

**UNIVERSIDAD DE ALCALÁ**

**Escuela Politécnica Superior**

**INGENIERÍA DE TELECOMUNICACIÓN**



**Trabajo Fin de Carrera**

**“Sistema de monitorización mediante visión artificial de zonas de riesgo potencial para la circulación ferroviaria y seguridad de las personas”**

David Jiménez Cabello  
2009



# UNIVERSIDAD DE ALCALÁ

Escuela Politécnica Superior

## INGENIERÍA DE TELECOMUNICACIÓN

Trabajo Fin de Carrera

**“Sistema de monitorización mediante visión artificial de zonas de riesgo potencial para la circulación ferroviaria y seguridad de las personas”**

Alumno: David Jiménez Cabello

Director: Dr. D. Daniel Pizarro Pérez

**Tribunal:**

**Presidente:** Dr. D. Santiago Cóbreces Álvarez.

**Vocal 1º:** Dr. D. Alfredo Gardel Vicente.

**Vocal 2º:** Dr. D. Daniel Pizarro Pérez.

Calificación: .....

Fecha: .....





Quisiera dedicar el presente proyecto final de carrera a mi familia,  
por inculcar en mí los valores necesarios para afrontar la vida  
con humildad y coherencia,  
y a mi chica Natalia, por estar a mi lado en todo momento  
y proporcionarme la comprensión y el apoyo incondicional  
que siempre he necesitado.  
A todos, muchas gracias: este trabajo es vuestro.

*“La utopía está en el horizonte.  
Me acerco dos pasos, ella se aleja dos pasos.  
Camino diez pasos y el horizonte se desplaza  
diez pasos más allá.  
Por mucho que camine, nunca la alcanzaré.  
Entonces, ¿para qué sirve la utopía?  
**Para eso: sirve para caminar.**”*

*Eduardo Galeano*



# Agradecimientos

*A mis compañeros del “Espacio Inteligente”, a los que están y a los que ya se fueron, por su ayuda y compañerismo que tanto me han aportado en este proyecto, así como al Dr. Manuel Mazo Quintas e integrantes del proyecto AVIZOR, por hacer posible que este trabajo salga adelante. En especial, quisiera agradecer a mi tutor, Dr. Daniel Pizarro Pérez, todo el apoyo y dedicación recibidos, sin el cual esta etapa no hubiera sido tan enriquecedora.*



# Índice general

|           |                                                                                |           |
|-----------|--------------------------------------------------------------------------------|-----------|
| <b>I</b>  | <b>Resumen</b>                                                                 | <b>19</b> |
| 1.        | Resumen                                                                        | 21        |
| <b>II</b> | <b>Memoria</b>                                                                 | <b>23</b> |
| 2.        | Introducción                                                                   | 25        |
| 2.1.      | Estado del arte . . . . .                                                      | 27        |
| 2.1.1.    | Sistemas de detección a ubicar en pasos a nivel . . . . .                      | 27        |
| 2.1.2.    | Sistemas de transmisión inalámbrica de vídeo . . . . .                         | 29        |
| 2.2.      | Objetivos propuestos . . . . .                                                 | 31        |
| 2.2.1.    | Especificaciones del proyecto . . . . .                                        | 32        |
| 2.3.      | Descripción de las tareas realizadas . . . . .                                 | 37        |
| 2.3.1.    | Tareas contempladas en el proyecto AVIZOR . . . . .                            | 37        |
| 2.3.2.    | Tareas desarrolladas en el presente proyecto . . . . .                         | 43        |
| 3.        | Diseño del módulo sensorial de visión a ubicar en los pasos a nivel            | 45        |
| 3.1.      | Introducción . . . . .                                                         | 45        |
| 3.2.      | Objetivos y definición del problema . . . . .                                  | 47        |
| 3.2.1.    | Resumen de la solución propuesta . . . . .                                     | 50        |
| 3.3.      | Diagrama general . . . . .                                                     | 52        |
| 3.4.      | Aspectos teóricos relacionados con la solución propuesta . . . . .             | 53        |
| 3.4.1.    | Geometría de formación de la imagen . . . . .                                  | 53        |
| 3.4.2.    | Espacios de color . . . . .                                                    | 73        |
| 3.4.3.    | Segmentación de fondo . . . . .                                                | 87        |
| 3.5.      | Solución propuesta . . . . .                                                   | 97        |
| 3.5.1.    | Diagrama modular del sistema de visión . . . . .                               | 97        |
| 3.5.2.    | Sistema de segmentación “KAM” . . . . .                                        | 100       |
| 3.5.3.    | Sistema de eliminación de sombras . . . . .                                    | 107       |
| 3.5.4.    | Sistema de actualización de fondo . . . . .                                    | 113       |
| 3.5.5.    | Sistema de conmutación día $\Leftrightarrow$ noche . . . . .                   | 120       |
| 3.5.6.    | Fusión con otras alternativas de segmentación . . . . .                        | 121       |
| 4.        | Desarrollo del sistema inalámbrico de banda ancha para transmisión de imágenes | 123       |
| 4.1.      | Introducción . . . . .                                                         | 123       |
| 4.2.      | Objetivos . . . . .                                                            | 124       |
| 4.3.      | Aspectos teóricos relacionados . . . . .                                       | 125       |

|            |                                                                             |            |
|------------|-----------------------------------------------------------------------------|------------|
| 4.3.1.     | Estudio de las tecnologías inalámbricas disponibles . . . . .               | 125        |
| 4.3.2.     | Vídeo vigilancia . . . . .                                                  | 134        |
| 4.3.3.     | Sistemas inalámbricos para vídeo IP . . . . .                               | 136        |
| 4.3.4.     | Aspectos teóricos relacionados con la solución propuesta: WiMAX . .         | 138        |
| 4.3.5.     | Solución probada: BreezeACCESS VL vídeo . . . . .                           | 146        |
| 4.4.       | Sistema de transmisión de vídeo en tiempo real al equipo receptor . . . . . | 151        |
| 4.4.1.     | Especificación de la interfaz externa . . . . .                             | 151        |
| <b>5.</b>  | <b>Implementación del sistema en tiempo real</b>                            | <b>157</b> |
| 5.1.       | Introducción . . . . .                                                      | 157        |
| 5.2.       | Servidor del módulo de tierra . . . . .                                     | 158        |
| 5.2.1.     | Subsistema de sincronización . . . . .                                      | 161        |
| 5.2.2.     | Subsistema servidor de imágenes . . . . .                                   | 161        |
| 5.2.3.     | Subsistema servidor de mensajes . . . . .                                   | 162        |
| 5.2.4.     | Subsistema de monitorización . . . . .                                      | 162        |
| 5.2.5.     | Subsistema de captura . . . . .                                             | 163        |
| 5.2.6.     | Subsistema de procesamiento . . . . .                                       | 164        |
| <b>6.</b>  | <b>Pruebas y resultados</b>                                                 | <b>165</b> |
| 6.1.       | Detección de obstáculos . . . . .                                           | 166        |
| 6.1.1.     | Segmentación de fondo mediante el algoritmo propuesto . . . . .             | 166        |
| 6.1.2.     | Segmentación en el espacio invariante a la iluminación . . . . .            | 171        |
| 6.1.3.     | Absorción de la problemática en entornos exteriores . . . . .               | 172        |
| 6.1.4.     | Fusión de alternativas de segmentación . . . . .                            | 174        |
| 6.1.5.     | Conmutación día $\Leftrightarrow$ noche . . . . .                           | 176        |
| 6.1.6.     | Tiempos de cómputo . . . . .                                                | 177        |
| 6.2.       | Servidor de imágenes y mensajes de control . . . . .                        | 178        |
| 6.2.1.     | Composición de la imagen transmitida y envío de mensajes de control .       | 178        |
| 6.2.2.     | Reproducción del video-streaming . . . . .                                  | 180        |
| 6.2.3.     | Análisis del ancho de banda transmitido . . . . .                           | 181        |
| 6.2.4.     | Tiempos de cómputo . . . . .                                                | 181        |
| <b>7.</b>  | <b>Conclusiones</b>                                                         | <b>183</b> |
| <b>III</b> | <b>Manual de Usuario</b>                                                    | <b>187</b> |
| <b>8.</b>  | <b>Manual de Usuario</b>                                                    | <b>189</b> |
| 8.1.       | Compilación de la solución software . . . . .                               | 189        |
| 8.2.       | Inicialización de la solución software . . . . .                            | 189        |
| 8.3.       | Preparación y colocación de los elementos hardware . . . . .                | 192        |
| <b>IV</b>  | <b>Pliego de condiciones</b>                                                | <b>193</b> |
| <b>9.</b>  | <b>Pliego de condiciones</b>                                                | <b>195</b> |
| 9.1.       | Hardware . . . . .                                                          | 195        |
| 9.2.       | Software . . . . .                                                          | 196        |

**V Presupuesto 197****10. Presupuesto 199**

|                                                                     |     |
|---------------------------------------------------------------------|-----|
| 10.1. Costes de ejecución material . . . . .                        | 199 |
| 10.1.1. Costes de equipamiento hardware . . . . .                   | 199 |
| 10.1.2. Costes de software . . . . .                                | 200 |
| 10.1.3. Costes por tiempo empleado . . . . .                        | 200 |
| 10.1.4. Coste total del presupuesto de ejecución material . . . . . | 200 |
| 10.2. Gastos generales y beneficio industrial . . . . .             | 201 |
| 10.3. Importe total del presupuesto . . . . .                       | 201 |

**VI Bibliografía 203**





# Índice de figuras

|                                                                                                                                                                                                                        |    |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 2.1. Arquitectura general del sistema en la zona vigilada. . . . .                                                                                                                                                     | 26 |
| 2.2. Diagrama modular del sistema. . . . .                                                                                                                                                                             | 31 |
| 2.3. Diagrama temporal de emisión de alarma y señal de vídeo al tren. . . . .                                                                                                                                          | 32 |
| 2.4. Estructura de conexión del MOVI. . . . .                                                                                                                                                                          | 33 |
| 3.1. Monitorización de elementos anómalos. . . . .                                                                                                                                                                     | 47 |
| 3.2. Diagrama inicial de la solución propuesta para la detección de obstáculos. . . .                                                                                                                                  | 51 |
| 3.3. Diagrama general de la solución propuesta. . . . .                                                                                                                                                                | 52 |
| 3.4. Representación de la formación de una imagen en el interior de una cámara pin-hole (izquierda) y el desplazamiento del plano imagen para evitar la inversión de la proyección (derecha). . . . .                  | 54 |
| 3.5. Modelado geométrico de la cámara pin-hole. Tener en cuenta que el plano imagen está desplazado en frente del centro óptico y por tanto la imagen no quedará invertida. . . . .                                    | 55 |
| 3.6. Modelado geométrico de la cámara pin-hole. Relación de triángulos. . . . .                                                                                                                                        | 56 |
| 3.7. Sistema de coordenadas de imagen $(u, v)$ y de la cámara $(u_{cam}, v_{cam})$ . El punto principal $\mathbf{p}$ tiene coordenadas $(u_0, v_0)$ . . . . .                                                          | 57 |
| 3.8. Aclaración de la notación empleada para el cambio del origen del sistema de referencia. . . . .                                                                                                                   | 58 |
| 3.9. Transformación necesaria para el cambio del sistema de referencia. . . . .                                                                                                                                        | 60 |
| 3.10. Relación de puntos entre una cámara proyectiva y un plano en el espacio tridimensional. . . . .                                                                                                                  | 65 |
| 3.11. Relación de puntos entre dos cámaras proyectivas y un plano en el espacio tridimensional. . . . .                                                                                                                | 66 |
| 3.12. De izquierda a derecha y de arriba a bajo, se muestra la convergencia del registro directo de imágenes en 8 iteraciones para dos situaciones distintas (diferente inicialización de puntos para la ROI). . . . . | 71 |
| 3.13. Espectro visible . . . . .                                                                                                                                                                                       | 73 |
| 3.14. Separación de los canales RGB (izquierda) y representación del espacio de color en forma de cubo (derecha). . . . .                                                                                              | 74 |
| 3.15. Representación del espacio de color HSI. . . . .                                                                                                                                                                 | 75 |
| 3.16. Representación del espacio de color HSV. . . . .                                                                                                                                                                 | 76 |
| 3.17. Una imagen junto con sus componentes Y, U, y V. . . . .                                                                                                                                                          | 76 |
| 3.18. Reflexión de una superficie lambertiana ideal. . . . .                                                                                                                                                           | 77 |
| 3.19. Diagrama de radiación de un cuerpo negro de distintas temperaturas. . . . .                                                                                                                                      | 78 |
| 3.20. Sensibilidad de la cámara. Se representan con unas deltas la respuesta de banda estrecha sobre la reales. . . . .                                                                                                | 78 |

|                                                                                                                                                                                                                                     |     |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 3.21. Representación de los valores de $\chi$ en 3 superficies, para diferentes temperaturas de color. . . . .                                                                                                                      | 80  |
| 3.22. Proyección de los puntos de una superficie iluminada a distintas temperaturas sobre la recta definida por el vector $\bar{e}^\perp$ . . . . .                                                                                 | 80  |
| 3.23. Representación de los logaritmos de los ratios de crominancia en 3 superficies, bajo diferentes tipos de iluminación. . . . .                                                                                                 | 81  |
| 3.24. Diagrama modular del proceso de obtención del espacio invariante a la iluminación. . . . .                                                                                                                                    | 82  |
| 3.25. Diagrama modular del proceso de calibración del ángulo $\theta$ . . . . .                                                                                                                                                     | 86  |
| 3.26. Ejemplos de sombra proyectada por una persona presente en la escena (izquierda) y las sombras proyectadas en la escena por varias personas que no se encuentran en su interior (derecha). . . . .                             | 90  |
| 3.27. Imagen en la que se observan sombras proyectadas en la cara de una persona y que no pertenecen al fondo. . . . .                                                                                                              | 90  |
| 3.28. Ejemplo del cambio de iluminación sufrido en una escena ante la llegada de un banco de nubes. . . . .                                                                                                                         | 90  |
| 3.29. Las dos primeras imágenes muestran el reflejo producido en una superficie mojada. La tercera y cuarta imagen representan los reflejos en espejo y suelo pulido, respectivamente. . . . .                                      | 91  |
| 3.30. Representaciones de diferentes GGD's en función de los parámetros $\vec{\lambda}$ . Los parámetros $\vec{\mu}$ y $\vec{\sigma}$ se fijan con valores (23, 24) y (7, 7) respectivamente. . . . .                               | 92  |
| 3.31. Ilustración de la posición de los centroides (puntos negros) al aplicar el algoritmo de k-medias (con k=2), durante varias iteraciones para una muestra compuesta de vectores bidimensionales (cruces grises). . . . .        | 94  |
| 3.32. Representaciones de la región de separación de dos clases según el criterio de la distancia de Mahalanobis (izquierda) y la dependencia de la función discriminante con la desviación típica de las clases (derecha). . . . . | 95  |
| 3.33. Diagrama modular de la solución propuesta para la detección de obstáculos. . .                                                                                                                                                | 99  |
| 3.34. Esquema general del proceso de clasificación mediante el algoritmo k-medias modificado. . . . .                                                                                                                               | 101 |
| 3.35. Diagrama modular del algoritmo k-medias modificado. . . . .                                                                                                                                                                   | 103 |
| 3.36. Diagrama modular del proceso de segmentación mediante distancia de Mahalanobis modificada. . . . .                                                                                                                            | 105 |
| 3.37. Imágenes capturadas de una escena por dos cámaras distintas (fila superior) y su correspondiente segmentación en una etapa previa a la eliminación de sombras (fila inferior). . . . .                                        | 108 |
| 3.38. Resultado del proceso de eliminación del plano del suelo. Redondeado en azul aparece la segmentación proveniente de CAM1 y en amarillo la de CAM2 transformada por la matriz de homografía $\mathcal{H}$ . . . . .            | 109 |
| 3.39. Resta en valor absoluto de las imágenes segmentadas a color. . . . .                                                                                                                                                          | 110 |
| 3.40. Zonas recuperadas de la eliminación del plano del suelo que pertenecen a los obstáculos detectados. . . . .                                                                                                                   | 110 |
| 3.41. Incorporación de la zona reconstruida en el paso 3 a la imagen resultante de la eliminación del plano de suelo. En rojo, aparece redondeadas las zonas donde principalmente se ha añadido la reconstrucción. . . . .          | 111 |
| 3.42. Resultado de la imagen segmentada tras el proceso de reconstrucción completo. .                                                                                                                                               | 111 |

|                                                                                                                                                              |     |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 3.43. Diagrama modular del proceso de eliminación de sombras empleando técnicas de visión estéreo (uso de la matriz de transformación u homografía). . . . . | 112 |
| 3.44. Diagrama modular del proceso de creación del mapa de disparidad para la cámara 1; se realizará un proceso dual para la otra cámara. . . . .            | 114 |
| 3.45. Diagrama modular del proceso de actualización del modelo probabilístico de fondo. . . . .                                                              | 117 |
| 3.46. Ejemplo del proceso de actualización con incremento/decremento del tiempo de vida ante una muestra actualizada. . . . .                                | 119 |
| 3.47. Esquema del proceso de fusión con 3 posible alternativas de segmentación. . .                                                                          | 121 |
| 4.1. Posicionamiento de las tecnologías inalámbricas . . . . .                                                                                               | 126 |
| 4.2. Estándares inalámbricos según su cobertura . . . . .                                                                                                    | 128 |
| 4.3. Comparativa de tecnologías inalámbricas . . . . .                                                                                                       | 130 |
| 4.4. Relación movilidad-ancho de banda entre tecnologías inalámbricas . . . . .                                                                              | 131 |
| 4.5. Relación cobertura-velocidad de transferencia . . . . .                                                                                                 | 131 |
| 4.6. FLASH OFDM apunta a usuarios móviles y fijos por igual . . . . .                                                                                        | 132 |
| 4.7. Evolución de los estándares 3G . . . . .                                                                                                                | 133 |
| 4.8. Formato de codificación de imágenes . . . . .                                                                                                           | 135 |
| 4.9. Requerimientos de ancho de banda para transmisión de vídeo . . . . .                                                                                    | 135 |
| 4.10. Diagrama de bloques del sistema de transmisión/recepción analógico . . . . .                                                                           | 138 |
| 4.11. Tipos de acceso (izquierda) y principales características (derecha) en las redes WiMAX . . . . .                                                       | 139 |
| 4.12. Estándar 802.16 y su integración con 802.11 . . . . .                                                                                                  | 140 |
| 4.13. Espectro de licencias disponibles . . . . .                                                                                                            | 142 |
| 4.14. Resumen de características WiMAX. . . . .                                                                                                              | 143 |
| 4.15. Comparativa en cobertura de las redes inalámbricas de banda ancha . . . . .                                                                            | 145 |
| 4.16. Elementos de la solución BreezeACCESS VL . . . . .                                                                                                     | 148 |
| 4.17. Pila de protocolos OSI . . . . .                                                                                                                       | 151 |
| 5.1. Esquema de los hilos empleados para el funcionamiento del sistema; procesado concurrente . . . . .                                                      | 159 |
| 6.1. Escenario empleado como fondo para las pruebas realizadas . . . . .                                                                                     | 165 |
| 6.2. Resultados obtenidos en la segmentación de fondo en un instante previo a la actualización. . . . .                                                      | 166 |
| 6.3. Separación del proceso de eliminación de sombras en cinco pasos . . . . .                                                                               | 167 |
| 6.4. Ejemplos del proceso de actualización del modelo de fondo . . . . .                                                                                     | 168 |
| 6.5. Resultado obtenido en un instante posterior a la actualización de fondo . . . . .                                                                       | 169 |
| 6.6. Ejemplos de las imágenes segmentadas tras la aplicación de una iteración del algoritmo completo . . . . .                                               | 170 |
| 6.7. Ejemplos de las imágenes segmentadas tras la aplicación de una iteración del algoritmo completo en el espacio invariante a la iluminación . . . . .     | 171 |
| 6.8. Ejemplos de eliminación de sombras proyectadas por objetos que no se encuentran presentes en el interior de la región de interés . . . . .              | 172 |
| 6.9. Resultado de la segmentación ante cambios en el modelo de fondo (evolución de sombras) . . . . .                                                        | 173 |
| 6.10. Resultado de la detección de obstáculos mientras el fondo varía . . . . .                                                                              | 173 |
| 6.11. Ejemplos de detección ante cambios de iluminación . . . . .                                                                                            | 174 |

|                                                                                                                                            |     |
|--------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 6.12. Los tres obstáculos son detectados por las distintas alternativas de segmentación                                                    | 175 |
| 6.13. La papelera amarilla es detectado por dos de las alternativas. Se mantiene la detección de los 3 obstáculos . . . . .                | 175 |
| 6.14. La papelera azul solo es detectada por uno de los algoritmos. Se mantiene la identificación de los 4 obstáculos . . . . .            | 176 |
| 6.15. Resultados obtenidos tras la conmutación del sistema a modo nocturno . . . .                                                         | 176 |
| 6.16. Imágenes compuestas para su envío al módulo receptor ubicado en el tren. Ejemplos para $N=0,2,3$ y 4 obstáculos detectados . . . . . | 178 |
| 6.17. Capturas realizadas en la verificación del correcto funcionamiento de la transmisión de vídeo empleando el reproductor VLC . . . . . | 180 |
| 8.1. Ejemplo de imagen transmitida al tren. . . . .                                                                                        | 191 |
| 8.2. Ejemplo de la correcta visualización del vídeo transmitido. . . . .                                                                   | 191 |
| 8.3. Esquema para la colocación y preparación de los elementos hardware. . . . .                                                           | 192 |

# Índice de tablas

|                                                                                                                                           |     |
|-------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 3.1. Requisitos del sistema de visión y soluciones propuestas. . . . .                                                                    | 49  |
| 3.2. Resumen del método de registro directo de imágenes . . . . .                                                                         | 72  |
| 3.3. Resumen del algoritmo K-medias dividido en 4 pasos . . . . .                                                                         | 94  |
| 4.1. Formato de la trama . . . . .                                                                                                        | 153 |
| 4.2. Identificación de la trama . . . . .                                                                                                 | 153 |
| 4.3. Estructura del paquete RTP. . . . .                                                                                                  | 154 |
| 4.4. Parámetros del mensaje de control . . . . .                                                                                          | 155 |
| 5.1. Parámetros supervisados en el subsistema de monitorización y medidas correc-<br>toras aplicadas . . . . .                            | 163 |
| 6.1. Tiempos de conmutación más relevantes en la implementación del sistema de<br>detección de obstáculos en ejecución diurna . . . . .   | 177 |
| 6.2. Tiempos de conmutación más relevantes en la implementación del sistema de<br>detección de obstáculos en ejecución nocturna . . . . . | 177 |
| 6.3. Ejemplo de mensaje de control enviado al módulo del tren para la última imagen<br>de la figura 6.16 . . . . .                        | 179 |
| 6.4. Análisis del ancho de banda . . . . .                                                                                                | 181 |
| 6.5. Tiempos de cómputo más relevantes en la implementación del servidor de imá-<br>genes y mensajes de control . . . . .                 | 181 |



# **Parte I**

## **Resumen**





# Capítulo 1

## Resumen

En la actualidad, las comunicaciones ferroviarias presentan a lo largo de su recorrido una serie de puntos considerados focos de alta siniestralidad. Los pasos a nivel son un ejemplo de estos puntos, en los que la seguridad se basa principalmente en la utilización de señales de tráfico y sistemas de barreras. Dichos sistemas tradicionales no son efectivos, produciendo en ocasiones atropellamientos y descarrilamientos del tren.

La finalidad del proyecto es el diseño de un sistema que contribuya a un aumento de la seguridad en el transporte por ferrocarril, centrándose en puntos de alta siniestralidad, y con la finalidad de reducir el número de accidentes. El objetivo es la inclusión de sensores en los puntos críticos que sean capaces de comunicar su situación al tren en el momento en el que éste dispone de suficiente distancia para realizar un frenado seguro.

El presente trabajo final de carrera se encuentra enmarcado dentro del proyecto de investigación “Sistema Global de Supervisión de Zonas de Riesgo Potencial para la Circulación Ferroviaria y Seguridad de las Personas”(AVIZOR), realizado conjuntamente por el Departamento de Electrónica de la Universidad de Alcalá y las empresas ADIF, RENFE, SEPSA y LOGYTEL.

El proyecto AVIZOR, y consecuentemente el presente trabajo, permitirá a las unidades de tren, en concreto al maquinista, disponer, de forma anticipada a la llegada al punto conflictivo, de la información necesaria que proporcione suficiente capacidad de reacción para evitar un posible accidente. Esta información consiste en imágenes e indicadores de alarma resultantes del análisis automático de imágenes tomadas por cámaras estáticas, dispuestas en los pasos a nivel ferroviarios. Dicho análisis tendrá como objetivo la obtención de aquellas áreas de la imagen que corresponden con objetos, personas u otro tipo de obstáculo que implique un peligro para la circulación del tren o la seguridad de las personas.

De este modo el sistema constará de un módulo de procesamiento de imágenes automático y un módulo de envío de vídeo en tiempo real a través de una red inalámbrica, que conecta el paso a nivel con el tren.

*Palabras clave: ferrocarril, paso a nivel, visión artificial, k-medias, Mahalanobis, transmisión.*



# **Parte II**

## **Memoria**



## Capítulo 2

### Introducción

Los trayectos ferroviarios presentan en sus recorridos multitud de puntos que representan zonas de riesgo potencial para la seguridad de las personas. El presente proyecto se basa en la investigación sobre sistemas de seguridad a implantar en el sector ferroviario con el objetivo de intentar reducir al mínimo los accidentes en puntos donde hay una mayor probabilidad de que se produzca una ocupación de las vía férreas (pasos a nivel, zonas de desprendimiento, zonas de tránsito de personas, etc).

Para cumplir con este objetivo se amplían los elementos que intervienen en la seguridad que hasta la actualidad han sido utilizados, principalmente, mecanismos de señalización y protección en las zonas próximas a las vías; añadiendo también información en tiempo real de la presencia de objetos en los pasos a nivel o similares y con la suficiente antelación como para que el maquinista pueda reaccionar a tiempo, aumentando con ello la seguridad del tráfico ferroviario.

Para que el maquinista pueda influir en la seguridad de una manera efectiva, necesita tener con suficiente antelación, debido a las grandes distancias de frenado que necesitan los trenes para su completa detención, información de los posibles riesgos o peligros que pueden generar un accidente, de manera que pueda realizar las acciones correctoras pertinentes para que el accidente no llegue a producirse. La información que éste va a disponer sobre el punto conflictivo será una imagen en tiempo real, y de manera continua, de lo que esté ocurriendo en dicho punto, por lo que podrá comprobar si existe una situación de riesgo o no.

De manera adicional, con el objetivo de añadir más seguridad a la circulación, se ha investigado e implementado un sistema de tratamiento inteligente de vídeo, cuya misión es la detección automática de posibles elementos que pueden provocar un accidente (con peligro tanto para el personal en tierra, como puede ser un atropellamiento, como en el tren, como puede ser un descarrilamiento). En el caso de que se produzca una alarma (presencia de objeto en la zona a vigilar) se avisa también al maquinista. Esta funcionalidad permitirá simplificar las acciones que deba de realizar el maquinista, así como eliminar posibles errores humanos debido a la no identificación del riesgo en la imagen, despistes, etc.

El presente proyecto final de carrera se encuentra enmarcado dentro del proyecto de investigación "Sistema de supervisión de zonas de riesgo potencial para la circulación ferroviaria y seguridad de las personas (AVIZOR)", financiado por el Ministerio de Fomento español y en el cual colaboran como entes investigadores, el Departamento de Electrónica de la Universidad de Alcalá y las empresas RENFE (Red Nacional de los Ferrocarriles Españoles), ADIF (Administrador de Infraestructuras Ferroviarias), SEPSA (Sistemas Electrónicos de Potencia) y LoGyTel (Logística y Telecomunicación S.L). El proyecto AVIZOR tiene como objetivos añadir sistemas inteligentes de sensado y procesamiento para el aumento de la seguridad en pasos a nivel ferroviarios. En AVIZOR se contemplan tareas de investigación y desarrollo orientadas a:

1. Diseñar un sistema, a ubicar en los puntos de tierra a vigilar, capaz de obtener información precisa para las diferentes condiciones climatológicas y de iluminación natural.
2. Un sistema de procesamiento de imágenes en tiempo real orientado a la detección de obstáculos que puedan suponer un peligro dentro de la zona a vigilar.
3. Un sistema de comunicaciones tierra-tren en tiempo real y con alta fiabilidad que garantice la transmisión de imágenes y posibles alarmas de detección de objetos a una distancia suficiente como para que el maquinista pueda reaccionar.

La figura 2.1 representa el diagrama general de la configuración a ubicar en los pasos a nivel:

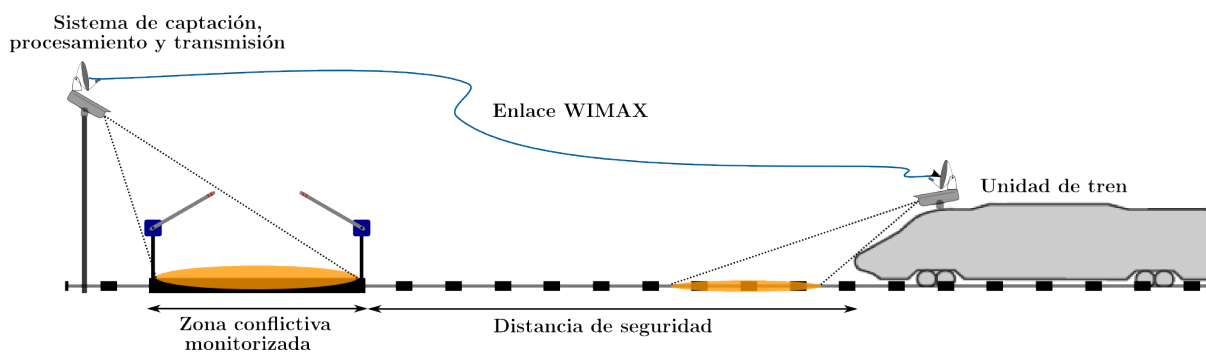


Figura 2.1: Arquitectura general del sistema en la zona vigilada.

Como se comentará en posteriores capítulos, el presente proyecto final de carrera se centra en ofrecer una solución en tiempo real a la detección de obstáculos en vía, resistente a cambios de iluminación y presencia de sombras, así como un canal de transmisión, basado en el estándar WiMAX, que permite comunicar los resultados y el vídeo en tiempo real a un sistema embarcado en un tren.

## 2.1. Estado del arte

### 2.1.1. Sistemas de detección a ubicar en pasos a nivel

Las cámaras de visión, junto a los sistemas de procesamiento digital de imágenes, constituyen uno de los sistemas más ampliamente utilizados en sistemas de vigilancia, ya que pueden aportar una información amplia sobre los elementos detectados e incluso avisar, de forma automática, de la presencia de nuevos objetos dentro de la escena.

Los sistemas de vídeo-vigilancia se pueden clasificar en tres tipos:

1. Aquellos destinados a suministrar información visual a un operador humano, el cual debe interpretar, según su criterio, el estado del entorno vigilado sin usarse ningún tipo de procesamiento.
2. Los sistemas totalmente automáticos, en los que un sistema inteligente de procesamiento analiza la imagen en cada instante de tiempo e infiere la información importante de la misma.
3. Los sistemas semiautomáticos o de ayuda, en los que un sistema inteligente envía el resultado de su procesamiento para su supervisión por un humano. Dichos sistemas de ayuda son importantes para reducir la atención del usuario, sin reducir su eficiencia.

El presente proyecto propone un sistema de ésta última clase, en el que la solución propuesta avisa al maquinista de un tren de los posibles obstáculos que aparecen en un paso a nivel cercano. En estos casos la supervisión humana es actualmente imprescindible, dada la complejidad de las situaciones que se pueden plantear y la experiencia necesaria para la toma de decisiones.

#### 2.1.1.1. Sistemas de visión artificial

Dentro de los sistemas de visión artificial destinados a la detección de la presencia de nuevos objetos dentro de una escena, existen numerosos proyectos y propuestas. En [Vazquez et al., 2004] se propone el uso de dos cámaras, y mediante el análisis de componentes principales (PCA) se detectan objetos en movimiento. Aplicando PCA se crea un modelo del fondo, de forma que las nuevas imágenes captadas de la escena se comparan con dicho modelo, detectando, así, la presencia de nuevos objetos en movimiento en el área analizada.

En [Ota, 2004] se propone el uso de dos cámaras para la detección de objetos en pasos a nivel, y mediante técnicas estéreo detectar todo tipo de obstáculo. Una de las ventajas de esta propuesta frente a los sistemas basados en radar (expuestos en el siguiente apartado), es la posibilidad de detectar objetos de pequeño tamaño. Aunque la cámara es el sensor más parecido al ojo humano, presenta algunos problemas que a veces hacen descartar su uso. Uno de ellos es el gran volumen de datos que se maneja, lo cual genera en ocasiones tiempos de proceso elevados, perdiéndose eficiencia si se desea una respuesta inmediata del sistema de vigilancia. Otro problema vinculado a este sensor es la necesidad de iluminación adecuada. Por otra parte,

se ve bastante afectado por la meteorología (lluvia, nieve, etc), sobre todo si ésta no permite ver adecuadamente.

Dadas las exigencias de fiabilidad del sistema de visión artificial, se deberá abordar la problemática existente con los cambios de iluminación y variación del modelo de fondo; temas desarrollados ampliamente con anterioridad. En [Finlayson et al., 2002] se emplea un método de eliminación de sombras basado en la obtención de un espacio unidimensional invariante a la iluminación. En [Allili et al., 2007a] se propone un método robusto de segmentación empleando un modelo finito de distribuciones gaussianas generalizado (MoGG). Dicho método tiene la flexibilidad de modelar el fondo del vídeo bajo la presencia de cambios de iluminación y sombras, ofreciendo un sistema eficiente para la segmentación de objetos.

El uso de geometrías estereoscópicas se hace necesario para determinar parámetros de distancia y altura respecto al plano de referencia (las vías del tren en este caso), esto incorpora en la medida del sistema un dato de ubicación 3D del objeto lo que puede resultar de interés si, por ejemplo, se desea tener una idea aproximada de la altura de los objetos.

### 2.1.1.2. Sistemas de seguridad de tráfico ferroviario

Existen numerosos grupos de trabajo en temas relacionados con la seguridad en el tráfico ferroviario, abarcando muy diversos aspectos del mismo y, sobre todo, incidiendo en sistemas a bordo. La propuesta más actual viene marcada por el proyecto italo-israelí **REOST** (“Railway Electro Optical System for Safe Transportation”) [REO, 2004], cofinanciado por la Unión Europea y algunas empresas privadas del sector ferroviario, con 3.851.247 euros. Según un informe de la propia Unión Europea [P. Bon, 2004], REOST ofrece un nuevo concepto en la seguridad en el ferrocarril, basado en la observación, niveles de alarma y decisión; proporcionando un sistema embarcado en tren para prevenir descarrilamientos u otro tipo de accidentes. En la misma línea que el anterior se enmarca el proyecto **KOMPAS**, financiado por el Fraunhofer Institute for Transportation Systems and Infrastructure Controls Engineering (IVI) [Ruder et al., 2003]. En este caso ubican tres cámaras en la cabina del tren: dos funcionando en estéreo para campo cercano, y otra cámara la usan para analizar el campo lejano. Además incorporan un radar infrarrojo (LADAR) capaz de detectar en un rango de 150 metros. Así mismo utilizan un GPS para informar de la ubicación y velocidad del tren. Según los autores pueden detectar objetos de 0.4x0.4 m de sección a 250 metros de distancia. En [Mockel et al., 2003] proponen la detección de obstáculos mediante cámaras y un LADAR embarcados. Según la información publicada, pueden detectar objetos a 400 metros del tren, y a una velocidad del mismo de 120 km/h. Mediante una cámara teleobjetivo, en [Ukai, 2004] pueden detectar objetos a una distancia de 600 metros del tren. En [Michael Tilleman, 1998] se propone exclusivamente el uso de un LADAR para la detección de objetos. Este es un sistema medio-embarcado, ya que el emisor láser va en el tren, pero usa prolongaciones del láser mediante fibra óptica, a 5 km de la posición del mismo.

En lo concerniente específicamente a la detección de obstáculos en vía, actualmente, los pasos a nivel son las zonas sometidas a los más minuciosos estudios [Safety, 2003] [Okutani et al., 2000], y las técnicas de detección usadas en los mismos se trasladan al resto de áreas. Por otro lado, diversas compañías nacionales e internacionales aportan algunas soluciones, casi siem-



pre basadas en un único tipo de sensores. Así, la compañía General Electric Transportation Rail [Transportation, 2005], propone un sistema radar para detectar objetos en pasos a nivel, basado en varias balizas emisoras y receptoras.

La empresa española “Logytel” propone un sistema basado en mallas de fibras ópticas, instaladas en la parte superior y lateral de los pasos elevados y los túneles. La rotura de un módulo de la malla provoca la rotura del cable de fibra óptica, y el corte de la continuidad óptica provoca la generación de alarma. Su principio de funcionamiento es similar a la detección basada en hilos conductores. Una de las principales ventajas de este sistema es su inmunidad a la climatología y a las interferencias electromagnéticas. Debido a que el principio de detección se debe a la rotura de la malla, cada vez que se detecta la caída de objeto se necesita realizar la reparación del sistema sensorial. Por otra parte, permite detectar que ha caído objeto, pero no que el mismo se encuentra sobre las vías.

### 2.1.2. Sistemas de transmisión inalámbrica de vídeo

Los sistemas de vídeo-vigilancia inalámbricos basados en IP han alcanzado un gran nivel de desarrollo gracias a su flexibilidad y facilidad de instalación. Además, la incorporación de protocolos IP hace que las imágenes puedan ser recibidas virtualmente en cualquier sitio que disponga de conectividad a Internet.

Tradicionalmente los sistemas analógicos de transmisión estaban constituidos de una cámara, un transmisor y un receptor. Cuando la distancia entre el emisor y receptor es muy grande se deben instalar repetidores que encarecen la solución. La conversión a sistemas digitales elimina estos problemas, siempre que se tenga al alcance una red de transmisión IP o se disponga de tecnología que permita desplegar redes de forma fácil y rápida, a un coste reducido (WiMAX o WiFi).

La mayoría de las redes inalámbricas IP para vídeo vigilancia disponibles en el mercado se basan en el estándar 802.11. Sin embargo su aplicación al problema planteado en este proyecto es poco factible debido a su limitado alcance (sobre 100 metros) y a su falta de capacidad para garantizar el ancho de banda necesario en cualquier circunstancia. Para resolver estas carencias se puede optar por los nuevos estándares 802.11n, por la tecnología WiMAX (802.16e,m) o por el empleo de redes celulares de banda ancha, algunas de las cuales ofrecen cobertura en las vías del tren: GPRS, UMTS, HSDPA, HSUPA [Goodman and Myers, 2005].

Sin embargo, antes de optar por una solución, es necesario analizar la velocidad máxima a la que el receptor puede moverse para recibir correctamente la información. En este sentido la tecnología 802.16e,m está limitada al entorno de los 250Km/h [WiMAX-Part, 2006], mientras que las redes celulares permiten velocidades entorno a los 200 Km/h [WiMAX, 2006b].

