir en el mismo instante en que ocurre. En concreto, la detección en tiempo real se produce cuando la respuesta que produce el algoritmo interactúa con la información de entrada de tal manera que el observador pueda identificar la detección de un objetivo sin retardos notables. El tiempo real tiene relación directa con la eficiencia que un algoritmo puede llegar a proporcionar.

3.2.3. El tiempo de procesamiento (T_p)

El tiempo de procesamiento se define como el período de tiempo que tarda un algoritmo en ejecutarse desde que se adquiere la información de las imágenes hasta que se produce el resultado deseado. El T_p es un parámetro que permite establecer la eficiencia computacional, relacionándolo con la precisión que ha alcanzado un método, en este trabajo de segmentación o detección, para establecer el rendimiento de mismo.

3.2.4. Índice de color específico (*ICE*) y general (*ICG*)

El índice de color representa una imagen de tonos grises, obtenida mediante operaciones algebraicas entre las componentes de color de una imagen. La denominación “índice” indica o resalta las ROI's que se desea separar mediante una segmentación por umbral. La ventaja del cálculo de un índice de color es que puede ser segmentada mediante un umbral único para separar las tonalidades deseadas de las demás. Para cada tonalidad se ha de diseñar un índice específico propio (*ICE*), basándose en las relaciones entre tonos cromáticos. En ocasiones puede ser

necesario combinar varios índices para aumentar el efecto de contraste entre dos tonalidades determinadas. Esta combinación (en este trabajo se trata de la sustracción de dos índices), se ha denominado índice de color general (*ICG*).

El capítulo 3.4 trata exhaustivamente del proceso de creación y aplicación de diferentes índices.

3.3. Métricas de evaluación

La tarea de diseñar o elegir una medida apropiada para conocer la efectividad de un método de segmentación es una tarea compleja. La métrica de evaluación que se escoja debe estar acorde al enfoque para el que se ha desarrollado el método de segmentación. En este caso, valorar la exactitud y la eficiencia de la segmentación y consiguiente detección de zonas irregulares, exige una amplia y completa evaluación. En la presente tesis se ha empleado todos estos criterios de evaluación. Cuando se habla de exactitud se refiere a la precisión con la que una técnica realiza la segmentación de tal forma que los resultados son coincidentes con la segmentación verdadera, de referencia o patrón que comúnmente es denominada *Ground Truth (GT)*.

La eficiencia tiene relación tanto con la rapidez con que se ejecuta un algoritmo como con la tasa de error en la segmentación o detección. Lo ideal es alcanzar un equilibrio entre el tiempo de procesado T_p y la precisión. El concepto de eficiencia es de mucha importancia en los sistemas de videovigilancia pues de esto dependerá la aplicación de un algoritmo en tiempo real. Cuando se alcanza altas tasas de eficiencia los objetivos podrán ser identificados en una secuencia de imágenes.

La metodología de comparación de un método de segmentación frente al GT es un procedimiento de uso general para la validación del algoritmo. El GT puede ser obtenida a través de una segmentación manual ejecutada por expertos, que se refiere al proceso de clasificación píxel por píxel realizado por el ser humano que debe cumplir con la norma de aproximación al 100% de la ROI, es decir a la segmentación ideal (exacta) de la ROI.

El GT puede ser una segmentación sintética, es decir generada por un método automático que se acerque al 100% de exactitud, es decir que previamente se ha realizado una evaluación de coincidencia entre el método automático de referencia y el GT. Como consecuencia de esta evaluación se obtiene alta tasa de coincidencia. El resultado de esta evaluación precisamente sirve de referencia para futuras comparaciones.

Los procedimientos para la evaluación de los algoritmos de segmentación, por norma general son dos y se clasifican en métricas geométricas y métricas estadísticas, según se evalúen propiedades relativas al contorno de la ROI o la región segmentada como tal (no únicamente el contorno). La evaluación a través de estas métricas es de uso frecuente y común para establecer la exactitud de un método de segmentación o detección (Grand-Brochier et al., 2015; X. Wang, Tang, Masnou, & Chen, 2015).

3.3.1. Métricas geométricas

Las métricas geométricas analizan distancias, variaciones de información, similitudes o disimilitudes que se pueden presentar entre los bordes de una segmentación automática frente al GT.

Cuando una segmentación muestra un exceso de detalles o regiones en relación a la segmentación vía GT, se presenta una sobre-segmentación. La sobre-segmentación genera ruido produciéndose falsos positivos durante la detección, además los tiempos de procesamiento aumentan. Por otra parte, cuando se presenta una segmentación con falta de detalles debido a una sub-segmentación (infra-segmentación) existe pérdida de información y por lo tanto las detecciones serán imprecisas y los falsos negativos se incrementarán.

En el presente trabajo se ha evaluado tanto la sobre-segmentación como la sub-segmentación utilizando la comparación de las segmentaciones frente a sus respectivos GT, para ello se ha empleado el método denominado “promedio de precisión en la comparación de segmentaciones” propuesta por D. Martin (Martin, Fowlkes, Tal, & Malik, 2001), que mide tanto la sub-segmentación como la sobre-segmentación a través de las métricas *Global Consistency Error (GCE)* y *Local Consistency Error (LCE)*. Las ecuaciones que describen las diferentes relaciones que definen *GCE* y *LCE* se presentan a continuación:

$$GCE = \frac{1}{n} \min\{\sum_i E(S_e, S_a, p_i), \sum_i E(S_a, S_e, p_i)\}, \quad (1)$$

$$LCE = \frac{1}{n} \sum_i \min\{E(S_e, S_a, p_i), \{E(S_a, S_e, p_i)\}\}, \quad (2)$$

donde S_e es la segmentación de expertos, S_a es la segmentación automática a través de umbrales, p_i la relación con los píxeles de interés, n el número de píxeles y E es el coeficiente de refinamiento de error local. Con referencia a (1) y (2), M. Polak et al., establece la métrica *Object Level Consistency Error (OCE)* (Polak, Zhang, & Pi, 2009), como sigue:

$$OCE(S_e, S_a) = \min(E_{e,a}, E_{a,e}). \quad (3)$$

OCE define una precisión del error, considerando los errores $E_{e,a}$ y $E_{a,e}$ basados en la penalización tanto de la sobre-segmentación como en la sub-segmentación. Es considerada una mejor segmentación cuando los valores de *GCE*, *LCE* y *OCE* se aproximan al valor cero.

También se ha establecido como métrica de evaluación el factor de *Random Index (RI)* que mide la similitud entre una segmentación y el GT. Este método fue desarrollado por William Rand (Rand, 1971). Para el uso de esta métrica, se han considerado los criterios propuestos por Sathya et al. (Sathya & Manavalan, 2011), por lo que el *RI* representa una comparación de las segmentaciones S_a (segmentación automática) y S_e (segmentación de expertos) valorando la coincidencia y no coincidencia de puntos n_t , n_u , n_v , n_w , de ambas segmentaciones. Esta comparación es valorada en relación al total de puntos evaluados n , con lo cual n_t representa el número de puntos que son coincidentes tanto en S_a como en S_e , n_u representa los puntos que son diferentes tanto en S_a como en S_e , n_v representa el número de puntos que son coincidentes en S_a y diferentes en S_e , n_w representa los puntos que son diferentes en S_a y son iguales en S_e . Por lo tanto, $n_t + n_u$ representa las coincidencias entre S_a y S_e y $n_v + n_w$ representan las no coincidencias entre las dos segmentaciones, la relación existente se presenta como en (4). El valor ideal para esta métrica es 1 ya que eso representa que efectivamente las dos segmentaciones son iguales.

$$R(S_a, S_e) = \frac{n_t + n_u}{n_t + n_u + n_v + n_w} \quad (4)$$

Finalmente se ha utilizado la métrica de variación de información (*Variation of Information- VOI*) (Meilã, 2007), la cual se define como la comparación de información mutua en relación con la entropía del sistema de tal manera que se pueda establecer la variación de la homogeneidad de la información. Se busca cuantificar la cantidad de información aleatoria que no puede ser explicada por una de las segmentaciones. En otras palabras, *VOI* representa la variación de la información entre el grupo de píxeles de la S_a y el grupo de píxeles de la S_e , un menor valor de *VOI* representa un mejor resultado. La expresión de *VOI* se muestra como sigue:

$$VOI(S_a; S_e) = H(S_a) + H(S_e) - 2I(S_a, S_e), \quad (5)$$

donde $H(S_a)$ es la entropía de S_a y $I(S_a, S_e)$ es información coincidente de S_a y S_e . La relación comparativa de información entre las dos segmentaciones permite obtener mayor certeza de información válida, es decir alcanzar mayor homogeneidad reduciendo la entropía.

3.3.2. Métricas estadísticas

Las métricas estadísticas evalúan la coincidencia entre grupos de píxeles que pertenecen a un algoritmo de segmentación frente al GT. Precisamente otra forma de establecer la calidad de la segmentación es evaluando la correspondencia entre regiones del método automático frente al GT, para ello se ha utilizado las coordenadas de la matriz de confusión. Este tipo de evaluación también es conocida como metodología de comparación de patrones y análisis de precisión, que básicamente se refiere al hecho de que la probabilidad de ocurrir un suceso se confirme o no.

En el presente trabajo, la matriz de confusión tiene cuatro coordenadas: la coordenada de verdaderos positivos (*TP*) representan los píxeles contenidos en una región segmentada mediante el método automático, que coinciden con los de la región equivalente segmentada en el GT. Los falsos positivos (*FP*) pertenecen a una región de píxeles de la segmentación automática que ha sido clasificada como parte del área segmentada en el GT pero que en realidad corresponde al fondo en la imagen. Los falsos negativos (*FN*) representan un grupo de píxeles que son clasificados como fondo por la segmentación automática cuando en realidad no lo son pues forman parte de la ROI en el GT. Finalmente, se presentan los verdaderos negativos (*TN*) que representan a un grupo de píxeles que han sido clasificados como parte fondo por el método de segmentación automático y que coincide con los píxeles del fondo que han sido seleccionados para el GT.

La evaluación de la calidad de la segmentación a través de la comparación de patrones es homologable para establecer la precisión de la detección, pues en ambos casos se evalúa la coincidencia o no de las ROI's. De esta manera se obtienen igualmente las coordenadas de la matriz de confusión que son utilizadas para establecer cuan efectiva ha sido la detección a través de un método en específico. En la Figura 3.1, se presenta un esquema gráfico del procedimiento de evaluación comparativa.

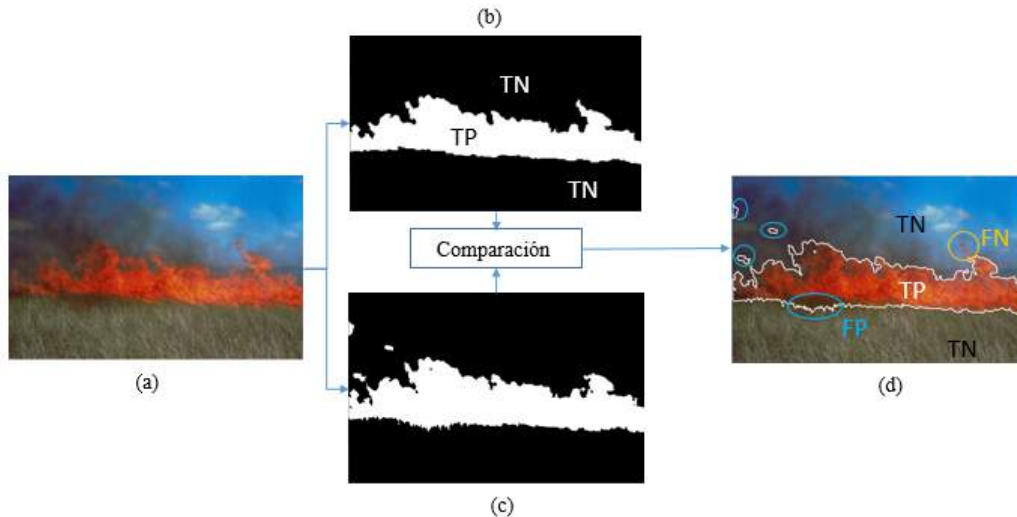


Figura 3.1. Evaluación comparativa entre la segmentación automática frente al GT. En (a) imagen original, (b) GT en binario, (c) binario de la segmentación automática, (d) clasificación de regiones por coordenadas de la matriz de confusión.