Otro aspecto a tener en cuenta es la disponibilidad de equipamiento en el mercado. Los equipos para los estándares WiMAX móvil están en plena fase de certificación, mientras que los equipos para redes celulares ya están disponibles.

Finalmente el despliegue de la red es otro aspecto clave. Las redes WiMAX se pueden desplegar de manera rápida y de forma propietaria empleando las bandas no licenciadas [Consulting, 2005b], lo que se ajusta perfectamente al problema planteado en el proyecto AVIZOR, [WiMAX, 2006a]. Por otro lado las redes celulares ya desplegadas pueden ofrecer coberturas deficientes en los trazados ferroviarios, lo que en ciertos pasos a nivel las haría inservibles. No obstante, en aquellos puntos donde la cobertura fuera suficiente permitirían una conexión tanto al tren como a un centro de supervisión en tierra, empleando las redes IP del proveedor de acceso.

## 2.2. Objetivos propuestos

En la figura 2.2 se muestra la arquitectura global del sistema propuesto en el proyecto AVI-ZOR. Se identifican cuatro grandes bloques o módulos funcionales.

**Módulo de visión (MOVI)** En él se incluyen todos los aspectos relacionados con la captación y procesamiento de imágenes (donde también interviene el módulo de procesamiento).

**Módulo de procesamiento (MOPRO)** Se encarga de realizar todos los procesos relacionados con el módulo de visión (procesamiento de imágenes, activación/desactivación de iluminación, etc.), gestión del módulo de comunicaciones tierra-tren y todos los aspectos relacionados con otros sensores periféricos (señal de subida/bajada de barreras, etc.).

**Módulo de comunicaciones tierra tren (COTIERRA)** Se encarga de todos los aspectos relacionados con el envío de información (imágenes y alarmas) vía radio entre tierra y tren, utilizando la norma de transmisión WiMAX (Worldwide Interoperability for Microwave Access - Interoperabilidad Mundial para Acceso por Microondas).

**Módulo de comunicaciones a bordo del tren (COTREN)** Es el encargado de recibir la información enviada por el módulo COTIERRA.

También se tendrá en consideración el **módulo de alimentación para equipos de tierra** que será el encargado de suministrar la energía que demanden los diferentes sistemas ubicados en tierra (paso a nivel). Se contempla la necesidad de incluir como entrada al sistema la información de inicio de bajada y subida de barreras.

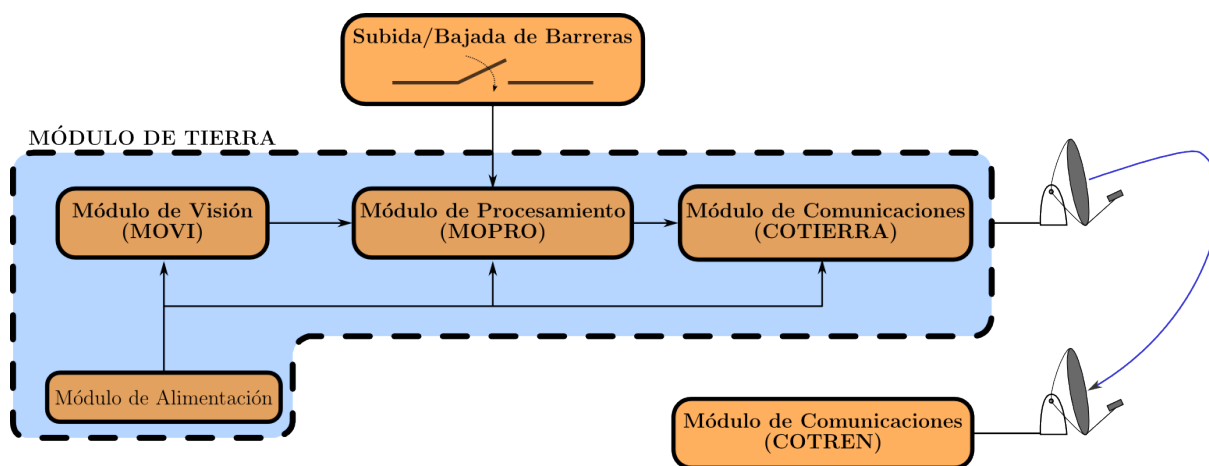


Figura 2.2: Diagrama modular del sistema.

El presente proyecto final de carrera aporta una solución a los 3 primeros módulos enmarcados dentro del sistema de tierra; proporcionando una solución para cada uno de ellos.

### 2.2.1. Especificaciones del proyecto

Las informaciones enviadas desde tierra al tren son básicamente dos, tal y como se muestra, de manera ilustrativa, en la figura 2.3:

1. La imagen del entorno a vigilar con resolución espacial de 1-CIF (Common Intermediate Format, corresponde a 352 x 288 píxeles) y velocidad de adquisición de 12,5 fps y una señal de alarma (coordenadas del centroide del objeto detectado y delimitación de la zona ocupada por el mismo en el plano imagen). El inicio del envío de imágenes de resolución 1-CIF y 12,5 fps coincidirá con la notificación (del equipo de comunicaciones embarcado al de tierra) de que el tren se encuentra dentro de la zona de cobertura, y se finalizará cuando el equipo de tierra detecte que el tren ya no se encuentra dentro de la zona de cobertura de la red de comunicaciones.
2. En lo que se refiere al envío de la información de detección de objeto (alarma) se realizará su envío durante el intervalo de tiempo que va desde que se inicia la bajada de barreras hasta que se inicie la subida de las mismas.

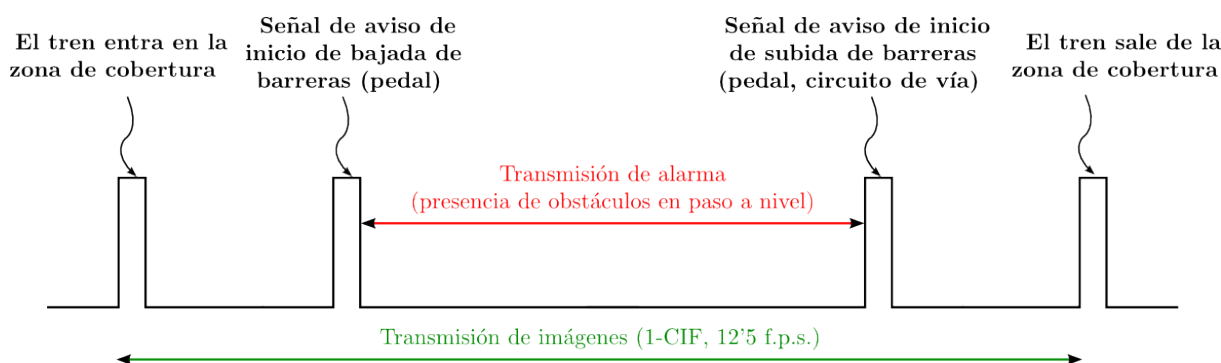


Figura 2.3: Diagrama temporal de emisión de alarma y señal de vídeo al tren.

La funcionalidad que deben cumplir los bloques de la figura 2.2 y que son tratados en el presente trabajo, se detalla a continuación.

**MÓDULO DE VISIÓN (MOVI)** Con carácter general indicar que el módulo de visión (MOVI) se encarga de captar y procesar las imágenes procedentes del entorno a vigilar.

Este módulo entregará una imagen del entorno con resolución espacial de 4-CIF y velocidad de adquisición de 25 fps, la misma imagen, con resolución 1-CIF y 12,5 fps, y las coordenadas píxelicas del objeto detectado (enmarcando si fuera necesario la zona del objeto).

El “módulo de visión” (MOVI) responde al diagrama de bloques mostrado en la figura 2.4 y tiene como objetivo captar información visual del paso a nivel, procesarla y detectar la presencia/ausencia de objetos dentro del área a vigilar. En el caso de detectar objetos

dentro de dicho área se generará el correspondiente aviso, indicando la zona de la imagen en la que se encuentra el objeto detectado. La resolución espacial de las imágenes captadas será de al menos 4-CIF y con 25 fps. Las imágenes a enviar al tren tendrán una resolución espacial de 1-CIF y una frecuencia de transmisión de 12,5 fps.

El MOVI contempla la eliminación de sombras en el suelo (que pudieran producir falsas alarmas) y la calibración necesaria para tal efecto que consiste en el cálculo de la homografía del suelo visto por dos cámaras.

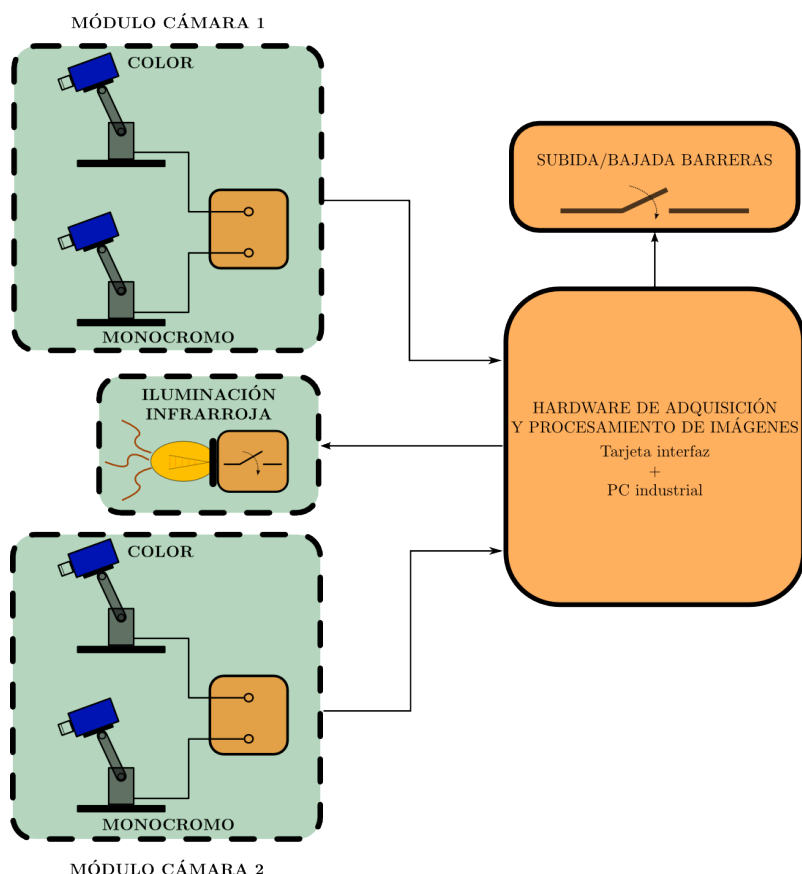


Figura 2.4: Estructura de conexión del MOVI.

Para llevar a cabo estos objetivos, el MOVI estará formado por al menos dos pares de cámaras (módulo cámara 1 y módulo cámara 2), un sistema de iluminación IR, una entrada de detección de activación de pedal y el correspondiente hardware de procesamiento de imágenes. La necesidad de utilizar en todo momento al menos dos cámaras por cada paso a nivel es con el objeto de facilitar la eliminación fiable de sombras.

Dado que las cámaras deben ser sensibles tanto al espectro visible (visión diurna) como al infrarrojo (IR) (visión nocturna), es necesario estudiar las diferentes alternativas comerciales que den respuesta satisfactoria a este requerimiento, teniendo además presente aspectos económicos, consumo, protocolos de comunicaciones, sincronismo externo, etc.

Por tanto, para implementar cada módulo de cámara, en principio caben tres alternativas básicas:

- Utilizar cámaras con doble CCD (uno sensible al IR y otro de color).
- Utilizar cámaras en blanco y negro (B/N), ya que éstas suelen tener una buena respuesta en la longitud de onda de los focos de iluminación IR comerciales (850 nm).
- Utilizar un par de cámaras de color y otro de B/N con buena sensibilidad al IR, para día y noche, respectivamente.

siendo esta última, la adoptada para la implementación del presente proyecto.

Con carácter general las cámaras deberán cumplir con los siguientes requisitos:

- La óptica asociada a cada cámara deberá ser elegida en función del tamaño de la escena a vigilar, ubicación física de las cámaras (altura y distancia al borde de la escena a vigilar). Por tanto, para cada instalación puede ser necesario realizar algún cambio, por lo que se debería disponer de algún método de ajuste (bien zoom/focus) o intercambio de lentes con rosca C/CS.
- Cámara día/noche (bien con doble CCD o CCD con buena sensibilidad al IR) que proporcione imágenes sin comprimir para poder procesarlas en un PC equipado con Linux (Ubuntu + OpenCV).
- El bus de comunicaciones deberá ser preferiblemente firewire (bien soportado en Linux y opencv), si bien también podrá ser GigE (Gigabit Ethernet).
- La resolución espacial debe ser como mínimo 4-CIF.
- La velocidad de adquisición debe ser como mínimo 25 fps.
- Las cámaras deben disponer de algún método de sincronismo, para sincronizar la captura de imágenes de ambas.
- Así mismo las cámaras deberán poder funcionar en condiciones ambientales adversas (IP65 y margen de temperatura de -25 a 60 C). Si fuera necesario se ubicarán dentro de carcasas debidamente acondicionadas para funcionamiento en exteriores.

Las características concretas de la óptica y cámara (tamaño de sensor, distancia focal, amplitud de campo FOV, etc ) serán fijadas a priori.

Dentro del MOVI se contempla también la medida de la intensidad luminosa ambiental, a partir de la información proporcionada por las cámaras, para la detección día/noche y en función de ello se desactivará/activará la iluminación IR y se procesará la información visual captada por el “módulo de cámara” que en cada caso proceda (caso de utilizar dos sensores CCD para cada módulo de cámara).

Se incluye dentro del MOVI un sistema de iluminación IR que se activará cuando las condiciones ambientales de iluminación así lo exijan, garantizando la correcta visión nocturna. La longitud de onda y la potencia del sistema de iluminación IR serán definidos en

función de las características de las cámaras seleccionadas.

La información proporcionada por los “módulos de cámaras” será enviada al “módulo de procesamiento” vía cable (si fuera necesario se incluirán los elementos necesarios que garanticen una correcta transmisión de las señales de vídeo entre ambos módulos). De igual forma se transmitirán las señales de sincronismo entre todas las cámaras y “módulo de procesamiento”.

**MÓDULO DE PROCESAMIENTO (MOPRO)** El módulo de procesamiento tiene como funciones básicas las siguientes:

- Realizar la captación y procesamiento de las imágenes procedentes de los dos módulos de cámaras.
- Generar el stream de vídeo a enviar al tren a partir de las imágenes con resolución espacial 1-CIF y 12,5 fps.
- Generar la señal de aviso de presencia de objeto (coordenadas en el plano imagen del centroide del objeto detectado).
- Control de iluminación IR en función de las condiciones de iluminación ambiental.
- Conexión al sistema de comunicaciones (COTIERRA).

**MÓDULO DE COMUNICACIONES TIERRA-TREN** El equipo en tierra (COTIERRA) está formado por una estación base y una antena, que se ubicarán en el paso a nivel. El equipo en el tren (COTREN) está constituido por una unidad de abonado, con antena integrada.

La estación base incluye una unidad interna y una externa. La interna se conecta con la red de tierra mediante una interfaz estándar Ethernet 10/100 BaseT (RJ-45) y la unidad externa, montada sobre un mástil, se conecta con la unidad interna mediante un cable CAT-5.

La unidad de abonado, aunque no se aborda en este documento, incluye una unidad interna con un único puerto de datos, que se conecta con la red del tren mediante una interfaz estándar Ethernet 10/100 BaseT (RJ-45), y una unidad externa de montaje sobre mástil y antena integrada, que se conecta con la unidad interna mediante un cable CAT-5.

Para la frecuencia de operación se ha elegido la banda no licenciada de 5.47 GHz. La tecnología inalámbrica es una solución basada en los estándares WiMAX, con ciertas modificaciones del fabricante, si bien, mantiene la compatibilidad con los futuros equipos WiMAX, todavía en fase de desarrollo.

Las prestaciones que proporciona este sistema son las siguientes:

- Transmitir las imágenes captadas por el sistema de visión y la información sobre las alarmas.
- En lo que se refiere a las imágenes transmitidas éstas deberán tener una resolución espacial de 1-CIF y con 12,5 fps.
- La velocidad máxima del tren será acorde con la tecnología (se estima una velocidad máxima de 200 km/h).



## 2.3. Descripción de las tareas realizadas

### 2.3.1. Tareas contempladas en el proyecto AVIZOR

En lo que sigue se hará una descripción de las distintas tareas en las que se divide el proyecto AVIZOR, en el cual se encuentra enmarcado el presente trabajo. La fijación de tareas se ha realizado atendiendo a los diferentes módulos que se pretenden desarrollar y de acuerdo a los objetivos anteriormente establecidos. Para cada tarea se detallan las actividades a realizar y la metodología que se empleada.

#### **Tarea 1: Estructuración del Proyecto y Análisis de requisitos del Sistema.**

##### Actividades:

- Análisis del Proyecto.
- Distribución de tareas y responsabilidades.
- Definición de plazos.
- Análisis de los requisitos que debe cumplir el Sistema.

##### Resultados previsibles:

Definición clara de las tareas que debe de realizar cada empresa/institución, fijación de los plazos para la realización de cada tarea, con el objetivo de finalizar el proyecto en el plazo de tiempo fijado inicialmente.

#### **Tarea 2: Sistema de Tierra. Definición de la arquitectura global y los requerimientos del sistema de vigilancia de pasos a nivel.**

##### Subtareas:

- Estudio de los requerimientos generales del sistema (atendiendo a las exigencias de los gestores de infraestructuras ferroviarias).
- Definición global del sistema y de los diferentes elementos que lo componen. Planteamiento de propuesta.
- Definición de las características que deben cumplir los diferentes subsistemas.
- Planificación de actividades y plan de documentación y gestión del proyecto.

Objetivos:

Definir la configuración general del proyecto, planificar las diferentes actividades, seguimiento y documentación a generar.

Resultados previsibles:

Descripción detallada del proyecto y planificación de las diferentes tareas.

**Tarea 3: Sistema de Tierra. Estudio y desarrollo del módulo sensorial de visión a ubicar en los pasos a nivel.**Actividades:

- Elección de cámara y óptica asociada. (Se evaluarán los diferentes tipos de sensores comerciales con un interfaz de gestión parametrizable externamente)
- Elección del sistema de iluminación infrarroja.
- Estudio de algoritmos de procesamiento de imágenes orientados a la detección de nuevos objetos dentro de la escena.
- Posicionamiento óptimo de los componentes (cámara, óptica, anillo de infrarrojos, ángulo de captación, ubicación más adecuada, etc).
- Realización de pruebas prácticas.

Metodología:

- Estudio teórico-práctico de las necesidades y exigencias de los distintos sistemas del sistema sensorial. Se determinará la cámara, óptica y sistema de iluminación infrarroja, así como el programa de adquisición de imágenes.
- Estudio de las diferentes alternativas de procesamiento de imágenes orientados a la detección de nuevos objetos dentro de una escena. Finalmente se realizarán pruebas prácticas que demuestren la validez de las propuestas realizadas.

Resultados previsibles:

Implementación de un sistema sensorial de visión capaz de detectar la presencia de nuevos objetos dentro de la escena de interés.

**Tarea 4: Sistema de Tierra. Estudio y diseño del sistema inalámbrico de banda ancha para transmisión de imágenes.**

En esta tarea se analizarán las tecnologías inalámbricas de banda ancha que posibiliten la transmisión al tren de las imágenes procedentes de la cámara instalada en el paso a nivel. Una vez completado el estudio, se escogerá la tecnología que ofrezca mejores prestaciones en alcance, velocidad del tren y coste y se diseñará el sistema de transmisión.

Actividades:

- Estudio de las tecnologías inalámbricas para instalación de redes propietarias en bandas no licenciadas: 802.11n y 802.11e, m.
- Estudio de las tecnologías inalámbricas basadas en redes celulares con cobertura en la línea de ferrocarril: GPRS, UMTS, HSDPA y HSUPA.
- Diseño del sistema de transmisión de señales de vídeo a un tren en marcha: subsistema de codificación de imágenes a MPEG, transmisión sobre IP y receptor.
- Implementación y pruebas.

Metodología:

Se procederá al análisis sistemático de las tecnologías descritas, valorando especialmente su capacidad para transmitir de forma segura y con los parámetros de calidad adecuados la señal digitalizada de vídeo. Se atenderá especialmente a la cobertura que ofrezcan y a la velocidad que permita el tren. Se continuará con el estudio de los equipos de mercado disponibles para su implantación y el coste que suponen. A la luz de los resultados se escogerá la solución óptima y se implantará un prototipo de pruebas.

Resultados previsibles:

Especificaciones y diseño de un sistema inalámbrico de transmisión de vídeo en entornos ferroviarios a trenes en circulación, con una cobertura que permita anticiparse a posibles problemas en los pasos a nivel.

**Tarea 5: Sistema de Tierra. Desarrollo Hardware del Prototipo.**

En esta tarea comprende el diseño hardware de todo el equipamiento necesario para el Sistema de Tierra, así como la realización del prototipo para poder realizar las pruebas de la funcionalidad del Sistema.

Actividades:

- Estudio de todas las tarjetas electrónicas disponibles, a partir de las cuales se formarán los equipos necesarios a instalar en Tierra, partiendo de los requisitos de las condiciones de trabajo (temperatura, etc).

- Diseño mecánico de los equipos.

#### Metodología:

Análisis de las tarjetas/equipos disponibles teniendo en cuenta los requisitos del Sistema, así como las características de los lugares donde van a ser instalados.

#### Resultados previsibles:

Disponer del equipamiento necesario para poder implementar la funcionalidad del Sistema.

### **Tarea 6: Sistema Embarcado. Análisis de requisitos.**

#### Actividades:

Análisis de los requisitos que debe cumplir el Sistema Embarcado.

#### Resultados previsibles:

Comprensión de toda la funcionalidad que debe de realizar el Sistema Embarcado, teniendo en cuenta su iteración con el Sistema de Tierra y el maquinista.

### **Tarea 7: Sistema Embarcado. Desarrollo Hardware del Prototipo.**

Esta tarea comprende el diseño hardware de todo el equipamiento necesario para el Sistema Embarcado, así como la realización del prototipo para poder realizar las pruebas de la funcionalidad del Sistema.

#### Actividades:

- Diseño electrónico de todas las tarjetas electrónicas, a partir de las cuales se formarán los equipos necesarios a instalar en el Sistema Embarcado, partiendo de los requisitos de las condiciones de trabajo (temperatura, vibraciones, etc).
- Diseño mecánico de los equipos, teniendo en cuenta los requisitos de espacio limitado que existen en trenes ya construidos.

#### Metodología:

Diseño de las tarjetas/equipos teniendo en cuenta los requisitos del Sistema, así como las características de los trenes donde van a ser instalados.

Resultados previsibles:

Disponer del equipamiento necesario para poder implementar la funcionalidad del Sistema.

**Tarea 8: Sistema Embarcado. Implementación de la funcionalidad (comunicaciones, visualización, grabación, etc).**

Esta tarea comprende el diseño de la funcionalidad (principalmente realizada por software) del Sistema Embarcado.

Actividades:

- Comunicación con el Sistema de Tierra.
- Diseño de la interfaz hombre-máquina.
- Diseño de algoritmos de grabación de vídeo.

Metodología:

- Definición de la información a mostrar al maquinista.
- Definir el modo de mostrar la información al maquinista, a través de un Terminal de Cabina, se centrará el diseño en conseguir una interfaz amigable y que requiera la mínima intervención del operador, maximizando la cantidad de información mostrada sin aumentar la carga de trabajo del operador del sistema.
- Estudio de las tecnologías electrónicas de almacenamiento, cumpliendo con los estándares de seguridad del entorno ferroviario.
- Simulación de las prestaciones del sistema (horas de grabación, calidad de imagen, etc).
- Realización de pruebas prácticas.

**Tarea 9: Sistema Embarcado. Implementación de la funcionalidad del Sistema Embarcado. Cámara Frontal.**Actividades:

Grabación de las imágenes recogidas por la cámara frontal instalada en el tren.

Metodología:

- Selección de la cámara y óptica que mejor se adapten a las condiciones de funcionamiento.
- Selección de las técnicas de compresión de vídeo.

- Simulación de las prestaciones del sistema (horas de grabación, calidad de imagen, etc).
- Realización de pruebas prácticas.

### **Tarea 10: Pruebas de Validación del Sistema completo en un Banco de Pruebas.**

#### Actividades:

Realizar las pruebas de validación del Sistema completo (Sistema de Tierra y Sistema Embarcado) en un Banco de Pruebas, verificando que cumple con los todos los requisitos marcados para el Proyecto.

#### Metodología:

Se realizarán las pruebas usando los medios necesarios para simular un entorno real de funcionamiento del Sistema en vía.

#### Resultados previsibles:

Comprobación del correcto funcionamiento del Sistema en un Banco de Pruebas.

### **Tarea 11: Pruebas de Validación del Sistema completo en Vía.**

#### Actividades:

Realizar las pruebas de validación del Sistema completo (Sistema de Tierra y Sistema Embarcado) en su entorno de funcionamiento real, verificando que cumple con los todos los requisitos marcados para el Proyecto.

#### Metodología:

Se realizarán las pruebas del Sistema en tren, para ello:

- Se instalará el Sistema de Tierra en un paso a nivel, se escogerá un paso a nivel que tenga unas características especiales donde se pueda comprobar el mayor número de casuística que pueda ocurrir en el día a día, permitiendo probar el Sistema en todo tipo de situaciones.
- Se instalará el Sistema Embarcado en una Unidad de Tren, la elección de Unidad de Tren se realizará en función de tipo de tren (facilidad de instalación del Sistema, velocidad del tren, etc).
- Se realizarán los trayectos por el paso a nivel necesarios para la validación completa del Sistema.

Resultados previsibles:

- Comprobación del correcto funcionamiento del Sistema en vía.
- Para la realización de estas pruebas es necesario disponer de una Unidad de Tren (aportada por RENFE), así como la infraestructura ferroviaria correspondiente (aportada por ADIF).

**Tarea 12: Conclusiones y expresión de resultados.**Actividades:

- Realizar las conclusiones sobre el Sistema durante toda la fase del Proyecto, desde el diseño, pasando por la implementación y terminando en las pruebas de validación.
- Realización de la documentación que expresen los resultados obtenidos.

**2.3.2. Tareas desarrolladas en el presente proyecto**

El presente proyecto comprende, fundamentalmente, las tareas 3 y 4 anteriormente citadas:

1. La concerniente al estudio y desarrollo del módulo sensorial de "Visión Artificial" para la detección de obstáculos.
2. Estudio y diseño del sistema inalámbrico de banda ancha para transmisión de imágenes y mensajes de control.

Los dos capítulos siguientes ahondan en sendas tareas, realizando un estudio exhaustivo de todas las actividades realizadas en el desarrollo y consecución de los objetivos propuestos.

**2.3.2.1. Estructura del documento**

La memoria del proyecto está estructurada en “6” capítulos. A continuación se describe brevemente su contenido:

- **Capítulo 2: Introducción** ⇒ en este capítulo se pretende dejar claros los principales conceptos del presente proyecto final de carrera, de manera que se pueda tener una visión general del proyecto en conjunto.
- **Capítulo 3: Diseño del módulo sensorial de visión a ubicar en los pasos a nivel** ⇒ se realiza un estudio exhaustivo tanto de los conceptos teóricos involucrados como de la solución propuesta para la detección de obstáculos en entornos exteriores.

- **Capítulo 4: Desarrollo del sistema inalámbrico de banda ancha para transmisión de imágenes**  $\Rightarrow$  este capítulo comienza con el estudio de las alternativas tecnológicas existentes para dotar de comunicación entre el módulo de procesamiento y el módulo receptor ubicado en la cabina del conductor del tren. A continuación, se explica de manera detallada y justificada cual ha sido la alternativa adoptada. Se concluye explicando el sistema de compresión de vídeo en tiempo real y el protocolo de transmisión al equipo receptor.
- **Capítulo 5: Implementación del sistema en tiempo real**  $\Rightarrow$  en él se especifica cual ha sido el método seguido en la implementación del sistema para cumplir los requisitos planteados.
- **Capítulo 6: Pruebas y resultados**  $\Rightarrow$  se muestran las pruebas y los resultados más relevantes de la puesta en funcionamiento de la solución implementada.
- **Capítulo 7: Conclusiones**  $\Rightarrow$  se presentan las conclusiones más importantes extraídas de la realización del proyecto final de carrera.



## Capítulo 3

# Diseño del módulo sensorial de visión a ubicar en los pasos a nivel

### 3.1. Introducción

La **Visión Artificial** es la ciencia que estudia, haciendo uso de computadoras, los procesos de obtención, caracterización e interpretación automática de la información procedente de imágenes tomadas del mundo real tridimensional. Por tanto, un **Sistema de Visión Artificial** es el encargado de actuar sobre una representación de la realidad que le proporciona información sobre brillo, colores, formas, etc..

Existe un gran abanico de aplicaciones en las que los sistemas de visión artificial cobran un papel muy importante. Una de las áreas de aplicación es la **Segmentación de Fondo**; proceso que consiste en dividir una imagen digital en regiones homogéneas con respecto a una o más características (como por ejemplo el brillo o el color) con el fin de facilitar un posterior análisis o reconocimiento automático. Localizar la cara de una persona dentro de la imagen de una fotografía o encontrar los límites de una palabra dentro de una imagen de un texto, constituyen ejemplos de problemas de segmentación. En lo que al proyecto se refiere, se emplea el proceso de segmentación para dar solución a la detección de obstáculos dentro de una región, objeto de monitorización. Se pueden distinguir dos líneas de desarrollo dentro de la segmentación de objetos empleando cámaras estáticas en función de la ubicación de las mismas:

- **Interiores** (“Espacios Inteligentes”) ⇒ Se trata de recintos cerrados dotados de “inteligencia” donde existe una interacción entre los usuarios y el propio entorno, con el fin de proporcionarles ayuda en algún tipo de tarea. Para que pueda existir la comunicación bilateral hombre-máquina se necesitan elementos sensores, como por ejemplos cámaras. Se trata de la línea de investigación que mayores avances a tenido en los últimos años debido a su ubicación en recintos cerrados y que permite disponer de un mayor control de las condiciones que afectan al objetivo final en desarrollo.
- **Exteriores** ⇒ Las variantes condiciones del medio externo a las que se encuentran sometidos este tipo de sistemas ha provocado que su uso se haya limitado, hasta un pasado muy reciente, a la seguridad pasiva en las inmediaciones de edificios, zonas de riesgo poten-

cial, etc.; método carente de cualquier tipo de “inteligencia” y dependiente directamente de la supervisión humana. La segmentación de manera robusta en este tipo de ambientes se presenta como un reto dentro de esta disciplina. El presente proyecto se centra en este tipo de entornos y proporciona alternativas viables para su implementación real.

Por tanto, la idea es llegar a un sistema que permita detectar la presencia de obstáculos en una zona delimitada de un entorno exterior. Se tendrá que tener garantía del cumplimiento de las siguientes características:

- **Robustez:** el sistema debe absorber vibraciones en las cámaras, cambios bruscos de iluminación, eliminación de sombras proyectadas no pertenecientes a objetos presentes en vía, ruidos, evolución del modelo de fondo, etc.. Además, el proceso debe adaptarse a los cambios de iluminación sufridos, tanto de día como de noche con iluminación infrarroja.
- **Eficiencia:** se debe soportar procesamiento en tiempo real.
- **Fiabilidad:** los resultados ofrecidos deben tener un alto porcentaje de acierto, haciendo especial énfasis en evitar falsos negativos (no se detecta obstáculo en un paso a nivel cuando en realidad existe).

La segmentación de objetos adquiere gran importancia en multitud de campos donde se aplica la visión artificial, entre los que se pueden citar:

- Sistemas de seguridad y vigilancia.
- Sector industrial  $\Rightarrow$  el proceso de segmentación de objetos puede cobrar gran importancia en líneas de producción automatizadas detectando, por ejemplo, la presencia de objetos en determinadas zonas de interés.
- Sistemas inteligentes de transporte y asistencia a la conducción  $\Rightarrow$  para poder mejorar la seguridad vial se desarrollan sistemas inteligentes capaces de anticiparse a numerosas situaciones de riesgo que pueden provocar accidentes de tráfico, por ejemplo, la detección de personas en pasos de peatones.
- Tecnologías aplicadas a la ayuda a la discapacidad, proporcionando sistemas que faciliten la ejecución de tareas a aquellas personas que tienen algún tipo de limitación.
- Domótica  $\Rightarrow$  se trata de sistemas capaces de automatizar una vivienda, aportando servicios de gestión energética, seguridad, bienestar y comunicación; integrándose la tecnología en el diseño inteligente de un recinto.

En la actualidad, se tiende a conseguir una mayor interacción entre el hombre y la máquina, tratando de independizar los procesos de la supervisión humana y desarrollando sistemas autónomos con capacidad de toma de decisiones y comunicación con el entorno que le rodea.

Este proyecto final de carrera se enmarca en el **área de sistemas inteligentes de transporte y asistencia a la conducción**. Se pretende aportar una solución viable para la detección de obstáculos en pasos a nivel haciendo uso de la información proporcionada por múltiples cámaras.

Los objetos a detectar deben cumplir una serie de requisitos en cuanto a tamaño y ubicación, además deberán ser detectados indistintamente si se encuentran estáticos o en movimiento. A parte de lo anterior, se toman como premisas que el algoritmo debe funcionar en tiempo real y adaptarse tanto a cambios de iluminación como a vibraciones y demás supuestos que engloban la compleja casuística de dichos entornos.

## 3.2. Objetivos y definición del problema

Una de las fases del proyecto desarrollado consiste en el diseño e implementación del “**Sistema de Tierra**”, ubicado en las zonas donde exista un riesgo potencial para la circulación ferroviaria y la seguridad de las personas, que permita la detección eficiente y fiable de obstáculos y la transmisión inalámbrica de dicha información al módulo embarcado en el tren.

El sistema de tierra debe contar con una estructura sensorial empleando cámaras de visión artificial, con ayuda de iluminación infrarroja externa para condiciones de funcionamiento en oscuridad o baja luminosidad. El campo de visión de las cámaras debe incluir las zonas de interés de los pasos a nivel a vigilar, de manera que permita obtener información fiable sobre el estado de los mismos y detectar de forma automática ( y con mínima dependencia de las condiciones atmosféricas) la presencia de objetos dentro de la región de interés. Dicha estructura se presenta en la figura 3.1 mostrada a continuación.

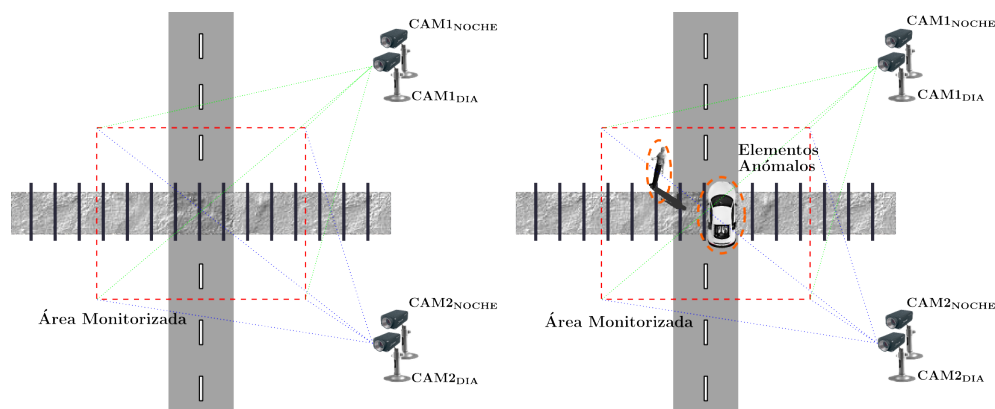


Figura 3.1: Monitorización de elementos anómalos.

Desde un punto de vista funcional del sistema de visión, se establecen los siguientes objetivos a cumplir:

- Tamaño mínimo del objeto detectable por el sistema para que el mismo genere una alarma, será del orden de 50x50x50cm.
- Tiempo de detección automático máximo de objetos se establece en menos de 1 segundo.
- Diseño de un sistema con mínimas tareas de mantenimiento.

En este proyecto se abordan los módulos relativos al sistema de Tierra, los cuales se amplían con detalle a continuación.

**Sistema sensorial.** Existirán cuatro cámaras de visión (dos para funcionamiento de día y dos para modo noche) cubriendo un área que incluya la zona vigilada en el entorno del paso a nivel. Los objetivos generales que se persiguen con este módulo, se pueden resumir en:

- a) Diseñar un sistema de apoyo a la seguridad en vías férreas, basado en técnicas de procesamiento digital de imágenes, que permita detectar si en los pasos a nivel existen objetos que puedan suponer un peligro para la circulación.
- b) Desarrollo de algoritmos que permitan detectar de forma fiable los posibles objetos, minimizando tanto los fallos de detección como la generación de falsas alarmas; y todo ello teniendo presentes las condiciones ambientales de funcionamiento, el tamaño mínimo de los objetos a detectar y las condiciones extremas de incorporación de éstos al entorno de las vías.

**Unidad de proceso.** Hay que tener en cuenta que el módulo sensor trabajará en exteriores, viéndose afectado por diferentes circunstancias: climatología adversa, sol, lluvia, suciedad en los sensores, etc.. Todas estas afectaciones del sistema de medida se verán minimizadas por el procesamiento de visión artificial realizado.

Los objetivos más concretos en torno al sistema de procesamiento de imágenes del entorno, están relacionados todos ellos con algoritmos de procesamiento digital de imágenes, orientados en este caso a la detección de nuevos objetos que se incorporan a una escena y que pueden permanecer estáticos o en movimiento, a saber:

1. Crear un modelo del fondo en ausencia de objetos y que refleja en la medida de lo posible las diferentes condiciones ambientales y de iluminación.
2. Diseñar una estrategia que, a partir del modelo del fondo, permita detectar la presencia de nuevos objetos. Esta estrategia debe permitir la detección de objetos y su ubicación aproximada dentro de la zona a vigilar.
3. Diseño de técnicas que permitan una segmentación adaptativa en función de las diferentes condiciones de funcionamiento del sistema.
4. Proponer técnicas de validación de candidatos a objetos, que minimicen los fallos de detección y la generación de falsas alarmas. Para ello se deberán proponer y fusionar estrategias que aporten información sobre las causas más importantes que pueden influir en la fiabilidad de la detección. Entre ellas se pueden indicar:
  - Redundancia en el sistema de detección y estimación aproximada de las coordenadas 3D de los candidatos, eliminación de candidatos en función del tamaño, aspectos relacionados con la evolución temporal de los objetos y zonas de entrada, más probables, de los objetos en el paso a nivel.
  - Validación experimental y en entornos exteriores (similares a los de las vías férreas) de todas las propuestas realizadas.
  - Por otra parte es importante tener presente las condicionantes que impone un sistema que debe trabajar en entornos exteriores, y por tanto estar sometido a

variaciones de iluminación y de condiciones climatológicas. Además habrá que prestar especial atención a la presencia de brillos, sombras, etc., que pueden tener incidencia en la fiabilidad de la detección.

La problemática existente en este tipo de entornos hace que el sistema inicialmente descrito se vea ampliado con otros mecanismos que traten de mitigar o absorber los efectos perniciosos que tienen los entornos exteriores sobre la detección final. La tabla 3.1 muestra, de manera resumida, un listado de los requisitos más destacables así como las soluciones planteadas.

| Requisitos                                                                   | Solución Propuesta                                                                                           | Resumen                                                                                                                                                                                                                                                                                                                             |
|------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Detección de obstáculos estáticos y en movimiento                            | Algoritmo k-medias adaptativo                                                                                | Modelo del fondo mediante una modificación del algoritmo k-medias para proporcionar el carácter adaptativo requerido.                                                                                                                                                                                                               |
| Eliminación de sombras proyectadas sobre el área de interés                  | Se aplican técnicas de visión estéreo o espacio invariante a la iluminación                                  | <p>Dos alternativas:</p> <ul style="list-style-type: none"> <li>■ Se eliminan las sombras dada su pertenencia al plano del suelo. La transformación de la imagen respecto al plano del suelo es conocida en base al cálculo previo de la homografía.</li> <li>■ Se trabaja sobre un espacio invariante a la iluminación.</li> </ul> |
| Independencia ante cambios de iluminación y variaciones en el medio exterior | Fusión del algoritmo k-medias adaptativo y visión estéreo o creación del espacio invariante a la iluminación | <p>Se plantean dos alternativas:</p> <ul style="list-style-type: none"> <li>■ Fusión de la propuesta de algoritmo k-medias adaptativo junto con las técnicas de visión estéreo.</li> <li>■ Empleo de un espacio invariante a la iluminación.</li> </ul>                                                                             |
| Mínimas tareas de mantenimiento                                              | Actualización de fondo y conmutación día-noche                                                               | Se plantea un sistema adaptativo mediante la inclusión de un mecanismo de actualización periódico del modelo de fondo y la conmutación entre cámaras de color y monocromáticas con iluminación infrarroja en función del modo de funcionamiento diurno o nocturno, respectivamente.                                                 |
| Ejecución en tiempo real                                                     | Optimización de código                                                                                       | Se aplican diversas alternativas para la optimización de código C y se plantea un esquema de procesamiento concurrente empleando múltiples hilos; específicos para cada tarea.                                                                                                                                                      |

Tabla 3.1: Requisitos del sistema de visión y soluciones propuestas.

### 3.2.1. Resumen de la solución propuesta

Con la información suministrada por las etapas de procesamiento, se realiza la validación de las medidas, se indica la existencia de objeto y se comunica al tren (u otros posibles nodos) la presencia de objetos dentro del área de detección. La estructura del sistema sensorial propuesto, la división en diferentes niveles de decisión del sistema y la inclusión de avanzadas técnicas de procesamiento, debe permitir detectar las situaciones mencionadas, y en definitiva evitar las falsas alarmas que otros sistemas con similares objetivos generan.

Para resolver el problema de la detección de obstáculos se propone adoptar como estrategia la obtención de un modelo probabilístico de fondo. Para ello se deberá someter el sistema a una etapa de inicialización o “entrenamiento” en la que se consigue, de manera aproximada y para cada píxel, la distribución estadística como una “Mezcla Multimodal Gausiana”. Se emplea el conocido algoritmo de agrupamiento o *clustering* “K-medias”; proceso mediante el cual se realiza una partición de las “n” observaciones de entrenamiento en “k” grupos o *clusters* de manera que cada observación pertenezca al *cluster* cuya “distancia” sea menor, [Hartigan and Wong, 1979]; se asemeja por tanto al método propuesto en [Allili et al., 2007a], en cuanto que ambos pretenden obtener, entre otros parámetros, el centroide de los *clusters*. Este método genera una clara dependencia con el entrenamiento al que se somete el sistema, por lo que la variabilidad de las situaciones que abarca es demasiado reducida como para ser una solución fiable en entornos exteriores.

Con el objetivo de independizar el sistema del entrenamiento inicial se modifica el algoritmo “k-medias” para que el entrenamiento se convierta en una primera fase de inicialización de clusters para el modelo de fondo, válido solo para un tiempo futuro relativamente corto. A continuación, y con la intención de incorporar un nuevo carácter adaptativo, se propone un mecanismo de actualización de las distribuciones gaussianas creadas en la fase de inicialización. Este método, ejecutado de manera periódica, elabora un mapa de actualización con criterio de decisión en torno a la pertenencia de cada píxel de la imagen de entrada al fondo modelado; permitiendo, en caso afirmativo, incluir dicho punto al *cluster* más oportuno y recalcular sus parámetros de manera adaptativa según varíe el fondo de la escena observada.

Como criterio de segmentación se establece la distancia de Mahalanobis, [Mahalanobis, 1936], al centro de los *cluster*, la cual permite obtener una distancia coherente en función del número y tipo de distribuciones gaussianas del que está formado el modelo probabilístico de fondo.

Se muestra en el diagrama 3.2 el esquema inicial de la solución computacional planteada.

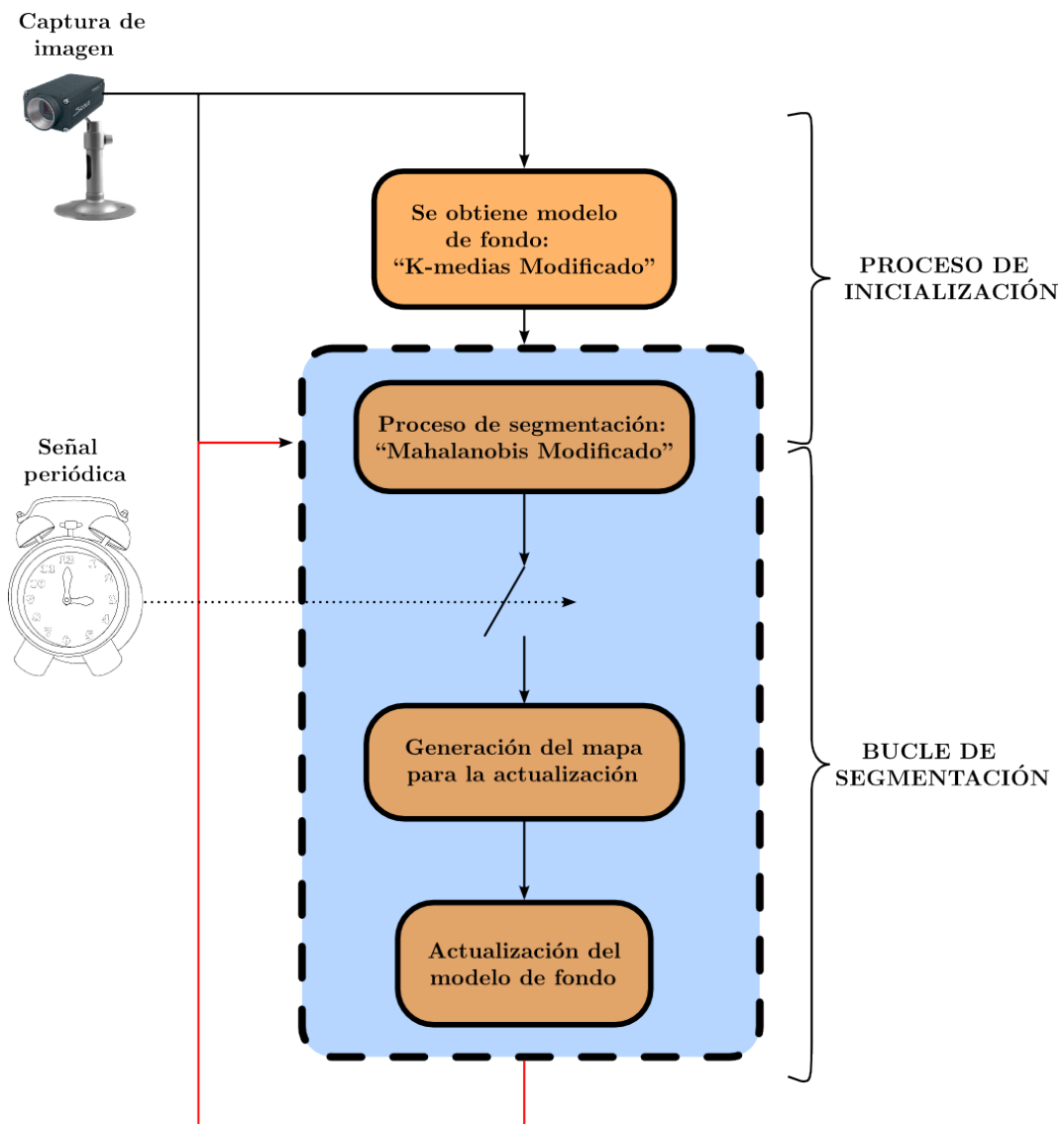


Figura 3.2: Diagrama inicial de la solución propuesta para la detección de obstáculos.

### 3.3. Diagrama general

Una vez analizado el problema en todos sus ámbitos, se propone una solución que pretende abarcar los supuestos más relevantes del proyecto. En la figura 3.3 se plasma el diagrama general de bloques.

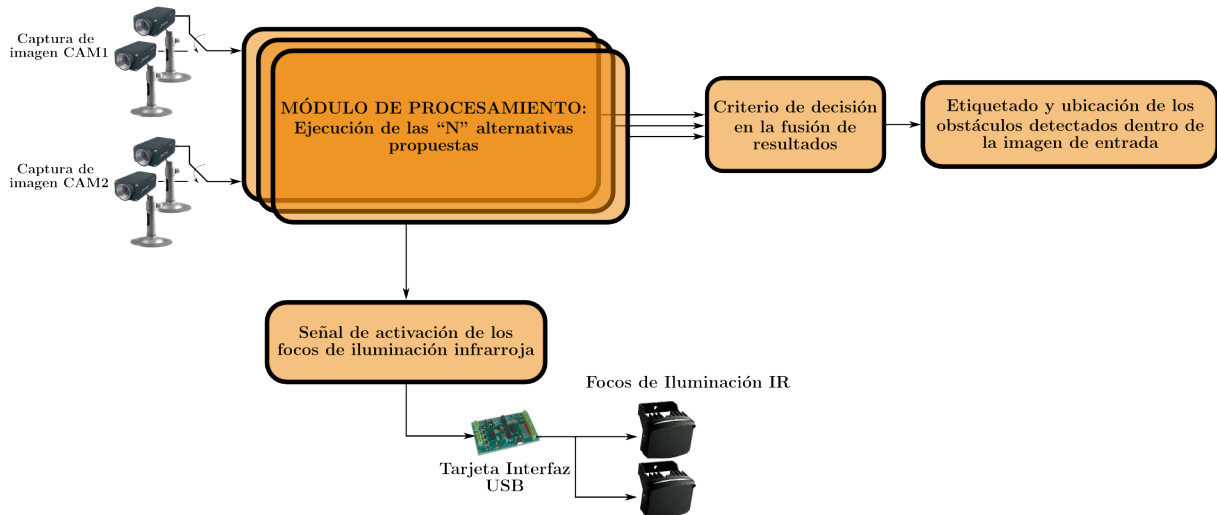


Figura 3.3: Diagrama general de la solución propuesta.

La fusión con otros algoritmos se plantea como una alternativa viable, y en ocasiones necesaria, para ofrecer mayor fiabilidad en los resultados. Se trata de independizar la identificación de obstáculos en varias alternativas y aplicar criterios de decisión cuando se arrojen los resultados.

Para mantener de manera robusta el funcionamiento del sistema, se implementa la conmutación entre el día y la noche:

- Modo diurno  $\Rightarrow$  se captura de las cámaras a color y se procesa con imágenes RGB.
- Modo nocturno  $\Rightarrow$  cuando se conmuta a este modo se deberán encender los focos de iluminación infrarroja y conmutar la captura de las cámaras de color a las monocromáticas.



## 3.4. Aspectos teóricos relacionados con la solución propuesta

En este capítulo se va a realizar un breve repaso a ciertos conceptos fundamentales de la visión por computador que han servido de base para la realización del proyecto y que irán apareciendo a lo largo del texto.

- **Geometría de formación de la imagen** - Se va a explicar el proceso de formación de una imagen en una cámara y como éste puede modelarse matemáticamente mediante una proyección en un plano del espacio 3D. Dicho proceso de correspondencia entre los puntos tridimensionales y los bidimensionales se representa mediante una matriz  $P$ . Además se detalla el método de registro de imágenes propuesto para la obtención de la matriz de homografía.
- **Espacios de color** - Se realizará una exposición de las diferentes bases matemáticas empleadas para representar la información luminosa suministrada por las cámaras. Así, se podrá interpretar y transformar dichas imágenes adecuadamente para la aplicación en cuestión.
- **Segmentación de fondo estático mediante enfoque estadístico** - Se presenta la alternativa adoptada para la detección de obstáculos, empleando un enfoque estadístico del fondo.

De la misma forma, este capítulo tiene como objetivo familiarizarse con la notación que se va a utilizar en el resto del documento.

### 3.4.1. Geometría de formación de la imagen

Antes de poder analizar una imagen, se debe conocer como ha sido formada. La información proporcionada por las cámaras es una representación bidimensional del mundo real 3D. Por tanto, queda patente la necesidad de encontrar la correspondencia geométrica entre puntos de la escena y puntos de la imagen. Frecuentemente, se han empleado dos modelos de óptica geométrica: el modelo pin-hole y el de lentes delgadas. En este trabajo, adoptaremos el primero de los modelos dada su sencillez matemática.

### 3.4.1.1. Cámara pin-hole

**INTRODUCCIÓN** El modelo de perspectiva proyectiva pin-hole (también conocido como perspectiva proyectiva central), propuesto por primera vez en el siglo XV (Brunelleschi), a pesar de su simplicidad, se trata de un modelo matemático conveniente y una buena aproximación en mucho de los casos.

Idealmente, se parte de la consideración de tener un agujero estrecho en un recinto cerrado dispuesto de tal manera que la única luz que lo atraviere alcance el plano imagen; superficie sensible a la luz donde se proyectará la escena. Si el agujero fuese infinitamente estrecho (lo cual es físicamente imposible), tan solo un rayo de luz atravesaría la “caja” para cada punto del plano imagen donde se proyecta la escena. La perspectiva proyectiva crea imágenes invertidas, por lo que es conveniente considerar una imagen virtual en un plano en frente del “pin-hole” a la misma distancia que el plano imagen, de manera que dicha imagen no quede invertida, figura 3.4.

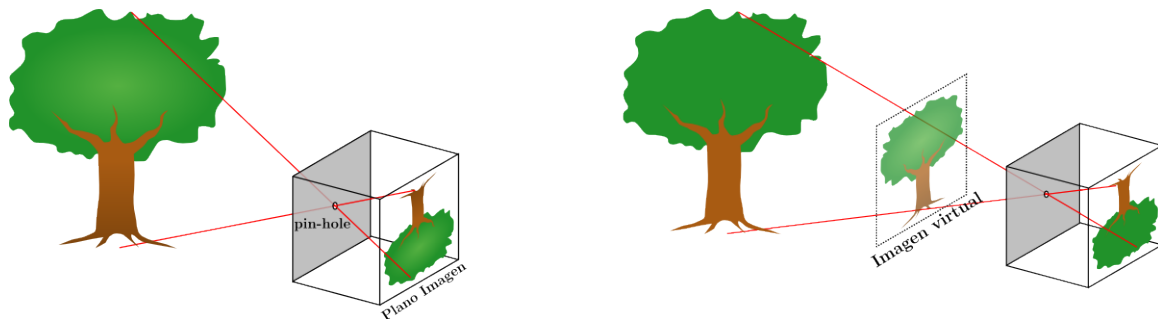


Figura 3.4: Representación de la formación de una imagen en el interior de una cámara pin-hole (izquierda) y el desplazamiento del plano imagen para evitar la inversión de la proyección (derecha).

Sin embargo, el agujero posee una anchura finita y cada punto en el plano imagen corresponderá con un cono de rayos, de manera que el modelo de geometría de la imagen no es estrictamente aplicable para todos los casos.

**EL MODELO PIN-HOLE** El modelo pin-hole [Hartley and Zisserman, 2003] parte de la consideración de la proyección central de los puntos en el espacio dentro de un plano. Se adopta de ahora en adelante la siguiente notación:

- **Centro Óptico -  $C$**   $\Rightarrow$  Se corresponde con el origen del sistema de coordenadas y hace referencia al centro de la cámara; punto donde convergen todos los rayos ópticos.
- **Plano Imagen o Focal**  $\Rightarrow$  Plano separado la distancia focal respecto del centro de la cámara  $C$ . Cada punto del escenario se comporta como foco emisor de rayos de luz que pasan por el centro óptico y atraviesan a dicho plano formando la imagen; si estuviera colocado detrás del centro, la imagen quedaría invertida.

- **Rayo o eje principal** -  $Z \Rightarrow$  Línea perpendicular al plano imagen que atraviesa el centro óptico  $C$ .
- **Punto principal** -  $p \Rightarrow$  Intersección del eje principal  $Z$  con el plano imagen.
- **Punto escena** -  $M_c \Rightarrow$  Punto cualquiera de la escena cuya proyección se representa en el plano imagen.
- **Punto imagen** -  $m_p \Rightarrow$  Proyección del punto escena  $M_c$  en el plano imagen.
- **Plano principal**  $\Rightarrow$  Se trata del plano paralelo al imagen que contiene al centro óptico  $C$  y cuyo contenido estará formado por todos los puntos del espacio cuyas proyecciones se correspondan con puntos en el infinito en la imagen.

En la figura 3.5 se muestra de manera gráfica la notación planteada.

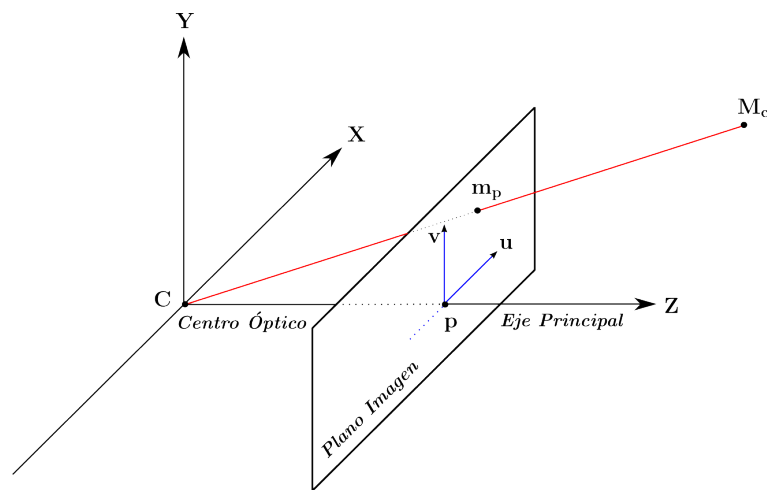


Figura 3.5: Modelado geométrico de la cámara pin-hole. Tener en cuenta que el plano imagen está desplazado en frente del centro óptico y por tanto la imagen no quedará invertida.

**MATRIZ PROYECTIVA** El modelo pin-hole fija un sistema de coordenadas proyectivas en el centro óptico  $C$ , coincidiendo el eje  $Z$  con el eje principal de la cámara y al que denotaremos como sistema de referencia de la cámara. En él, se sitúa el plano imagen a una distancia  $Z = f$  respecto del centro óptico. De esta manera se define la siguiente notación:

- **$f$  - Distancia focal**  $\Rightarrow$  Distancia a la que se encuentra separado el plano imagen respecto el origen de coordenadas.
- $M_c = (X_c, Y_c, Z_c)^T \Rightarrow$  Coordenadas de un punto cualquiera del espacio respecto al sistema de coordenadas de la cámara.
- $m_p \Rightarrow$  Coordenadas en el plano imagen del punto perteneciente al espacio tridimensional.

Por tanto, se puede afirmar que, bajo el modelo de cámara pin-hole, un punto en el espacio tridimensional  $M_c = (X_c, Y_c, Z_c)$  es proyectado en el plano imagen en las coordenadas  $m_p = (f \frac{X_c}{Z_c}, f \frac{Y_c}{Z_c})$ . La correspondencia entre un punto cualquiera del espacio y su proyección en el plano imagen se puede modelar mediante relación de triángulos, tal y como se muestra en la figura 3.6.

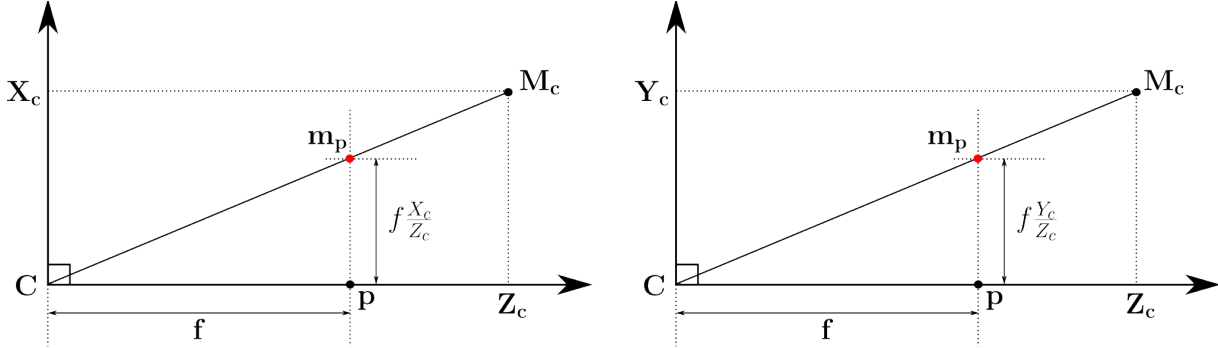


Figura 3.6: Modelado geométrico de la cámara pin-hole. Relación de triángulos.

De esta manera se genera un espacio bidimensional  $\mathbb{R}^2$  a partir de otro tridimensional  $\mathbb{R}^3$  en el que se pierde información acerca de la profundidad:

$$(X_c, Y_c, Z_c)^T \mapsto (f \frac{X_c}{Z_c}, f \frac{Y_c}{Z_c})^T \quad (3.1)$$

Haciendo uso de coordenadas homogéneas (en la notación se especifica mediante el superíndice  $\hat{\cdot}$ ) podemos expresar la relación anterior en forma matricial:

$$\begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} f X_c \\ f Y_c \\ Z_c \end{pmatrix} = \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix} \quad (3.2)$$

Pudiéndose expresar de manera abreviada de la siguiente forma:

$$P = \text{diag}(f, f, 1)[I|0] \implies \hat{m}_p = P \hat{M}_c \quad (3.3)$$

donde:

- $\text{diag}(f, f, 1)$  es una matriz diagonal.
- $[I|0]$  representa la concatenación de una matriz identidad de tamaño  $3 \times 3$  y un vector nulo de tamaño  $3 \times 1$ .
- $\hat{M}_c$  es el punto del espacio tridimensional expresado en coordenadas homogéneas como un vector columna de dimensión 4 respecto al sistema de referencia de la cámara.

- $\hat{m}_p$  es el punto de la imagen expresado en coordenadas homogéneas como un vector columna de dimensión 3.
- $P$  hace referencia a la **matriz homogénea de proyección de la cámara**.

**CAMBIO DEL SISTEMA DE REFERENCIA** Hasta ahora, se ha asumido que el origen del sistema de coordenadas en el plano imagen se situaba en el punto principal  $p$ . Sin embargo, en los casos prácticos emplearemos un cambio del sistema de coordenadas de manera que éste quede situado en uno de los extremos, véase figura 3.7.

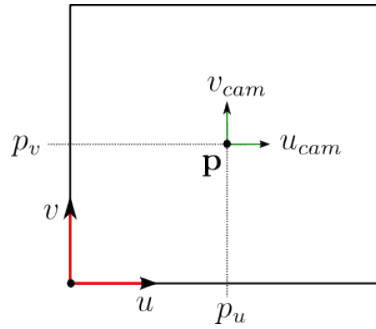


Figura 3.7: Sistema de coordenadas de imagen  $(u, v)$  y de la cámara  $(u_{cam}, v_{cam})$ . El punto principal  $p$  tiene coordenadas  $(u_0, v_0)$

Dicha transformación se modela sumando a la coordenada  $(u, v)$  el valor de las coordenadas del punto principal  $(p_u, p_v)$ :

$$M_c = (X_c, Y_c, Z_c)^T \mapsto m_k = \left( f \frac{X_c}{Z_c} + p_u, f \frac{Y_c}{Z_c} + p_v \right)^T \quad (3.4)$$

Haciendo uso de coordenadas homogéneas, se puede expresar de manera cómoda en forma matricial:

$$\begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix} \mapsto \hat{m}_k = \begin{pmatrix} fX_c + Z_cp_u \\ fY_c + Z_cp_v \\ Z_c \end{pmatrix} = \begin{pmatrix} f & p_u & 0 \\ f & p_v & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix} \quad (3.5)$$

O lo que es lo mismo:

$$\hat{m}_k = K[\mathbf{I}|\mathbf{0}]\hat{M}_c \quad (3.6)$$

donde:

- $m_k \Rightarrow$  es el punto proyectado en el plano imagen desde el nuevo origen del sistema de referencia. Notar que tanto  $m_p$  como  $m_k$  son el mismo punto del plano imagen, pero referidos a distintos orígenes. Se recuerda que la notación  $\hat{m}_k$  indica su expresión en coordenadas homogéneas. En la figura 3.8 queda plasmado este hecho.

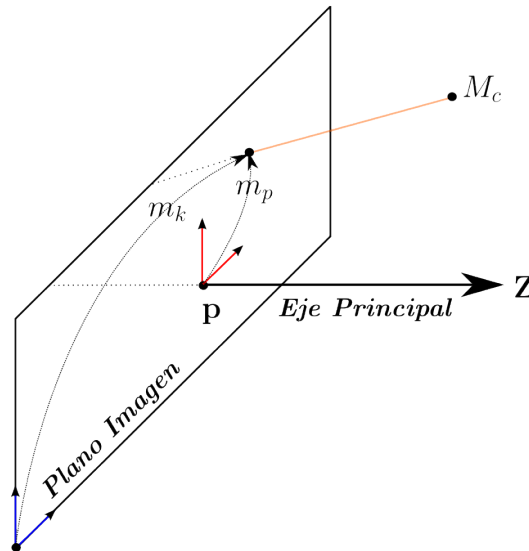


Figura 3.8: Aclaración de la notación empleada para el cambio del origen del sistema de referencia.

- $K \Rightarrow$  esta matriz recibe el nombre de **matriz de calibración de la cámara** y queda definida de la siguiente manera:

$$K = \begin{pmatrix} f & p_u \\ f & p_v \\ 1 \end{pmatrix} \quad (3.7)$$

Con esta notación, podemos expresar la ecuación 3.5 como:

$$\hat{m}_k = K[\mathbf{I}|\mathbf{0}]\hat{M}_c \quad (3.8)$$

El modelo de cámara pin-hole visto hasta ahora asume que los ejes del sistema de coordenadas poseen la misma escala. En el caso de muchas cámaras CCD, los píxeles no son cuadrados, por lo que si las coordenadas se miden en píxeles, se tendrá que tener en consideración el efecto de introducir un factor de escala desigual en cada eje.

Si se considera el número de píxeles por unidad de longitud en los ejes del plano imagen como  $\frac{1}{du}$  y  $\frac{1}{dv}$  (ejes  $u$  y  $v$  respectivamente) se obtiene la transformación de las coordenadas del espacio tridimensional a las coordenadas en píxeles, multiplicando por la izquierda por un factor

extra  $\text{diag}(\frac{1}{du}, \frac{1}{dv}, 1)$ ; a esto se conoce como **expresión general de la matriz de calibración de la cámara**.

$$K = \begin{pmatrix} \frac{1}{du} & & \\ & \frac{1}{dv} & \\ & & 1 \end{pmatrix} \begin{pmatrix} f & p_u \\ f & p_v \\ & 1 \end{pmatrix} = \begin{pmatrix} \alpha_u & & u_0 \\ & \alpha_v & v_0 \\ & & 1 \end{pmatrix} \quad (3.9)$$

donde:

- $\frac{1}{du} \Rightarrow$  es el número de píxeles por unidad de distancia en el eje  $u$  del sistema de referencia del plano imagen.
- $\frac{1}{dv} \Rightarrow$  es el número de píxeles por unidad de distancia en el eje  $v$  del sistema de referencia del plano imagen.
- $\alpha_u = f \frac{1}{du} \Rightarrow$  representa la distancia focal de la cámara en píxeles en la dirección  $u$ .
- $\alpha_v = f \frac{1}{dv} \Rightarrow$  representa la distancia focal de la cámara en píxeles en la dirección  $v$ .
- $(u_0, v_0) = (\frac{1}{du}p_u, \frac{1}{dv}p_v) \Rightarrow$  hace referencia al punto principal expresado en píxeles.

Por último, para terminar de proporcionar carácter general a la matriz de calibración de la cámara, se añade un parámetro que referencia la no perpendicularidad de los ejes. Esto obliga a introducir un factor, conocido como “skew”, que refleja la deformación en función del ángulo que forman.

$$K = \begin{pmatrix} \alpha_u & s & u_0 \\ & \alpha_v & v_0 \\ & & 1 \end{pmatrix} \quad (3.10)$$

donde:

- $s \Rightarrow$  es el factor de “no perpendicularidad” o “skew”. En la mayoría de las cámaras los ejes formarán un ángulo recto y éste parámetro tomará el valor 0.

**ROTACIÓN Y TRASLACIÓN DEL SISTEMA DE REFERENCIA** El sistema de coordenadas de los puntos del espacio 3D se ha situado con origen en el centro óptico  $C$ , pero en ocasiones es preferible situar el sistema de referencia en otro punto, dando lugar a lo que se conoce como “**sistema de coordenadas del mundo**”, cuyo origen se denota como  $O_w$ . La conversión entre ambos sistemas de coordenadas se obtiene a través de una matriz de **rotación** ( $R$ ) y otra de **traslación** ( $t$ ), tal y como se muestra en la figura 3.9.

De esta manera, se puede expresar la relación entre ambos sistemas como:

$$M_c = R(M_w - C_w) \quad (3.11)$$

donde:

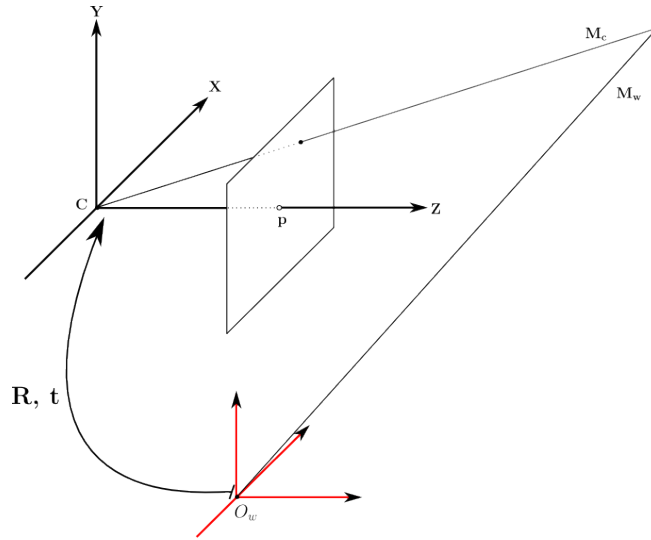


Figura 3.9: Transformación necesaria para el cambio del sistema de referencia.

- $M_c \Rightarrow$  son las coordenadas del punto del espacio referidas al sistema de referencia de la cámara.
- $M_w \Rightarrow$  son las coordenadas no homogéneas del mismo punto respecto al nuevo sistema de referencia del mundo. Es un vector columna de dimensión 3.
- $O_w \Rightarrow$  representa el origen del sistema de referencia del mundo.
- $C_w \Rightarrow$  es el centro del sistema de referencia de la cámara pero expresado desde el sistema de referencia del mundo.
- $R \Rightarrow$  denominada matriz de rotación, es de dimensión  $3 \times 3$  y representa el cambio de orientación del sistema de referencia de la cámara para realizar la conversión entre ambos sistemas.

Si se escribe la ecuación 3.11 en términos de coordenadas homogéneas se obtiene la siguiente expresión en forma matricial:

$$\hat{M}_c = \begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix} = \begin{pmatrix} R & -RC_w \\ 0 & 1 \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} = \begin{pmatrix} R & -RC_w \\ 0 & 1 \end{pmatrix} \hat{M}_w \quad (3.12)$$

que de manera abreviada la expresión 3.6 resulta:

$$\hat{m}_k = KR[I - C_w]\hat{M}_w \quad (3.13)$$

Dado que en ocasiones no se conoce el punto de proyección central de la cámara  $C_w$  se introduce el concepto de **matriz de traslación**,  $t$ , de manera que:

$$\hat{M}_c = R\hat{M}_w + t \quad \longmapsto \quad \hat{M}_c = [R|t]\hat{M}_w \quad (3.14)$$



siendo:

$$\mathbf{t} = -\mathbf{R}\mathbf{C}_w \quad (3.15)$$

De esta manera, podemos expresar la ecuación 3.13 a partir de 3.8 y 3.14 como:

$$\hat{m}_k = K[R|t]\hat{M}_w \quad (3.16)$$

mostrando de manera compacta una expresión que relaciona las coordenadas del punto en el plano imagen  $-m_k-$  (con origen en un extremo) con un punto del espacio 3D referido a cualquier sistema de coordenadas del mundo. Se llega así a obtener la expresión general de la matriz de proyección de la cámara como:

$$P = K[R|t] \longmapsto P = \begin{pmatrix} \alpha_u & s & u_0 \\ & \alpha_v & v_0 \\ & & 1 \end{pmatrix} \begin{pmatrix} R_{11} & R_{12} & R_{13} & t_x \\ R_{21} & R_{22} & R_{23} & t_y \\ R_{31} & R_{32} & R_{33} & t_z \end{pmatrix} \quad (3.17)$$

A las cámaras con una matriz de proyección como ésta se las denominan de **proyección finita**; se representa como una matriz homogénea de dimensiones  $3 \times 4$ , rango 3 y 11 grados de libertad a falta de definir un factor de escala.

De la expresión general de la matriz proyectiva se extraen los siguientes conceptos:

- **Parámetros intrínsecos de la cámara**  $\Rightarrow$  determinan las características internas de la cámara y son parámetros propios del sistema de adquisición. Se definen como 5 parámetros provenientes de la matriz de calibración de la cámara:  $(\alpha_u, \alpha_v, \mathbf{u}_0, \mathbf{v}_0, s)$ .
- **Parámetros extrínsecos de la cámara**  $\Rightarrow$  se trata de una serie de parámetros externos a la cámara. Dependen del marco de referencia definido y están relacionados con la orientación y desplazamiento del sistema de referencia de la cámara respecto al del espacio. Se definen como 12 parámetros provenientes, 9 de la matriz de rotación  $\mathbf{R}$  ( $R_{11} \dots R_{33}$ ) y 3 de la matriz de traslación  $\mathbf{t}$  ( $t_x, t_y, t_z$ ).

### 3.4.1.2. Concepto de calibración y homografía

Calibrar una cámara consiste en estimar los valores de sus parámetros intrínsecos ( $K$ ) y extrínsecos ( $R, t$ ) y así obtener la relación entre los puntos del espacio 3D y los proyectados en el plano bidimensional de la cámara. En la mayoría de los casos podremos distinguir dos fases en la calibración de una cámara:

1. Estimación de la matriz proyectiva  $P$ .
2. Estimación de los parámetros intrínsecos y extrínsecos a partir de la matriz  $P$  obtenida, en caso de ser necesario.

## ESTIMACIÓN DE LA MATRIZ PROYECTIVA

- a) Se asume conocida la correspondencia entre una serie de puntos del espacio 3D  $\hat{M}_w^i$  y su proyección en el plano imagen  $\hat{m}_k^i$ . La idea es encontrar la matriz proyectiva  $P$  de dimensión  $3 \times 4$  de manera que:

$$\hat{m}_k^i = P \hat{M}_w^i, \forall i \quad \longrightarrow \quad \hat{m}_k^i = \begin{pmatrix} P^{1T} \hat{M}_w^i \\ P^{2T} \hat{M}_w^i \\ P^{3T} \hat{M}_w^i \end{pmatrix} \quad (3.18)$$

donde:

- $P^{jT} \Rightarrow$  es un vector fila de dimensión 4 que se corresponde con la fila  $j$  de la matriz proyectiva  $P$ .
- b) Se ha de tener en cuenta que se está trabajando con una ecuación que involucra vectores homogéneos, por tanto, los vectores  $\hat{m}_k^i$  y  $P \hat{M}_w^i$  no son iguales, poseen la misma dirección pero difieren en magnitud por un factor de escala no nulo. La ecuación anterior se expresa como un producto vectorial:

$$\hat{m}_k^i \times P \hat{M}_w^i = 0 \quad (3.19)$$

Esta forma permite obtener una solución lineal sencilla de  $P$  de manera que pueda ser derivada.

Sea  $\hat{m}_k^i = (x_k^i, y_k^i, z_k^i)$ , entonces se puede expresar el producto vectorial anterior de manera explícita como:

$$\begin{pmatrix} y_k^i P^{3T} M_w^i - z_k^i P^{2T} M_w^i \\ z_k^i P^{1T} M_w^i - x_k^i P^{3T} M_w^i \\ x_k^i P^{2T} M_w^i - y_k^i P^{1T} M_w^i \end{pmatrix} = 0 \quad (3.20)$$

Además, dado que el producto  $P^{iT} M_w^i = M_w^i P^i, \forall i \in [1, 2, 3]$  se obtiene un sistema de 3 ecuaciones donde los elementos de la matriz  $P$  son las incógnitas:

$$\begin{pmatrix} 0^T & -z_k^i M_w^{iT} & y_k^i M_w^{iT} \\ z_k^i M_w^{iT} & 0^T & -x_k^i M_w^{iT} \\ -y_k^i M_w^{iT} & x_k^i M_w^{iT} & 0^T \end{pmatrix} \begin{pmatrix} P^1 \\ P^2 \\ P^3 \end{pmatrix} = 0 \quad (3.21)$$

De estas 3 ecuaciones solo 2 son linealmente independientes. Se toman solo las dos primeras ecuaciones de 3.21 resultando:

$$\begin{pmatrix} 0^T & -z_k^i M_w^{iT} & y_k^i M_w^{iT} \\ z_k^i M_w^{iT} & 0^T & -x_k^i M_w^{iT} \end{pmatrix} \begin{pmatrix} P^1 \\ P^2 \\ P^3 \end{pmatrix} = 0 \quad (3.22)$$

De una serie de  $n$  correspondencias de puntos  $M_w^i \leftrightarrow m_k^i, \forall i \in [1, 2, \dots, n]$ , se obtiene una matriz de dimensión  $2n \times 12$  a partir de la expresión 3.22:

$$\begin{pmatrix} 0^T & -z_k^1 M_w^{1T} & y_k^1 M_w^{1T} \\ z_k^1 M_w^{1T} & 0^T & -x_k^1 M_w^{1T} \\ \vdots & \vdots & \vdots \\ 0^T & -z_k^n M_w^{nT} & y_k^n M_w^{nT} \\ z_k^n M_w^{nT} & 0^T & -x_k^n M_w^{nT} \end{pmatrix} \begin{pmatrix} P^1 \\ P^2 \\ P^3 \end{pmatrix} = 0, \quad (3.23)$$

donde el sistema anterior toma la forma  $\mathbf{A}p = 0$ , siendo  $p$  un vector con las incógnitas de la matriz proyectiva  $P$ .