El procedimiento de evaluación basado en la comparación con el GT puede llegar a ser una tarea muy compleja y que demanda mucho tiempo. En ese sentido, para alcanzar una mayor eficacia en las evaluaciones, en la presente tesis doctoral se ha implementado un sistema de evaluación automática de la segmentación y la detección, utilizando computación paralela (Cruz, Eckert, Meneses, & Martinez, 2017). A través de este sistema se ha alcanzado evaluaciones mucho más precisas y sobre todo en menor tiempo, de esta manera se garantiza que las evaluaciones sean lo más fiables posible.

3.3.3. Métricas estadísticas de precisión, sensibilidad y rendimiento para evaluar la detección

Las coordenadas de la matriz de confusión son consideradas como el número de predicciones de detección en relación a lo que sucede en realidad, de tal manera que es posible contabilizar aciertos y errores. En la detección, la matriz de confusión permite establecer el rendimiento del clasificador considerando detecciones ciertas o no. Las métricas que evalúan precisión, sensibilidad y especificidad, precisamente utilizan los valores de estas coordenadas para validar el rendimiento de un sistema de detección.

La exactitud y la sensibilidad

En particular, la precisión o exactitud (p) y la sensibilidad o *recall* (r) son dos variables estadísticas fundamentales que permiten establecer cuan efectivo ha sido un algoritmo de segmentación o detección. La precisión mide proximidad entre el resultado de la detección que ha arrojado un algoritmo frente al patrón de comparación o el resultado esperado de detección, en términos generales es el porcentaje de aciertos. La relación que describe este parámetro se expresa como sigue:

$$p = \frac{TP}{TP+FP} \quad . \quad (6)$$

Por otra parte, la sensibilidad permite establecer el nivel de rendimiento del algoritmo. Para ello se evalúa el número de detecciones efectivas frente al número total de detecciones que debieron realizarse. La sensibilidad se calcula a través de la siguiente expresión:

$$r = \frac{TP}{TP+FN} \quad . \quad (7)$$

Otra métrica importante que se ha considerado es el *Jaccard Index (JI)* debido a que es una medida para establecer el grado de similitud entre el algoritmo propuesto y el GT. Este índice permite conocer cuál es la relación de las detecciones acertadas frente a los fallos. El *JI* se representa como sigue:

$$JI = \frac{TP}{TP+FP+FN} \quad . \quad (8)$$

El coeficiente Dice como medida de precisión global

El coeficiente *Dice* o *Dice Index (DI)* es otra métrica que mide similitud entre un algoritmo de segmentación o detección frente al GT. El *DI* establece en forma general cual ha sido el rendimiento del algoritmo al considerar tanto la precisión y la sensibilidad en la evaluación. El *DI* también es conocido como coeficiente refinamiento de la precisión (*F-measure* o *F-score*) y su expresión se muestra como sigue:

$$Fm = 2 * \frac{p \cdot r}{p+r} \quad . \quad (9)$$

El *Fm* establece la precisión global de un sistema en este caso muestra el grado de precisión de la segmentación o detección, cuando el valor de *Fm* se acerca a 1 (100%) existe un mayor grado de refinamiento de un método o algoritmo.

Rendimiento del sistema

El rendimiento *R* es la relación que presenta la precisión *p* frente al T_p . Esta medida es fundamental para saber si un algoritmo es funcional, mientras más alto sea el valor de *R* el rendimiento será mejor. La relación matemática se presenta como:

$$R = \frac{p}{T_p} \quad . \quad (10)$$

El parámetro *p* puede ser reemplazado por el *Fm*, si lo que se desea es comprobar la eficacia en relación al rendimiento global del sistema. El rendimiento del sistema ideal puede ser análogo a la tasa de cuadros por segundo para reproducción en tiempo real, pues si el valor de *Fm* alcanza el valor máximo (la unidad) y T_p es igual a 0,04 s, entonces se dispondrá de un valor adimensional de 25, que representaría la tasa de reproducción de las imágenes en un segundo.

El Índice Manhattan o indicador de exactitud

Otro indicador de similitud es el *Manhattan Index (MI)*, también conocido como el indicador de exactitud. Este índice analiza el grado de coincidencia entre píxeles, evaluando la mínima distancia entre segmentación y GT, su formulación se presenta como sigue:

$$MI = \frac{TP+TN}{TP+TN+FP+FN} \quad . \quad (11)$$

El coeficiente o índice MI da como máximo valor 1 (100%) cuando la segmentación está completamente contenida dentro del GT o cuando en una detección las regiones etiquetadas coinciden con el GT. El siguiente apartado refleja el desarrollo de los índices de color como base teórica para las contribuciones que se presentan en el presente trabajo.

3.4. Metodología de detección de objetivos con formas irregulares por índices de color

En esta sección se plantea el desarrollo de una metodología de detección de objetivos con formas irregulares que consta de tres etapas y en donde el uso de los índices de color cumple un papel preponderante.

Precisamente lo que se busca es proponer una estrategia que sea efectiva en términos de precisión de detección y eficiente buscando un tiempo de respuesta inmediato con una reducción en el coste computacional. Bajo estas premisas el planteamiento de la estrategia de detección utilizando nuevos índices de color, deben cumplir con los siguientes requerimientos:

- El método a proponer debe permitir detecciones de objetivos en primeros planos con formas irregulares o poco predecibles y en fondos ambiguos, es decir que no siguen un patrón de forma o textura específico.
- El método a desarrollar debe ser aplicable en el espacio RGB por ser el modelo base en colorimetría y que permite que el método pueda ser adaptado a otros modelos tricromáticas.
- El método desarrollado pueda ser empleado en sistemas de videovigilancia en especial del tipo medioambiental que presenta una complejidad de escenario (primeros planos irregulares y fondos ambiguos).
- El algoritmo computacional producto del método a proponer, debe alcanzar un compromiso entre la precisión de la detección y el tiempo de procesamiento de tal manera que permita su uso en tiempo real.
- La respuesta de detección debe permitir generar realidad aumentada.

El cumplimiento de estos requisitos, está supeditado a que se plantea una estrategia que incluye las siguientes etapas:

Etapas 1: Extracción de uno o varios ICE basados en las características de color y las relaciones entre las componentes cromáticas de los píxeles contenidos en el área de interés (regiones u objetos a detectar). Cálculo del ICG como sustracción de los ICE 's. Extracción del histograma del ICG .

Etapas 2: Cálculo del umbral T_{ICG} y aplicación sobre el ICE para crear la segmentación binaria. Creación del mapa de objetos mediante etiquetado.

Etapas 3: Clasificación de las regiones obtenidas y detección de las regiones y objetos relevantes. Extracción y cálculo de información de interés (p.ej. superficie) para la generación de contenido representable como Realidad Aumentada.

La descripción gráfica de esta estrategia con todos los procesos involucrados se presenta en la Figura 3.2. Estos procesos se exponen detalladamente en los siguientes apartados.

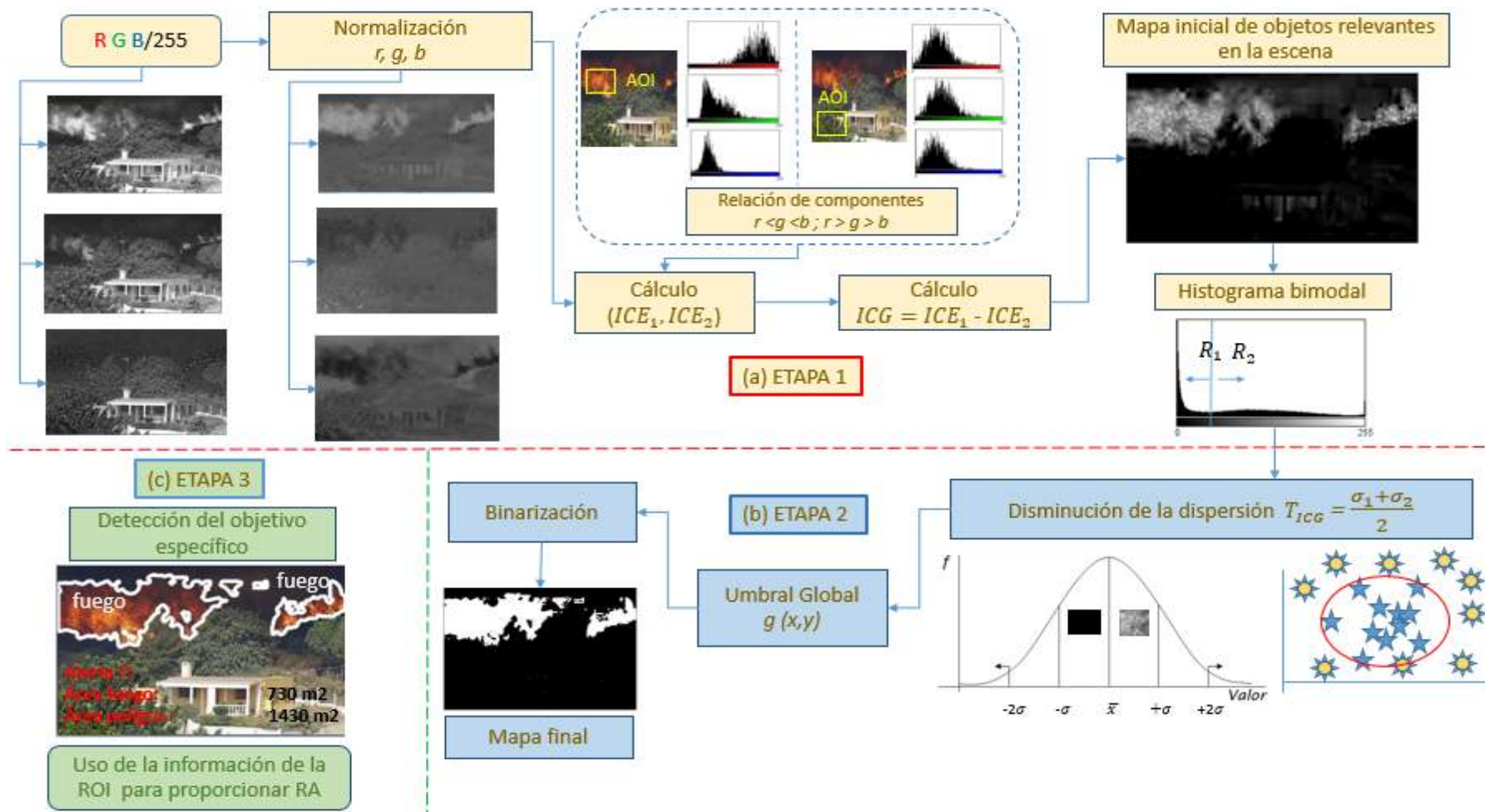


Figura 3.2. Detección de un objetivo específico con formas irregulares basado en ICE: (a) Etapa 1. Extracción del ICE e ICG en la imagen. (b) Etapa 2. Refinamiento de la detección. (c) Etapa 3. Uso de la información de la ROI para generar Realidad Aumentada.

3.4.1. Etapa 1: Creación de Índices de Color Específicos

El proceso principal en la cadena de detección de ROI's en imágenes está ubicado en la fase de la segmentación. La segmentación por definición es un proceso mediante el cual se separan regiones de píxeles que tienen características comunes del resto de píxeles. En el caso ideal, el grupo de píxeles separados corresponde a las regiones de interés.

Los principales métodos que segmentan la imagen en regiones están basados en métodos como la umbralización, el crecimiento de regiones y en la división y fusión de regiones. En el presente trabajo se utilizan los valores de tonalidades de color de los píxeles y la división en regiones se realiza a través de la umbralización por índices de color. Un índice de color es una imagen de tonos grises, creada procesando la información de color contenida en la imagen original. En el resultado, los tonos grises con valores altos corresponden a los colores de interés, mientras los tonos de niveles bajos corresponden al resto de colores. La segmentación de esta imagen de índice de color, realizada a continuación mediante un umbral óptimo, permite separar los objetos relevantes en la escena. El objetivo es generar un índice de color con mucho contraste entre las tonalidades de color buscadas y el resto para obtener los mejores resultados en la fase de segmentación.

Los índices de color tienen su base en el estudio colorimétrico para establecer la cantidad de luz que es reflejada o transmitida. Los principales índices son el Exceso de Verde (ExG), el exceso de rojo (ExR), el exceso de azul (ExB), el exceso de verdes menos rojos ($ExGR$), el índice normalizado de la diferencia entre el verde y el rojo (NDI), el índice de color en la vegetación ($CIVE$), el índice de vegetación (VEG), etc. Un estudio más detallado sobre la evolución de los índices de color se ha efectuado en la revisión del estado tecnológico actual (sección 2.7.1). En especial dos índices de color específico como son el ExG y ExR han sido la base para el desarrollo y la propuesta de nuevos índices que se presenta en esta Tesis Doctoral, por lo cual su estudio tiene una especial importancia.

En general, los índices de color son funcionales en escenas donde se presentan contrastes entre tonalidades de color, en donde unas tonalidades representan el primer plano y otras el fondo, fuera de este escenario se muestran limitaciones en la detección. Un índice será mejor tanto en cuanto abarque la mayor cantidad de tonalidades a ser detectadas y que estas pertenezcan al objetivo específico.

El Exceso de Verdes (ExG)

Este índice representa un ICE, el ExG tiene una connotación especial en el presente trabajo pues es la base teórica de los nuevos índices que a posteriori se presentarán, este índice fue presentado por Woebbecke et al., con la finalidad de identificar especies de plantas y hojas (Woebbecke et al., 1995). Para realizar el cálculo del ExG , Woebbecke realizó los siguientes pasos:

1. Análisis de áreas de interés (*Area of Interest - AOI*), Woebbecke define áreas de 100 píxeles.
2. Obtención de valores promedios de R, G, B, (R_c, G_c, B_c) de las AOI.
3. Obtención de coordenadas cromáticas r, g y b a partir de R_c, G_c, B_c . Combinando los valores de R_c, G_c, B_c , las ecuaciones normalizadas serán las siguientes:

$$r = \frac{R_c}{R_c + G_c + B_c} \quad , \quad (12)$$

$$g = \frac{G_c}{R_c + G_c + B_c} , \quad (13)$$

$$b = \frac{B_c}{R_c + G_c + B_c} , \quad (14)$$

Otra opción que presenta Woebbecke para obtener r , g y b es la relación entre R_c, G_c, B_c frente a los máximos valores que se presentan en cada AOI, de esta manera las ecuaciones normalizadas pueden ser obtenidas como:

$$r = \frac{R_c}{R_m} ; g = \frac{G_c}{G_m} ; b = \frac{B_c}{B_m} . \quad (15)$$

Donde R_m, G_m, B_m , representan el máximo valor de los píxeles en cada uno de los espacios de color en la AOI, los valores de los máximos pueden llegar a ser 255.

Posteriormente, Woebbecke establece la relación cromática entre las hojas y el suelo a través de pruebas de reflectancia cuyo resultado muestra un diagrama de cromaticidad donde se comprueba que la focalización de la energía espectral de g se encuentra en la longitud de onda entre 550 nm y 560 nm y para r entre 620 y 640 nm.

La suma de las coordenadas normalizadas de r, g, b es la unidad, es decir representa al punto de concentración de la energía que corresponde al blanco:

$$r + g + b = 1 . \quad (16)$$

A partir de esta premisa, Woebbecke realiza operaciones de adición y sustracción para definir o resaltar ciertas tonalidades específicas en una imagen. El objetivo del *ExG* en este caso es encontrar la mejor opción de contraste entre los valores colorimétricos de las plantas, en este sentido Woebbecke plantea la sustracción de g de r ($r-g$), lo que se interpreta como el fondo de la escena ya que el suelo y otros elementos disponen de mayor cantidad de rojos, al quitar los verdes se intensifica el valor de r . Mientras que para resaltar el verde de las plantas se sustrae el azul b del verde g ($g-b$), lo que representaría el primer plano. De esta manera se plantea la sustracción del fondo del primer plano a través de:

$$ExG = g - b - r + g = 2g - r - b . \quad (17)$$

La adquisición del *ExG* presenta una opción importante para la extracción de objetivos relevantes o predominantes con formas irregulares en una escena a partir del color y con bajas tasas de procesamiento.

Si se considera un fondo ambiguo como un contexto desordenado o caótico que no guarda un patrón simétrico como puede ser un mosaico, el conseguir distinguir un objeto en este tipo de escenarios es una tarea que involucra mucha complejidad. Precisamente Woebbecke demostró que el *ExG* es capaz de detectar eficientemente la vegetación en escenarios de fondo que no guardan una uniformidad o patrón determinado.

El Exceso de Rojos (ExR)

Otro índice que es de interés para este trabajo, es el exceso de rojos (*Red Excess "ExR"*), este es otro *ICE* que fue presentado por G Meyer et al. (G. E. Meyer, Hindman, & Laksmi, 1999b), este índice es el resultado de potenciar el valor de los rojos frente a verdes y azules. Meyer establece

que obteniendo la información del fondo se puede clasificar de mejor forma la información del primer plano, para ello Meyer utiliza la distribución humana de los conos de la retina para la percepción tricromática propuesta por Murch (Murch, 1984) con un 64% para rojo, 32% para verde y un 4% para azul. El porcentaje del rojo se potencia en el doble de tal manera que se obtiene $2(0.64)r = 1.28r \sim 1.3r$, con ello se extrae todos los valores de rojos en la escena, filtrando además los valores de g y anulando el valor b , esta relación se presenta como sigue:

$$ExR = 1.3r - g \quad . \quad (18)$$

Luego de pruebas de validación, Camargo Neto ajusta el valor de rojo a $1.4r$, modificando la ecuación planteada en (18) como $ExR = 1.4r - g$ (Camargo Neto, 2004b).

El ExR permite resaltar suelo y restos de vegetación seca, índice arroja una baja tasa de precisión sobre todo clasificar erróneamente como suelo a vegetación con tonalidades más opacas de verde, sombras, etc. (Guijarro et al., 2011a; G. E. Meyer & Neto, 2008a; G. E. Meyer, 2011).

El Exceso de verde en relación al rojo (ExGR)

El $ExGR$ representa un ICG que fue propuesto por C. Neto et al., con el objetivo de mejorar la detección de vegetación (plantas y hojas), este índice relaciona el ExR con el ExG para obtener la diferencia del exceso de verde con el rojo $ExGR$, con ello se logra resaltar el verde de la planta y se mejora el contraste en relación al suelo, residuos sólidos y vegetación muerta (Camargo Neto, 2004b). Sin embargo la precisión de detección de este índice está condicionado a escenarios más bien focalizados es decir donde las plantas están rodeadas únicamente por suelo o residuos sin tonalidades verdes. Cuando se presentan un escenario más diverso, extenso y más real en donde existen malas hierbas, hojarasca verde, pequeños matorrales con tonalidades verduzcas e inclusive sombras, el $ExGR$ clasifica erróneamente como parte de la planta o cultivo a estas malezas y/o sombras, esto es debido a la baja precisión que presenta el ExR . El $ExGR$ se formula como sigue:

$$ExGR = ExG - ExR \quad . \quad (19)$$

La aplicación de los índices de color en las imágenes permite que se resalten ciertas regiones relevantes del primer plano en relación al fondo de la escena, este proceso es parte del pre-procesamiento de la imagen. La imagen resultante está representada en diversos niveles de gris, en donde el objeto de interés tiende al valor máximo (255) y el resto al valor mínimo (0).

Extracción genérica de un índice de color en la imagen

A raíz del estudio de los índices de color mencionados hasta aquí, se ejecuta un conjunto de procedimientos que permite la creación de cualquier índice de color, basándose en las características de color de la región buscada.

El proceso debe iniciarse con la división de la imagen en sus tres componentes R , G y B , y su normalización para disminuir la influencia de diferentes iluminaciones. Esto permite estabilizar los valores de la reflectancia de tal manera que se puedan distinguir mejor los valores de las tonalidades al no existir variaciones. La Figura 3.3 muestra un ejemplo en el cual se extrae la componente R de la imagen. Si se comparan los histogramas de la componente original y su

versión normalizada, se puede observar que sólo el histograma de r tiende a ser bimodal², lo que permitiría la aplicación de un único umbral para separar regiones.

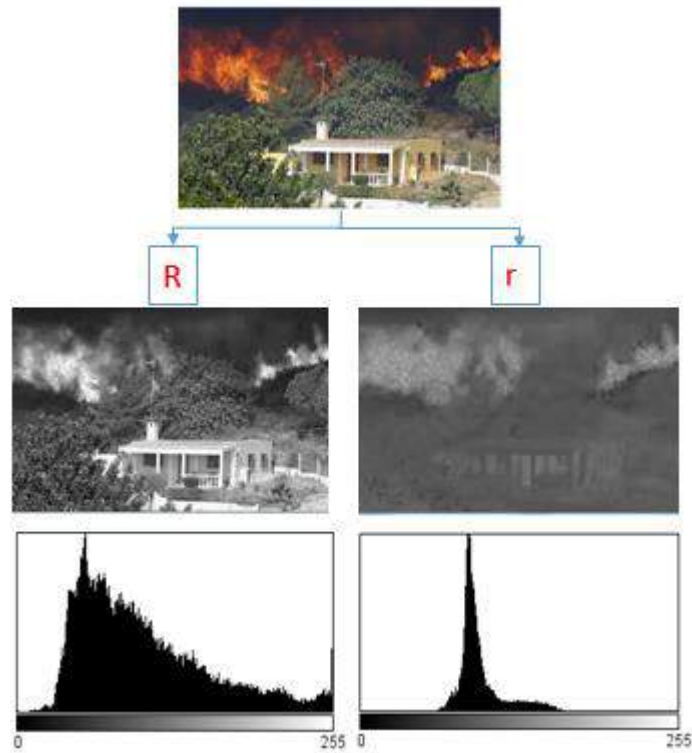


Figura 3.3. Extracción de la componente R y r de una imagen: En (a) imagen de R y su respectivo histograma. En (b) imagen de r con su histograma.

Para ejecutar la normalización, se sigue el criterio empleado por Woebbecke, en el espacio RGB la normalización se ajusta a valores r , g y b , en el rango de valor de $[0,1]$ de nivel de gris. Esta normalización surge de la relación del valor que dispone cada componente de color en RGB de una imagen en relación al máximo valor que puede tener cada componente, es decir $2^8=256$ para imágenes de 24 bits. De esta forma, los componentes R , G y B se normalizan en base a las fórmulas (12) – (15).