La matriz  $P$  tiene 12 elementos, pero si se ignora el factor de escala, se tienen 11 grados de libertad. Por tanto, se necesitan 11 ecuaciones linealmente independientes para poder estimar  $P$ . Puesto que cada correspondencia  $M_w^i \leftrightarrow m_k^i$  proporciona 2 ecuaciones, como mínimo se necesitan conocer 6 puntos con sus correspondientes proyecciones en el plano imagen.

Si se conociese con exactitud la posición del par de puntos de cada correspondencia, únicamente necesitaríamos el mínimo número de ecuaciones para obtener la solución exacta de la matriz  $P$ . Sin embargo, los datos no son precisos debido a imprecisiones en la medida de los puntos del espacio y sus proyecciones, presencia de ruido, etc.. Por tanto, se suele trabajar con sistemas sobre-determinados, donde  $n \geq 6$ . En este caso, la solución no será exacta y se tendrá que aproximar el valor de la matriz  $P$  de manera que se minimice el error algebraico.

**ESTIMACIÓN DE LOS PARÁMETROS INTRÍNSECOS Y EXTRÍNSECOS DE LA CÁMARA A PARTIR DE LA MATRIZ  $P$**  Una vez estimada la solución para la matriz proyectiva  $P$ , podemos obtener los parámetros intrínsecos y extrínsecos de la cámara. Tomando como ecuación de partida 3.17 y teniendo en cuenta las propiedades de toda matriz de rotación ortonormal:

- La norma de cada vector fila de la matriz de rotación vale 1.
- El producto vectorial de dos vectores filas da como resultado el tercer vector fila.
- El producto escalar de un vector fila consigo mismo es 1.
- El producto escalar de dos vectores fila distintos es 0, pues son ortogonales.

se obtienen los parámetros intrínsecos y extrínsecos ( $\mathbf{K}, \mathbf{R}, \mathbf{t}$ ) como:

$$P = K[R|t]$$

$$\begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{pmatrix} = \begin{pmatrix} \alpha_u & s & u_0 \\ & \alpha_v & v_0 \\ & & 1 \end{pmatrix} \begin{pmatrix} R_{11} & R_{12} & R_{13} & t_x \\ R_{21} & R_{22} & R_{23} & t_y \\ R_{31} & R_{32} & R_{33} & t_z \end{pmatrix} \quad (3.24)$$

$$\begin{cases} p_{1i} = \alpha_u R_{1i} + s R_{2i} + u_0 R_{3i}, & \forall i \in (1, 2, 3) \\ p_{14} = \alpha_u t_x + s t_y + u_0 t_z \\ p_{2i} = \alpha_v R_{2i} + v_0 R_{3i}, & \forall i \in (1, 2, 3) \\ p_{24} = \alpha_v t_y + v_0 t_z \\ p_{3i} = R_{3i}, & \forall i \in (1, 2, 3) \\ p_{34} = t_z \end{cases} \quad (3.25)$$

de esta manera, y empleando las propiedades citadas anteriormente llegamos a obtener los parámetros buscados:

$$\begin{cases} R_{3i} = p_{3i}, & \forall i \in (1, 2, 3) \\ T_z = p_{34} \\ u_0 = p_{1i} \cdot p_{3i}, & \forall i \in (1, 2, 3) \\ v_0 = p_{2i} \cdot p_{3i}, & \forall i \in (1, 2, 3) \\ \alpha_u = \|p_{1i} - s R_{2i} - u_0 R_{3i}\| \\ R_{1i} = \frac{p_{1i} - s R_{2i} - u_0 R_{3i}}{\alpha_u} \\ \alpha_v = \|p_{2i} - v_0 R_{3i}\| \\ R_{2i} = \frac{p_{2i} - v_0 R_{3i}}{\alpha_v} \\ t_y = \frac{p_{24} - v_0 t_z}{\alpha_v} \\ t_z = \frac{p_{14} - s t_y - u_0 t_z}{\alpha_u} \end{cases} \quad (3.26)$$

**CONCEPTO DE HOMOGRAFÍA** La geometría 2D proyectiva estudia las propiedades del plano proyectivo, denominado  $\mathbb{P}^2$ , que son invariantes bajo un grupo de transformaciones conocidas como transformaciones proyectivas u homografías. El plano proyectivo describe propiedades de puntos expresados en coordenadas homogéneas, como puede ser el punto  $\hat{m}_k$ , resultado de la proyección mediante una cámara “pin-hole”. Se dice que una correspondencia  $h : \mathbb{P}^2 \mapsto \mathbb{P}^2$  es una proyectividad si existe una matriz no singular  $\mathcal{H}$ , de dimensiones  $3 \times 3$ , tal que, para cualquier punto  $\hat{m}_k \in \mathbb{P}^2$ , se cumple que  $h(\hat{m}_k) = \mathcal{H}\hat{m}_k$ . A dicha matriz se la denomina como matriz de homografía. Por tanto, una proyectividad u homografía es una correspondencia lineal entre vectores expresados en coordenadas homogéneas de 3 dimensiones:

$$\hat{m}_k = \lambda \mathcal{H} m_p \quad \lambda \in \mathbb{R} \quad \mapsto \quad \mathcal{H} = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix} \quad (3.27)$$

Puesto que la matriz de homografía  $\mathcal{H}$  es no singular, es invertible y su inversa es a su vez una proyectividad. Además, si multiplicamos la matriz  $\mathcal{H}$  por un factor no nulo, la transformación proyectiva no varía puesto que, al estar trabajando con coordenadas homogéneas, puntos que difieren en un factor de escala no nulo son equivalentes. Por tanto, si normalizamos la matriz  $\mathcal{H}$  de forma que  $h_{33} = 1$ , dicha matriz tendrá 8 grados de libertad en el caso más general, a falta de definir el factor de escala.

Mediante una homografía se pueden modelar determinadas relaciones geométricas entre imágenes tomadas por cámaras “pin-hole”, entre las que destacan las siguientes:

- **Dos cámaras proyectivas que comparten el origen de proyección:** En este caso se tienen dos cámaras, descritas por sus respectivas matrices  $P_1$  y  $P_2$ , en las que el origen de proyección coincide en ambas, y por tanto los vectores  $t_1$  y  $t_2$  son coincidentes. En dicha configuración, existe una matriz de homografía  $\mathcal{H}$  que relaciona, de manera biyectiva, la imagen vista por una cámara con la otra. Es decir que, si se considera un punto  $\hat{m}_k^1$  visto por la cámara  $P_1$  y un punto  $\hat{m}_k^2$  visto por la cámara  $P_2$ , ambos correspondientes a la proyección en cada cámara del mismo punto  $M_w$ , entonces se puede encontrar una homografía  $\mathcal{H}$  tal que cumpla la siguiente expresión:

$$\hat{m}_k^1 = \lambda \mathcal{H} \hat{m}_k^2, \quad \lambda \in \mathbb{R} \quad (3.28)$$

- **Una cámara proyectiva y un plano en el espacio tridimensional:** Dada una cámara  $P_1$  y un plano en el espacio tridimensional  $\Pi : ax + by + cz + d = 0$ , existe una homografía  $\mathcal{H}$  que relaciona puntos definidos en un origen local al plano  $m_\Pi$  (en coordenadas homogéneas), con dichos puntos  $\hat{m}_k^1$  vistos en el plano imagen de la cámara. En la figura 3.10 se representa la transformación generada por la matriz de homografía,  $\mathcal{H}$  en este último caso.

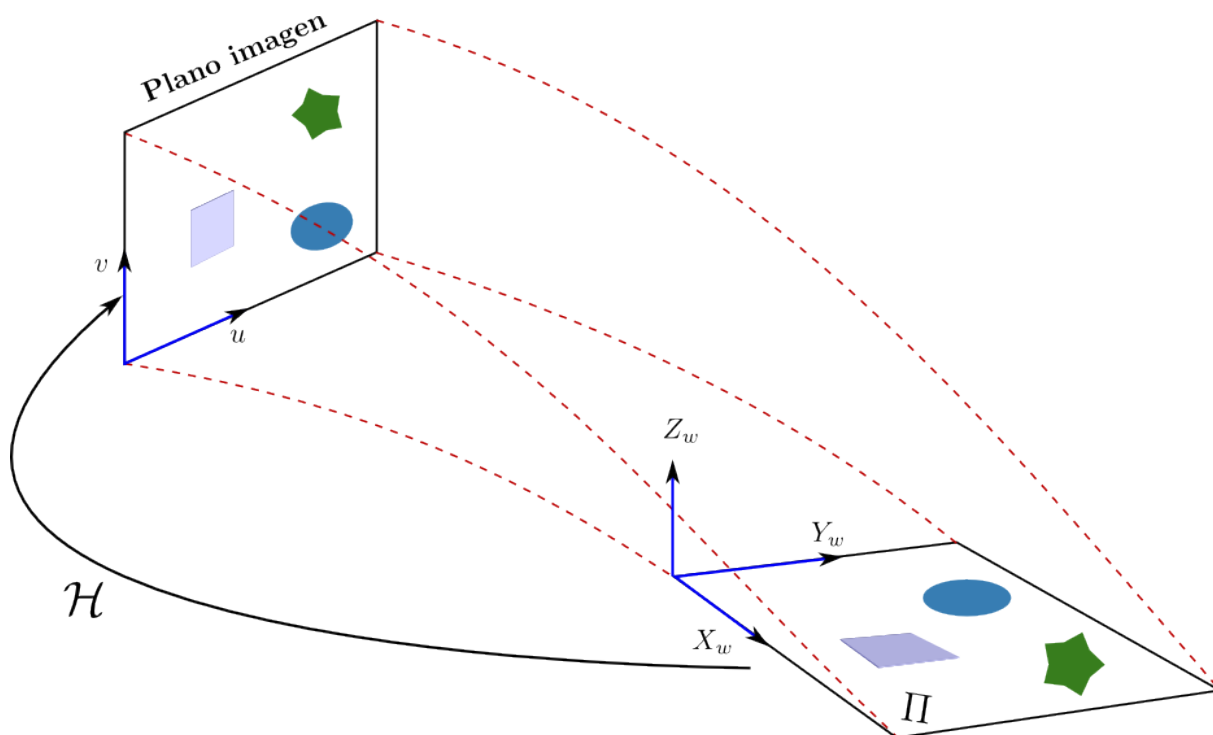


Figura 3.10: Relación de puntos entre una cámara proyectiva y un plano en el espacio tridimensional.

- **Dos cámaras proyectivas y un plano en el espacio tridimensional:** Como extensión del punto anterior, si se tienen dos cámaras  $P_1$  y  $P_2$  y un plano en el espacio tridimensional  $\Pi : ax + by + cz + d = 0$ , existen dos homografías,  $\mathcal{H}_1$  y  $\mathcal{H}_2$ , que relacionan el plano con cada una de las dos cámaras. Es decir que dados dos puntos  $\hat{m}_k^1, \hat{m}_k^2$  vistos por cada cámara y correspondientes al punto  $m_p$  del plano, se cumple que:

$$\hat{m}_k^1 = \lambda_1 \mathcal{H}_1 m_p \quad \hat{m}_k^2 = \lambda_2 \mathcal{H}_2 m_p \quad \lambda_1, \lambda_2 \in \mathbb{R} \quad (3.29)$$

Puesto que tanto  $\mathcal{H}_1$  como  $\mathcal{H}_2$  son no singulares, se puede asegurar que existe una homografía  $\mathcal{H}_{12}$  que relaciona puntos del plano imagen de la cámara  $P_1$  con puntos vistos por la cámara  $P_2$ , resultando  $\mathcal{H}_{12}$  de la siguiente expresión:

$$\mathcal{H}_{12} = \mathcal{H}_2 \mathcal{H}_1^{-1} \quad \hat{m}_k^2 = \lambda_{1,2} \mathcal{H}_{12} \hat{m}_k^1 \quad \lambda_{1,2} \in \mathbb{R} \quad (3.30)$$

En la figura 3.11 se muestra esta última situación, la cual será relevante para el algoritmo de eliminación de sombras, mediante dos cámaras, diseñado en este proyecto; para el plano del suelo.

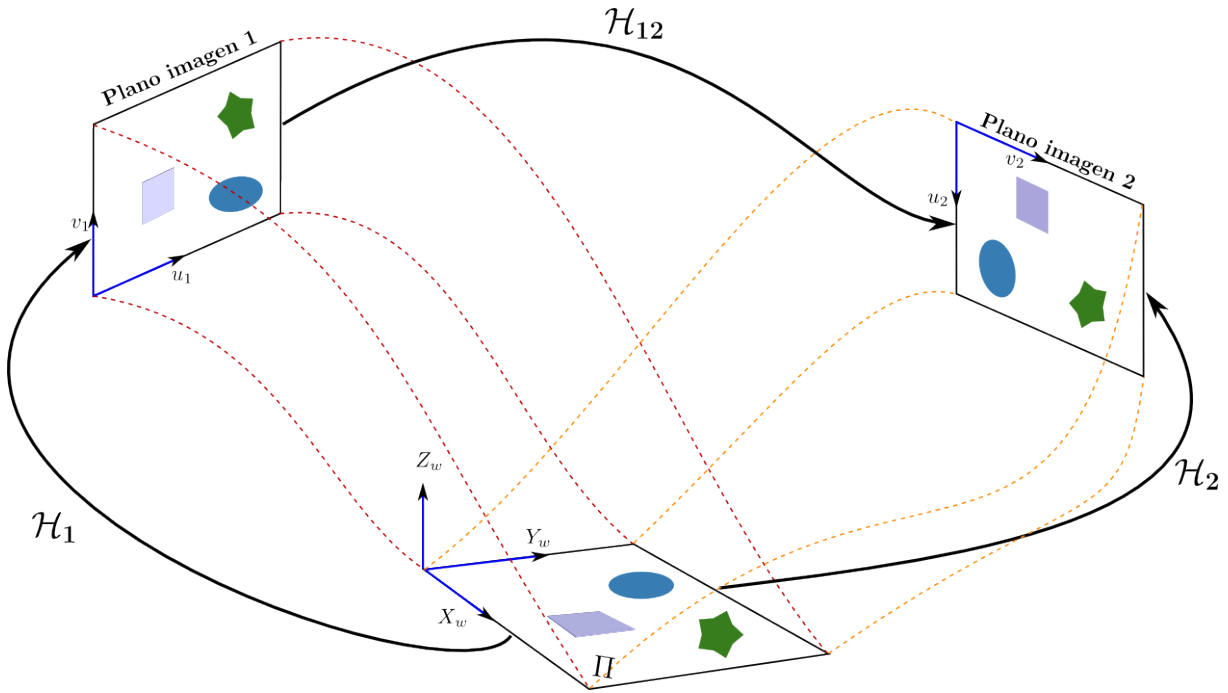


Figura 3.11: Relación de puntos entre dos cámaras proyectivas y un plano en el espacio tridimensional.

### 3.4.1.3. Métodos de estimación de la matriz de homografía

En este apartado se comentan dos métodos diferentes para la estimación de una matriz de homografía genérica  $\mathcal{H}$ :

1. Estimación de la homografía a partir de correspondencias de puntos: en este método se utilizará únicamente información geométrica, correspondiente a puntos en dos planos proyectivos, relacionados mediante la matriz  $\mathcal{H}$ . El método propuesto se resuelve mediante un procedimiento de mínimos cuadrados que generaliza la estimación para situaciones en los que se tiene un conjunto de correspondencias superior a los grados de libertad que admite  $\mathcal{H}$ .

2. Estimación de la homografía mediante el registro directo de imágenes: en este método se utiliza la información de intensidad de dos imágenes (p. ej. dos cámaras observando un plano común) para el cálculo de la matriz  $\mathcal{H}$ . Este tipo de métodos requieren de una solución iterativa, que se describe en esta sección. El objetivo de este método es refinar la solución basada en el método anterior, de modo que, en conjunto, se pueda disponer de un sistema semiautomático de ajuste de la homografía buscada.

## 1. ESTIMACIÓN DE LA MATRIZ DE HOMOGRAFÍA A PARTIR DE CORRESPONDENCIAS DE PUNTOS

Se parte de la consideración de disponer de un conjunto de puntos,  $m_p^i \in \mathbb{P}^2$  definidos en un sistema de coordenadas local a un plano tridimensional conocido (ver figura 3.10), del que se conocen sus correspondencias en el plano imagen  $m_k^i \in \mathbb{P}^2$ . La idea es obtener la transformación proyectiva, en forma de una matriz  $\mathcal{H}$ , que relaciona los dos conjuntos de puntos entre sí:

$$\hat{m}_k^i = \lambda \mathcal{H} m_p^i \quad \lambda \in \mathbb{R} \quad \Rightarrow \quad \hat{m}_k^i = \lambda \begin{pmatrix} h^1 m_p^i \\ h^2 m_p^i \\ h^3 m_p^i \end{pmatrix} \quad (3.31)$$

- $h^j \Rightarrow$  se trata de un vector fila de dimensión 3 que se corresponde con la fila  $j$  de la matriz de homografía  $\mathcal{H}$ .

Como objetivo se plantea la determinación de  $\mathcal{H}$  a partir del conjunto de ecuaciones mostrados en (3.31). Puesto que una homografía relaciona vectores homogéneos,  $\hat{m}_k^i$  y  $\mathcal{H}m_p^i$  no son iguales término a término; en cambio, tienen la misma dirección y difieren en magnitud, a falta de un factor de escala no nulo y desconocido  $\lambda$ . Para la eliminación de dicho factor existe la alternativa de dividir cada uno de los vectores por su tercera coordenada, recuperando así las posiciones 2D en el plano imagen. Esta alternativa deriva en un sistema de ecuaciones no lineal con respecto a los términos que componen  $\mathcal{H}$ , requiriendo un método de optimización iterativo para encontrar su solución. Por el contrario, la alternativa propuesta se basa en expresar la ecuación 3.31 en términos de un producto vectorial de manera que se pueda obtener una solución de  $\mathcal{H}$  eliminando la indeterminación de escala y conservando la linealidad del problema:

$$\hat{m}_k^i \times \mathcal{H}m_p^i = 0 \quad (3.32)$$

Sea  $\hat{m}_k^i = (x_k^i, y_k^i, z_k^i)$ , entonces se puede expresar el producto vectorial de manera explícita como:

$$m_k^i \times \mathcal{H}m_p^i = \begin{pmatrix} y_k^i h^3 m_p^i - z_k^i h^2 m_p^i \\ z_k^i h^1 m_p^i - x_k^i h^3 m_p^i \\ x_k^i h^2 m_p^i - y_k^i h^1 m_p^i \end{pmatrix} \quad (3.33)$$

A partir de la expresión 3.33 se obtiene un sistema de 3 ecuaciones lineales donde el vector de incógnitas  $\Phi$  se compone de los elementos de la matriz  $\mathcal{H}$ :

$$\begin{pmatrix} 0^T & -z_k^i m_p^i & y_k^i m_p^i \\ z_k^i m_p^i & 0^T & -x_k^i m_p^i \\ -y_k^i m_p^i & x_k^i m_p^i & 0^T \end{pmatrix} \Phi = 0, \quad \Phi = \begin{pmatrix} (h^1)^T \\ (h^2)^T \\ (h^3)^T \end{pmatrix} \quad (3.34)$$

Aunque se han obtenido 3 ecuaciones, solo dos de ellas son linealmente independientes. Si se toman las dos primeras ecuaciones de 3.34 se puede expresar el sistema final de forma matricial, para un conjunto de  $n$  puntos, de la siguiente forma:

$$A\Phi = 0, \quad A = \begin{pmatrix} 0^T & -z_k^i m_p^i & y_k^i m_p^i \\ z_k^i m_p^i & 0^T & -x_k^i m_p^i \\ \vdots & \vdots & \vdots \\ 0^T & -z_k^i m_p^i & y_k^i m_p^i \\ z_k^i m_p^i & 0^T & -x_k^i m_p^i \end{pmatrix} \quad (3.35)$$

Estas ecuaciones son de la forma  $A_i h = 0$ , donde  $A_i$  es una matriz de dimensión  $2n \times 8$ .

Dado que, de cada correspondencia se obtienen dos ecuaciones linealmente independientes (cada punto 2D dispone de 2 grados de libertad, uno por cada componente del sistema de referencia) y que la matriz de homografía tiene 8 grados de libertad, se requiere en teoría un mínimo conjunto de  $n = 4$  correspondencias  $m_p^i \leftrightarrow \hat{m}_k^i$  para poder estimar la matriz  $\mathcal{H}$ .

A pesar de que los grados de libertad de la matriz  $\mathcal{H}$  son 8 se ha elegido una parametrización mediante sus 9 elementos y por lo tanto se requiere añadir un criterio para eliminar un grado de libertad del vector  $\Phi$ . Una alternativa común es forzar a que algún elemento de la homografía sea igual a la unidad, convirtiendo el sistema de la ecuación 3.35 en un sistema lineal no homogéneo  $A'\Phi' = B$ , siendo  $\Phi'$  el vector con 8 elementos libres de  $\mathcal{H}$ ,  $A'$  la matriz del sistema resultando y  $B$  el vector término independiente de tamaño  $2n \times 1$ . En dicho sistema, el cual admite la solución trivial para  $n = 4$  correspondencias o una solución de mínimos cuadrados en caso de que  $n > 4$ , se tiene el problema de que se debe saber, a priori, que el elemento de  $\mathcal{H}$  elegido para ser la unidad no es nulo o cercano a ser lo, ya que en ese caso el resultado del sistema daría con una solución degenerada o errónea. En este proyecto se propone incluir en el sistema de ecuaciones expuesto en 3.35, la restricción  $|\Phi| = 1$ , de modo que no sea necesario averiguar que elemento de la matriz  $\mathcal{H}$  es no nulo. Mediante una sencilla demostración, la cual se puede encontrar detallada en la literatura [Hartley and Zisserman, 2003], existe una solución directa al problema planteado incluyendo la nueva restricción, la cual consiste en elegir  $\Phi$  como el autovector de la matriz  $A^T A$  asociado a su autovalor nulo.

La descomposición de  $A^T A$  resulta en la siguiente matriz de autovectores  $P$  y autovalores  $\sigma_1, \dots, \sigma_9$ , ordenados de mayor a menor.

$$A^T A = P \left( \text{diag}(\sigma_1, \dots, \sigma_9) \right) P^T, \quad \sigma_i > \sigma_{i+1} \geq 0, \quad \forall i \quad P = \begin{pmatrix} v_1 & v_2 & v_3 \end{pmatrix} \quad (3.36)$$

eligiendo  $\Phi = v_9$ , autovector asociado al menor autovalor  $\sigma_9$ , se asegura que  $|\Phi| = 1$ , ya que por definición cualquier autovector de una matriz simétrica cumple la restricción.



Si las correspondencias elegidas para calcular  $A$  carecen de error o  $n = 4$ , el menor autovalor  $\sigma_9$  será idénticamente igual a cero. En sistemas sobredimensionados ( $n > 4$ ), el rango de  $A^T A$  será en general completo y por tanto  $\sigma_9 > 0$ . Eligiendo el vector  $v_9$  como solución corresponde en este caso a la solución de mínimo error cuadrático y por tanto la solución aportada es igualmente válida para este tipo de sistemas.

## 2. ESTIMACIÓN DE LA MATRIZ DE HOMOGRAFÍA MEDIANTE MÉTODOS DE REGISTRO DIRECTO DE IMÁGENES

En este apartado se detalla el método de obtención de una homografía entre dos imágenes mediante un método, denominado de registro directo, el cual minimiza la desigualdad en los niveles de intensidad entre una imagen  $\mathcal{T}$  y otra imagen  $\mathcal{I}$  transformada por una función geométrica general ( en este caso la homografía  $\mathcal{H}$  ). Los métodos de registro directo son muy utilizados en visión artificial para diversidad de aplicaciones, incluyendo modelos geométricos más complejos que una homografía, como pueden ser funciones de geometría deformable y bases de apariencia.

El problema del registro geométrico reside en la minimización de una función de error no lineal de mínimos cuadrados, dada por la discrepancia entre la intensidad de los píxeles del patrón  $\mathcal{T}$  y la imagen  $\mathcal{I}$ , la cual es abatida en el patrón mediante la transformación geométrica objeto de ser estimada. La transformación geométrica, denotada como  $\mathcal{G}$ , mapea un píxel  $\mathbf{q}$  en la región de interés  $\mathcal{R}$ , definida en el patrón, a su correspondiente píxel  $\mathcal{G}(\mathbf{q}; \mathbf{g})$  en la imagen. El vector  $\mathbf{g}$  encapsula sus parámetros. Se espera que dado un vector de parámetros  $\mathbf{g}$  idóneo,  $\mathcal{T}[\mathbf{q}]$  esté cercano a  $\mathcal{I}[\mathcal{G}(\mathbf{q}; \mathbf{g})]$ ,  $\forall \mathbf{q} \in \mathcal{R}$ . Formalmente, el problema del registro directo de imágenes se expresa como:

$$\min_{\mathbf{g}} \sum_{\mathbf{q} \in \mathcal{R}} (\mathcal{T}[\mathbf{q}] - \mathcal{I}[\mathcal{G}(\mathbf{q}; \mathbf{g})])^2 \quad (3.37)$$

El algoritmo propuesto en este proyecto para encontrar el mínimo de la función 3.37, se basa en el algoritmo iterativo propuesto por Lucas-Kanade en [Baker and Matthews, 2004], en el cual, mediante un método de Gauss-Newton se obtiene una regla de actualización iterativa del vector de parámetros  $\mathbf{g}$ , partiendo de una estimación inicial del mismo. El algoritmo de Gauss-Newton se basa en linealizar la función  $\mathcal{I}(\mathcal{G}(\mathbf{q}; \mathbf{g}))$  en torno a un incremento  $\delta_{\mathbf{g}}$  en los parámetros  $\mathbf{g}$ , obteniendo así un problema de mínimos cuadrados en el vector  $\delta_{\mathbf{g}}$ .

A continuación se detalla el proceso completo:

1. Se parte de un valor para los parámetros  $\mathbf{g}$ , bien de una estimación inicial del mismo o de la iteración anterior.
2. Mediante un desarrollo de Taylor de primer orden de la función  $\mathcal{I}[\mathcal{G}(\mathbf{q}; \mathbf{g} + \delta_{\mathbf{g}})]$  en torno a  $\delta_{\mathbf{g}} = 0$ , se obtiene la siguiente función de coste equivalente en función de  $\delta_{\mathbf{g}}$ :

$$\min_{\mathbf{g}} \sum_{\mathbf{q} \in \mathcal{R}} \left( \mathcal{T}[\mathbf{q}] - \mathcal{I}[\mathcal{G}(\mathbf{q}; \mathbf{g})] + \nabla \mathcal{I}[\mathcal{G}(\mathbf{q}; \mathbf{g})]^T \frac{\partial \mathcal{G}}{\partial \mathbf{g}} \bigg|_{\mathbf{q}; \mathbf{g}} \delta_{\mathbf{g}} \right)^2 \quad (3.38)$$

donde  $\nabla \mathcal{I}[\mathbf{q}]^T$  corresponde con el gradiente de la imagen  $\mathcal{I}$ , utilizando para ello una aproximación de la derivada mediante restas de píxeles vecinos. La matriz  $\frac{\partial \mathcal{G}}{\partial \mathbf{g}}$  corresponde con la derivada de la función de transformación geométrica en función de sus parámetros.

3. Derivando la expresión 3.38, con respecto a  $\delta_{\mathbf{g}}$ , e igualando a 0 para encontrar el mínimo se obtiene el siguiente sistema de ecuaciones en  $\delta_{\mathbf{g}}$ :

$$H\delta_{\mathbf{g}} = J \quad (3.39)$$

siendo  $H$  y  $J$ :

$$H = \sum_{\mathbf{q} \in \mathcal{R}} \left( \nabla \mathcal{I}[\mathcal{G}(\mathbf{q}; \mathbf{g})]^T \frac{\partial \mathcal{G}}{\partial \mathbf{g}} \right)^T \nabla \mathcal{I}[\mathcal{G}(\mathbf{q}; \mathbf{g})]^T \frac{\partial \mathcal{G}}{\partial \mathbf{g}} \quad (3.40)$$

$$J = \sum_{\mathbf{q} \in \mathcal{R}} \left( \nabla \mathcal{I}[\mathcal{G}(\mathbf{q}; \mathbf{g})]^T \frac{\partial \mathcal{G}}{\partial \mathbf{g}} \right)^T (\mathcal{T}[\mathbf{q}] - \mathcal{I}[\mathcal{G}(\mathbf{q}; \mathbf{g})]) \quad (3.41)$$

El sistema de ecuaciones resultante en la ecuación 3.39 se denominan las ecuaciones normales de Gauss-Newton, siendo  $H$  el Hessiano de la transformación y  $J$  el Jacobiano de la misma. La solución de  $\delta_{\mathbf{g}}$  se obtiene mediante simple inversión de la matriz  $H$ .

4. Actualización del vector de parámetros  $\mathbf{g} = \mathbf{g} + \delta_{\mathbf{g}}$ .

En este caso la función  $\mathcal{G}(\mathbf{q}; \mathbf{g})$  corresponde con la expresión no-homogénea de la homografía definida por sus 9 elementos y  $\mathbf{q} = (q_u, q_v)^T$  las coordenadas en el plano imagen de  $\mathcal{I}$ :

$$\mathcal{G}(\mathbf{q}; \mathbf{g}) = \frac{1}{h_{31}q_u + h_{32} + h_{33}} \begin{pmatrix} h_{11}q_u + h_{12}q_v + h_{13} \\ h_{21}q_u + h_{22}q_v + h_{23} \end{pmatrix} \quad (3.42)$$

El proceso se itera hasta la convergencia, determinada mediante la umbralización de la norma de  $\delta_{\mathbf{g}}$  en  $\epsilon = 10^{-8}$ , o cuando la actualización incremente el error. En el último caso, la actualización será cancelada antes de finalizar la iteración.

Los resultados obtenidos, en dos pruebas estándar, se muestran en la figura 3.12.

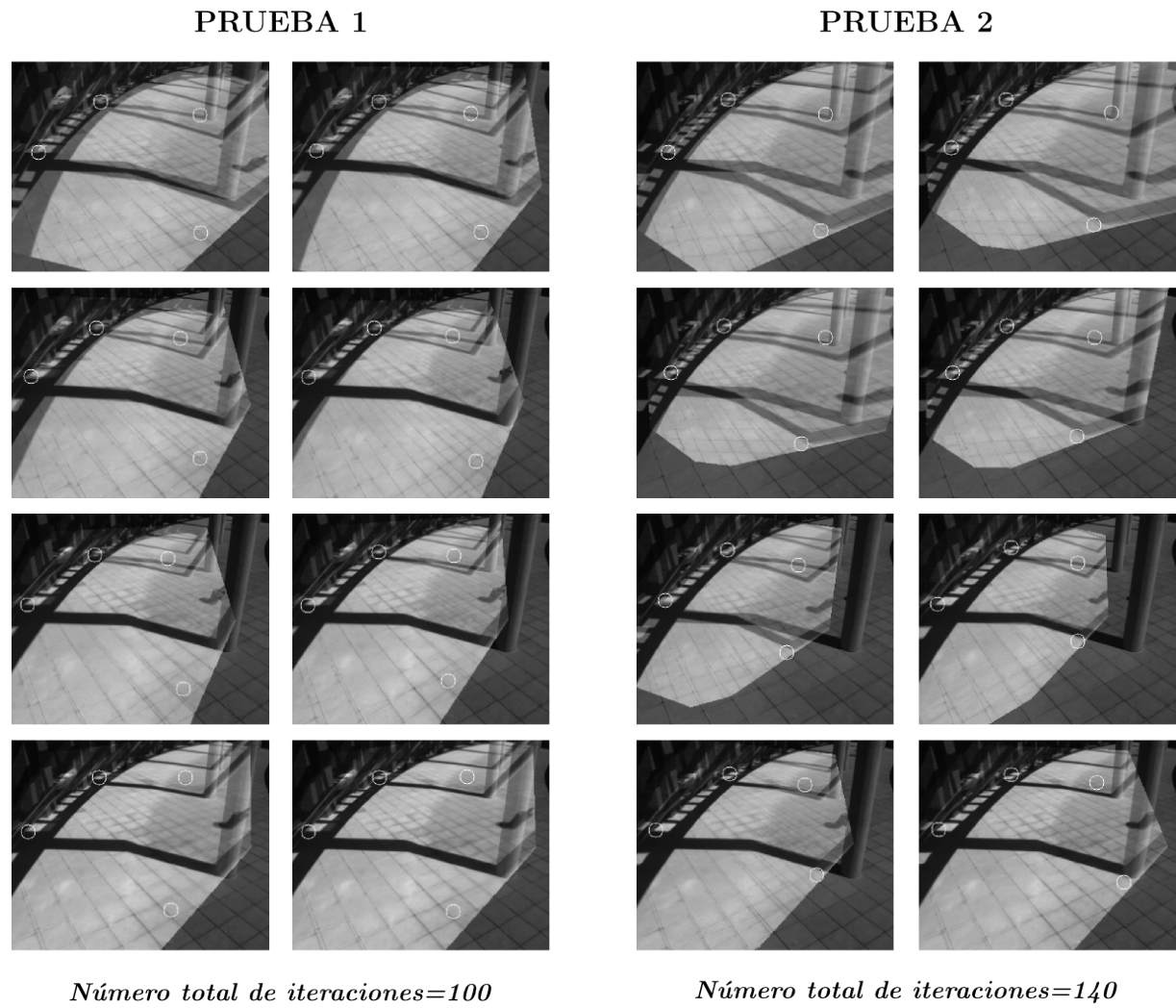


Figura 3.12: De izquierda a derecha y de arriba a bajo, se muestra la convergencia del registro directo de imágenes en 8 iteraciones para dos situaciones distintas (diferente inicialización de puntos para la ROI).

La tabla 3.2 resume el algoritmo anteriormente expuesto.

---

## MÉTODO DE REGISTRO DIRECTO DE IMÁGENES

---

### Objetivo

---

Registrar una imagen  $\mathcal{I}$  con un patrón  $\mathcal{T}$  mediante el cálculo de los parámetros  $\mathbf{g}$  de la transformación geométrica  $\mathcal{G}(\mathbf{q}; \mathbf{g})$ . Para ello se resuelve el problema de la minimización de una función de error de la discrepancia entre la intensidad de los píxeles. Otras entradas del algoritmo son la región de interés ( $\mathcal{R}$ ) en el patrón y el valor inicial dado a  $\mathbf{g}$ .

---

### Descripción del algoritmo

---

#### Cálculos previos

1. Cálculo de la función de coste equivalente mediante un desarrollo de Taylor de primer orden de la función  $\mathcal{I}[\mathcal{G}(\mathbf{q}; \mathbf{g} + \delta_{\mathbf{g}})]$  en torno a  $\delta_{\mathbf{g}} = 0$ . De esta manera se obtiene 3.38.
2. Cálculo del mínimo de la función de coste, derivando respecto  $\delta_{\mathbf{g}}$  e igualando a 0. Se obtienen así las ecuaciones normales de Gauss-Newton  $H\delta_{\mathbf{g}} = J$ , donde:
  - $H \Rightarrow$  es el Hessiano de la transformación (ecuación 3.40).
  - $J \Rightarrow$  es el Jacobiano de la transformación (ecuación 3.41).
3. Cálculo de la solución  $\delta_{\mathbf{g}}$  mediante inversión de la matriz  $H$ .

#### Iteraciones

1. Se deforma la imagen  $\mathcal{I}$  empleando los parámetros de la transformación  $\mathbf{g}$ .
  - Cálculo de la función  $\mathcal{G}(\mathbf{q}; \mathbf{g})$ :

$$\mathcal{G}(\mathbf{q}; \mathbf{g}) = \frac{1}{h_{31}q_u + h_{32} + h_{33}} \begin{pmatrix} h_{11}q_u + h_{12}q_v + h_{13} \\ h_{21}q_u + h_{22}q_v + h_{23} \end{pmatrix}$$

2. Actualización de parámetros  $\mathbf{g} = \mathbf{g} + \delta_{\mathbf{g}}$ .
- 

Tabla 3.2: Resumen del método de registro directo de imágenes

### 3.4.2. Espacios de color

El color según [Pajares and De la Cruz, 2003] es una de las propiedades de los materiales que depende de las longitudes de onda que el cuerpo es capaz de absorber o reflejar en el espectro visible (véase figura 3.13), de tal manera que, en un objeto blanco ninguna componente frecuencial es absorbida así como en un objeto negro ninguna es reflejada. Un color intermedio es fruto de la reflexión de ciertas longitudes de onda que al combinarse lo generan. Por ejemplo, los objetos verdes reflejan la luz con longitudes de onda en el rango de 500 a 570 nm y absorben gran cantidad de energía en otras.

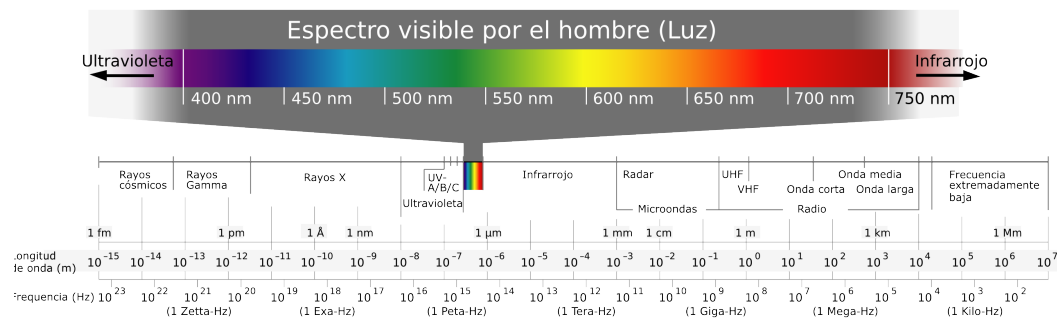


Figura 3.13: Espectro visible

En muchas ocasiones no es práctico manejar el color especificando las longitudes de onda reflejadas y absorbidas, con lo que se debe buscar otro método para su representación; esto es precisamente lo que se resuelve mediante los espacios de color. Según [Quintas et al., 1996], se puede definir espacio de color como las diferentes bases matemáticas que pueden ser útiles para representar información luminosa. Su principal función es la de facilitar la especificación de colores de una forma estandarizada.

Los espacios de color proporcionan un método para especificar, ordenar y manipular colores. Estas representaciones se corresponden con vectores de  $n$  componentes que representan las sensaciones de color. Los colores se representan mediante puntos en estos espacios. Existen numerosos espacios de color en la actualidad. La gran mayoría de ellos se han desarrollado para aplicaciones específicas, aunque todos parten de un mismo concepto: la teoría tricromática de colores primarios rojo, verde y azul.