El procedimiento de extracción de la información del color se basa en adquirir por separado muestras de la región de interés a detectar así como de su entorno. Estas muestras pueden ser bloques de píxeles (cuadrados o rectangulares) que conforman las AOI. Para efectos de simplificar la deducción de fórmulas, los bloques de píxeles se han renombrado como n , como se muestra en la Figura 3.4.

Para establecer la relación cromática se evalúa una serie de imágenes (I_1 a I_N) como se muestran en la Figura 3.4. Tomando como muestra la primera imagen (I_1), se extraen un número determinado de bloques de la región de interés (bloques n_1 hasta el n_n). De cada bloque se calculan los histogramas para cada componente de color (RGB), de esta manera se obtienen los valores medios de R , G , B o sea \bar{R} , \bar{G} , \bar{B} . Los valores medios así como los valores de la varianza

² Un histograma bimodal describe una curva en donde se define exactamente dos picos y un valle, lo cual representa que en la imagen existe dos contrastes claramente definidos lo que facilita extraer las ROI's con mayor precisión.

(σ^2) y la desviación estándar (σ) se calculan a partir de los valores máximos (V_{max}) y mínimos (V_{min}) de la curva.

Si se realiza la media de todos los valores de $\bar{R}, \bar{G}, \bar{B}$ desde el área de interés n_1 hasta n_N en la imagen I_1 , la representación de estas valoraciones puede ser traducida como:

$$\bar{R}_{I_1} = \frac{1}{n} \sum_{k=1}^n \bar{R}_k, \quad \bar{G}_{I_1} = \frac{1}{n} \sum_{k=1}^n \bar{G}_k, \quad \bar{B}_{I_1} = \frac{1}{n} \sum_{k=1}^n \bar{B}_k, \quad (20)$$

donde $\bar{R}_k, \bar{G}_k, \bar{B}_k$ representan las medias de la k-ésima región evaluada, en este caso de la imagen I_1 . Esta valoración se repite en un set de imágenes de tal manera que un mayor conjunto de datos avale una generalización de la relación encontrada, en el presente caso esta valoración se ha realizado considerando el set de 80 imágenes en conjunto, que han servido para implementar tanto el *NFDI* como el *FFDI*.

En base a la relación (21) se formula los valores medios totales de cada componente de color ($\bar{R}_T, \bar{G}_T, \bar{B}_T$) para un set de N imágenes (I_1 a I_N), para el cual se considera la media de los valores medios parciales de cada imagen de tal forma que $\bar{R}_K, \bar{G}_K, \bar{B}_K$ representan la K-ésima media aritmética de todas las imágenes (de I_1 a I_N).

$$\bar{R}_T = \frac{1}{N} \sum_{K=1}^N \bar{R}_K, \quad \bar{G}_T = \frac{1}{N} \sum_{K=1}^N \bar{G}_K, \quad \bar{B}_T = \frac{1}{N} \sum_{K=1}^N \bar{B}_K \quad . \quad (21)$$

Un procedimiento similar se ha efectuado para obtener los valores de las medias de las componentes normalizadas ($\bar{r}, \bar{g}, \bar{b}$) así como el promedio de las desviaciones estándar ($\sigma_R, \sigma_G, \sigma_B$) tanto de los canales de color como de sus respectivos normalizadas ($\sigma_r, \sigma_g, \sigma_b$). El objetivo de todas estas medidas es establecer si las relaciones que se presentan son consistentes y genéricas, es decir se aplican en todos los casos.

De esta forma, a partir de los valores encontrados, se establecen las relaciones para establecer la predominancia de un componente sobre otro en una determinada ROI, por ejemplo si consideramos $\bar{R}_T, \bar{G}_T, \bar{B}_T$ las relaciones serán las siguientes:

$$\frac{\bar{R}_T}{\bar{G}_T}, \frac{\bar{R}_T}{\bar{B}_T}, \frac{\bar{G}_T}{\bar{B}_T} \quad (22)$$

De idéntica forma se ha procedido con valores de desviación estándar σ de las componentes RGB así como de sus normalizadas r, g, b .

En la Tabla 3. 1 se presenta un ejemplo de análisis cromático utilizando histogramas sobre diferentes regiones en donde se incluyen valoraciones de las coordenadas *RGB, rgb* y luminancia (Y), de tal manera que se obtienen sus valores máximos (V_{max}), la media aritmética (\bar{x}) y la desviación estándar (σ). Para comprobar que los valores de las componentes normalizadas obtenidos son verdaderos se utiliza la relación mostrada en (16), verificando de esta forma que la suma de ellos es igual a la unidad. En la Tabla 3. 2 se muestra la relación entre las componentes cromáticas y sus correspondientes normalizadas, lo que permite establecer la relación de predominancia entre ellos.

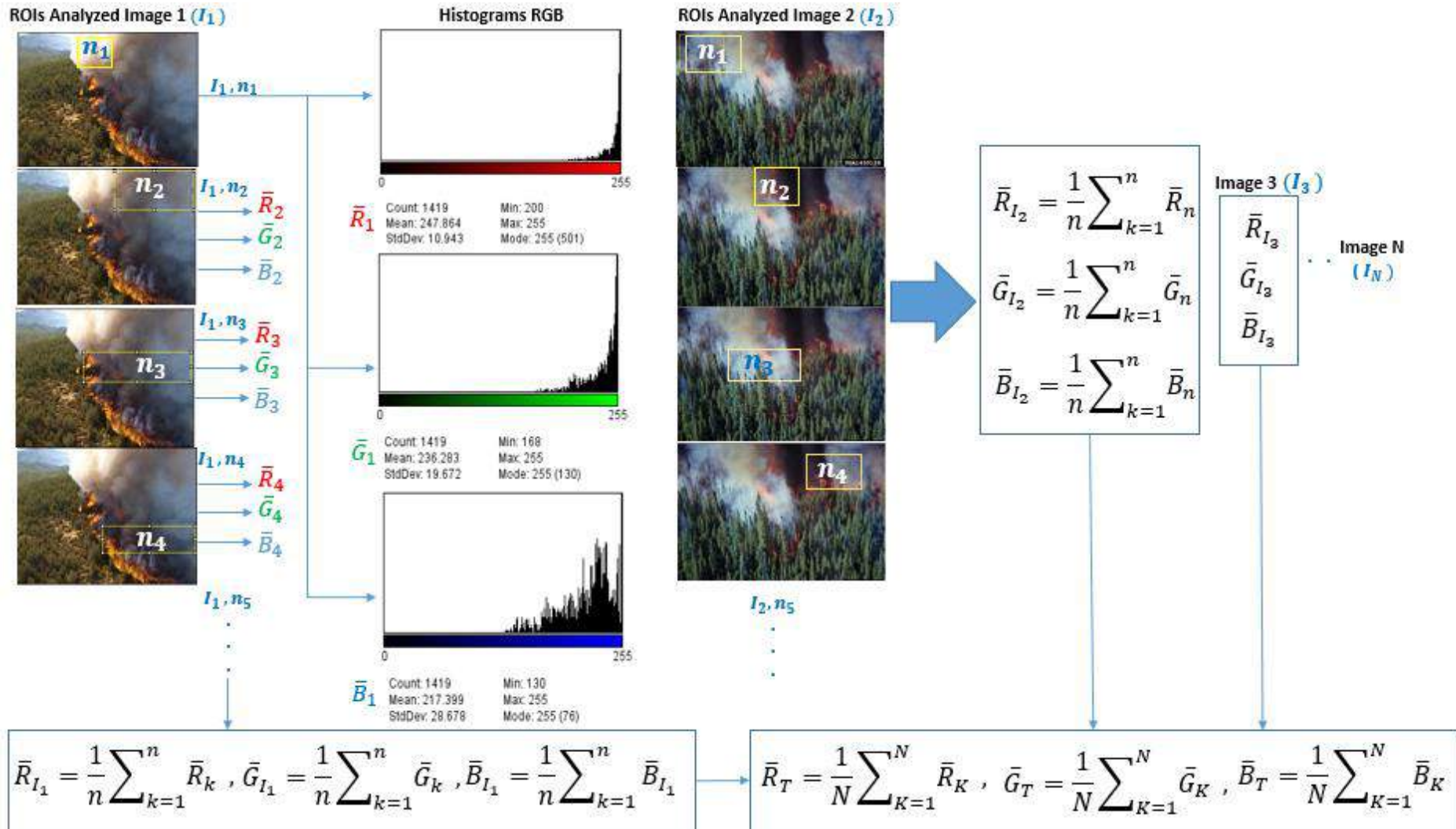


Figura 3.4. Evaluación de regiones de píxeles en imágenes.

Tabla 3. 1. Muestra del análisis cromático sobre 7 regiones.















| Muestra | | R | G | B | Y | R' | G' | B' | r | g | b | Σ rgb |
|---|-----------|--------|--------|--------|--------|--------|---------|--------|--------|--------|--------|--------------|
|  n1 | Vmax | 239 | 224 | 204 | 222 | 0,9371 | 0,8781 | 0,8004 | 0,3582 | 0,3351 | 0,3052 | 1 |
| | \bar{x} | 194,16 | 180,48 | 157,21 | 177,29 | 0,7613 | 0,7072 | 0,6164 | 0,3653 | 0,3398 | 0,2953 | 1 |
| | σ | 16,79 | 15,49 | 16,80 | 16,17 | 0,0654 | 0,0615 | 0,0652 | 0,3425 | 0,3152 | 0,3424 | 1 |
|  n2 | Vmax | 255 | 251 | 241 | 249 | 1,0010 | 0,9847 | 0,9454 | 0,3416 | 0,3363 | 0,3224 | 1 |
| | \bar{x} | 198,92 | 189,14 | 174,98 | 187,63 | 0,7809 | 0,74163 | 0,6869 | 0,3537 | 0,3357 | 0,3105 | 1 |
| | σ | 24,43 | 21,68 | 24,23 | 23,18 | 0,0952 | 0,0858 | 0,0957 | 0,3472 | 0,3084 | 0,3443 | 1 |
|  n3 | Vmax | 255 | 255 | 255 | 255 | 1,0004 | 1,0003 | 1,0002 | 0,3334 | 0,3336 | 0,3337 | 1 |
| | \bar{x} | 180,28 | 164,5 | 158,57 | 168,01 | 0,7067 | 0,6454 | 0,6213 | 0,3580 | 0,3265 | 0,3143 | 1 |
| | σ | 25,96 | 26,29 | 28,49 | 29,95 | 0,1013 | 0,1039 | 0,1114 | 0,3211 | 0,3251 | 0,3522 | 1 |
|  n4 | Vmax | 214 | 194 | 187 | 198 | 0,8392 | 0,7603 | 0,7337 | 0,3597 | 0,3262 | 0,3141 | 1 |
| | \bar{x} | 168,32 | 147,06 | 136,63 | 150,70 | 0,6604 | 0,5760 | 0,5354 | 0,3724 | 0,3255 | 0,3020 | 1 |
| | σ | 24,61 | 21,94 | 22,83 | 22,95 | 0,0968 | 0,0860 | 0,0898 | 0,3545 | 0,3163 | 0,3293 | 1 |
|  n5 | Vmax | 206 | 194 | 174 | 190 | 0,8072 | 0,7602 | 0,6822 | 0,3581 | 0,3377 | 0,3031 | 1 |
| | \bar{x} | 179,53 | 162,60 | 145,11 | 162,39 | 0,7041 | 0,6374 | 0,5694 | 0,3683 | 0,3334 | 0,2979 | 1 |
| | σ | 11,06 | 10,8 | 10,11 | 10,60 | 0,0431 | 0,0426 | 0,0391 | 0,3462 | 0,3377 | 0,3162 | 1 |
|  n6 | Vmax | 249 | 231 | 219 | 233 | 0,9762 | 0,9057 | 0,8585 | 0,3564 | 0,3323 | 0,3131 | 1 |
| | \bar{x} | 203,67 | 186,57 | 172,85 | 187,73 | 0,7984 | 0,7312 | 0,6771 | 0,3616 | 0,3312 | 0,3063 | 1 |
| | σ | 27,74 | 22,36 | 24,79 | 24,80 | 0,1086 | 0,0871 | 0,097 | 0,3703 | 0,2985 | 0,3304 | 1 |
|  n7 | Vmax | 255 | 251 | 240 | 247 | 1,0003 | 0,9844 | 0,9412 | 0,3410 | 0,3361 | 0,3212 | 1 |
| | \bar{x} | 205,16 | 195,12 | 183,05 | 194,39 | 0,8042 | 0,7655 | 0,7174 | 0,3512 | 0,3342 | 0,3135 | 1 |
| | σ | 22,75 | 21,39 | 23,28 | 22,31 | 0,0897 | 0,0837 | 0,0916 | 0,3375 | 0,3174 | 0,3459 | 1 |