Aunque el color es una de las características más importantes que definen a los objetos, hasta hace poco tiempo se le prestaba poca atención debido al coste computacional y a la memoria necesaria para procesar imágenes a color. Con el desarrollo del hardware esto ha cambiado. Algunas definiciones básicas para comprender los espacios de color son:

- **Brillo:** sensación que indica si un área está más o menos iluminada.
- **Tono:** sensación que indica si un área parece similar al rojo, amarillo, verde o azul o a una proporción de dos de ellos.
- **Coloración:** sensación por la que un área tiene un mayor o menor tono.

- **Luminosidad:** brillo de una zona respecto a otra blanca en la imagen.
- **Saturación:** la relación entre la coloridad y el brillo

siendo los parámetros psíquicos de la percepción del color la luminosidad, el tono y la saturación.

### 3.4.2.1. El espacio RGB

La descripción RGB (del inglés Red, Green, Blue; “rojo, verde, azul”) de un color hace referencia a la composición del color en términos de la intensidad de los colores primarios con que se forma: el rojo, el verde y el azul. Es un modelo de color basado en la síntesis aditiva, con el que es posible representar un color mediante la mezcla por adición de los tres colores primarios. Indicar que el modelo de color RGB no define, por sí mismo, lo que significa exactamente rojo, verde o azul, razón por la cual los mismos valores RGB pueden mostrar colores notablemente diferentes en diferentes dispositivos que usen este modelo de color. Aunque utilicen un mismo modelo de color, sus espacios de color pueden variar considerablemente.

Para indicar con qué proporción mezclamos cada color, se asigna un valor a cada uno de los colores primarios, de manera que, por ejemplo, el valor 0 significa que no interviene en la mezcla, y a medida que ese valor aumenta, se entiende que aporta más intensidad a la mezcla. Aunque el intervalo de valores podría ser cualquiera (valores reales entre 0 y 1, valores enteros entre 0 y 37, etc.), es frecuente que cada color primario se codifique con un byte (8 bits). Así, de una manera estándar, la intensidad de cada una de las componentes se mide según una escala que va del 0 al 255. Por lo tanto, el rojo se obtiene con (255,0,0), el verde con (0,255,0) y el azul con (0,0,255), obteniendo un color resultante monocromático. La ausencia de color (color negro) se obtiene cuando las tres componentes son 0.

En la figura 3.14 se muestra, de manera gráfica, cómo el conjunto de todos los colores se puede representar en forma de cubo; cada color es un punto de la superficie o del interior de éste. La **escala de grises** es la diagonal que une al blanco y al negro, ya que las tres componentes son iguales.

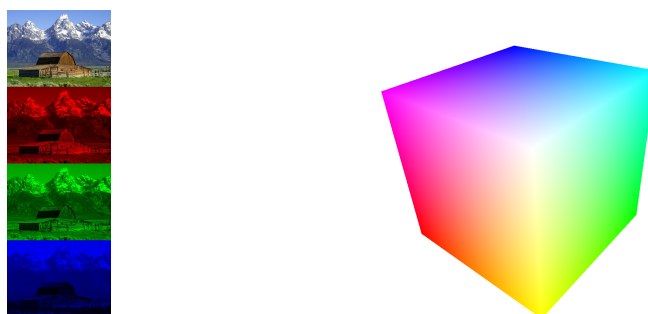


Figura 3.14: Separación de los canales RGB (izquierda) y representación del espacio de color en forma de cubo (derecha).

### 3.4.2.2. Otros espacios de color

**EL ESPACIO HSI** El espacio de color HSI se basa en el modo de percibir los colores que tenemos los humanos. Dicho sistema caracteriza el color en términos de tono o tinte (Hue), saturación o cromatismo (Saturation) y brillo (Intensity); componentes que se muestran favorables de cara a realizar segmentaciones de la imagen en atención al tono o tinte del color. Así, de dos fuentes de luz de mismo espectro, aquella que tenga mayor intensidad aparecerá más brillante. La saturación distingue la blancura de un color (el rojo tiene saturación máxima y el rosa mínima).

Las transformaciones matemáticas que permiten el paso del espacio RGB al HSI se muestran a continuación:

$$I = \frac{R+G+B}{3}$$

$$H = \arctan \left( \frac{\sqrt{3}(G-B)}{(R-G)(R-B)} \right) \quad (3.43)$$

$$S = 1 - \frac{\min(R,G,B)}{I}$$

Así como el espacio RGB se representa por un cubo, el HSI lo forman dos pirámides unidas por su base. Dependiendo del valor de la intensidad se tendrá un corte con las pirámides obteniéndose un triángulo. Dentro del triángulo la componente H viene definida por su orientación y la saturación la indica la lejanía del centro del círculo, figura 3.15.

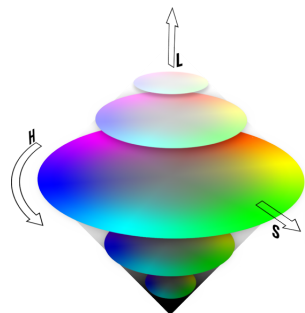


Figura 3.15: Representación del espacio de color HSI.

Dentro de las variaciones de este espacio de color, se destaca, por su utilidad, el modelo HSV. Éste difiere del modelo HSI en que está formado por un único cono en lugar de dos, tal y como se muestra en la figura 3.16.

**EL ESPACIO YUV** Estos espacios se basan en la obtención de la luminancia (luminosidad) y la croma (color). Es utilizado con frecuencia en algunas cámaras de color.

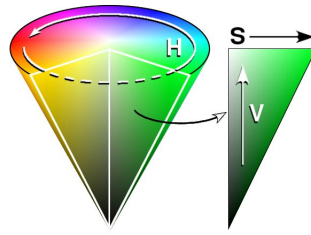


Figura 3.16: Representación del espacio de color HSV.

A partir de la elección de los colores primarios y del nivel de blanco de referencia, se puede obtener la expresión fundamental de la luminancia monocroma como:

$$Y = 0,30R + 0,59G + 0,11B \quad (3.44)$$

Las señales simples de crominancia se obtienen al restar, a las señales primarias, la señal de luminancia Y. Las ecuaciones son:

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0,299 & 0,587 & 0,114 \\ 0,147 & -0,289 & 0,436 \\ 0,615 & -0,515 & -0,10 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (3.45)$$

La figura 3.17 muestra una representación del espacio de color YUV.

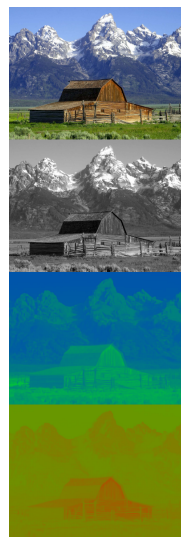


Figura 3.17: Una imagen junto con sus componentes Y, U, y V.



### 3.4.2.3. El espacio invariante a la iluminación

**INTRODUCCIÓN** Las diversas condiciones de iluminación pueden causar problemas en muchos de los algoritmos empleados en visión. Por ejemplo, cambios en el color o la intensidad de la iluminación de una escena pueden causar errores en algoritmos de segmentación, seguimiento o reconocimiento. Un efecto de la iluminación que puede causar problemas en estos algoritmos es la creación de sombras. En el presente proyecto se emplea el método propuesto por [Finlayson et al., 2002] para obtener una imagen unidimensional (o en niveles de gris) invariante a los cambios de iluminación y por tanto carente de sombras, de manera que se mantenga la información proporcionada por el resto de la imagen.

Desafortunadamente, este método requiere de una etapa previa de calibración de las cámaras, en la que es necesario la captura de más de una imagen. Más adelante se verá como se puede conseguir un buen método de calibración simplemente capturando imágenes de una escena a lo largo del curso de un día de manera que la iluminación varíe considerablemente (especialmente la posición de las sombras). Una vez que se ha concluido la fase de calibración, tan solo será necesario una imagen de entrada para la eliminación de sombras.

**FORMACIÓN DE LA IMAGEN:** se parte de una serie de suposiciones bajo las cuales se podrá obtener un método rápido y abreviado para obtener dicho espacio invariante a la iluminación:

1. Se consideran superficies lambertianas.
2. Se toma una aproximación de fuente de luz Planckiana.
3. El dispositivo de captura debe disponer de sensores de banda estrecha, pudiéndose representar como funciones  $\delta$  (solo responden a una longitud de onda, aproximadamente).

donde los términos más relevantes se explican brevemente a continuación:

- **Superficie lambertiana**  $\Rightarrow$  es un reflector perfectamente difuso, donde el brillo aparente es igual en todas las direcciones de vista. Bajo esta suposición, la radiancia recibida por el sensor se asume que es proporcional al coseno del ángulo de incidencia ( $\cos i$ ), que es el ángulo entre la normal a la superficie y el vector solar ( $L_{\mu} \cos i$ ). Tal superficie se denomina lambertiana porque responde a la ley de Lambert [Brown, 1965]. Este efecto se puede observar en la figura 3.18.

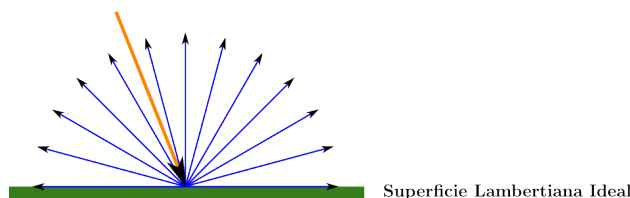


Figura 3.18: Reflexión de una superficie lambertiana ideal.

- **Fuente de luz Planckiana o cuerpo negro**  $\Rightarrow$  radiador térmico que absorbe completamente todas las radiaciones que inciden sobre él, cualquiera que sea su longitud de onda, dirección de incidencia o de polarización. Este radiador posee, para cualquier longitud de onda, la concentración espectral máxima de excitancia radiante a determinada temperatura. La radiación del cuerpo negro está definida por la ley de radiación de Planck. Muestra el incremento de la radiación que se convierte en espectros visible e infrarrojo en función del incremento de la temperatura, de ahí el concepto de temperatura de color. En la figura 3.19 se muestra las diferentes distribuciones espectrales para distintas temperaturas.

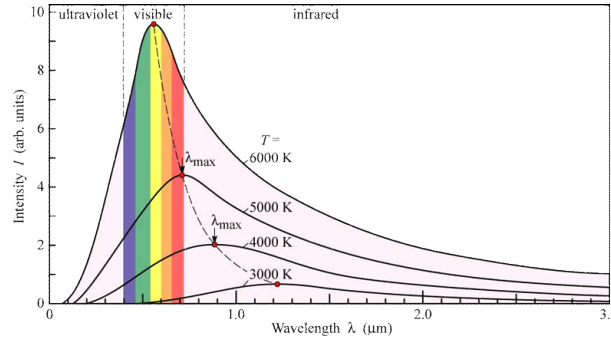


Figura 3.19: Diagrama de radiación de un cuerpo negro de distintas temperaturas.

- **Cámara de banda estrecha**  $\Rightarrow$  aquella que centra la absorción en ciertas longitudes de onda con un ancho de banda infinitamente estrecho, es decir, su respuesta frecuencial se puede modelar por una serie de deltas; véase figura 3.20.

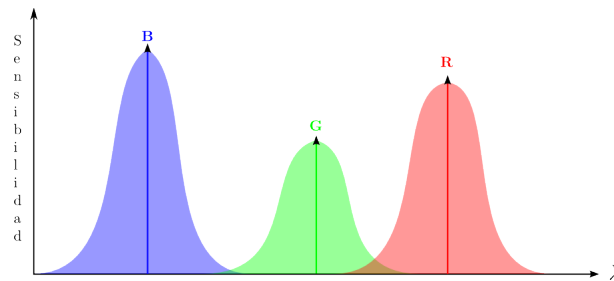


Figura 3.20: Sensibilidad de la cámara. Se representan con unas deltas la respuesta de banda estrecha sobre la reales.

Por tanto, considerando que todas las superficies son lambertianas, que la iluminación sigue el modelo planckiano y que el sensor de la cámara es de banda estrecha, se puede modelar el color RGB obtenido en un píxel por el siguiente modelo físico.

$$\rho_k = \sigma S(\lambda_k) E(\lambda_k, T) Q_k \Delta(\lambda - \lambda_k), \quad k = 1, 2, 3 \quad (3.46)$$

donde:

- $\sigma S(\lambda_k) \Rightarrow$  representa la función de reflectancia espectral de la superficie lambertiana.
- $Q_k \Delta(\lambda - \lambda_k) \Rightarrow$  representa la función de respuesta espectral del sensor para cada canal de color  $k$  (R,G,B) centrado en la longitud de onda  $\lambda_k$ .
- $E(\lambda_k, T) \Rightarrow$  es la distribución espectral de potencia de la luz en el modelo de Planck para un rango muy amplio de temperaturas de color  $T = [2500^\circ, 10000^\circ]$ . Se puede modelar según la siguiente expresión:

$$E(\lambda, T) = I_{c_1} \lambda^{-5} e^{\left(\frac{-c_2}{T\lambda}\right)} \quad (3.47)$$

donde:

- $I \Rightarrow$  representa la intensidad de luz global.
- $c_1, c_2 \Rightarrow$  constantes de valor  $3,7418310^{-16} W m^2$  y  $1,438810^{-2} mK$ , respectivamente.

De acuerdo con el modelo, el valor obtenido por la cámara en un píxel cualquiera  $\rho_k$  se obtiene directamente de la expresión 3.48.

$$\rho_k = \sigma I c_1 (\lambda_k)^{-5} e^{\left(\frac{-c_2}{T\lambda_k}\right)} S(\lambda_k) Q_k \quad (3.48)$$

**CONCEPTOS TEÓRICOS** La transformación que permite la creación de una imagen invariante está basada en el trabajo original de [Finlayson et al., 2002] en el que se desarrolla un método para obtener un espacio invariante a la iluminación a partir de la captura de imagen de color. Este método se sustenta en el modelo de formación de la imagen ya explicado, en el que se asumen superficies lambertianas, sensores de banda estrecha e iluminación planckiana.

Sean los tres canales de color  $\rho = (\rho_1, \rho_2, \rho_3)$  descritos en 3.48, se forman los logaritmos de los ratios de crominancia, en forma de coordenadas  $(\chi_1, \chi_2)$ , como:

$$\chi_1 = \log\left(\frac{\rho_1}{\rho_2}\right) = \log\left(\frac{s_1}{s_3}\right) + \frac{e_1 - e_3}{T}, \quad \chi_2 = \log\left(\frac{\rho_2}{\rho_3}\right) = \log\left(\frac{s_2}{s_3}\right) + \frac{e_2 - e_3}{T} \quad (3.49)$$

donde:

- $e_k = -\frac{c_2}{\lambda_k} \Rightarrow$  solo depende de la respuesta espectral de la cámara y no de la superficie considerada.
- $s_k = c_i (\lambda_k)^{-5} S(\lambda_k) Q_k \Rightarrow$  no depende de la temperatura de color  $T$ .

Estas dos coordenadas  $(\chi_1, \chi_2)$  caen en una línea con dirección  $\bar{e} = (e_1 - e_3, e_2 - e_3)$ . De este modo, una superficie iluminada a distintas temperaturas de color  $T$  genera una recta con dirección  $\bar{e}$  en un espacio de dos dimensiones con ejes  $(\chi_1, \chi_2)$ , véase figura 3.21.

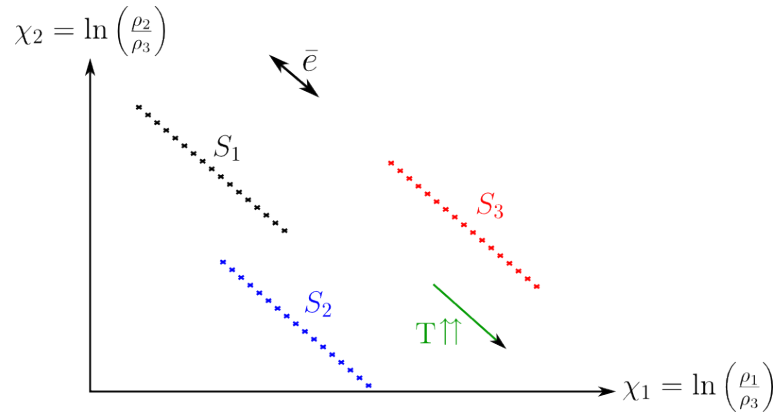


Figura 3.21: Representación de los valores de  $\chi$  en 3 superficies, para diferentes temperaturas de color.

Se puede formar un espacio invariante a la iluminación proyectando cualquier vector  $\chi$  en la línea ortogonal definida por  $\bar{e}^\perp = (\cos(\theta), \sin(\theta))$ . Así, dos píxeles provenientes de una misma superficie pero vistos bajo diferentes tipos de iluminación serán proyectados en el mismo punto; tal y como se muestra en la figura 3.22.

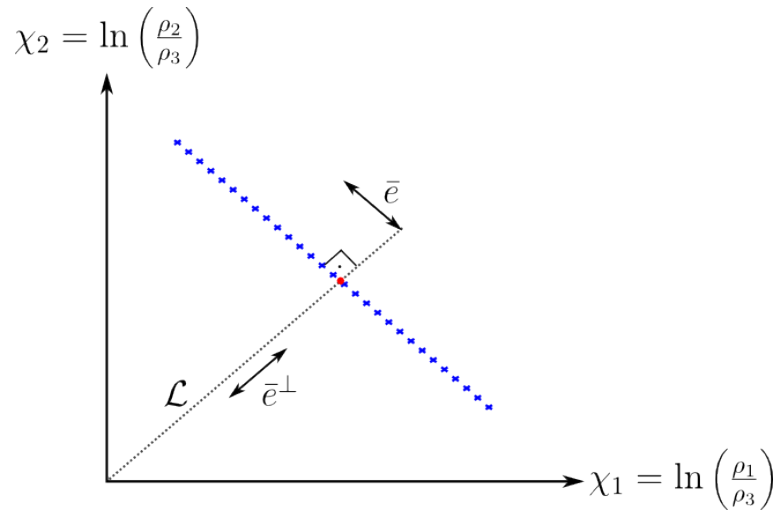


Figura 3.22: Proyección de los puntos de una superficie iluminada a distintas temperaturas sobre la recta definida por el vector  $\bar{e}^\perp$ .

En la figura 3.23 se representan la respuesta de 3 tipos de superficies ( $S_1, S_2, S_3$ ) bajo diferentes tipos de iluminación, así como su proyección sobre la recta de dirección  $\bar{e}^\perp$ . Se observa claramente como, para una misma superficie bajo diferentes tipos de iluminación, se obtiene el

mismo valor para la transformación al espacio invariante ( $\mathcal{L}_i$ ); cambiando dicho valor al variar el tipo de superficie.

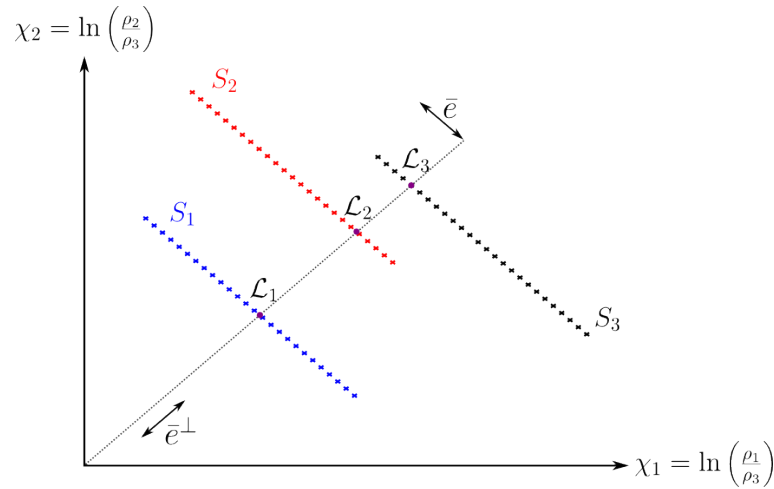


Figura 3.23: Representación de los logaritmos de los ratios de crominancia en 3 superficies, bajo diferentes tipos de iluminación.

Para reducir la elección aleatoria de los ratios de crominancia, se puede emplear un método en el que se obtiene la media geométrica de los tres canales de color  $\sqrt[3]{\rho_1 \rho_2 \rho_3}$  como denominador. Se obtiene un vector de 3 coordenadas linealmente dependientes.

Eligiendo la descomposición apropiada, se obtiene un vector bidimensional equivalente ( $\chi$ ) en el interior de la recta invariante, caracterizada por el ángulo de su pendiente ( $\theta$ ):

$$\mathcal{L}(\rho, \theta) = \chi_1(\rho) \cos(\theta) + \chi_2(\rho) \sin(\theta) \quad (3.50)$$

Esta proyección equipara dos colores correspondientes a puntos vistos bajo diferentes tipos de iluminación. De este modo, se relaciona un color  $\rho$  con su correspondiente representación invariante a la iluminación.

Si se transforma el valor de cada píxel de una imagen a color  $\mathcal{I}$  como una entrada en 3.50, el resultado de  $\mathcal{L}(\mathcal{I}, \theta)$  es una imagen unidimensional invariante a sombras. Por lo tanto, la transformación es global e independiente de la posición del píxel  $q \in \mathbb{R}^2$ ; solo depende del valor de su color.

**CALIBRACIÓN DEL PARÁMETRO  $\theta$**  Como se ha expuesto con anterioridad, el uso adecuado del espacio invariante requiere de una etapa previa de calibración de manera que se obtenga el ángulo  $\theta$  del vector  $\vec{e}^\perp$ ; es decir, el ángulo de la recta perpendicular a las descritas en el apartado previo. En la figura 3.24 se muestra un diagrama modular de los procesos seguidos en la obtención del espacio invariante.

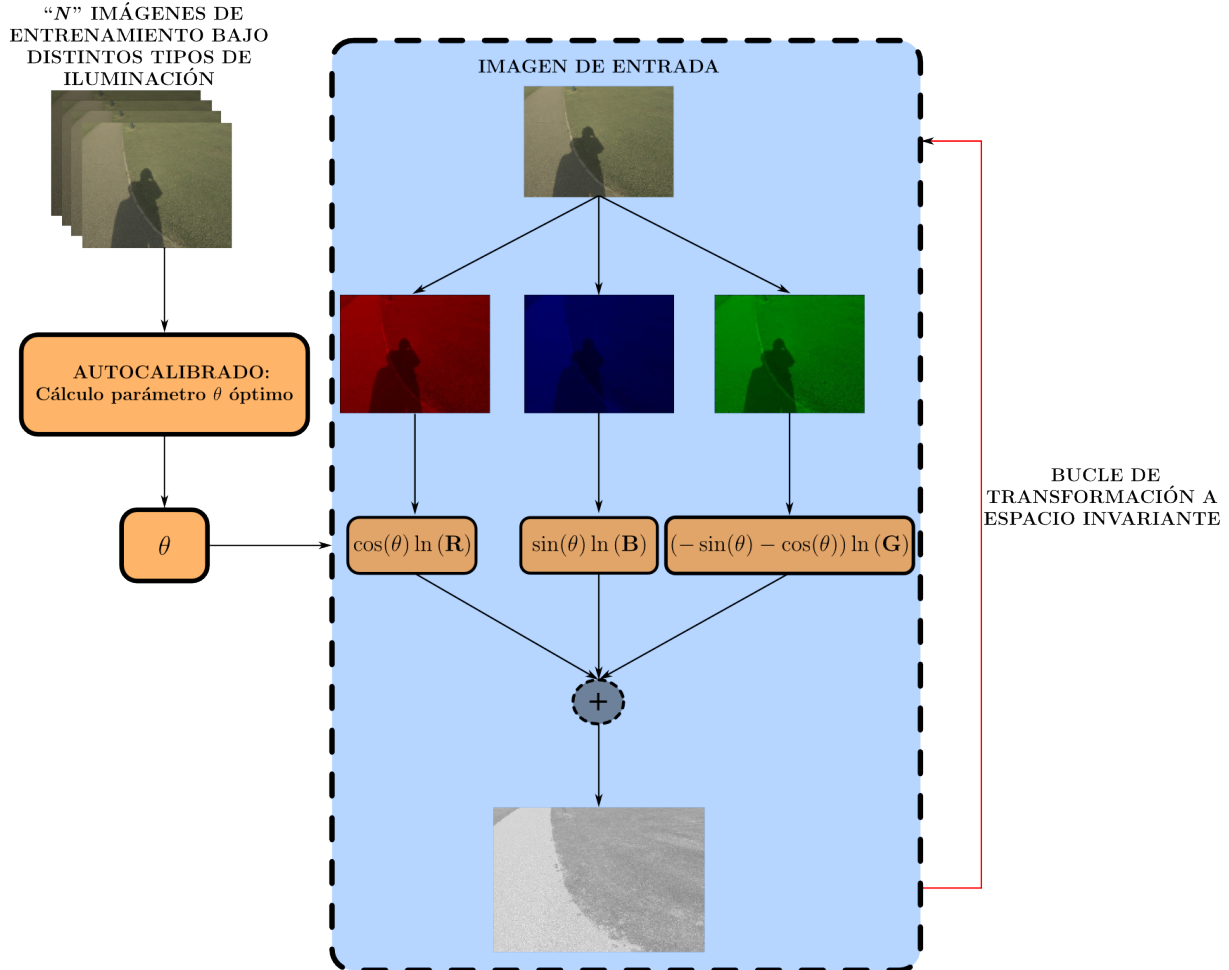


Figura 3.24: Diagrama modular del proceso de obtención del espacio invariante a la iluminación.

El proceso de calibración del parámetro  $\theta$  se basa en su resolución por mínimos cuadrados; cuyos pasos se indican a continuación:

1. Inicialmente, se parte de la disposición de un conjunto de "N" imágenes, capturadas a partir de una cámara ubicada en un emplazamiento fijo, que se ha sometido a diferentes tipos de iluminación; preferiblemente en la que se pueda observar la evolución de las sombras en la escena.

$$\mathcal{L}(I_i(u, v), \theta) = \log \left( \frac{\rho_1^i(u, v)}{\rho_3^i(u, v)} \right) \cos(\theta) + \log \left( \frac{\rho_2^i(u, v)}{\rho_3^i(u, v)} \right) \sin(\theta), \quad (3.51)$$

$$\begin{cases} i = 1, 2, \dots, N \\ u = 0, 1, \dots, ancho - 1 \\ v = 0, 1, \dots, alto - 1 \end{cases}$$

donde:

- $ancho$  y  $alto \Rightarrow$  hacen referencia a la resolución de la cámara en los ejes  $u$  y  $v$  respectivamente.
- $\rho_k^i(u, v) \Rightarrow$  referencia al modelado del color RGB obtenido en un píxel de coordenadas  $(u, v)$  para la imagen “ $k$ ”. En este caso práctico se propone la siguiente distribución:
  - $\rho_1 \Rightarrow$  corresponde el espacio de color R (rojo).
  - $\rho_2 \Rightarrow$  corresponde el espacio de color B (azul).
  - $\rho_3 \Rightarrow$  corresponde el espacio de color G (verde).

Se ha de tener en cuenta que, de ahora en adelante, se hará referencia al proceso seguido para todas las coordenadas de la imagen y para un mismo píxel de cada una de las “ $N$ ” de entrenamiento; de esta manera, se evita tener que especificar en la notación las coordenadas  $(u, v)$  para cada una de las imágenes.

2. Bajo la suposición de que la transformación al espacio invariante de todas las imágenes es la misma:

$$\mathcal{L}(I_1, \theta) = \mathcal{L}(I_2, \theta) = \dots = \mathcal{L}(I_N, \theta) \quad (3.52)$$

y tomando como imagen de referencia la media de las  $N$  imágenes empleadas:

$$I_{ref} = \frac{1}{N} \sum_{i=0}^N I_i \quad (3.53)$$

se puede encontrar el valor de  $\theta$  que minimiza las relaciones expresadas en 3.54; dado que todos los términos  $I_i$  en el espacio invariante son iguales, la media de todos ellos es igual a cualquiera por separado, es decir, la media de los valores en el espacio invariante para un mismo píxel en “ $N$ ” imágenes debe ser igual a cualquiera de sus valores individualmente:

$$\begin{cases} \mathcal{L}(I_{ref}, \theta) - \mathcal{L}(I_1, \theta) = 0 \\ \mathcal{L}(I_{ref}, \theta) - \mathcal{L}(I_2, \theta) = 0 \\ \vdots \\ \mathcal{L}(I_{ref}, \theta) - \mathcal{L}(I_N, \theta) = 0 \end{cases} \quad (3.54)$$

donde:

- $\theta \Rightarrow$  referencia al parámetro objeto de la fase de calibración; es el ángulo que forma el vector  $\bar{e}^\perp$ .
- $I_i \Rightarrow$  referencia a la imagen de color  $i$  empleada en la calibración del parámetros  $\theta$ ; al tratamiento para cada uno de sus píxeles.

- $I_{ref} \Rightarrow$  es la imagen de referencia calculada como la media de las  $N$  imágenes de entrenamiento; para un mismo píxel genérico en todas.
- $\mathcal{L}(I_x, \theta) \Rightarrow$  referencia la transformación al espacio invariante de un píxel genérico de una imagen  $x$  para un ángulo  $\theta$ .

3. Si se expresa las relaciones anteriores en modo matricial se obtiene que:

$$\begin{pmatrix} \left[ \log \left( \frac{R_{ref}}{G_{ref}} \right) \cos(\theta) + \log \left( \frac{B_{ref}}{G_{ref}} \right) \sin(\theta) \right] - \left[ \log \left( \frac{R_1}{G_1} \right) \cos(\theta) + \log \left( \frac{B_1}{G_1} \right) \sin(\theta) \right] \\ \vdots \\ \left[ \log \left( \frac{R_{ref}}{G_{ref}} \right) \cos(\theta) + \log \left( \frac{B_{ref}}{G_{ref}} \right) \sin(\theta) \right] - \left[ \log \left( \frac{R_N}{G_N} \right) \cos(\theta) + \log \left( \frac{B_N}{G_N} \right) \sin(\theta) \right] \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \quad (3.55)$$

aplicando la propiedades de los logaritmos se llega a 3.56:

$$\begin{pmatrix} \cos(\theta) \left( \log \left( \frac{R_{ref}}{G_{ref}} \right) - \log \left( \frac{R_1}{G_1} \right) \right) + \sin(\theta) \left( \log \left( \frac{B_{ref}}{G_{ref}} \right) - \log \left( \frac{B_1}{G_1} \right) \right) \\ \vdots \\ \cos(\theta) \left( \log \left( \frac{R_{ref}}{G_{ref}} \right) - \log \left( \frac{R_1}{G_1} \right) \right) + \sin(\theta) \left( \log \left( \frac{B_{ref}}{G_{ref}} \right) - \log \left( \frac{B_1}{G_1} \right) \right) \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \quad (3.56)$$

La expresión 3.56 se puede expresar como:

$$\begin{pmatrix} \log \left( \frac{R_{ref} G_1}{G_{ref} R_1} \right) & \log \left( \frac{B_{ref} G_1}{G_{ref} B_1} \right) \\ \vdots & \vdots \\ \log \left( \frac{R_{ref} G_N}{G_{ref} R_N} \right) & \log \left( \frac{B_{ref} G_N}{G_{ref} B_N} \right) \end{pmatrix} \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \quad \mapsto A\Phi = 0 \quad (3.57)$$

donde:

- $R_i, B_i, G_i \Rightarrow$  corresponden a los canales  $R, G, B$  de una imagen  $i$  respectivamente; para cada píxel.
- $A \Rightarrow$  representa la matriz de dimensiones  $N \times 2$  empleada en la función de minimización.
- $\Phi \Rightarrow$  es el vector de incógnitas objeto de minimización:

$$\Phi = \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix} \quad (3.58)$$



4. El último paso consiste en obtener el vector de incógnitas  $\Phi$  mediante el método de mínimos cuadrados.

Por efecto del ruido y no idealidad de las cámaras, la parte derecha de la ecuación 3.57 no es exactamente un array de ceros, sino un array de valores próximos a cero, ecuación 3.59.

$$A\Phi = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{pmatrix} = \bar{\varepsilon}, \quad \varepsilon_i \Rightarrow 0 \quad (3.59)$$

Los términos  $\varepsilon_i$  dependen de las incógnitas buscadas,  $\Phi$ . En concreto, las incógnitas deben minimizar el sumatorio de los cuadrados de los valores  $\varepsilon_i$ , ecuación 3.60.

$$\sum_i \varepsilon_i^2 = \bar{\varepsilon}^T \cdot \bar{\varepsilon} \quad (3.60)$$

Ahora bien, dado que  $\bar{\varepsilon}^T \cdot \bar{\varepsilon} = \Phi^T A^T A \Phi$ , entonces el proceso se centra en minimizar  $\Phi^T A^T A \Phi$ .

Bajo la aproximación  $\|\Phi\| = 1$  y aplicando el método de Lagrange, se llega a la siguiente ecuación:

$$A^T A \Phi = \lambda \Phi \quad (3.61)$$

donde  $\lambda$  es conocido como multiplicador de Lagrange.

Según la definición de autovector, en la ecuación 3.61,  $\lambda$  corresponde al autovalor asociado al autovector  $\Phi$  de la matriz  $A^T A$ . Es decir, las incógnitas buscadas son justamente el autovector de la matriz  $A^T A$ . Puesto que dicha matriz es de dimensión  $2 \times 2$ , entonces existen dos autovectores:  $\Phi_1, \Phi_2$ , así como sus dos autovalores asociados:  $\lambda_1, \lambda_2$ , respectivamente. Las incógnitas buscadas son el autovector asociado al menor autovalor.

En la figura 3.25 se muestra el proceso de calibración del ángulo  $\theta$  óptimo.

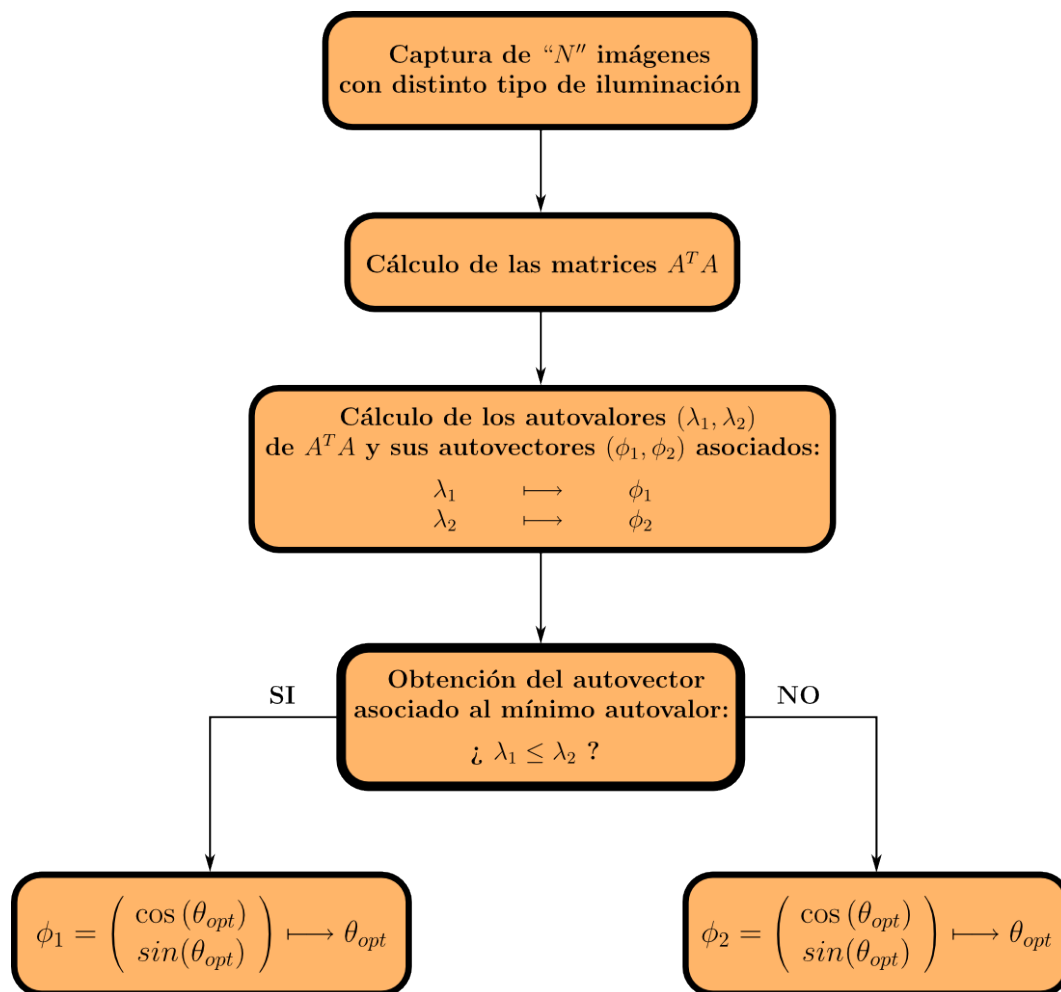


Figura 3.25: Diagrama modular del proceso de calibración del ángulo  $\theta$

### 3.4.3. Segmentación de fondo

#### 3.4.3.1. Introducción

Según [Velez et al., 2003], se entiende por segmentación el proceso de reconocimiento mediante el cual se pretende identificar los elementos u objetos que existen en una escena, separarlos del medio en el que se encuentran y distinguirlos entre sí. Se trata, por tanto, de dividir una imagen digital en regiones homogéneas con respecto a una o más características (brillo, color, etc..) con el fin de facilitar un posterior análisis o reconocimiento automático. Localizar un objeto dentro de una escena o encontrar los límites de una palabra dentro de una imagen de un texto, constituyen ejemplos de problemas de segmentación.

En este capítulo se estudiarán diferentes enfoques para realizar el proceso de segmentación hasta llegar a la propuesta definitiva. En la práctica se demuestra que la segmentación no tiene reglas estrictas a seguir, y dependiendo del problema en cuestión, puede ser necesario idear técnicas a medida.

Debe verse como un proceso que a partir de una imagen, produce otra en la que cada píxel tiene asociada una etiqueta distintiva del objeto al que pertenece. Así, una vez segmentada una imagen, se podría formar una lista de objetos consistentes en las agrupaciones de los píxeles que tengan la misma etiqueta.

La segmentación termina cuando los objetos extraídos de la imagen se corresponden unívocamente con las distintas regiones disjuntas a localizar en la misma. En este caso se habla de segmentación completa de la escena o imagen y en el caso contrario, de segmentación parcial. En una escena compleja, el resultado de la segmentación podría ser un conjunto de regiones homogéneas superpuestas y en este caso, la imagen parcialmente segmentada deberá ser sometida después a un tratamiento posterior con el fin de conseguir una segmentación completa.

En general, el proceso de la segmentación suele resultar complejo debido, por un lado, a que no se tiene una información adecuada de los objetos a extraer y, por otro, a que en la escena a segmentar aparece normalmente ruido. Este hecho se ve agravado cuando la escena de fondo se encuentra emplazada en entornos exteriores; factores como cambios de iluminación, aparición de sombras o pequeñas vibraciones en las cámaras hacen que la casuística de este tipo de entornos aumente considerablemente. Es por esto que el uso de conocimiento sobre el tipo de imagen a segmentar o alguna otra información de alto nivel puede resultar muy útil para conseguir la segmentación de la imagen.