Tabla 3. 2. Relación entre coordenadas cromáticas y sus correspondientes normalizadas en 7 regiones.

| Muestra | $\sigma R/\sigma G$ | \bar{R}/\bar{G} | $\sigma R/\sigma B$ | \bar{R}/\bar{B} | $\sigma r/\sigma g$ | \bar{r}/\bar{g} | $\sigma r/\sigma b$ | \bar{r}/\bar{b} | $\sigma G/\sigma B$ | \bar{g}/\bar{b} |
|---|---------------------|-------------------|---------------------|-------------------|---------------------|-------------------|---------------------|-------------------|---------------------|-------------------|
| n1  | 1,0669 | 1,0757 | 1,1715 | 1,2350 | 1,0669 | 1,0757 | 1,1715 | 1,2350 | 1,0980 | 1,1480 |
| n2  | 1,0159 | 1,0517 | 1,0580 | 1,1368 | 1,0159 | 1,0517 | 1,0580 | 1,1368 | 1,0414 | 1,0809 |
| n3  | 1,0002 | 1,0954 | 1,0007 | 1,1368 | 1,0043 | 1,0954 | 1,0007 | 1,1368 | 1,0005 | 1,0377 |
| n4  | 1,1030 | 1,1445 | 1,1443 | 1,2318 | 1,1030 | 1,1445 | 1,1443 | 1,2319 | 1,0374 | 1,0763 |
| n5  | 1,0618 | 1,1041 | 1,1839 | 1,2372 | 1,0618 | 1,1041 | 1,1839 | 1,2372 | 1,1149 | 1,1205 |
| n6  | 1,0779 | 1,0916 | 1,1369 | 1,1783 | 1,0779 | 1,0916 | 1,1369 | 1,1783 | 1,0547 | 1,0793 |
| n7  | 1,0159 | 1,0514 | 1,0625 | 1,1208 | 1,0159 | 1,0514 | 1,0625 | 1,1208 | 1,0458 | 1,0659 |
| \bar{x} | 1,0488 | 1,0878 | 1,1082 | 1,1824 | 1,0488 | 1,0878 | 1,1082 | 1,1824 | 1,0560 | 1,0870 |

A partir de las expresiones presentadas en (22) y sus correspondientes para r , g , b , se puede determinar la relación existente entre las componentes de una región determinada. Para el caso de las muestras planteadas en las Tabla 3. 1 y Tabla 3. 2, la relación que se presenta es $R > G > B$ y sus coordenadas normalizadas como $r > g > b$. Este concepto permite plantear ecuaciones algebraicas con la finalidad de encontrar regiones de píxeles en las que coinciden píxeles con la misma tonalidad de color.

En el siguiente paso se determina qué tipo de ecuación permite identificar este tipo de píxeles de interés, para ello se plantea el cálculo de un índice de color específico (ICE). El ICE es el resultado de la relación entre tonalidades de color y depende del balance cromático de la región de interés.

Es importante precisar que, para ejecutar la segmentación en la imagen, el estudio de la relación cromática no solo es de la región de interés sino que también se debe hacer con el fondo. Por lo tanto se relaciona el ICE parcial de las ROI's (ICE_1) frente al ICE parcial del fondo (ICE_2) del tal manera que se pueda encontrar ICG que sienta la base para el umbral óptimo que separa las dos regiones (primer plano del fondo). La relación establecida guarda similitud con la relación descrita en (19) y que en forma genérica puede ser de la siguiente manera:

$$ICG = ICE_1 - ICE_2 \quad . \quad (23)$$

La diferencia entre los dos ICE permite la substracción del fondo del primer plano, con ello se dispone de un mapa de regiones relevantes en la escena, proporcionando un histograma bimodal bastante bien logrado como se podrá observar en secciones posteriores.

3.4.2. Etapa 2: Segmentación del objetivo específico

Los métodos que proponen el uso de la umbralización como técnica de segmentación, buscan alcanzar un umbral óptimo. Para conseguir este objetivo se han planteado el uso de umbrales globales, locales, dinámicos o la combinación de estos. Está claro que mientras más umbrales se utilicen o se combinen mejorará la segmentación o la detección de las ROI's, sin embargo esta mejora va en detrimento de la disminución del coste computacional, lo que contrasta con el objetivo de alcanzar eficiencia considerando el empleo en sistemas de videovigilancia. Es por ello que el objetivo primordial es alcanzar umbrales óptimos que generen un compromiso entre la precisión de detección pero con bajos tiempos de procesamiento.

La umbralización es una técnica de entre las más importantes que permite la segmentación o separación de las ROI's del resto (Gonzalez & Woods, 2008). Esta técnica puede resultar muy útil a la hora de conseguir rapidez en el procesamiento y clasifica cierto grupo de píxeles en ROI's de acuerdo a características intrínsecas del mismo. Muchos autores definen a la umbralización como una técnica de segmentación más bien finita, ya que una región (R) pertenece a un set de regiones. Esto se lo puede representar matemáticamente como sigue:

$$R = \bigcup_{i=1}^S R_i \ ; \ R_i \cap R_j = 0 \ ; \ i \neq j \quad . \quad (24)$$

Donde R , pertenece al conjunto de regiones que van desde la R_1 hasta la R_S . De las revisiones realizadas, se puede inferir que la umbralización muestra un alto rendimiento cuando los contrastes que se presentan en una imagen son claramente distinguibles. Es por ello que en imágenes representadas en niveles de gris, esta técnica resulta de mucha utilidad y alcanza niveles altos de precisión. La umbralización es considerada como un método de segmentación orientado

a píxeles, ya que se realiza una comparación de los valores que disponen los píxeles frente a un patrón de referencia.

Las técnicas de umbralización se apoyan frecuentemente en herramientas de análisis como son los histogramas, estos son gráficos que permiten visualizar el punto de inflexión (valle) en donde el umbral resulta ideal. El gráfico de histogramas muestra la relación existente de los valores de los píxeles en relación a la frecuencia con los que aparecen en una imagen.

Los histogramas permiten conocer el comportamiento de los píxeles considerando valores tanto del color como del nivel de gris y su empleo depende de la funcionalidad y el criterio de la aplicación. La Figura 3.5 muestra el histograma sobre la distribución del color RGB en una imagen y su correspondiente nivel de gris.

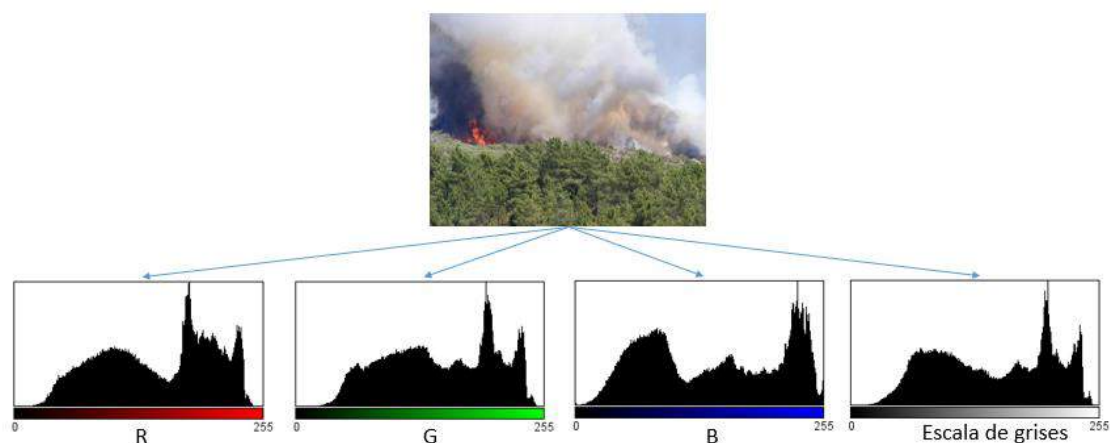


Figura 3.5. Histograma de la distribución de color para RGB y en escala de grises.

En los histogramas de la Figura 3.5, el eje de abscisas x representa el valor de color correspondiente a los píxeles en la imagen en R, G, B en un rango entre 0 y 255. En el histograma de escala de grises el eje x representa el valor normalizado del color. Por otra parte, el eje y representa el número de píxeles que corresponden a cierto valor de color o más bien dicho la frecuencia con la que se presentan en cada componente de color que conforma la imagen.

La mejor umbralización es aquella que clasifica correctamente los píxeles de interés del resto sin que afecten condiciones externas como la iluminación, así como características propias de la imagen como son tonalidades de color, textura, fondos abstractos y/o primeros planos irregulares. Es por ello que encontrar un umbral único que consiga la detecciones de ROI's y que abarque todos estos factores, es una tarea difícil de conseguir. Una umbralización denota simplicidad cuando la curva del histograma define claramente la transición entre una región de interés y otra.

Una técnica frecuentemente aplicada en la segmentación por umbralización es la propuesta por N. Otsu (Otsu, 1975). Este es un método no paramétrico, en donde el umbral óptimo se obtiene maximizando el valor estadístico de la varianza entre los valores del primer plano y el fondo. Los umbrales de clasificación binaria llegan a ser más precisos asignando en forma adecuada valores de 0 al fondo y 1 al primer plano. La aplicación del método de umbralización por Otsu combinado a los índices de color fue propuesto por Meyer y Neto (G. E. Meyer & Neto, 2008b). Una detallada descripción analítica y matemática de este método se realiza en el Anexo B, pues el criterio empleado por Otsu es de utilidad en el refinamiento de la detección a través de los nuevos índices.

En la clasificación se define la probabilidad de pertenencia de píxeles en diversos grupos, precisamente al disponer de regiones más separadas o con mayor diferencia entre sus valores garantiza una mejor agrupación entre cada región. La asignación de valores máximos y mínimos a cada región se puede conseguir a través de la clasificación no supervisada. Este tipo de clasificaciones minimizan las desviaciones entre las observaciones que pertenecen al mismo grupo de píxeles y maximizan las distancias entre los centros de los grupos. En términos del trabajo de investigación realizado, la clasificación permite dividir a las regiones en dos grupos, un grupo de píxeles que pertenecen a la ROI a los cuales se asigna el valor máximo (1) y el resto que es parte del fondo que se asigna el valor mínimo (0), este procedimiento de asignación se conoce como binarización.