Algunos ejemplos típicos de procesos de segmentación son:

- Extraer los vehículos que aparecen en una imagen de una carretera.
- Tratar de separar los caracteres que forman una palabra dentro de una imagen de un texto.
- Detectar ciertos tipos de células en imágenes médicas.

Los diferentes objetos que aparecen en una imagen pueden ser localizados atendiendo a aspectos como: sus contornos o su textura. Cada una de las técnicas existentes atienden a alguna de estas características, y para su estudio han sido englobadas en tres grupos: técnicas basadas en umbralización, basadas en detección de los contornos de los objetos y técnicas basadas en propiedades locales de las regiones.

### 3.4.3.2. Métodos de segmentación

A continuación se exponen algunas de las técnicas más empleadas:

- **Segmentación basada en umbralización**  $\Rightarrow$  la umbralización es un proceso que permite convertir una imagen de niveles de gris o de color en una imagen binaria, de tal forma que los objetos de interés se etiqueten con un valor distinto al de los píxeles del fondo. Se trata de una técnica de segmentación rápida, que tiene un coste computacional bajo y que puede ser realizada en tiempo real durante la captura de la imagen usando un computador personal de propósito general. Por otro lado, el histograma de una imagen no tiene en cuenta la información espacial sino solamente la distribución de grises en la imagen. Por ello, dos imágenes muy diferentes pueden tener el mismo histograma. Esto hace que, los métodos de segmentación basados en la umbralización, como único medio de segmentación, resulten limitados en muchos problemas reales. Aunque sí se usan con frecuencia como complemento de otros métodos.
- **Técnicas basadas en la detección de contornos**  $\Rightarrow$  este método agrupa un gran número de técnicas que usan la información proporcionada por las fronteras de los objetos que aparecen en una imagen. Puesto que se desea encontrar los objetos individuales presentes en una imagen, parece lógico que si se encuentran las fronteras de tales objetos con el fondo se podría segmentar los objetos de la escena general.
- **Técnicas basadas en crecimiento de regiones**  $\Rightarrow$  determinan zonas dentro de una imagen basándose en criterios de similitud y proximidad entre los píxeles de la misma. En estas técnicas la homogeneidad (o falta de homogeneidad) entre regiones adyacentes es el criterio utilizado para unir (o dividir) regiones de la imagen. Dicha homogeneidad se puede definir a partir de criterios como: el nivel de gris medio, el color, la forma, etc.. El resultado de la segmentación es una partición de la imagen en regiones homogéneas.

Dada la complejidad de los entornos exteriores, los métodos de segmentación anteriormente descritos presentan múltiples inconvenientes; alejándose de poder proporcionar una solución robusta y fiable para la identificación de objetos en la escena. El método propuesto plantea un esquema de segmentación basándose en un enfoque estadístico. Dicho enfoque está basado, como se detalla en [Allili et al., 2007a], en modelar la escena de fondo como una mezcla de distribuciones multimodales gaussianas, de manera que ante la aparición de un objeto se pueda discernir, a nivel de píxel, si pertenece o no al modelo de fondo.

### 3.4.3.3. Problemática asociada a entornos exteriores

La segmentación de objetos para cualquier tipo de entorno y condición es un problema complicado de resolver de forma general. Por ejemplo, en exteriores, donde la iluminación, las sombras, los reflejos y las condiciones atmosféricas pueden llegar a ser muy extremas, se requiere un tipo de procesamiento muy diferente al que es necesario en entornos de interior, en los que pueden darse unas condiciones más controladas. Aun así, en este último caso tampoco se trata de un problema trivial. Además, se tendrá en cuenta la aparición de ruido en los mismos píxeles de imágenes estáticas capturadas en diferentes instantes de tiempo.

El **ruido** presente en la mayoría de las imágenes provoca una oscilación en el valor de sus píxeles alrededor de su valor real. Así, una cámara que enfocará a una imagen estática en dos instantes diferentes de tiempo presentaría diferencias en el valor de sus píxeles, cuya magnitud variará en función de la agresividad del ruido y de la calidad de la cámara empleada. Este ruido se modela, en la mayoría de los casos, como gaussiano y por tanto el valor de un píxel en la imagen vendrá dado por la siguiente función densidad de probabilidad:

$$f(p) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(p-m)^2}{2\sigma^2}}$$

donde:

- $m \Rightarrow$  corresponde al valor medio.
- $\sigma \Rightarrow$  hace referencia a la desviación típica.

Asimismo, se hace importante tener presente las condicionantes que impone un sistema que debe trabajar en entornos exteriores, escenario del sistema de visión desarrollado, y por tanto estar sometido a **variaciones de iluminación y de condiciones climatológicas**. Enmarcado dentro de este tipo de fenómenos, se tendrá que prestar especial atención a la presencia de brillos, reflejos, sombras, etc., ya que pueden tener incidencia en la fiabilidad de la detección.

Se pueden distinguir dos tipos de **sombras**, cuyo tratamiento debe ser completamente diferente:

1. Sombras proyectadas  $\Rightarrow$  hace referencia a áreas en el fondo proyectadas por objetos en la dirección de los rayos de luz, produciendo siluetas distorsionadas del objeto en cuestión. Puesto que representan una modificación sobre el fondo, estas zonas serán claras candidatas a ser detectadas como objetos, cuando en realidad debieran estar contenidas en el fondo. En la figura 3.26 se clarifica el fenómeno producido por las sombras proyectadas por un objeto.
2. Sombras propias  $\Rightarrow$  se refieren a zonas pertenecientes a objeto que no han sido iluminadas. Una buena propuesta de eliminación de sombras no debería incluirlas como parte del fondo, ya que pertenecen a la silueta del objeto a detectar, véase la figura 3.27.



Figura 3.26: Ejemplos de sombra proyectada por una persona presente en la escena (izquierda) y las sombras proyectadas en la escena por varias personas que no se encuentran en su interior (derecha).



Figura 3.27: Imagen en la que se observan sombras proyectadas en la cara de una persona y que no pertenecen al fondo.

De manera análoga a las sombras, un ***cambio de iluminación*** puede suponer que la imagen completa cambie sin haber ningún objeto presente. Las causas que pueden provocar este tipo de fenómenos son múltiples y en ocasiones acontecen súbitamente (p. ejem. ante la oclusión del sol por la presencia de nubes). Además, este efecto se produce de forma natural con el paso de las horas; del día a la noche. En la figura 3.28 se observa el efecto producido en la imagen ante la llegada repentina de un banco de nubes.



Figura 3.28: Ejemplo del cambio de iluminación sufrido en una escena ante la llegada de un banco de nubes.

Quizás, el efecto menos pernicioso, debido a la ubicación del sistema de detección, sea el provocado por las posibles *reflexiones*. Este efecto genera como resultado una imagen especular de la original que, al igual que las sombras, pueden provocar una detección errónea. Normalmente, son en las superficies pulidas o mojadas y los espejos donde se presenta el efecto de la reflexión, de ahí que solo cuando el paso a nivel (zona a ubicar el sistema de detección) presente su superficie mojada, será cuando este efecto pueda llegar a aparecer. En la figura 3.29 se muestran varios ejemplos del resultado de trabajar con este tipo de superficies.



Figura 3.29: Las dos primeras imágenes muestran el reflejo producido en una superficie mojada. La tercera y cuarta imagen representan los reflejos en espejo y suelo pulido, respectivamente.

#### 3.4.3.4. Distribuciones multimodales gaussianas

Los modelos finitos de mezcla de distribuciones gaussianas (MoG: “Mixture of Gaussian distribution”) han sido y son ampliamente utilizados en diversos campos de visión por computador y procesamiento digital de imágenes. El empleo de este modelo basado en una aproximación a las técnicas de agrupamiento o “clustering” presenta ventajas importantes: la selección del número de clases o la evaluación de la validez del modelo se pueden expresar de una manera formal. Sin embargo, es conocido que las distribuciones gaussianas presentan una serie de inconvenientes, como la rigidez de sus formas, impidiendo ser una buena aproximación en los casos reales. Por este motivo, muchas investigaciones comienzan a emplear distribuciones Gaussianas Generalizadas (GGD: Generalized Gaussian Density) por su flexibilidad a modelar dichas distribuciones de diferentes formas. En las aplicaciones orientadas a la segmentación de imágenes, donde los datos se concentran en varios “clusters”, se requiere una representación del modelo probabilístico con múltiples componentes como una mezcla de distribuciones, tal y como se muestra en [Allili et al., 2007b].

El modelo de mezcla de distribuciones Gaussianas generalizadas, denotado según la literatura más común como MoGG, posee mayor flexibilidad en el ajuste de la morfología de las distribuciones a los datos disponibles, permitiendo disponer de un modelo robusto y efectivo para el modelado de datos ante la presencia de ruido. Se entiende por robustez, a la habilidad del modelo de representar con precisión la forma de los datos con menor sensibilidad y no sobre dimensionar el número de clusters ante la presencia de ruido.

**MoGG: Mezcla de GGD** Una distribución GGD unidimensional para una variable  $X \in \mathbb{R}^2$  se define como sigue:

$$p(X/\mu, \sigma, \lambda) = A(\lambda) e^{(-B(\lambda) \left| \frac{X-\mu}{\sigma} \right|^\lambda)} \quad (3.62)$$

donde:

- $\mu \Rightarrow$  es la media de la función densidad de probabilidad (fdp).
- $\sigma \Rightarrow$  es la desviación típica de la función densidad de probabilidad.
- $\lambda \geq 1 \Rightarrow$  controla la cola de la fdp así como la anchura de la misma; cuanto mayor sea éste parámetro más ancha será la fdp.

Se ha de tener en cuenta que las distribuciones gaussianas y laplaciana son un caso particular de las GGD donde  $\lambda = 1$  o  $2$ , respectivamente. En la figura 3.30 se muestran diferentes GGD's para distintos valores de  $\vec{\lambda}$ .

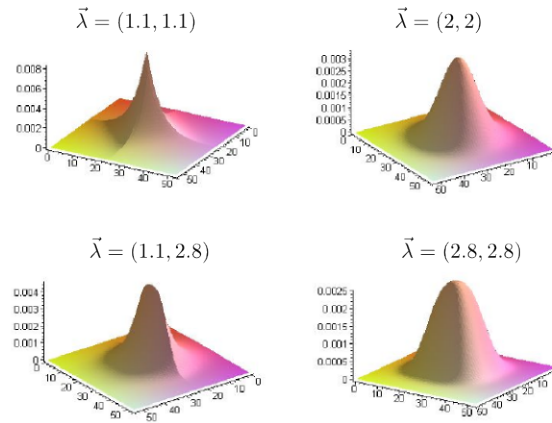


Figura 3.30: Representaciones de diferentes GGD's en función de los parámetros  $\vec{\lambda}$ . Los parámetros  $\vec{\mu}$  y  $\vec{\sigma}$  se fijan con valores  $(23, 24)$  y  $(7, 7)$  respectivamente.

Una generalización de 3.62 para el caso multi-dimensional no es nada trivial. Para poder seguir proporcionando la flexibilidad existente en los modelos probabilísticos GGD, se ha de encontrar una solución que permita, de la misma manera, modelar cada dimensión con una morfología diferente. En la práctica, sin embargo, es impracticable si los datos están correlados. Para mantener las propiedades de su forma, se parte de la suposición de la independencia entre dimensiones. Sea un vector de  $n$  dimensiones  $\vec{X} = (X_1, \dots, X_n)$ , entonces la probabilidad del vector  $\vec{X}$  en una GGD se expresa como 3.63:

$$p(\vec{X}/\vec{\mu}, \vec{\sigma}, \vec{\lambda}) = \prod_{k=1}^n A(\lambda_k) e^{(-B(\lambda_k) \left| \frac{X_k - \mu_k}{\sigma_k} \right|^{\lambda_k})} \quad (3.63)$$

donde  $\vec{\mu} = (\mu_1, \dots, \mu_n)$ ,  $\vec{\sigma} = (\sigma_1, \dots, \sigma_n)$  y  $\vec{\lambda} = (\lambda_1, \dots, \lambda_n)$  son vectores  $n$ -dimensionales.



### 3.4.3.5. Algoritmo de clasificación o “clustering”

Existen diferentes algoritmos que permiten clasificar los elementos que aparecen dentro de una escena para poder entenderla; éstos reciben el nombre de algoritmos de clasificación. Su objetivo es el de distinguir entre objetos diferentes de un conjunto predefinido llamado universo de trabajo. Normalmente, el universo de trabajo se considera dividido en una colección de “ $K$ ” clases  $(\alpha_1, \dots, \alpha_k)$ , perteneciendo o no, los diferentes objetos, a algunas de estas clases.

Dentro de los algoritmos de clasificación se destacan los algoritmos de agrupación de clases; estos se suelen utilizar cuando no existe conocimiento a priori de las clases en que se pueden distribuir los objetos, cuando las clases no son interpretables por un humano, o cuando el número de clases es muy elevado para un procesamiento no automático. Entre los algoritmos más empleados cabe destacar:

- Algoritmo de distancias encadenadas.
- El algoritmo MaxMin.
- El algoritmo de las “K-medias”.
- Los dendrogramas o algoritmos de clustering jerárquico, ineficaces en problemas en los que la muestra tiene un tamaño grande.

En el presente trabajo se opta por emplear el algoritmo k-medias. Este algoritmo goza de una amplia difusión debido a su simplicidad, a la fiabilidad de sus resultados, y a que no depende de ningún umbral heurístico. Sin embargo, es preciso conocer de antemano el número de clases en que se divide la muestra; aunque esto, en las ocasiones en que se dispone de tal dato, es más una ventaja que un inconveniente. También debe observarse que la solución obtenida puede ser dependiente de la elección que se haga de los patrones iniciales, por lo que debe ponerse cuidado en este punto. Esta variación del resultado dependiendo de los patrones iniciales ocurre sobre todo cuando la distinción entre clases no está muy clara, o cuando no se estima adecuadamente el número de clases.

El algoritmo K-medias, propuesto en [Hartigan and Wong, 1979], permite determinar la posición de  $k$  centroides que distribuyan de manera equitativa un conjunto de patrones. En la figura 3.31 se muestra un ejemplo de aplicación del algoritmo de k-medias sobre un conjunto de patrones formados por dos características (por lo que se pueden presentar en un plano). El problema es tal que inicialmente no se conoce qué patrones pertenecen a cada una de las clases, pero se sabe a priori que la muestra está dividida en dos clases, por lo que se aplica el algoritmo de k-medias con  $k=2$ , tomándose, como centroides, dos patrones al azar en el instante  $t=1$ . En la figura se muestra en gris la zona en la que están los patrones que están más próximos a ese centroide. Serán los patrones de cada una de las zonas los que se usen para calcular el nuevo centroide en la siguiente iteración. En los instantes sucesivos se aprecia como los centroides van viajando hasta su ubicación definitiva. En la tercera iteración los centroides alcanzan una posición estable que no cambiará en sucesivas iteraciones, por lo que el algoritmo finaliza.

La tabla 3.3 resume el funcionamiento del algoritmo k-medias.

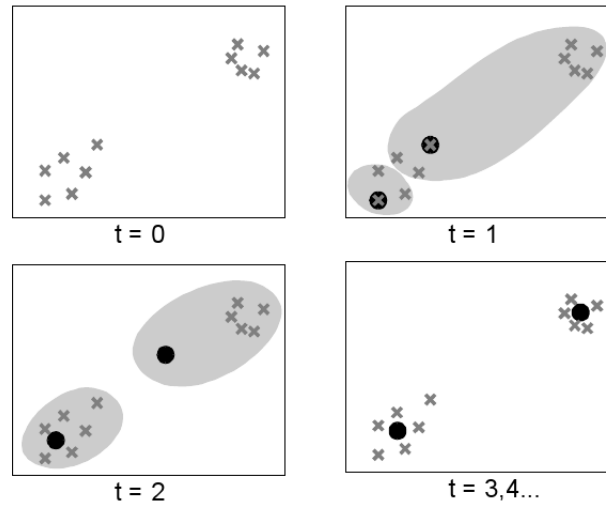


Figura 3.31: Ilustración de la posición de los centroides (puntos negros) al aplicar el algoritmo de k-medias (con  $k=2$ ), durante varias iteraciones para una muestra compuesta de vectores bidimensionales (cruces grises).

### ALGORITMO K-MEDIAS

Paso 1.- Se hace  $t = 1$ . Se toman al azar  $k$  vectores de los  $N$  existentes y se convierten en centroides de cada una de las  $k$  clases respectivamente.

$$\begin{aligned} m_1(1) & \text{ de } \alpha_1 \\ & \vdots \\ m_k(1) & \text{ de } \alpha_k \end{aligned} \quad (3.64)$$

Paso 2.- Se distribuyen las  $N$  muestras entre las  $k$  clases. Se asigna cada vector  $x$  a la clase cuyo centroide esté más próximo.

$$x \in \alpha_j \Leftrightarrow |x - m_j(t)| < |x - m_i(t)| \quad \forall i = 1, 2, \dots, k \quad / \quad i \neq j \quad (3.65)$$

Paso 3.- Se calculan los centroides de las clases como la media ponderada de los vectores de cada clase  $\alpha_j$ .

Paso 4.- Si alguno de los  $k$  centroides  $m_k(t)$  es distinto de los nuevos centroides  $m_k(t+1)$  hacer  $t = t + 1$  e ir al paso 2; en otro caso finalizar el algoritmo.

Tabla 3.3: Resumen del algoritmo K-medias dividido en 4 pasos

### 3.4.3.6. Segmentación mediante distancia estadística

En las situaciones en que los vectores de alguna clase presenten una dispersión significativa respecto a la media, o en aquéllas en las que no existe posible separación lineal entre las clases, puede ofrecer mejores resultados la sustitución de la distancia euclídea por la **distancia de Mahalanobis**. Esta medida, que tiene en cuenta la desviación típica de los vectores de características de los patrones de la muestra, puede proporcionar regiones de separación entre clases que sigan curvas cónicas. Por ello, este clasificador ofrece más garantías al tratar de separar patrones para los que no se encuentra una separación lineal. Además, proporciona probabilidades de pertenencia a las clases, adquiriendo un carácter no determinista.

La distancia de Mahalanobis ( $d_{maha}$ ) de un punto  $X$  a una clase  $\alpha_k$  viene dada por la expresión 3.66:

$$d_{maha} = (X - m_k) \Sigma_k^{-1} (X - m_k) \quad (3.66)$$

donde:

- $m_k \Rightarrow$  corresponde a la media de la clase  $k$ .
- $\Sigma_k \Rightarrow$  hace referencia a la matriz de covarianzas de la clase  $k$ .

Cuando existe diferente dispersión de los valores de una característica en dos o más clases, una medida de la distancia que tenga en cuenta la desviación típica de la clase ofrecerá mejores resultados que otra que sólo tenga en cuenta la distancia euclídea entre los centroides de las clases. La figura 3.32 ejemplifica la disposición de los patrones de dos clases respecto de una característica particular. En la figura se aprecia que la función discriminante no equidista de los centroides de  $\alpha_1$  y  $\alpha_2$ , sino que está más cerca del centroide de  $\alpha_2$  porque su desviación típica es menor.

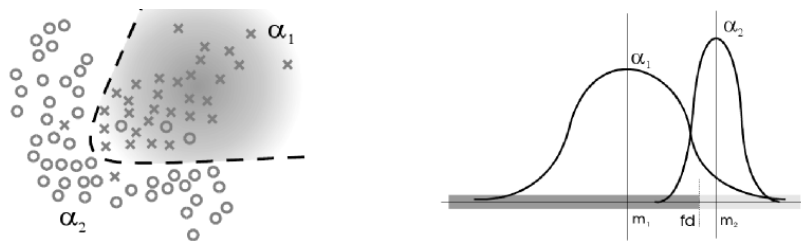


Figura 3.32: Representaciones de la región de separación de dos clases según el criterio de la distancia de Mahalanobis (izquierda) y la dependencia de la función discriminante con la desviación típica de las clases (derecha).

El teorema de Bayes puede ayudar a decidir la función discriminante ( $fd$ ); función que separa el espacio perteneciente a varias clases:

$$P(\alpha_i/X) = \frac{P(X/\alpha_i)P(\alpha_i)}{P(X)} \quad (3.67)$$

donde:

- $P(\alpha_i) \Rightarrow$  es la probabilidad de que un patrón al azar sea de la clase  $\alpha_i$ .
- $P(X) \Rightarrow$  es la probabilidad de que se presente exactamente un patrón con el vector de características  $X$ . Cumpliéndose que:

$$P(X) = \sum_{k=1}^K P(X/\alpha_k)P(\alpha_k)$$

- $P(X/\alpha_i) \Rightarrow$  es la probabilidad de que sabiendo que el patrón a clasificar pertenece a la clase  $\alpha_i$  sea precisamente aquel que tiene el vector de características  $X$ . En otras palabras es la función de densidad de la clase  $\alpha_i$  y se denomina *probabilidad a priori*.
- $P(\alpha_i/X) \Rightarrow$  se conoce como *probabilidad a posteriori* y corresponde a la probabilidad de que un  $X$  conocido pertenezca a la clase  $\alpha_i$ .

Además, aunque el clasificador estadístico sea no determinista, cuando se precise una respuesta concreta puede transformarse en determinista añadiendo la siguiente regla:

$$X \in \alpha_i \Leftrightarrow P(\alpha_i/X) > P(\alpha_j/X) \quad \forall i \neq j, \quad j = 1, 2, \dots, N$$

Como  $P(X)$  es un término constante para todos los cálculos de  $P(\alpha_i/X)$ , el cálculo puede simplificarse eliminando ese factor. Con esto la función discriminante para la clase  $\alpha_i$  es:

$$fd_i(X) = P(X/\alpha_i)P(\alpha_i) \quad \forall i = 1, 2, \dots, N$$

Y se dice que:

$$X \in \alpha_i \Leftrightarrow fd_i(X) > fd_j(X) \quad \forall i \neq j, \quad j = 1, 2, \dots, N$$

Así, la **la regla de decisión** “escoger la clase de la máxima  $fd_i(X)$ ” se puede interpretar como:

**“escoger la clase cuya distancia estadística (Mahalanobis) a la muestra  $X$  sea mínima”**

## 3.5. Solución propuesta

### 3.5.1. Diagrama modular del sistema de visión

La solución propuesta a la segmentación de obstáculos parte de realizar una serie de modificaciones a las técnicas anteriormente planteadas de manera que se adapte a los objetivos propuestos en este tipo de escenarios. La figura 3.33 representa el diagrama general del esquema de segmentación adoptado, en el que se muestran todas las fases necesarias para obtener una imagen binaria con las áreas ocupadas por posibles obstáculos en los pasos a nivel monitorizados. Se distinguen 3 procesos en función del estado en el que se encuentre el sistema:

- 1.- **Proceso de Inicialización Asistido:** esta fase requiere de la intervención de un operario para marcar, en las imágenes provenientes de las cámaras, la región de interés; los puntos seleccionados para ambas cámaras deberán ser aproximadamente los mismos puntos de la escena tridimensional. Se distinguen dos etapas:
  - a) Equilibrado del color de ambas cámaras  $\Rightarrow$  debido a una posible disparidad en la configuración de las cámaras, se realiza una etapa de igualación de parámetros de manera que el valor medio en la captura de ambas cámaras sea el mismo.
  - b) Cálculo de la homografía respecto al plano del suelo  $\Rightarrow$  a partir de los 4 puntos capturados en el punto anterior, se ejecuta un algoritmo de registro directo de imágenes de manera que se minimice el error en la matriz proyectiva, u homografía, que transforma la imagen de una de las cámaras en otra, siendo los puntos pertenecientes al plano del suelo coincidentes.
- 2.- **Proceso de Inicialización No Asistido:** de ahora en adelante, el sistema se ejecuta de manera automática. Se trata de una fase de inicialización de parámetros y modelos previos.
  - Obtención del modelo de fondo; “K-medias modificado (KAM)”  $\Rightarrow$  el modelo de segmentación propuesto está basado en el algoritmo k-medias, ampliamente utilizado como técnica de “*clustering*”; proceso de organización de objetos dentro de grupos cuyos miembros poseen propiedades comunes. Éste requiere de una etapa de inicialización en la que se obtendrá, píxel a píxel, el modelo probabilístico del fondo de la escena como una distribución estadística caracterizada por sus “k” *cluster* (media, covarianza y probabilidad a priori), pudiendo así identificar si la imagen de entrada pertenece o no al fondo. Se denota como k-medias “adaptativo”, basándose en el hecho de que no se parte de un número fijo de *clusters*, sino que se van creando dinámicamente en función de las exigencias de cada píxel de la imagen.
- 3.- **Bucle de Segmentación:** esta fase se encarga de realizar el proceso de segmentación de objetos en la región de interés. Dada la complejidad de los entornos exteriores, se implementan varias etapas que absorben la problemática subyacente:
  - Proceso de segmentación mediante distancia de Mahalanobis: tal y como se definió en 3.4.3.6, se trata de una medida de la distancia en la que, a diferencia de la distancia euclídea, tiene en cuenta tanto la varianza de cada variable como la correlación

entre las variables que entren en juego. De esta forma, se permite comprobar, de manera fiable, como de cerca o lejos se encuentra la observación del vector de medidas de cada una de las clases. Se emplea, por tanto, como criterio de decisión en la función de segmentación. Para eliminar posibles errores de carácter impulsivo se realiza una modificación de manera que, “*clusters*” designados con solo muestras de ruido, y consecuentemente con valores de probabilidad a priori reducidos, se fuercen a situarse en zonas de fondo, no siendo interpretadas como objeto. De esta manera, surge el término “distancia de Mahalanobis Modificada”, tal y como se verá más detalladamente en 3.5.2.2.

- Eliminación de sombras empleando técnicas de visión estéreo: el principal inconveniente en entornos exteriores reside en los múltiples efectos provocados por los cambios de iluminación; siendo las sombras uno de sus efectos más perjudiciales. La eliminación de sombras pasa por realizar la transformación proyectiva (homografía) a una de las imágenes de manera que los puntos pertenecientes al suelo en ambas sean coincidentes. A partir de este punto se realiza un proceso de supresión de sombras y efectos debidos a cambios de iluminación.
- Generación del mapa para la actualización: se elabora un mapa de disparidad en el que se muestra qué píxeles de la imagen se han de incorporar al modelo de fondo y cuales no.
- Actualización del modelo de fondo: a partir del mapa de actualización se implementa un sistema de actualización del modelo estadístico de fondo de manera que vaya adaptándose a las variaciones sufridas en el mismo.

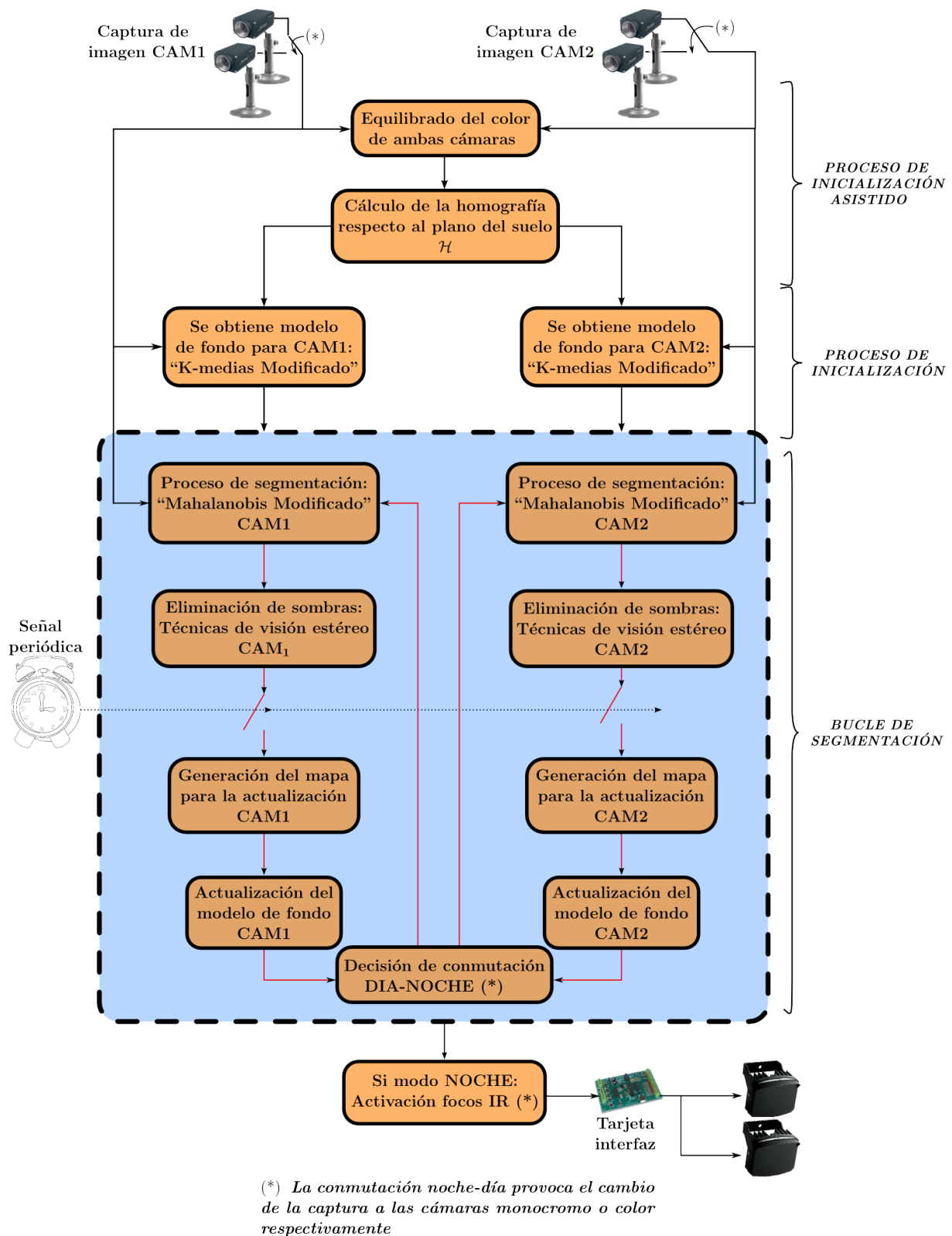


Figura 3.33: Diagrama modular de la solución propuesta para la detección de obstáculos.

### 3.5.2. Sistema de segmentación “KAM”

Siguiendo la línea de investigación propuesta en [Pappas, 1992] y [Chinrungrueng and Sequin, 1995], se desarrolla en el presente trabajo un sistema de segmentación basado en una modificación del algoritmo k-medias, empleando la distancia de Mahalanobis como criterio para la obtención del valor que actúa sobre la función discriminante para diferenciar entre fondo y objeto. A este nuevo algoritmo se le denomina “**K-medias Adaptativo con distancia de Mahalanobis**” (“KAM”).

#### 3.5.2.1. K-medias adaptativo

El algoritmo k-medias adaptativo parte del concepto inicial de k-medias, al cual se añaden una serie de modificaciones de manera que se ajuste a los objetivos planteados para entornos exteriores. El concepto de **adaptabilidad** surge como idea para romper con la rigidez del algoritmo inicial, en el que se parte de un número fijo de clases definidas a priori.

Se parte, por tanto, de una sola clase, cuyo centroide viene dado por la primera muestra del conjunto de patrones. Si la distancia euclídea de alguna de las muestras al centroide de la nueva clase es menor que un determinado umbral, definido como “DISTM”, dicha muestra se incorpora a la clase; en caso contrario se marca como candidata a generar una nueva, con centroide el valor de la muestra. Una vez asignada cada muestra a una clase determinada, se recalculan los centroides y se repite el proceso de asignación a clases hasta que todas las muestras convergen a una clase determinada. Por último, y teniendo constituidos todos los cluster, se obtienen los parámetros característicos de los “K” clusters; media ( $m_k$ ), probabilidad a priori ( $\omega_k$ ) y matriz de covarianza ( $\Sigma_k$ ). La figura 3.34 muestra un breve esquema de la solución planteada.



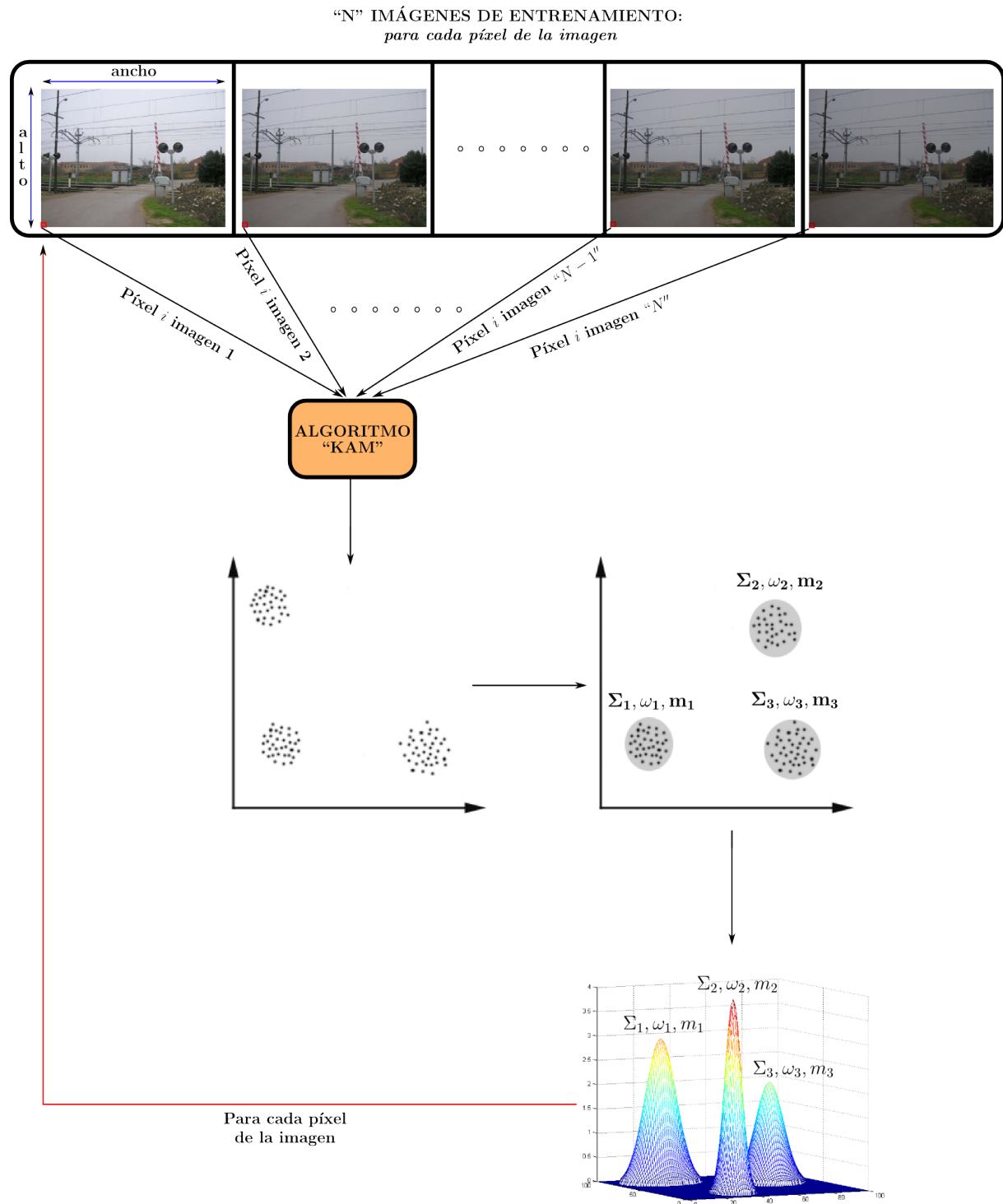


Figura 3.34: Esquema general del proceso de clasificación mediante el algoritmo k-medias modificado.

En la figura 3.35 se muestra el diagrama modular del algoritmo “k-medias modificado” cuyas fases se detallan a continuación.

1. Se define un conjunto de “ $N$ ” imágenes de entrenamiento a partir de las cuales se generará, para cada píxel, la distribución multimodal gaussiana que caracteriza el modelo de fondo calculado.
2. En la modificación implementada se comienza con, tan solo, una clase, cuyo centroide viene dado por el primer punto del patrón o imagen de entrenamiento. En la clase generada con dicho punto como centroide, se situarán todos aquellos puntos cuya distancia euclídea esté por debajo de un determinado umbral identificado por “**DISTM**”. Con las muestras asignadas a cada *cluster* se recalcula sus centroides. Si existen muestras no asignadas a las clases ya creadas, se crea una nueva y se repite el proceso; éste continuará hasta que los puntos converjan a un número finito de clases.
3. Se calculan los parámetros estadísticos propios de cada *cluster*, obteniéndose así la matriz de covarianza ( $\Sigma_k$ ), la probabilidad a priori ( $\omega_k$ ) y la media ( $m_k$ ) de cada una de las clases.

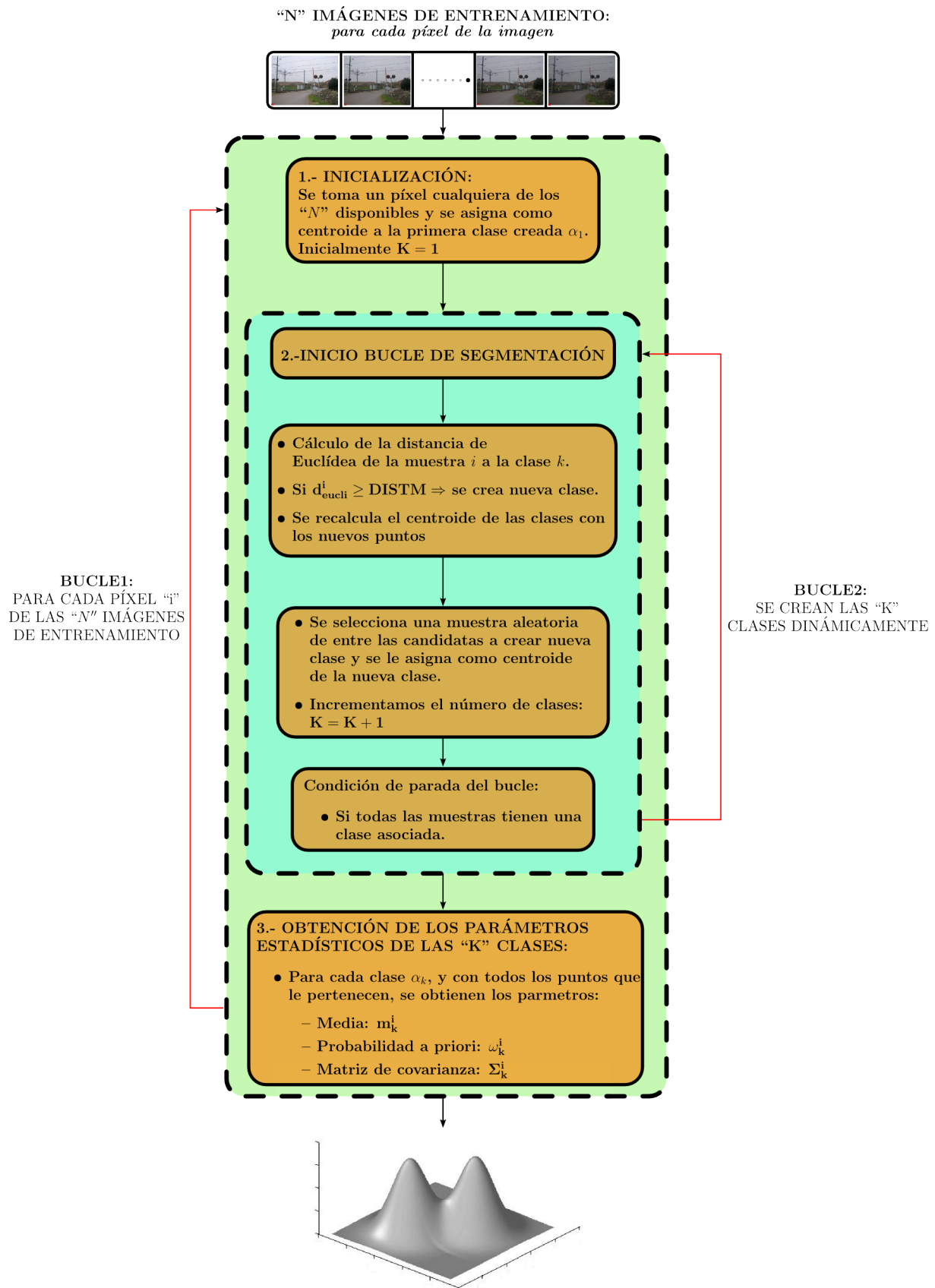


Figura 3.35: Diagrama modular del algoritmo k-medias modificado.