Para reconocer estas regiones como tales, es necesario proceder a identificarlas, para ello se realiza el etiquetado. A partir de la imagen binarizada (0,1) se agrupan los píxeles considerando sus regiones conexas de tal forma que se logre identificar una sola región. Previo al etiquetamiento se pueden aplicar filtros de eliminación de objetos pequeños que disponen de una mínima cantidad de píxeles con la finalidad de reducir ruidos y evitar problemas de sobre-etiquetamiento.

En definitiva, el etiquetado consiste en agrupar píxeles con las mismas características y que pertenecen a una misma región asignándoles un identificador llamado etiqueta que puede ser un valor numérico. La mayor parte de técnicas de etiquetado utilizan conectividad de 4 o de 8 píxeles. La propagación de las etiquetas puede ser ejecutada desde arriba hacia abajo, de izquierda a derecha o viceversa. La combinación de formas de propagación es considerada como iterativa.

La clasificación de ROI's se hace en función de la densidad de probabilidad $p(x)$, es decir, la probabilidad de que cierto grupo de píxeles pertenezca a una región u otra. Cuando en una imagen se logra definir claramente las regiones de interés, la función de densidad es la unidad, es decir $p(x)=1$, la $p(x)$ es producto de la adición de probabilidades parciales.

En las imágenes la adición de ruido gaussiano es un hecho, es por ello que la función de la densidad de probabilidad debe incluir este valor. El estimar las densidades de probabilidad en forma balanceada permite establecer un umbral óptimo. Si se considera que un umbral T divide una imagen en dos regiones claramente distinguibles (regiones claras y regiones oscuras), la función densidad de probabilidad de pertenencia a estas regiones es la resultante de la combinación de las probabilidades parciales relacionadas con su respectivo ruido aditivo gaussiano (Gonzalez & Woods, 2008; Martinsanz & de la Cruz García, Jesús Manuel, 2001), matemáticamente esta relación se expresa como:

$$p(x) = P_1 p_1(x) + P_2 p_2(x) \quad . \quad (25)$$

Donde P_1 , y P_2 resultan las probabilidades a priori y $p_n(x)$ el ruido gaussiano. La función del ruido blanco se representa como sigue:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad . \quad (26)$$

Donde μ es la media, σ representa la desviación estándar, σ^2 la varianza de los valores de la muestra. Sabiendo que la suma de las probabilidades no pueden sobrepasar la unidad, la relación a priori será $P_1 + P_2 = 1$.

Si se busca un umbral óptimo, las dos regiones de píxeles estarán claramente definidas (primer plano y fondo) y estas deberán formar parte del conjunto global, acorde a la relación presentada en (25). De todas formas, para que estas regiones estén claramente definidas, deberá existir una variación marcada entre un determinado nivel de brillo, intensidad, color, etc. entre una región de la imagen y otra. Si consideramos el nivel de gris como parámetro de medición, entonces la variación del nivel de gris en las regiones estará dado por $u_1 > u_2$. Esto indicaría que una región posee mayor nivel de gris que la otra pero considerando un valor de referencia (punto de equilibrio) en donde se produce el cambio que en nuestro caso será el umbral (T).

Para analizar cuál es la probabilidad de que un píxel se clasifique erróneamente, se establece un solo conjunto de probabilidades erróneas (E), de las cuales se derivan básicamente dos subconjuntos: uno formado por la probabilidad de que un píxel que pertenezca al primer plano se clasifique como parte del fondo y otra al contrario. Si la función densidad de probabilidad se presenta como una curva, entonces los valores que pertenezcan a cada grupo estará dada por el área bajo la curva en los límites definidos para cada uno de ellos, en donde T es uno de los puntos límite en ambos casos. Esto puede ser representado matemáticamente a través de las ecuaciones 27 y 28.

$$E_1(T) = \int_{-\infty}^T p_2(x) dx \quad , \quad (27)$$

$$E_2(T) = \int_T^{\infty} p_1(x) dx \quad . \quad (28)$$

Considerando la ecuación proporcionada en (25), se puede establecer que el conjunto de probabilidades erróneas estará dado por:

$$E(T) = P_2 \int_{-\infty}^T p_2(x) dx + P_1 \int_T^{\infty} p_1(x) dx \quad , \quad (29)$$

o bien:

$$E(T) = P_2 E_1(T) + P_1 E_2(T) \quad . \quad (30)$$

Para encontrar el punto en donde se produce el mínimo error se deriva $E(T)$ con respecto al umbral T y se iguala a cero .

Por otra parte también si se considera que $u_1 > u_2$ es la relación de clasificación en el primer plano o en el fondo; entonces el punto que correspondería al punto de equilibrio en donde se presentaría la mínima variación de error sería x (valor aleatorio en la escala), lo que significaría que las densidades de probabilidades son iguales, esto puede ser representado como:

$$P_1 p_1(x) = P_2 p_2(x) \quad . \quad (31)$$

Si se considera a x como T , entonces tenemos:

$$P_1 p_1(T) = P_2 p_2(T) \quad . \quad (32)$$

Aplicando a (32) la densidad gaussiana y realizando las simplificaciones correspondientes se obtendrá la siguiente ecuación cuadrática:

$$AT^2 + BT + C = 0 \quad . \quad (33)$$

Donde:

$$A = \sigma_1^2 - \sigma_2^2; B = 2(\mu_1\sigma_1^2 - \mu_2\sigma_1^2); C = \mu_2^2\sigma_1^2 - \mu_1^2\sigma_2^2 + 2\sigma_2^2\sigma_1^2 \ln\left(\frac{\sigma_2 P_1}{\sigma_1 P_2}\right). \quad (34)$$

Como se puede observar, la solución a la ecuación (34) dará como resultado dos valores de umbral, con lo cual si se asume una igualdad en las varianzas entonces se podría disponer de un único umbral. En resumen, considerando la probabilidad de clasificar erróneamente un punto de un objeto, es decir clasificar un punto del fondo como parte de un objeto y establecer una relación entre un error mínimo frente a un umbral T determinado, la obtención de un umbral óptimo estaría dada por la siguiente relación:

$$T = \frac{\mu_1 + \mu_2}{2} + \frac{\sigma^2}{\mu_1 - \mu_2} \ln\left(\frac{P_2}{P_1}\right) \quad . \quad (35)$$

Donde μ_1 y μ_2 representan la media de los dos niveles de brillo (intensidad, niveles de gris, tonalidades etc.), σ^2 resulta de la igualdad de las variancias de μ_1 y μ_2 respectivamente. Cuando $P_1 = P_2$ entonces $\ln\left(\frac{P_2}{P_1}\right) = 0$ por lo que el umbral óptimo correspondería a la media de μ_1 y μ_2 .

La tarea de umbralización en imágenes en color resulta mucho más compleja debido a factores de cambios de tonalidad por variación en la incidencia de la luz, la variabilidad de reflectancia y confusiones que se producen por sombras y áreas ocultas que no permiten un efectivo etiquetado.

La normalización de imágenes de color a escala de grises facilita la segmentación por regiones, pues estabiliza la imagen y permite agrupar los píxeles en dos niveles de intensidad, considerando que los objetos estarán más o menos iluminados en relación al fondo. Cuando se presenta esta relación se dice que la imagen tiende a describir un histograma bimodal.

El buen rendimiento de un método de umbralización depende de la capacidad del mismo para obtener un histograma bimodal. Es por ello que cuando se aplica la umbralización en fuentes de información basadas en escala de grises, el proceso de detección resulta más preciso. La literatura introduce dos términos para definir esta binarización, uno es el primer plano de la imagen (*foreground*) y otro es el fondo (*background*). En la Figura 3.6 se muestra la gráfica de un histograma bimodal frente a otro claramente no bimodal. En Figura 3.6a, se muestra el punto del umbral T donde claramente se puede distinguir la tendencia del histograma para dividirse en dos regiones de interés (R_1 y R_2). En contraste, lo mostrado en Figura 3.6b con múltiples regiones de interés (R_n) y una cantidad finita de umbrales (T_n).

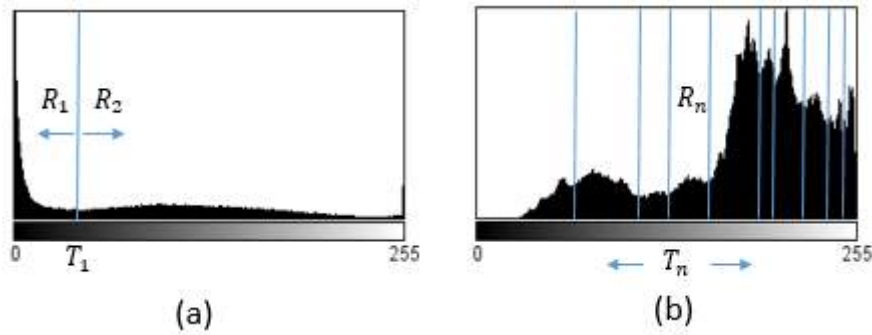


Figura 3.6. Histogramas en escala lineal empleados en la umbralización: En (a) Histograma bimodal o ideal. En (b) Histograma multimodal.

Refinamiento de la detección

La principal ventaja que dispone este método es la agilidad para realizar cálculos, aligerando la carga computacional y mejorando los tiempos de procesamiento. La principal desventaja radica en la alta sensibilidad que presenta frente a cambios de luminosidad lo cual genera falsos positivos. El objetivo final de la aplicación del umbral a través de Otsu en los sistemas de visión artificial es conseguir una imagen binaria con dos valores (1, 0) que representan el blanco y el negro.

En el caso específico de la presente tesis doctoral, el refinamiento de la detección ha sido ejecutado siguiendo el criterio de reducción de la dispersión pero relacionando la desviación estándar (σ). La σ permite aglutinar la mayor cantidad de píxeles en relación a la media del valor de cada grupo de píxeles (ICE_1, ICE_2). Esta forma de clasificación ha permitido generar mejores resultados en los casos de uso de *NFDI* y *FFDI*, los resultados de precisión de detección se observarán en secciones posteriores. En la Figura 3.7, se muestra el histograma encontrado a partir de la ejecución del $ICG=NFDI$.

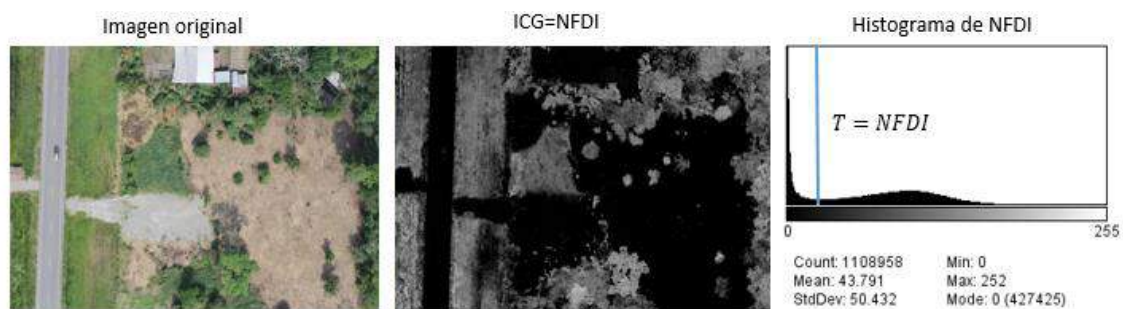


Figura 3.7. Histograma encontrado a partir de la aplicación de un ICG en la imagen.

Como se puede observar en la Figura 3.7, la respuesta de clasificación mejora frente al histograma de producto de la normalización (Figura 3.3), prácticamente por si solo el *ICG* produce una respuesta de alta precisión, sin embargo en la búsqueda continua de mejorar la clasificación de los píxeles se propone el uso de la \bar{x} y σ , estas variables estadísticas han sido utilizadas por diversos autores como se ha mostrado en secciones anteriores. Específicamente se ha tomado como referencia el trabajo presentado por N. Otsu quien maximiza la diferencia de la razón de

σ^2 entre grupos de píxeles distintos y minimiza la diferencia entre propios para alcanzar el umbral ideal.

En el presente trabajo de tesis no se ha utilizado el umbral propuesto por Otsu como tal, sino más bien se ha seguido su criterio para establecer una variante, pues se ha podido verificar en base a pruebas empíricas que el rendimiento del modelo Otsu en combinación con los índices de color no en todos los casos presentan los mejores resultados. Esta apreciación coincide con evaluaciones desarrolladas por otros autores (Guijarro et al., 2011b).

En términos de realce, lo que se busca es relacionar la \bar{x} , que representa el promedio del nivel de gris en una imagen con la σ que es una medida promedio del contraste. La \bar{x} y σ globales se aplican sobre la imagen del mapa inicial de regiones relevantes obtenidas a través del ICG, resaltando una región determinada de píxeles. Considerando el punto de vista de reducción de dispersión representado en las ecuaciones del Anexo B, además asumiendo la igualdad en las probabilidades de clasificación errónea, es decir $P_1 = P_2$. Según la ecuación (35) se propone el uso de la media de las desviaciones como un umbral óptimo T y se representa como sigue:

$$T_{ICG} = \frac{\sigma_{ICE1} + \sigma_{ICE2}}{2} , \quad (36)$$

donde el umbral de un índice en general es T_{ICG} y los valores de σ_{ICE1} y σ_{ICE2} son las desviaciones de ICE_1 e ICE_2 respectivamente. La desviación estándar nos presenta la razón de variación de valores existentes frente a la media del valor de gris en cada grupo de píxeles. Posteriormente se aplica un último umbral global, en donde se relaciona el ICG es decir los niveles de gris del mapa inicial frente al T_{ICG} , el etiquetado (0,1) se lleva a cabo a través de la siguiente relación:

$$g(x,y) = \begin{cases} ICG \geq T_{ICG} = 1 \\ ICG < T_{ICG} = 0 \end{cases} . \quad (37)$$

Al aplicar el umbral global se obtiene dos regiones claramente definidas, una que representa el primer plano en blanco y el fondo en negro, es decir la detección se ha llevado a cabo produciéndose un mapa final binarizado en donde se separa claramente la ROI (que representa al objetivo específico) del fondo.

En conjunto, la estrategia implementada ha permitido alcanzar detecciones precisas de objetivos con primeros planos irregulares, con fondos más bien ambiguos y sobre todo guardando un compromiso entre la precisión y el tiempo de ejecución. A continuación, y siguiendo la metodología, se describen los índices desarrollados y su aplicación en la detección de incidentes en vigilancia ambiental.

3.4.3. Etapa 3: Extracción de información y uso para generación de Realidad Aumentada

Una vez que se dispone de la ROI, se puede proceder a ejecutar operaciones de filtrado y etiquetado lo que facilita la identificación del objeto en la escena. Las etapas preliminares 1 y 2, permiten la detección precisa del objetivo y de esta manera obtener información clave de la ROI para producir RA. Este tipo de información está en relación con disponer de valores del objeto sobre el área en píxeles, centro geométrico, ejes mayores y menores, valores de la tonalidad de color, ubicación espacial en la escena, bordes del objeto, etc. Como se ha podido analizar en la

sección 2.4, la RA está directamente relacionada con la producción de información virtual del objetivo detectado y que es utilizada en soluciones aplicativas.

En la presente Tesis se han empleado dos técnicas para producir RA, una es a través de la combinación de las características espaciales de la ROI con relaciones geométricas de la escena y otra a partir de la combinación de la información de valores de la ROI con diversos sensores de los dispositivos inteligentes (acelerómetro, GPS, magnetómetro, etc.). Un detalle de la generación de RA se presentará en el apartado de aplicaciones y prototipos (Capítulo IV). La información virtual de RA es especialmente visual, sin embargo también puede ser audible a través de alarmas y mensajes audibles e incluso sensitiva a través de la vibración de los dispositivos. La principal información virtual de tipo visual que se puede generar a partir de la ROI puede ser:

- Bordes en color para identificar los objetivos específicos del resto de la escena.
- Perímetro del objetivo específico.
- Área aproximada del objetivo específico.
- Área del fondo de la escena.
- Mensajes visuales de alerta sobre la detección realizada.
- Ubicación de la ROI en base a coordenadas de posicionamiento geográfico.
- Fecha y hora de la detección.
- Velocidad aproximada del objetivo específico.
- Temperaturas considerando valores de las tonalidades de color de la ROI.

Si se considera en el análisis una secuencia de cuadros de video, entonces es posible obtener información del desplazamiento de la ROI, para ello se utiliza algoritmos auxiliares como flujo óptico y variación de intensidad entre píxeles (de tipo gaussiano).

En las siguientes secciones se presentan contribuciones al estado del arte tanto en cuanto se presentan nuevos índices de color que permiten la detección de fuego forestal (*FFDI*) (Cruz, Eckert, Meneses, & Martínez, 2016b) y zonas deforestadas (*NFDI*) (Cruz, Eckert, Meneses, & Martínez, 2016a). Estas detecciones cumplen con los requerimientos de detección de formas irregulares y no predecibles y en fondos ambiguos, pues la simetría, estructura y textura de los mismos son variables es decir no siguen un patrón determinado. Además se expondrá cual ha sido la respuesta de precisión frente al tiempo de procesamiento, y como estas respuestas han permitido obtener detecciones en tiempo real.

3.5. Proceso general de detección de zonas deforestadas

El proceso en general de la detección de zonas no forestadas incluye las tres etapas que se proponen en la sección 3.4, la principal diferencia representa la formulación y aplicación del índice de detección de zonas no forestadas o *Non Forest Detection Index (NFDI)* por sus siglas en inglés. El proceso en general de la detección de zonas no forestadas incluye la extracción y normalización de los componentes cromáticos, el cálculo del *NFDI*, aplicación del umbral óptimo T_{NFDI} y global $g(x,y)$ para conseguir la imagen binaria de la detección y posterior solapamiento de bordes en la imagen resultado. La Figura 3.8, muestra en forma gráfica el proceso en forma general.

El *NFDI* es una propuesta original que permite la detección y obtención de zonas deforestadas a través del cálculo del exceso de marrón (*ExBr*) y aplicación del *ExG*. Para la obtención de este

índice se ha seguido exactamente la metodología planteada en las secciones 3.4.1 y 3.4.2, considerando las siguientes particularidades del escenario de aplicación:

- Relación cromática particular entre las zonas no forestales y la vegetación.
- El cálculo del *ICE* como *ExBr* y del *ICG* como *NFDI*, en base a las relaciones cromáticas particulares encontradas.
- Cálculo del umbral global de detección considerando el escenario forestal y no forestal en específico (T_{NFDI}).

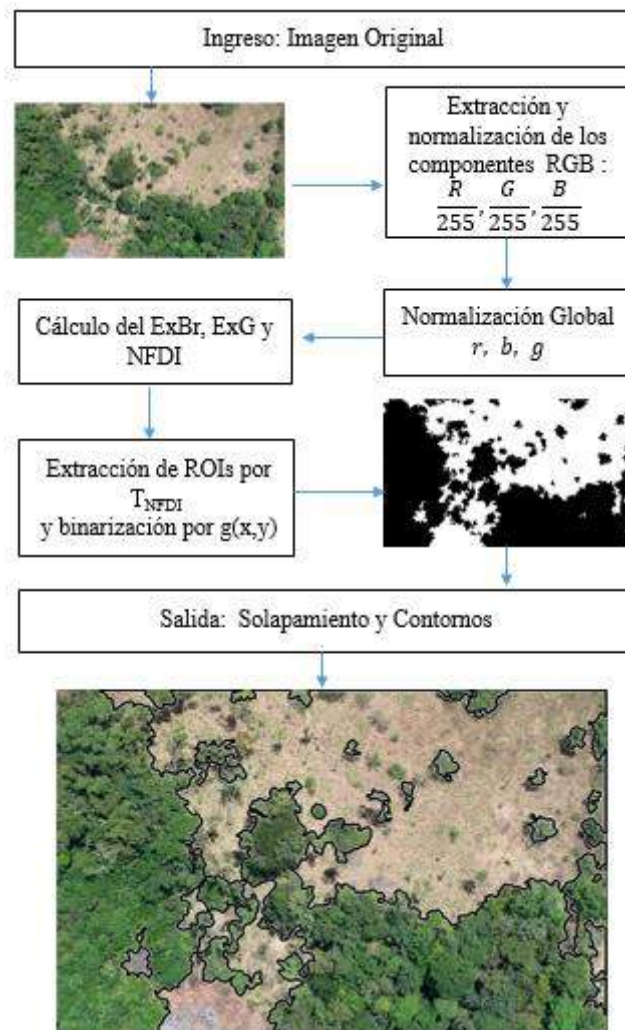


Figura 3.8. Esquema general de extracción de zonas no forestadas a través del NFDI

3.5.1. Relación entre componentes cromáticas para zonas deforestadas

Considerando la metodología de análisis y relación cromática, se realiza la correspondiente evaluación de muestras de zonas deforestadas así como de la vegetación en un variado set de imágenes. Específicamente para este caso se ha empleado alrededor de 500 imágenes de la selva ecuatoriana pertenecientes a la base de datos del Centro de Investigación y Desarrollo de la Fuerza Aérea Ecuatoriana (CIDFAE), cuyo detalle se presenta en la Anexo C.1. El propósito de ejecutar este procedimiento es obtener la relación entre los canales *r*, *g*, *b* y saber cuánto existe de cada

canal. La relación cromática encontrada para zonas deforestadas en imágenes de la selva es la siguiente:

$$r > g > b. \quad (38)$$

Para las masas forestales se presenta como:

$$g > r > b. \quad (39)$$

3.5.2. Extracción del Exceso de Marrón (*ExBr*)

A partir de (38) y (39), se pueden establecer las relaciones en las componentes con la finalidad de resaltar ciertas tonalidades de color en la imagen que corresponde a los píxeles de interés. En este caso, al conocer que la componente r es mayor en las regiones pertenecientes al suelo, se resta los valores de verde y azul de la imagen a través de las siguientes relaciones $r-b$, $r-g$ y $g-b$, a fin de obtener un índice de detección exclusivo del suelo, que se lo ha denominado como Exceso de Marrón (*ExBr*). Se lo ha denominado de esta forma pues este índice abarca tonalidades de marrón que son representativos de cierto tipo de suelos (arenosos, arcillosos y francos) además porque el índice es resultado aditivo de la componente r con muy poco de verde y azul. La relación matemática para la extracción del *ExBr* se presenta como sigue:

$$ExBr = (r - b) + (r - g) - (g - b) = 2(r - g). \quad (40)$$

El *ExBr* muestra un excelente resultado cuando existe un contraste marcado entre las tonalidades de la vegetación, este tipo de escenarios se presentan en huertos agrícolas, surcos y cultivos. Pero en imágenes en donde la vegetación es más copiosa como es el caso de las selvas tropicales, en donde no existe un marcado contraste entre el suelo y la masa forestal, el *ExBr* no tiene un buen rendimiento pues no es capaz de generar mapas de regiones del suelo (zonas no forestadas). El principal motivo de la reducción de eficacia se debe a confusiones por no reconocimiento de tonalidades del suelo debido al crecimiento de las hierbas, matorrales y/o pequeña vegetación. En la Figura 3.9 se muestra cómo actúa el *ExBr* en dos tipos de imágenes, en (a) el contraste de las tonalidades entre los surcos y la regiones cultivadas son claramente identificables, con lo cual se genera un mapa con las regiones que forman parte del suelo, además su histograma tiende a ser bimodal. En su defecto en (b), la imagen muestra una zona selvática deforestada en donde no existe un contraste bien definido con lo cual el rendimiento del *ExBr* es pobre y así lo demuestra también su histograma que tiende a ser plano.