Esta fase es considerada como una etapa de inicialización o “entrenamiento”, ya que en ella se obtiene el modelo estadístico inicial de fondo. Para acometer dicho proceso, se plantea obtener imágenes a lo largo de uno o varios días en la región de interés, de manera que se tenga una buena aproximación de la variabilidad del fondo. Idealmente, se deben contemplar en esta fase todos los fenómenos que sufrirá la escena de fondo, de manera que cuanto mayores sean los cambios observados en las imágenes capturadas, las clases del modelo de fondo contendrán mayor información de la posible casuística de la escena. Por tanto, es aconsejable que dichas imágenes presentasen cambios de iluminación, evolución de sombras proyectadas y, por supuesto, que no incluya ningún objeto o persona que no pertenezca al fondo. Sin embargo, esta etapa se aleja demasiado de una situación real. Debido a los requerimientos de autonomía y mínimo coste de mantenimiento del sistema, y la imposibilidad de reflejar toda la casuística de dichos entornos en tan solo unas pocas capturas, no se muestra viable el proceso de realizar el almacenado previo de imágenes de la escena. Surge así la necesidad de idear un sistema que actualice dinámicamente el modelo de fondo, evitando tener que modelar desde un principio toda la fenomenología de dichos entornos y proporcionando mayor robustez y fiabilidad en los resultados, ya que la actualización se referencia a un presente inmediato y no a un pasado, cada vez más remoto.

### 3.5.2.2. Distancia de Mahalanobis Modificada

La distancia de Mahalanobis modificada al centroide de cada clase es el criterio empleado para decidir si una nueva muestra pertenece o no a una determinada clase y, por tanto, al modelo estadístico de fondo generado en la etapa previa. La modificación planteada en la distancia de Mahalanobis tradicional consiste en la inserción de un término aditivo que resta el logaritmo neperiano de la probabilidad a priori  $\omega_k$ . Con esto, se pretende que, *clusters* designados con solo muestras de ruido, al tener valores de probabilidad a priori muy pequeños, sean ignorados al generar valores de distancia muy elevados. Así, se reduce el efecto de la creación de nuevas clases con solo muestras de ruido de carácter impulsivo. En la expresión 3.68 se muestra la modificación planteada:

$$\hat{d}_{maha} = (X - m_k) \Sigma_k^{-1} (X - m_k) - \ln(\omega_k) \quad (3.68)$$

De esta manera, el proceso de segmentación consiste en:

***“Para cada píxel de la imagen, si su distancia al centro de alguna de las clases es menor que un umbral fijo, se estará en el caso en el que pertenece al modelo fondo, representándose como un píxel en negro dentro de la imagen segmentada. No obstante, si presenta una distancia a todas las clases mayor que dicho umbral, el nuevo píxel no pertenece al modelo de fondo, representándose dentro de la imagen en blanco.*”**

En la figura 3.36 se muestra el diagrama funcional del proceso de segmentación propuesto.

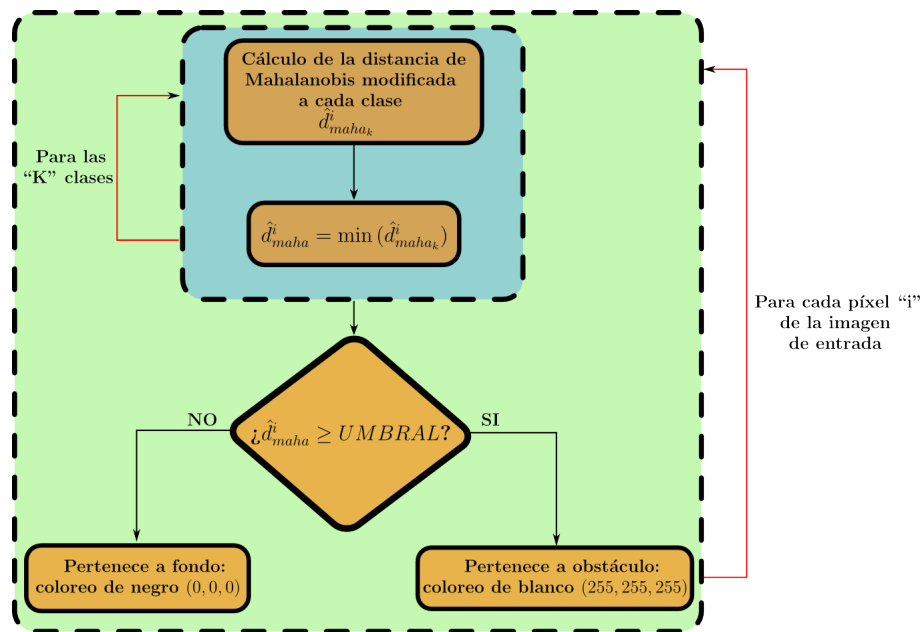


Figura 3.36: Diagrama modular del proceso de segmentación mediante distancia de Mahalanobis modificada.

### 3.5.2.3. Problemática asociada al sistema “KAM”

Una vez queda analizada la solución propuesta en “KAM”, se observan una serie de deficiencias; éstas serán resueltas en etapas posteriores mediante el sistema de eliminación de sombras y de actualización de fondo. Se muestra a continuación una breve explicación de las posibles carencias del sistema de segmentación:

- En la etapa de inicialización o entrenamiento del algoritmo k-medias adaptativo, véase 3.5.2.1, se debe reflejar toda la casuística a la que estará sometido el fondo de la escena monitorizada. Existen 2 motivos por los cuales este método queda descartado para el modelado unívoco y definitivo del fondo:
  1. En primer lugar, se deberían capturar imágenes que contemplaran evolución de sombras, cambios de iluminación y del modelo de fondo, etc., lo cual requiere un espacio de tiempo muy grande para la etapa de inicialización. La imposibilidad de reflejar toda la casuística de dichos entornos en unas pocas capturas, hace que no sea viable el proceso de almacenado previo de imágenes.
  2. En segundo lugar, las variantes condiciones climatológicas hace que sea imposible asegurar, con total seguridad, que las imágenes de entrenamiento, por muchas que se hayan almacenado, son representativas de todas las condiciones que podrán acontecerse en el futuro.

El *sistema de actualización de fondo* explicado más adelante, permitirá absorber este efecto y proporcionar una alternativa viable que permita ir incorporando al modelo de fondo los cambios temporales o permanentes sufridos en la escena.

- La evolución de las condiciones climatológicas con el paso del tiempo, que empiezan a hacerse patentes a lo largo de un día, provocan que el modelo de fondo se vea alterado. El progreso de una sombra a lo largo de un día, la inclusión de sombras provenientes de elementos no presentes en la escena, etc. pueden provocar, con frecuencia, la emisión de falsos positivos.

El *sistema de eliminación de sombras*, explicado a continuación, es el encargado de eliminar estos defectos, proporcionando mayor fiabilidad al sistema de detección.

### 3.5.3. Sistema de eliminación de sombras

El principal inconveniente de los escenarios en entornos exteriores se centra en los diversos fenómenos generados por los cambios de iluminación. Las sombras proyectadas y los cambios bruscos de iluminación son, sin lugar a duda, fuente de la mayor parte de falsos positivos generados por este tipo de sistemas. Con el objetivo de mitigar estos efectos se implementa un mecanismo de eliminación de sombras basado en técnicas de visión estéreo.

Al proyectar un punto en una imagen se pierde la información de profundidad ya que las imágenes son una representación bidimensional del mundo 3D. La utilización de dos cámaras permite extraer la información tridimensional del entorno; a esta técnica se la conoce como visión estereoscópica y requiere de métodos capaces de poner en correspondencia la información vista por dos cámaras (denotadas como CAM1 y CAM2). En general dicho proceso de correspondencia es muy complejo cuando se desconoce la estructura tridimensional de la escena vista por las cámaras. En este proyecto se propone una técnica estéreo restringida a un plano de la escena (plano del suelo) donde se supone que se proyectan las sombras de los objetos. Como se ha descrito con anterioridad, mediante una matriz de homografía es posible hacer corresponder de manera biyectiva las posiciones de puntos pertenecientes al plano mencionado en cada una de las cámaras. En la figura 3.38 se puede observar la transformación generada por la matriz de homografía.

Con la homografía calculada se construye una rutina que, a partir de las imágenes segmentadas en ambas cámaras, obtiene, a su salida, la imagen binarizada en la que las sombras han sido suprimidas. A continuación se detalla el proceso seguido para la eliminación de sombras:

- Se parte de la disposición de la imagen segmentada (binaria) de ambas cámaras en un mismo instante. Se aplica la transformación de perspectiva proporcionada por la matriz de homografía ( $\mathcal{H}$ ) a la imagen proveniente de CAM2, obteniéndose una nueva imagen donde los puntos pertenecientes al plano del suelo coinciden con los de la imagen proveniente de la otra cámara (CAM1) sin transformar. Por tanto, la diferencia en valor absoluto entre ambas imágenes suprime todos los puntos pertenecientes al plano del suelo; en el que se incluyen las sombras y el área de los objetos coincidentes con dicho plano, además de las zonas coincidentes al superponer ambas imágenes.
- Hasta este punto, la aparición de un objeto en la escena genera ciertas deficiencias en la detección:
  - Si el objeto presenta una silueta fina y estirada (x.e. una persona), los resultados arrojados serían satisfactorios; se eliminaría la sombra y un pequeño área en la base del mismo; la coincidente con el plano del suelo.
  - Si, por el contrario, el objeto presenta una mayor superficie de contacto con el plano del suelo (x.e. un coche), la eliminación de dicho plano degeneraría en la supresión de gran parte del objeto; la resta de ambas imágenes provoca la eliminación de zonas que en realidad corresponden a objeto y no al plano de suelo. Esto se produce en aquellas situaciones en las que la imagen segmentada, proveniente de una de las cámaras, tiene zonas comunes con la resultante del cambio de perspectiva realizada en la otra cámara.

Se hace necesario, por tanto, un **proceso de reconstrucción** de aquellas zonas que, perteneciendo o no al plano de suelo, son parte de los obstáculos detectados en la escena y han sido eliminadas en esta etapa.

- El mecanismo de reconstrucción implementado se basa en averiguar las zonas que, no siendo sombra, han sido eliminadas erróneamente. Un objeto, al ser observado por cámaras diferentes separadas una cierta distancia, presenta en su textura pequeñas diferencias; no apreciables en la sombra que el objeto proyecta en el suelo. Dicho efecto es aprovechado para recuperar las zonas de objeto eliminadas en base a dos causas:
  1. Por pertenecer al plano de suelo.
  2. Por coincidir en la superposición de la imagen proveniente de CAM1 y la imagen de CAM2 transformada por la homografía.

obteniendo, a la salida, la imagen segmentada en la que tan solo las sombras han sido eliminadas.

A continuación, se muestra el proceso explicado en cinco pasos y haciendo uso de imágenes reales:

- \* En un principio se dispone de la imagen capturada por ambas cámaras y su correspondiente segmentación; tal y como se muestra en la figura 3.37.

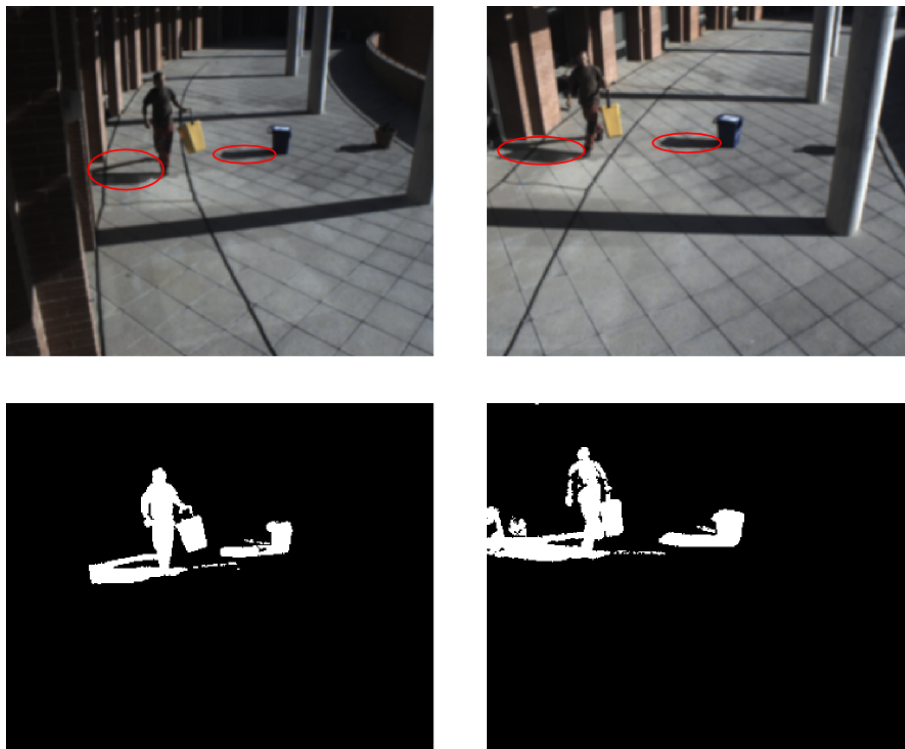


Figura 3.37: Imágenes capturadas de una escena por dos cámaras distintas (fila superior) y su correspondiente segmentación en una etapa previa a la eliminación de sombras (fila inferior).



1. Eliminación del plano del suelo mediante la diferencia en valor absoluto entre la segmentación proveniente de CAM1 y la de CAM2 transformada mediante  $\mathcal{H}$ . El resultado se muestra en la figura 3.38 donde se observa que existen zonas que, no perteneciendo al plano del suelo, son eliminadas debido a intersecciones entre la segmentación de ambas imágenes o por pertenecer explícitamente al plano del suelo.

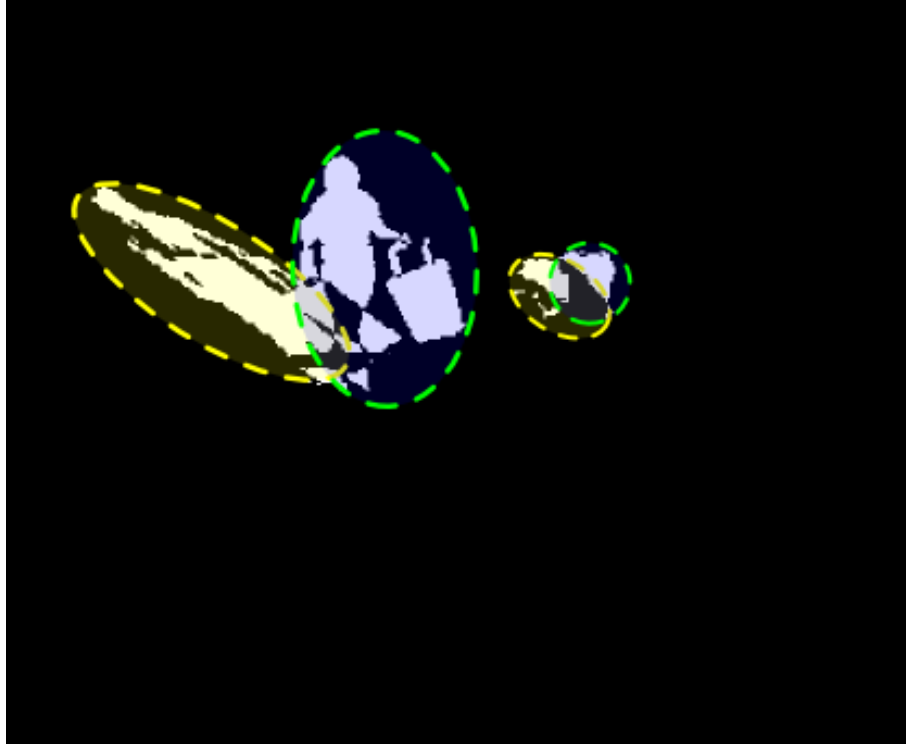


Figura 3.38: Resultado del proceso de eliminación del plano del suelo. Redondeado en azul aparece la segmentación proveniente de CAM1 y en amarillo la de CAM2 transformada por la matriz de homografía  $\mathcal{H}$ .

Obsérvese cómo las sombras quedan totalmente eliminadas por el mero hecho de pertenecer al plano del suelo.

2. Resaltado de textura mediante la diferencia en valor absoluto entre la segmentación a color proveniente de CAM1 y la de CAM2 transformada mediante  $\mathcal{H}$ . Como se puede observar en la figura 3.39, las zonas eliminadas de manera indebida en el punto anterior aparecen en este caso, aunque en un nivel muy reducido, aprovechando las diferencias en las texturas al ser observado por dos cámaras situadas en posiciones distintas.



Figura 3.39: Resta en valor absoluto de las imágenes segmentadas a color.

3. Zona reconstruida por presentar pequeñas diferencias en textura o diferencias en el valor de la captura entre ambas cámaras, figura 3.40.



Figura 3.40: Zonas recuperadas de la eliminación del plano del suelo que pertenecen a los obstáculos detectados.

4. Incorporación de la zona reconstruida a la imagen de la figura 3.38 de manera que se corrija la eliminación de zonas no pertenecientes a suelo.

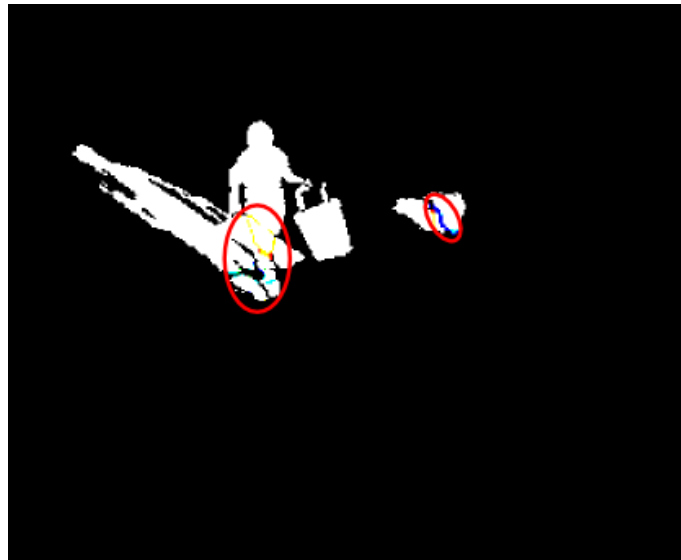


Figura 3.41: Incorporación de la zona reconstruida en el paso 3 a la imagen resultante de la eliminación del plano de suelo. En rojo, aparece redondeadas las zonas donde principalmente se ha añadido la reconstrucción.

5. Segmentación final. Se calcula la intersección de la imagen proveniente de la figura 3.41 con la imagen segmentada original. De esta forma se obtiene la segmentación final en la que la sombra queda totalmente eliminada.

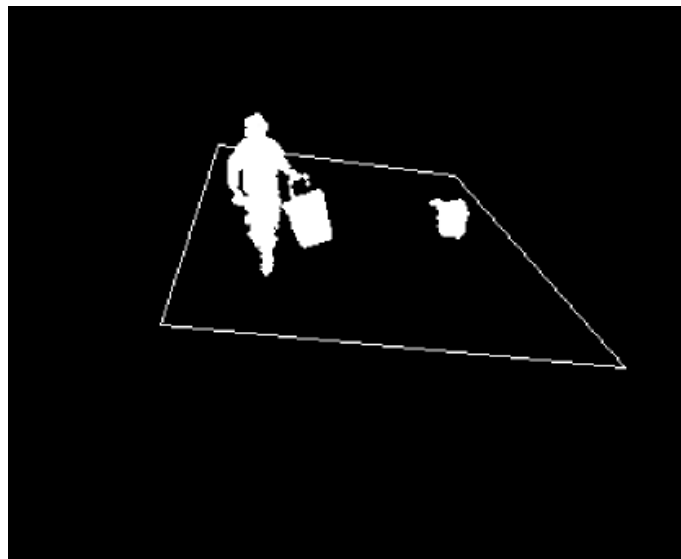


Figura 3.42: Resultado de la imagen segmentada tras el proceso de reconstrucción completo.

En la figura 3.43 se muestra el flujograma del proceso a seguir para la eliminación de sombras y la posterior reconstrucción de las zonas indebidamente eliminadas. Las imágenes que aparecen en su interior hacen referencia a las mostradas en el apartado previo.

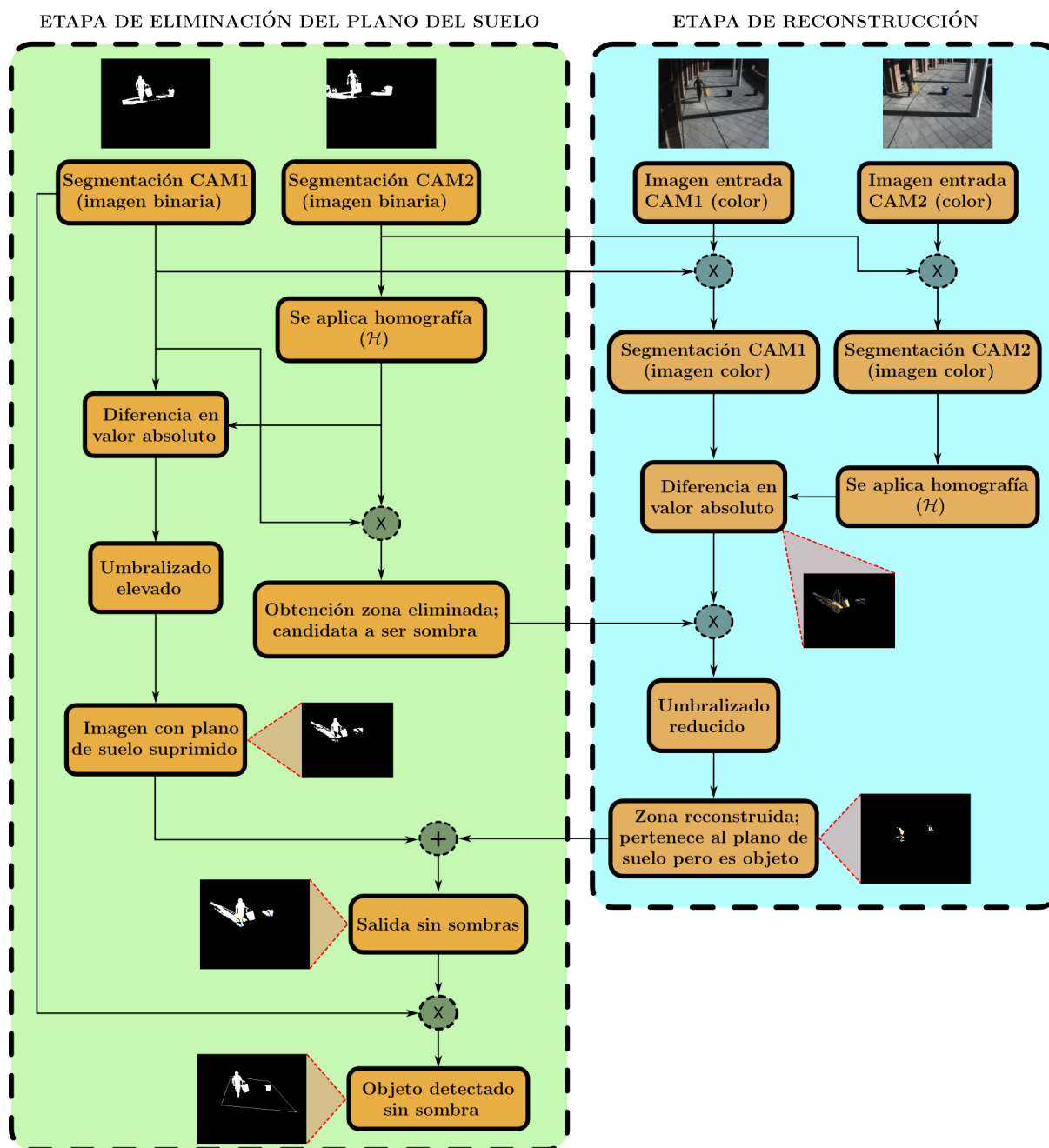


Figura 3.43: Diagrama modular del proceso de eliminación de sombras empleando técnicas de visión estereó (uso de la matriz de transformación u homografía).

### 3.5.4. Sistema de actualización de fondo

Con el objetivo de proporcionar mayor robustez y adaptabilidad, y debido a las deficiencias reveladas en el algoritmo “KAM” (véase 3.5.2.3), se propone la inclusión de un modelo de actualización de fondo dinámico. Con ello, se consigue que el sistema se pueda adaptar a las variaciones permanentes que se producen en el modelo de fondo de la escena. Se estudia como un complemento que permite, en caso de ser requerido, poder analizar cual ha sido la naturaleza del fenómeno que ha provocado un desplazamiento en el centroide de las clases (cambio de iluminación, sombra permanente...); método que nos permitiría obtener mayor cantidad de información acerca de lo que sucede ante un evento determinado.

Su ejecución se realiza de manera periódica, en el flujo principal del programa, y a continuación del proceso de eliminación de sombras. De esta manera, como las sombras han sido eliminadas de la segmentación, éstas se incorporarán al modelo de fondo, absorbiéndolas como parte de la escena observada.

El sistema de actualización se plantea dividido en dos bloques funcionales:

- 1.- Decisión de los píxeles candidatos a ser insertados en el modelo de fondo; se crea un mapa de disparidad o actualización.
- 2.- Actualización del modelo estadístico de fondo para aquellos píxeles marcados para ser actualizados en el bloque anterior.

A continuación se detallan cada uno de los procesos contenidos en los dos bloques anteriormente expuestos.

#### 3.5.4.1. Obtención del mapa de actualización o disparidad

Dentro del bloque encargado de discernir qué píxeles de la imagen deben ser actualizados y, por tanto, incorporados al modelo de fondo creado previamente, se destaca la generación de un mapa de disparidad. Este mapa parte de un concepto sencillo y pretende marcar, como candidatos a ser actualizados, todos aquellos puntos de la imagen pertenecientes al plano del suelo (dejando libre de actualización aquellas zonas de la imagen que sean tridimensionales).

En la figura 3.44 se muestra el diagrama empleado en el proceso de decisión de los píxeles a ser actualizados; su descripción se detalla a continuación.

- 1.- La resta de la imagen capturada por CAM1 y la imagen de CAM2 transformada por la matriz de homografía, provoca la “reducción” del valor de los píxeles pertenecientes al plano del suelo, dejando todas aquellas partes tridimensionales de la imagen visibles.
- 2.- La distancia euclídea al origen será el parámetro empleado para comparar con un umbral reducido y decidir si se ha de actualizar o no el píxel bajo análisis. Se genera un mapa

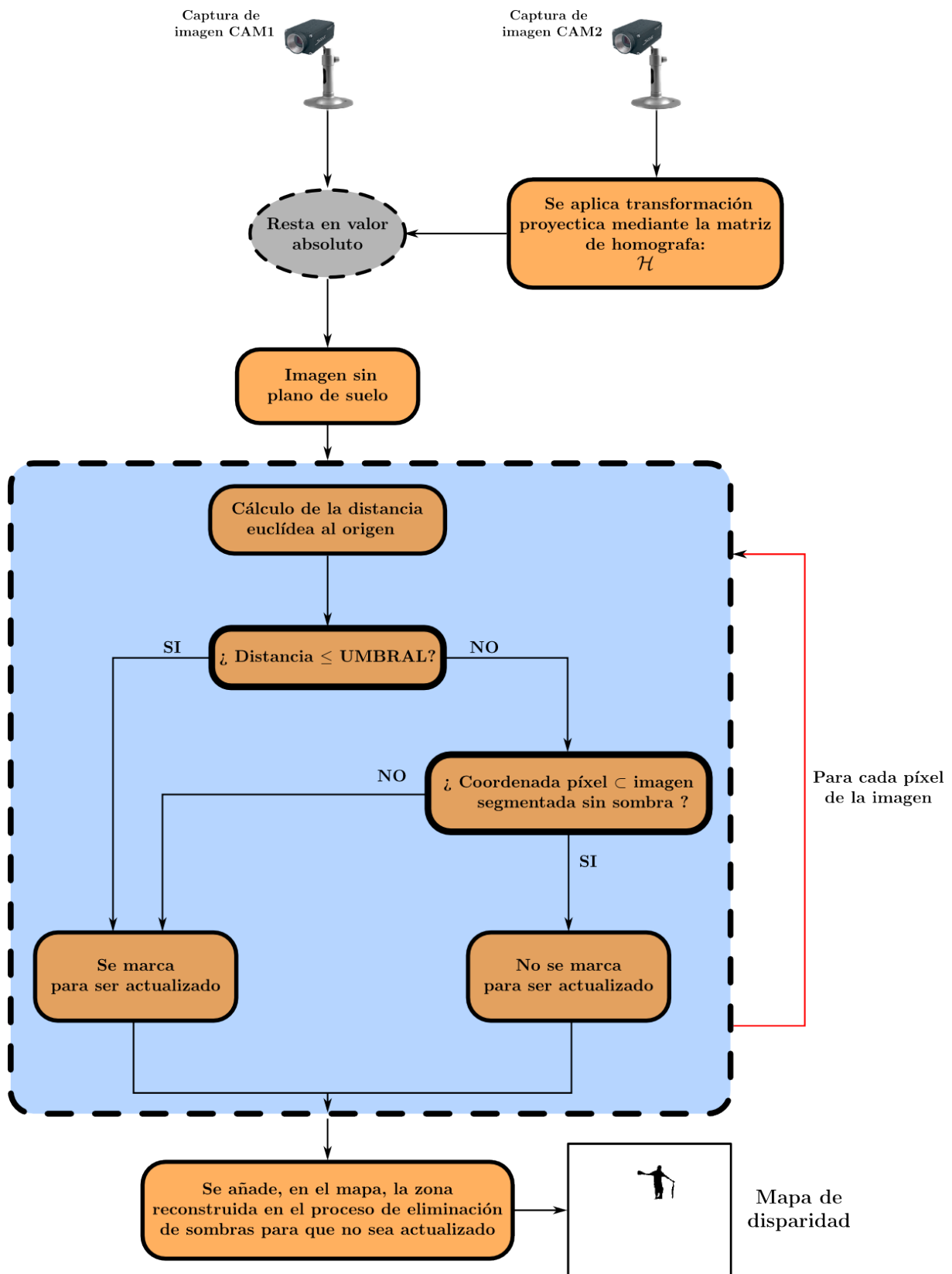


Figura 3.44: Diagrama modular del proceso de creación del mapa de disparidad para la cámara 1; se realizará un proceso dual para la otra cámara.

de disparidad para cada cámara, en el que solo se deberán excluir de ser actualizados los obstáculos tridimensionales vistos por cada una. Para ello, se implementa la siguiente tabla de verdad:

- Si la distancia euclídea al origen es menor que el umbral, el píxel se marca como una muestra a ser actualizada.
- En caso contrario:
  - Si el píxel no está contenido en la imagen segmentada sin sombra, para la cámara en cuestión, se marca para su actualización.
  - Si el píxel está contenido en la segmentación, se deduce que no ha de ser incorporado en el modelo de fondo por pertenecer a un obstáculo.

3.- En el proceso de reconstrucción del obstáculo detectado, enmarcado en el apartado de eliminación de sombras, se obtuvo la zona de los obstáculos que habían sido eliminadas erróneamente al suprimir el plano de suelo. El proceso realizado en el apartado 1 es equivalente, por lo que emplearemos dicha reconstrucción para especificar en el mapa de actualización que dichas zonas no han de ser incorporadas al modelo de fondo.

#### 3.5.4.2. Actualización del modelo estadístico de fondo

Una vez obtenido el mapa de disparidad, se incorpora, para el modelo de fondo existente en cada píxel, la muestra marcada para actualizarse. Este proceso es el encargado de generar el dinamismo de las clases en función de la casuística observada en el modelo y de proporcionar la memoria deseada a las clases, pudiendo definir la capacidad de retención de la información a priori.

Se detalla a continuación el proceso seguido; el cual se ilustra, a modo de resumen, en las siguientes dos figuras, 3.45 y 3.46.

- 1.- Para evitar que el modelo pueda llegar a degenerarse con el paso del tiempo y las sucesivas actualizaciones, se establecen dos valores máximos:
  - Máximo número de clases ( $MAX_{clases}$ )  $\Rightarrow$  acotando el número de clases por píxel de la imagen, se evita la convergencia del modelo a un espacio multimodal con infinitud de clases que abarquen todo el espectro; este efecto puede provocar que no se detecte objeto alguno, ya que sus muestras siempre estarían contenidas en el modelo de fondo.
  - Máximo número de muestras ( $MAX_{muestras}$ )  $\Rightarrow$  su valor será indicativo de la memoria que se dispone de la escena. Así, cuanto mayor sea el número de muestras que pueda contener una clase, mayor será la información temporal almacenada.
- 2.- Si el píxel bajo análisis está marcado en el mapa de disparidad para ser actualizado, se itera el algoritmo “K-medias adaptativo”, modificado para cumplir con el objetivo de la actualización del modelo de fondo. Se han de tener en cuenta todas las posibilidades a la hora de incorporar una muestra. Se implementa un **mecanismo de sustitución** para los

casos en los que la muestra se asigne a una clase que ha alcanzado su máximo número de puntos; y un **mecanismo de eliminación de clases** basado en el consumo de los tiempos de vida asignados a priori a cada clase. Se contemplan las siguientes posibilidades, donde se tendrá en cuenta que la asignación de una muestra a una determinada clase vendrá dada por la **distancia euclídea** existente entre la misma y el centroide de la clase (umbralizado con “DISTM”):

- Si la muestra es asignada a una clase y el número de muestras es inferior al máximo estipulado, ésta se incorpora directamente a la clase.
  - Si la muestra es asignada a una clase y el número de muestras ya ha alcanzado su máximo, ésta sustituye a otra elegida de manera aleatoria dentro de la clase a la que ha sido asignada.
  - Si la muestra no es asignada a ninguna de las clases existentes, y el número máximo de clases y muestras todavía no ha sido superado, se genera una nueva clase; tomándose como centroide el valor de la muestra.
  - Si la muestra no es asignada a ninguna de las clases existentes, cuando ya se ha alcanzado el máximo número de muestras pero no el de clases, se genera una nueva clase, con centroide el valor de la muestra, y se elimina una muestra elegida de manera aleatoria de una clase también elegida aleatoriamente.
  - Si la muestra no es asignada a ninguna de las clases existentes, cuando ya se ha alcanzado el máximo número de clases pero no el de muestras, se asigna la muestra a la clase más cercana.
  - Si la muestra no es asignada a ninguna de las clases existentes, cuando ya se ha alcanzado el máximo número de clases y de muestras, se asigna a la clase más cercana y se elimina una muestra de una clase, ambas elegidas de manera aleatoria.
- 3.- A continuación, se recalculan los centroides de las clases y los parámetros probabilísticos de las clases afectadas (media, probabilidad a priori y matriz de covarianza); realizándose una nueva iteración que permita reubicar muestras que presenten menor distancia a la nueva clase creada (este paso puede generar clases vacías).
- 4.- Con el objetivo de evitar ambigüedades, el algoritmo de actualización está continuamente supervisado por rutinas encargadas de reordenar y eliminar clases vacías generadas por la reubicación de muestras en la creación de nuevas clases.



La figura 3.45 ilustra los pasos seguidos hasta este punto.

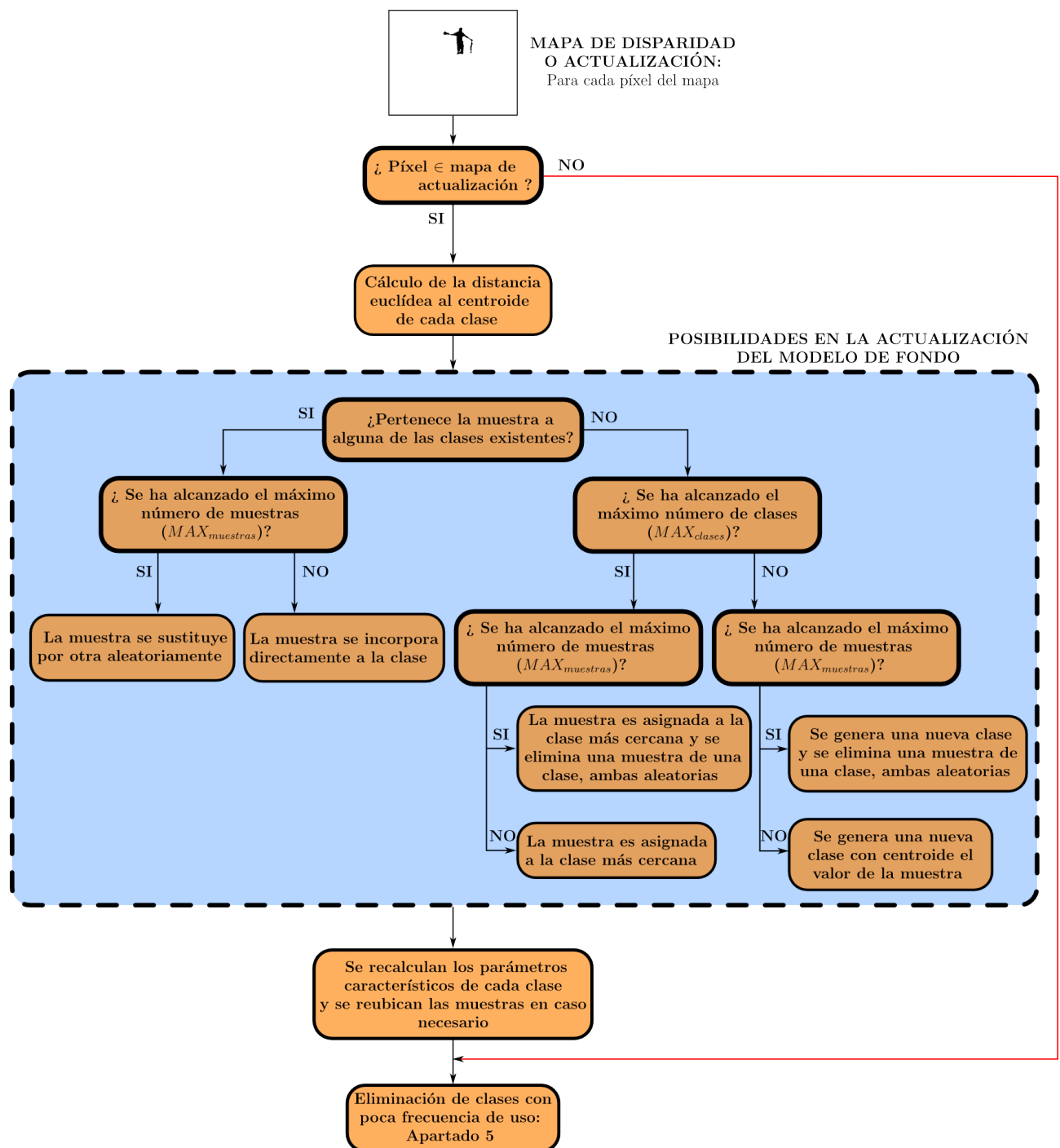


Figura 3.45: Diagrama modular del proceso de actualización del modelo probabilístico de fondo.