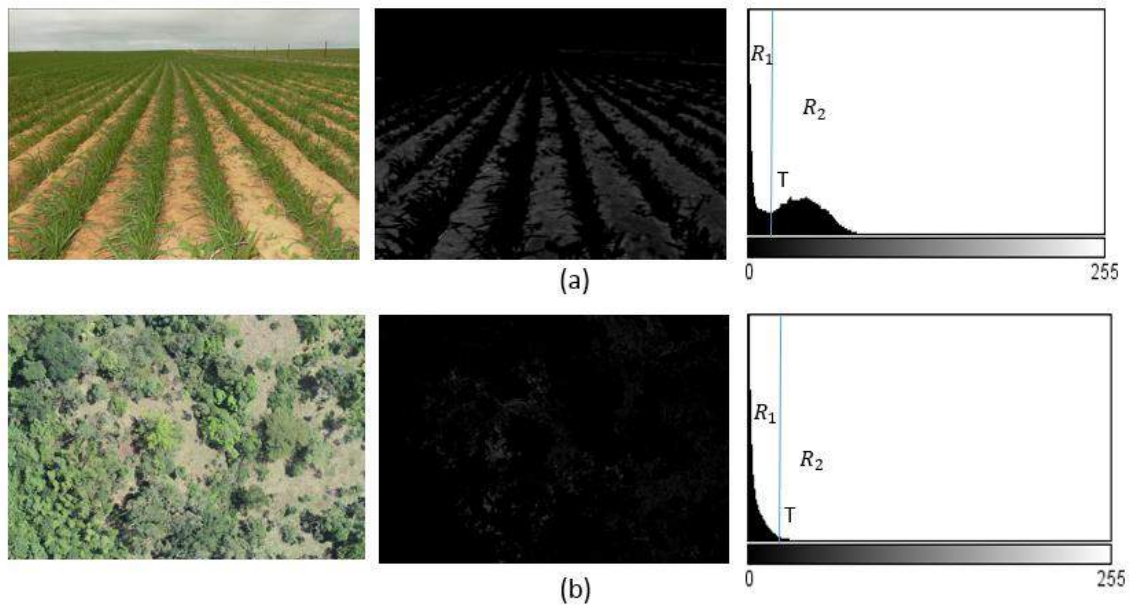


Figura 3.9. Aplicación del *ExBr* en (a) imagen con contrastes bien definidos entre el suelo y la vegetación, en (b) imagen con mayor cantidad de vegetación y contrastes poco definidos.

3.5.3. Extracción del índice de zonas no forestadas (*NFDI*)

Para conseguir una mejora en la adquisición de los mapas de regiones no forestadas evitando las confusiones por tonalidades que se presenta en especial en bosques y foresta de densa vegetación, se propone el uso del *NFDI*. Este nuevo índice es el resultado de la sustracción del *ExBr* del Exceso de Verdes (*ExG*) como se presenta en (17), siguiendo además el procedimiento planteado para obtener el *ICG* y revisado en la sección 3.4.1. El resultado se presenta como sigue:

$$NFDI = ExG - ExBr = 4g - 3r - b \quad . \quad (41)$$

Al aplicar el *ExG* en la escena, se extraen todos los píxeles que conforman la vegetación, es decir que llevan información mayoritariamente de la componente *g*. Por otra parte, al aplicar el *ExBr*, se extraen los píxeles que conforman el suelo y otros componentes cuya información procede en su mayoría de la componente *r* (vegetación muerta, seca o quemada, residuos sólidos, sombras, etc.). Al substraer la información de *ExBr* del *ExG* se potencia los valores de verdes en la imagen de tal manera que un mayor número tonalidades de verde es admitido dentro de nuevo umbral (*NFDI*). De esta manera el *NFDI* muestra un mapa general de la vegetación que abarca la mayor parte de la masa forestal o vegetativa en la imagen. En la Figura 3.10 se puede observar el proceso de obtención del *NFDI* así como el mapa de la región forestal obtenido.

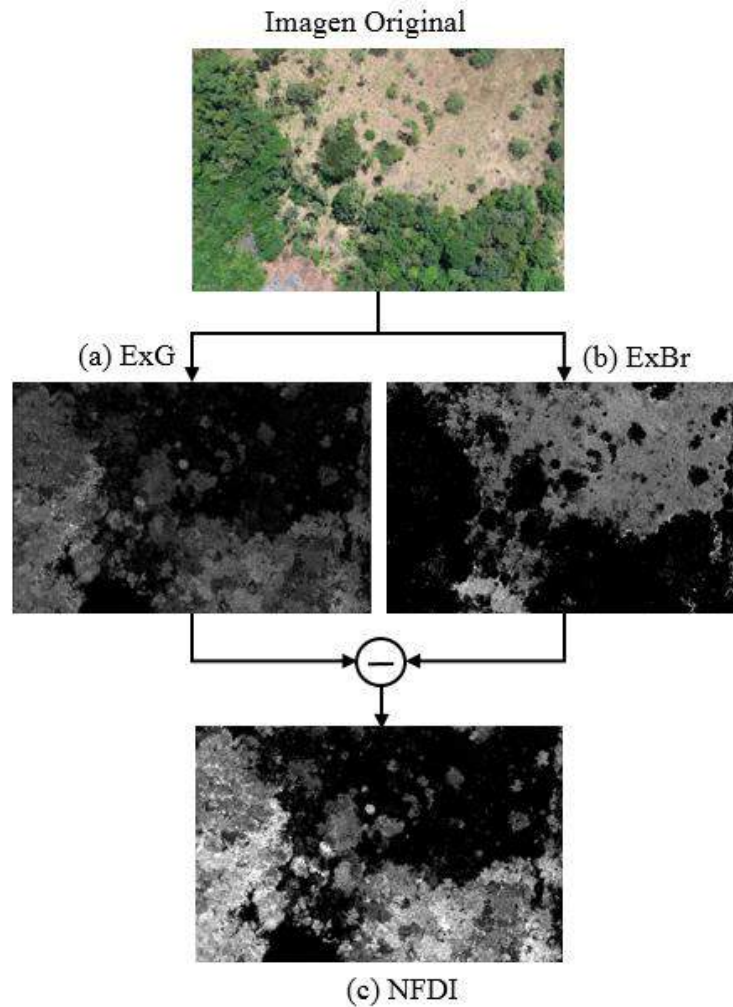


Figura 3.10. Proceso de obtención del NFDI en (a) mapa de la vegetación a través del ExG en (b) mapa de las zonas deforestadas encontradas a partir del ExBr en (c) se muestra el resultado de la detección a través del NFDI.

Como se ha revisado en secciones anteriores, el índice *ExGR* permite también la extracción de los píxeles que conforman la vegetación, sin embargo este índice muestra confusiones y deficiencias en la detección. La segmentación a través del *ExGR* no se ajusta a los contornos de la masa forestal lo que impide que se delimite correctamente las dos regiones (suelo y vegetación), además presenta múltiples falsos positivos debido a confusiones de tonalidad dentro de la misma masa forestal. En la Figura 3.11 se presenta precisamente una muestra comparativa del rendimiento del *ExGR* frente al *NFDI*, en especial el histograma del *NFDI* es bimodal, lo que decanta al final en una reducción de falsos positivos y falsos negativos, como se puede observar en la imagen el *NFDI*, delimita claramente la región no forestada de la vegetación. En posteriores secciones se realizará un análisis pormenorizado del rendimiento del *NFDI* frente a otros algoritmos de segmentación.

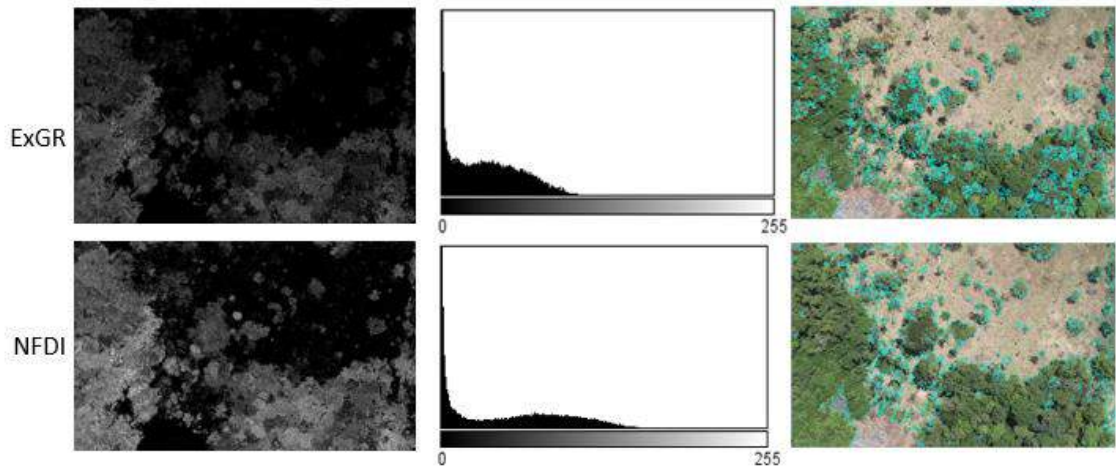


Figura 3.11. Análisis gráfico comparativo del rendimiento del NFDI frente al ExGR.

3.5.4. Umbral global para la detección de zonas no forestadas (T_{NFDI})

Como se ha obtenido un mapa de regiones relevantes con un claro contraste, la tendencia es a que se genere un histograma bimodal cada vez más exacto, es importante aprovechar esta ventaja para alcanzar a posteriori una mejor clasificación. Precisamente se plantea la opción de reducir la dispersión de los píxeles que tienen las mismas características de color, agrupándolos en dos regiones (vegetación y zona no forestada). Para cumplir con este propósito, se aplica la propuesta de umbral óptimo planteada en (36), pero adaptada para el requerimiento de detección actual. De esta manera, se establece un umbral denominado *Non Forest Detection Threshold* (T_{NFDI}), el mismo que actúa como un clasificador sin supervisión y se adquiere a partir de la media de las desviaciones estándar de ExG y $ExBr$ como:

$$T_{NFDI} = \frac{\sigma_{ExBr} + \sigma_{ExG}}{2} . \quad (42)$$

Donde σ_{ExBr} y σ_{ExG} son las desviaciones estándar de $ExBr$ y ExG respectivamente. La extracción de la imagen binaria $g(x, y)$ se plantea como:

$$g(x, y) = \begin{cases} NFDI < T_{NFDI} = 0 \\ NFDI \geq T_{NFDI} = 1 \end{cases} . \quad (43)$$

La Figura 3.12, muestra el resultado final de la binarización del $NFDI$ a través del T_{NFDI} . Aquí se puede ver el ajuste preciso de los contornos irregulares de las ROI's, evitando además confusiones por similitud de tonalidades, sombras y oclusiones. Si se considera que se ha efectuado un etiquetado binario, se puede establecer que se cumple el requerimiento de una detección de objetivos específicos con primeros planos irregulares (zonas no forestadas), con formas no definidas y en entornos con un fondo que no presenta una textura uniforme o patrón determinado (es decir es ambiguo).