5.- Por último, se ha diseñado un mecanismo de eliminación de clases con pocos puntos, de manera que se supriman aquellas que han caído en desuso. Para poder monitorizar el uso de las clases existentes, se implementa una rutina de consumo de tiempos de vida, consistente en:

- Se asigna una marca de tiempo (TTL: Time To Live) a cada clase creada.
- Cuando una muestra se incorpora a una clase determinada, ésta ve incrementado su tiempo de vida un valor constante, al que nos referiremos como  $INC_{TTL}$ ; mientras que al resto de clases, las cuales no han sufrido modificación alguna, se les decrementa otro valor constante denominado  $DEC_{TTL}$ . Cuando el valor del tiempo de vida de alguno de los clusters llegue a cero, éste será eliminado junto con las muestras que le pertenecían.
- Los parámetros de incremento/decremento del tiempo de vida son fijados de antemano en función de la prioridad que se quiera dar a la desaparición de las clases con poca frecuencia de uso. Así, como ejemplo, si se aumenta el valor de la constante  $INC_{TTL}$   $\uparrow\uparrow$  y disminuimos el valor de  $DEC_{TTL}$   $\downarrow\downarrow$  se estará dando mayor facilidad a que clases, con poca frecuencia de uso, se mantengan en el modelo de fondo. Esto es debido a que la inserción de una muestra provoca un aumento elevado de su tiempo de vida, mientras que las sucesivas faltas de uso hará que su tiempo de vida se decremente lentamente.

La figura 3.46 ejemplifica el proceso de actualización de una muestra en el modelo de fondo, teniendo en cuenta el incremento/decremento de los tiempos de vida para la eliminación de clases con poca frecuencia de uso.

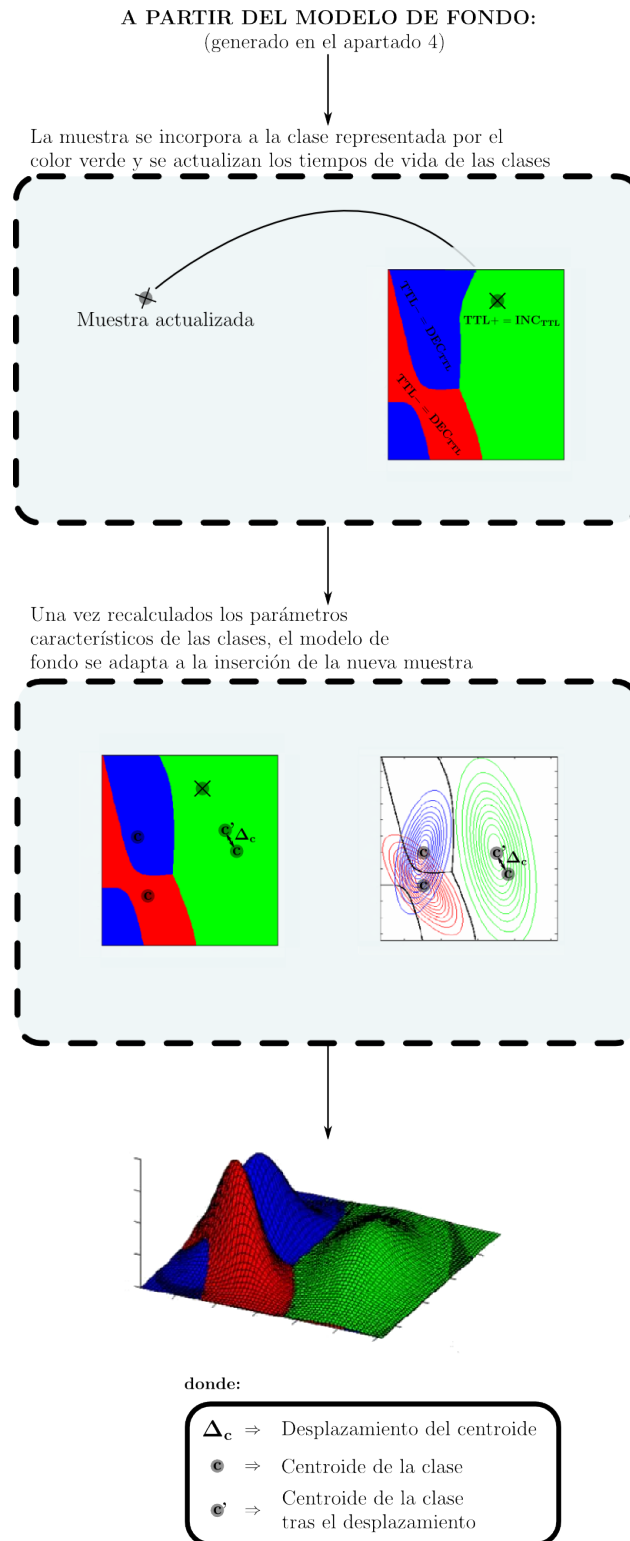


Figura 3.46: Ejemplo del proceso de actualización con incremento/decremento del tiempo de vida ante una muestra actualizada.

### 3.5.5. Sistema de conmutación día $\Leftrightarrow$ noche

Según los requisitos planteados en el proyecto AVIZOR, el sistema debe estar en funcionamiento durante largos periodos de tiempo y, por tanto, estar preparado para afrontar la mayor cantidad de eventualidades posibles. La mayoría de las cámaras a color no poseen la suficiente sensibilidad como para realizar capturas durante la noche. Así, se evidencia la necesidad de realizar las capturas mediante cámaras monocromáticas, con mayor sensibilidad al infrarrojo (IR), apoyadas con focos de iluminación dentro de ese espectro.

Con este fin, se implementa la conmutación día $\leftrightarrow$ noche. Como se ha explicado, el paso al modo nocturno implica el cambio de la captura de las cámaras a color a las cámaras monocromáticas y la activación de los focos de iluminación infrarroja. El criterio para decidir si se está bajo iluminación diurna o nocturna pasa por la observación de las cámaras a color:

***La conmutación a modo nocturno viene dada por la incapacidad de las cámaras a color de capturar las imágenes con un valor medio de intensidad luminosa que se sitúe alrededor de un umbral, el cual se define a priori. De manera análoga, cuando las cámaras a color vuelvan a poder capturar imágenes en torno a dicho umbral, se volverá al modo diurno.***

Por tanto, se ejecuta, de manera periódica, una rutina que supervisa el valor medio de la zona monitorizada y, en caso de que se excedan los límites establecidos en torno al umbral, se reajustan los parámetros de las cámaras. Mediante esta comprobación se puede verificar el criterio explicado con anterioridad y consecuentemente decidir si se ha de conmutar o no.

Empleando la tarjeta interfaz USB Velleman K8055 y el software de acceso, de licencia GPL, se activa el encendido y apagado de los focos según las circunstancias descritas.

### 3.5.6. Fusión con otras alternativas de segmentación

Existen infinidad de aplicaciones en las que la seguridad y fiabilidad de los resultados es objetivo prioritario y crucial para la implementación del sistema. Bajo estas premisas, es frecuente encontrar sistemas en los que se fusionan varias alternativas, teniendo, así, disponible mayor cantidad de información a su salida. En el presente proyecto se plantea la inserción de 3 alternativas para la detección de obstáculos, aunque queda abierto a la implementación paralela de tantas alternativas como la fiabilidad del sistema requiera o la tecnología permita. De esta manera, a la salida se dispondrá de 3 veces más información acerca de los objetos presentes en la escena.

Debido a los requisitos del sistema:

*máxima fiabilidad en cuanto a la detección de obstáculos en vía, intentando reducir al mínimo la emisión de falsos negativos (existen obstáculos en vía pero el sistema no los ha detectado). Se tendrá en consideración que la restricción en cuanto a la emisión de falsos positivos (se indica que existe obstáculo cuando en realidad no es cierto) es más relajada, a cambio de una mayor rigidez en la fiabilidad relativa a falsos negativos.*

se adopta una estrategia conservativa a la hora de incorporar las “N” alternativas de manera paralela. La estrategia adoptada consiste en mostrar un aviso de alarma siempre y cuando alguna de las alternativas advierta de la presencia de, al menos, un obstáculo en el paso a nivel. De esta manera, se minimiza la posibilidad de emitir falsos negativos al módulo receptor ubicado en la cabina del tren.

El esquema de la figura 3.47 muestra la fusión de 3 posibles alternativas para la detección de obstáculos en vía.

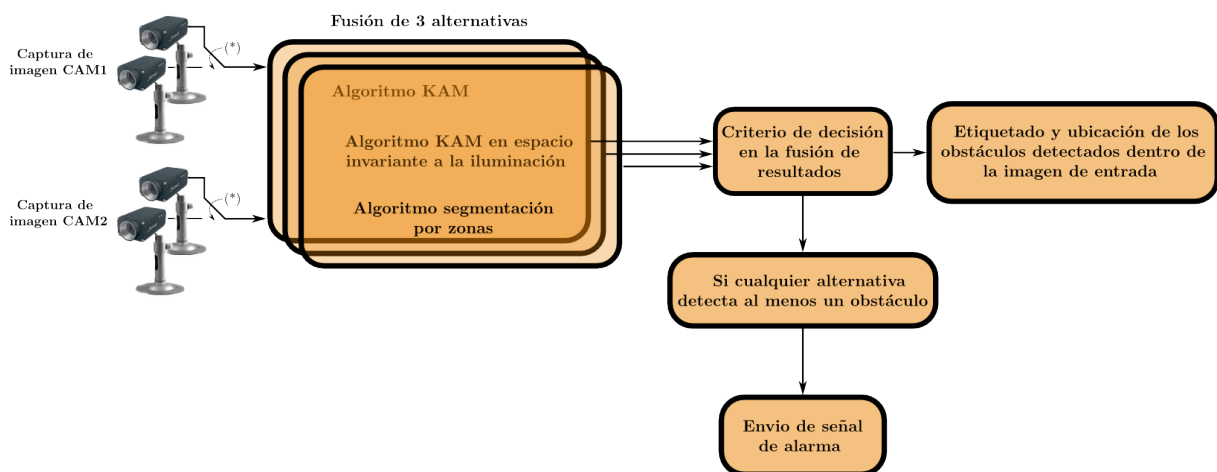


Figura 3.47: Esquema del proceso de fusión con 3 posible alternativas de segmentación.

donde:

- **Algoritmo KAM**  $\Rightarrow$  se trata del algoritmo de segmentación “K-medias Adaptativo con distancia de Mahalanobis” explicado con anterioridad.
- **Algoritmo KAM en espacio invariante a la iluminación**  $\Rightarrow$  hace referencia al algoritmo “KAM” modificado para emplear imágenes transformadas en el espacio invariante a la iluminación.
- **Algoritmo de segmentación por zonas**  $\Rightarrow$  tercera alternativa implementada para detección de obstáculos.

Existen otras alternativas de segmentación que pueden ser implementadas como suplemento de fiabilidad al sistema de detección. Entre ellas cabe citar el trabajo realizado en [Wang et al., 2006], el cual propone un modelo DCRF (“Dynamic Conditional Random Field”) para la segmentación de objetos y sombras en la escena. En [Cucchiara et al., 2003] se propone un método de propósito general que combina procedimientos estadísticos con la información proporcionada por objetos en movimiento. Por último, [Wang et al., 2005] presenta un método de segmentación de sombras y objetos en secuencias de imágenes capturadas en entornos interiores y empleando una sola cámara; en éste, los modelos de fondo y la información de bordes y sombras son iniciados y actualizados de manera adaptativa. Todos los trabajos citados se presentan como alternativas fiables y de fácil incorporación al sistema sensorial diseñado.

## Capítulo 4

# Desarrollo del sistema inalámbrico de banda ancha para transmisión de imágenes

### 4.1. Introducción

Este capítulo presenta una solución para el sistema de comunicación propuesto entre un paso a nivel monitorizado y un tren. En lo que se refiere a la transmisión de imágenes, y los posibles mensajes de detección de objetos en el paso a nivel, se aborda el estudio de una alternativa para la transmisión de información de vídeo (y señales de alarma), que garantice alta fiabilidad en la transmisión y niveles de calidad de servicio adecuados. La técnica de transmisión será inalámbrica y apoyada en el estándar de comunicaciones Vídeo-IP. Las tecnologías disponibles que ofrecen el ancho de banda necesario para el envío de imágenes codificadas en MPEG, permiten, bien la construcción de redes propias en bandas no licenciadas (WiMAX móvil, WiFi avanzado en su estándar 802.11n) o el uso de redes celulares públicas de banda ancha, que ofrecen cobertura en las líneas de ferrocarril, como UMTS, HSDPA y HSUPA. Por tanto se estudiarán todas las opciones, prestando especial atención a la velocidad del tren a la que es posible una correcta recepción, la cobertura que ofrecen y el coste de la solución.

Una vez determinada la mejor alternativa tecnológica, se implantará el sistema de transmisión que atenderá, esquemáticamente, a la siguiente secuencia:

1. Despliegue de la red: estación base y clientes inalámbricos en el tren y el sistema de procesamiento.
2. Codificación de la señal de vídeo a MPEG-4.
3. Generación y envío de las tramas MPEG-4 IP.
4. Recepción de las tramas y reconstrucción de la señal de vídeo en el tren.

En el presente proyecto se realiza un estudio de las alternativas existentes, justificando debidamente la elección adoptada. Así mismo, se desarrollan los apartados 2 y 3 anteriormente

descritos; integrando en el software, para la detección de obstáculos, la codificación y envío de las tramas MPEG-4 IP.

## 4.2. Objetivos

El objetivo de esta tarea es diseñar un sistema de comunicaciones inalámbrico que permita la transmisión de señales de vídeo desde el paso a nivel hasta los trenes que circulan en un entorno mínimo de 2 km. La velocidad del tren puede llegar a ser de hasta 180 km/h, aunque será la tecnología la que defina la velocidad máxima posible del tren para recepción de vídeo.

Las comunicaciones enlace tren-tierra se realizará empleando bandas NO LICENCIADAS. El sistema de comunicación únicamente tendrá la limitación de no interferir en la banda UHF, entre las frecuencias 447 - 875 MHz, ya que dicha banda está reservada para aplicaciones concretas en infraestructura ferroviaria de pasos a nivel.

La resolución de la imagen transmitida de la parte de tierra a la parte embarcada será como mínimo CIF, con 12'5 imágenes por segundo. En cualquier caso, permitirá la identificación clara de objetos del tamaño mínimo definido.

La transmisión de vídeo y alarmas debe cumplir con unos niveles de servicio. Este nivel de servicio se exige para cada paso a nivel y no como promedio de todos los pasos que se pudieran instalar.

### Vídeo

- Pérdidas sustanciales o cortes en la información de vídeo de 2 segundos o menores:

$$\frac{\sum \text{cortes de video} \leq 2 \text{ segundos}}{\sum \text{tiempo de servicio}} \leq 0,05$$

- Pérdidas sustanciales o cortes en la información de vídeo mayores de 2 segundos:

$$\frac{\sum \text{cortes de video} > 2 \text{ segundos}}{\sum \text{tiempo de servicio}} \leq 0,02$$

- El tiempo de servicio es el tiempo desde que el sistema embarcado entra en el área de cobertura del paso a nivel y empieza a recibir la señal de vídeo hasta el final de la transmisión.

### Alarmas

- Cortes en la información de las alarmas de 2 segundos o menores:

$$\frac{\sum \text{cortes de alarmas} \leq 2 \text{ segundos}}{\sum \text{tiempo de servicio}} \leq 0,02$$



- Cortes en la información de las alarmas mayores de 2 segundos:

$$\frac{\sum \text{cortes de video} > 2 \text{ segundos}}{\sum \text{tiempo de servicio}} \leq 0,005$$

En todo caso, el interfaz con el conductor del sistema embarcado indicará si la transmisión está activa y si la imagen/indicación de alarma es fiable.

Es deseable que la probabilidad de intrusión no detectada (alarma no detectada), sea lo menos posible; dando prioridad máxima a la fiabilidad del sistema.

La señal transmitida debe estar protegida suficientemente para evitar su suplantación y para impedir que pueda ser recibida por sistemas no autorizados.

## 4.3. Aspectos teóricos relacionados

### 4.3.1. Estudio de las tecnologías inalámbricas disponibles

Se entiende por redes inalámbricas al conjunto de tecnologías que permiten una comunicación entre dos o más dispositivos, mediante la transmisión de ondas electromagnéticas en el espectro radioeléctrico y por medio aéreo.

Las distintas tecnologías inalámbricas se distinguen, básicamente, por los siguientes aspectos:

- Cobertura.
- Velocidad de transferencia que permiten.
- Calidad de servicio.
- Movilidad que permiten al usuario.
- Sometida a estándares o propietaria.
- Rango de frecuencia en las que operan.
- Eficiencia espectral (bits que pueden transmitir por cada Hz de ancho de banda).
- Técnicas de modulación.

La combinación de los aspectos anteriores ha dado lugar a un sinnúmero de tecnologías cada una de aplicación en un ámbito diferente y que, hoy día, tienden a converger en lo que se ha dado en llamar 4G y que se espera proporcione acceso de banda ancha a todo tipo de usuarios en un futuro cercano, [Goodman and Myers, 2005].

Antes de presentar de forma esquemática las distintas tecnologías y dado que el aspecto clave en la evolución actual de las redes inalámbricas, radica en su capacidad para dar cobertura de alta velocidad y servicios IP a los usuarios, independientemente de su ubicación, se hace necesario definir qué se entiende por acceso fijo, nómada, portátil y móvil, [Consulting, 2005b].

- **Acceso fijo:** el equipo con el que el usuario accede a la red se encuentra fijo en un determinado lugar geográfico. Por ejemplo, el caso habitual de acceso a través del ordenador de la oficina o de casa.
- **Acceso nómada:** el equipo está fijo en un lugar geográfico determinado, al menos durante el tiempo que dura la sesión de conexión con la red. El usuario puede cambiar de ubicación geográfica y, posteriormente, iniciar un nuevo acceso a la red. Por ejemplo, los usuarios que disponen de ordenador portátil y que se conectan en la oficina, en casa o en un hotel. Es importante destacar que durante la conexión, el equipo se encuentra estático.
- **Acceso portátil:** el equipo debe permanecer conectado cuando el usuario se desplaza a velocidades de peatón, dentro de determinada área de cobertura. Es el caso del acceso con un teléfono móvil, con el usuario paseando.
- **Acceso móvil:** el equipo se conecta a la red desde un vehículo que se desplaza a velocidades altas. Es el caso de un usuario que se conecta con su ordenador portátil mientras viaja en el tren o en un automóvil (se entiende que como pasajero).

En la figura 4.1 se puede apreciar el posicionamiento general de las tecnologías inalámbricas.

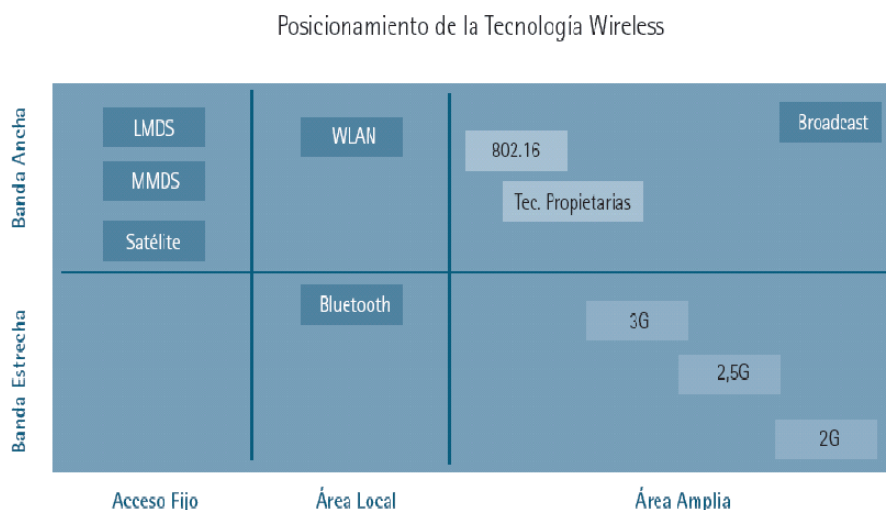


Figura 4.1: Posicionamiento de las tecnologías inalámbricas

Dentro de las tecnologías inalámbricas de acceso fijo de gran capacidad, se han venido utilizando durante los últimos años las variantes xMDS, que en lo fundamental, permiten cubrir la llamada última milla (desde la central hasta la ubicación del usuario) sin necesidad de cablear.

Entre estas tecnologías cabe destacar MMDS (Multichannel Multipoint Distribution Service) o LMDS (Local Multipoint Distribution Services). Ambos sistemas se basan en el uso de frecuencias microondas, por lo que se requiere que no exista ningún tipo de obstáculo físico entre la antena emisora y la receptora. La transmisión es punto-multipunto, por lo que una única antena transmite la señal a varias antenas receptoras dentro de una zona geográfica determinada.

**LMDS (Local Multipoint Distribution Services)** Esta tecnología permite, en un radio de cobertura limitado, transmitir información a alta velocidad desde un punto (la estación base) a muchos puntos (los clientes) y viceversa.

Utiliza bandas de alta frecuencia cuyo uso está regulado y que requieren el pago de la correspondiente licencia. En España las bandas de frecuencias son 3'5 y 26 GHz. Sus principales desventajas son las siguientes:

- El ancho de banda es compartido por los usuarios, por tanto las prestaciones disminuyen a medida que aumenta el número de usuarios.
- Se requiere visión directa (LOS: Line Of Sight) entre las antenas para efectuar la transmisión de datos.

**MMDS (Multichannel Multipoint Distribution Services)** El funcionamiento de esta tecnología es muy similar a la anterior, teniendo ambas las mismas desventajas. Las principales diferencias son:

- Utiliza una banda de frecuencia diferente, también regulada, generalmente inferior a 2'5 GHz.
- La distancia entre la estación base y los clientes puede ser mayor a 10 km, mientras que en LMDS no suele superar los 5 km.

Las tecnologías xMDS deben superar la restricción motivada por la necesidad de disponer de visión directa entre antenas para tener impacto en el mercado. En este sentido, WiMAX fijo (variante 802.16-2004) se establece como principal alternativa permitiendo enlaces sin visión directa (NLOS) entre estación base y cliente.

#### 4.3.1.1. Estándares inalámbricos de banda ancha

En cualquier tipo de tecnología de comunicaciones los estándares son clave para promover grandes volúmenes de producción, y de este modo, reducir costes y posibilitar un aumento de la cuota de mercado y el acceso de gran número de usuarios a dicha tecnología. Adicionalmente, la estandarización simplifica los procesos de prueba y evaluación de productos, a la vez que reduce los tiempos de desarrollo y de implantación.

En la figura 4.2 se pueden ver estos estándares según su cobertura.

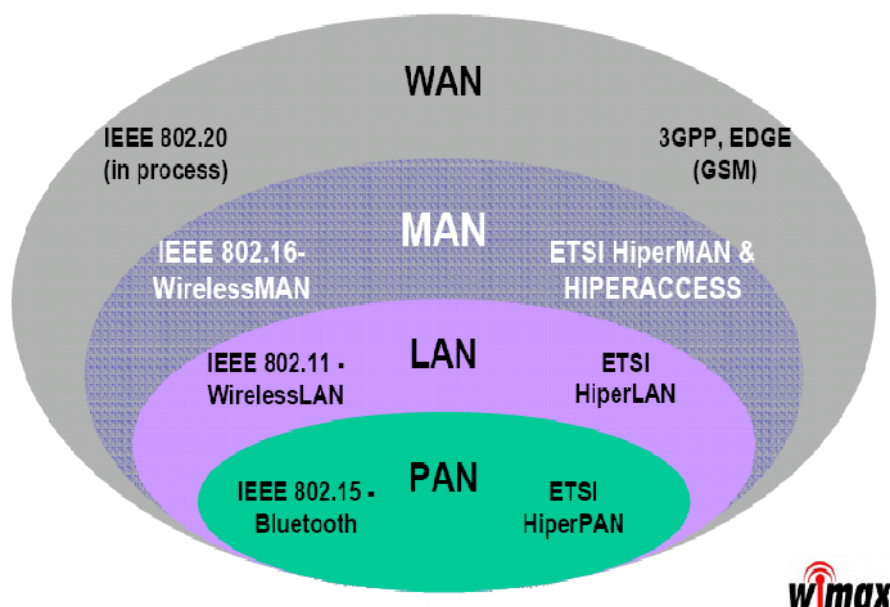


Figura 4.2: Estándares inalámbricos según su cobertura

Como se puede apreciar, IEEE ha establecido una jerarquía de estándares inalámbricos complementarios entre ellos: IEEE 802.15 para Redes de Área Personal (PAN), IEEE 802.11 para Redes de Área Local (LAN), 802.16 para Redes de Área Metropolitana, y el propuesto IEEE 802.20 para Redes de Área Amplia (WAN). Cada estándar mencionado representa una tecnología optimizada para un mercado y modelo de uso distinto y está diseñado para complementar los otros.

**IEEE 802.15 - Personal Area Network (PAN)** Las tecnologías inalámbricas basadas en la familia de estándares IEEE 802.15 (bluetooth, zigbee, etc) proporcionan, por un bajo coste y una baja potencia, un enlace radio de corto alcance para dispositivos móviles. De esta manera permite establecer conexiones de tiempo real de voz y datos mediante la banda libre de 2'400 - 2'483 GHz, alcanzando tasas de transferencia de hasta 2 Mbps.

**IEEE 802.11 - Wireless Local Area Network (WLAN)** El protocolo IEEE 802.11 (WiFi) es un estándar de protocolo de comunicaciones de la IEEE que define el uso de los dos niveles más bajos de la arquitectura OSI, especificando sus normas de funcionamiento en una WLAN.

La familia 802.11 incluye seis técnicas de transmisión por modulación que utilizan los mismos protocolos. Actualmente, el estándar ha ido generando nuevas publicaciones, siendo la versión 802.11n la última.

**IEEE 802.16 - Metropolitan Area Network (MAN)** Se trata de una especificación para las redes de acceso metropolitanas inalámbricas de banda ancha.

El estándar IEEE Std 802.16-2004 cubre únicamente los enlaces estáticos (WiMAX fijo). Hay una extensión denominada 802.16e para sistemas móviles, aprobada posteriormente, [WiMAX-Part, 2006].

802.16 y las actuales tecnologías propietarias cercanas al 802.16 se posicionan con mayor ancho de banda que las tecnologías de redes móviles, aunque con menor alcance.

También puede apreciarse, que los nuevos estándares del IEEE en tecnología inalámbrica de banda ancha poseen una eficiencia espectral muy alta. Así, el estándar 802.11a alcanza los 27 bps/Hz trabajando a velocidades de 54 Mbps. El estándar 802.16a alcanza los 5 bps/Hz trabajando a 70 Mbps, mientras que tecnologías como el EDGE, o el CDMA2000 se sitúan en 1.9 y 1.6 bps/Hz respectivamente.

**Mobile-Fi (IEEE 802.20)** El estándar móvil del IEEE 802.20 o Mobile-Fi está específicamente pensado para transportar tráfico IP nativo para un acceso totalmente móvil de banda ancha, ofreciendo tiempos de latencia de 10 ms (milisegundos), incluso cuando el vehículo se desplaza a gran velocidad, comparado con los tiempos de latencia de 500 ms de la tecnología 3G.

La misión del 802.20 consiste en el desarrollo de un interfaz aéreo, basado en paquetes y optimizado para transportar servicios IP, que, según sus impulsores: “permitirá el despliegue factible de redes inalámbricas de banda ancha, a nivel mundial, “always-on”, interoperables, que cumplan con las necesidades de los mercados residencial y de negocios”.

El estándar está previsto para trabajar en bandas licenciadas de 3.5 GHz y promete soportar un mayor número de usuarios simultáneos que los sistemas móviles, proporcionando una mayor eficiencia espectral y menor latencia. Se ha declarado que Mobile-Fi tendrá el doble de eficiencia espectral que los sistemas celulares, ofreciendo baja latencia y calidad de servicio, de modo que proporcionará una experiencia de usuario similar a la de las conexiones cableadas.

Un resumen comparativo de todos los estándares anteriores, con sus características más importantes, se muestra en la tabla 4.3.

T2. COMPARACIÓN DE TECNOLOGÍAS DE ACCESO INALÁMBRICO ESTÁNDARES IEEE

|                                | 802.11                                      | 802.16                                         | 802.16a                                         | 802.16e                                    | 802.20                       |
|--------------------------------|---------------------------------------------|------------------------------------------------|-------------------------------------------------|--------------------------------------------|------------------------------|
| Status                         | Complete                                    | Dec. 2001                                      | Jan. 2003                                       | Jan. 2004                                  | ETA 05-06                    |
| Target App.                    | LAN                                         | MAN                                            | MAN                                             | MAN                                        | WAN                          |
| Range                          | Up to 300ft.<br>optimized for<br>indoor LAN | Up to 5 miles<br>Average Cell<br>Radius 1-3 mi | Up to 25 miles<br>Average Cell<br>Radius 4-6 mi | Average Cell<br>Radius 1-3 mi              |                              |
| Channel<br>Conditions          | LOS when<br>outdoors                        | LOS                                            | nLOS                                            | nLOS                                       | nLOS                         |
| Spectrum                       | 2.4 GHz & 5 GHz-<br>Unlicensed              | 10-66 GHz<br>Licensed                          | 2-11 GHz<br>Licensed                            | 2-6 GHz<br>Licensed and<br>Unlicensed      | <3.5 GHz<br>Licensed         |
| Mobility<br>Support<br>Roaming | Portable-<br>Local Roaming                  | Fixed                                          | Fixed                                           | Pedestrian<br>Mobility-                    | Vehicular<br>Mobility-Global |
| Channelization                 | 20 MHz                                      | Scalable 1.5-<br>20 MHz                        | Scalable 1.5-<br>20 MHz                         | Scalable 1.5-<br>5 MHz w/<br>sub- channels | 1.25 or 5 MHz                |
| Spectral Efficiency            | <2.7 bps/Hz                                 | <4.8 bps/Hz                                    | <3.75 bps/Hz                                    | <3 bps/Hz                                  | <1.25 bps/Hz                 |
| Bit Rate                       | 54 Mbps<br>(20 MHz BW)                      | <134 Mbps<br>(20 MHz BW)                       | <75 Mbps<br>(20 MHz BW)                         | 15 Mbps<br>(20 MHz BW)                     | <6 Mbps<br>(20 MHz BW)       |

Fuente: Intel Wireless Division.

Figura 4.3: Comparativa de tecnologías inalámbricas

En la figura 4.4 se puede apreciar la relación entre los estándares descritos y las tecnologías de acceso celular (3G) por lo que respecta al ancho de banda disponible para el usuario y el soporte para movilidad.

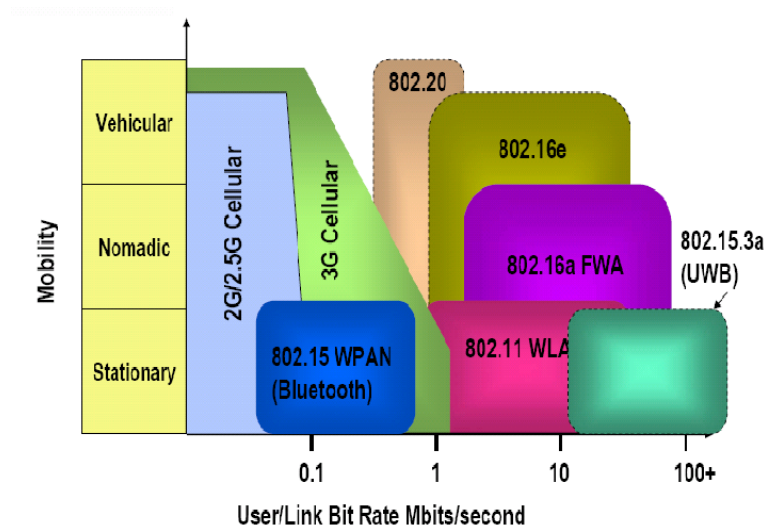


Figura 4.4: Relación movilidad-ancho de banda entre tecnologías inalámbricas

La relación ancho de banda-cobertura, se puede apreciar en la figura 4.5.

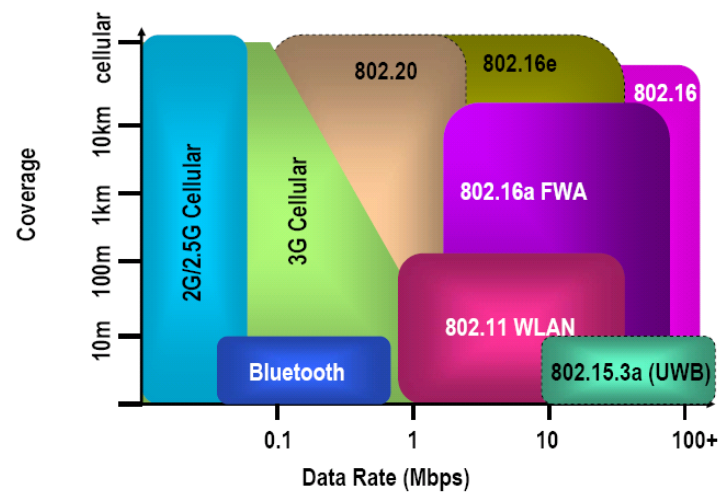


Figura 4.5: Relación cobertura-velocidad de transferencia

#### 4.3.1.2. Otras alternativas

**FLASH OFDM: Una alternativa de banda ancha inalámbrica para 450 MHz** FLASH OFDM (Fast, Low-latency Access with Seamless Handoff OFDM) es una alternativa económica de banda ancha inalámbrica para operadores con acceso a la banda de 450 MHz. Diseñada para transmitir múltiples señales simultáneamente sobre un enlace de transmisión, FLASH OFDM dispersa los datos sobre numerosas portadoras separadas en frecuencias precisas y ortogonales entre ellas.

Esta solución está diseñada para servir a los operadores de redes móviles que buscan brindar servicios de datos de alta velocidad en bandas licenciadas y con una gran área de cobertura.

Gracias a que FLASH OFDM trabaja en la frecuencia de 450 MHz, el número de estaciones base para cubrir un área geográfica determinada se reduce en un 90 % comparado con las tecnologías en la banda de 2 GHz.

Otra de las características principales de FLASH OFDM es su baja latencia, inferior a 50 ms, comparado con otros estándares. Esto permite que el usuario experimente un servicio de mayor calidad, similar a los DSL.

En cuanto a la velocidad de acceso, es de 1.5 Mbps en el Downlink y 500 Kbps en el Uplink, con picos de 5.3 y 1.8 Mbps respectivamente. Otra ventaja de esta tecnología es que, a diferencia de WiMAX por ejemplo, puede brindar conectividad para usuarios incluso a velocidades vehiculares (300 Km/h) (figura 4.6).

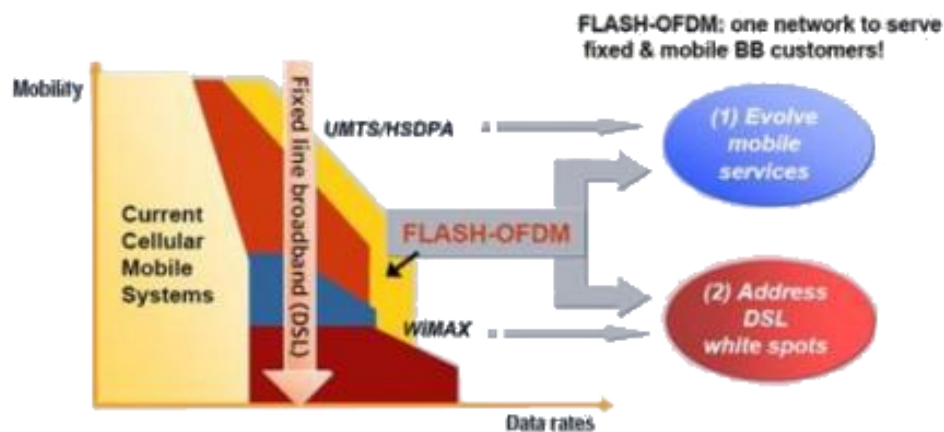


Figura 4.6: FLASH OFDM apunta a usuarios móviles y fijos por igual

Puramente basado en IP, Flash OFDM soporta cualquier tipo de aplicación o contenido basado en este protocolo, además de proveer una integración fluida con cualquier otra infraestructura o equipos de redes IP.



Puede decirse que el principal talón de Aquiles de esta tecnología es que no es una evolución como UMTS lo es a GSM, sino que es totalmente nueva, por tanto el equipamiento es escaso y los precios elevados.

**Banda ancha en redes celulares (evolución de 3G)** El 3GPP y 3GPP2 son los organismos encargados de la definición de los estándares de 3G, de modo que sobre las redes desplegadas actualmente, se puedan mejorar las prestaciones de ancho de banda ofrecidas a los clientes y permitir el despegue de aplicaciones basadas en IP [WiMAX, 2006b].

Para ello, se han definido las siguientes extensiones sobre WCDMA y CDMA2000:

- 1xEVDO (Evolution Data Optimized), también conocido como HRPD (High Rate Packet Data) es una evolución de CDMA2000 desarrollado por 3GPP2. Sobre un canal de 1.25 MHz, 1xEVDO ofrece máximos de transferencia de datos de 2.4 Mbps (Rev 0) y 3.1 Mbps (Rev A) en el enlace descendente (DL) y 153.6 kbps (Rev 0) y 1.8 Mbps (Rev A) en el de subida (UL). EVDO-Rev B añade más capacidad en el enlace DL, hasta los 4.9 Mbps en un canal de 1.25 MHz. Este estándar está siendo aplicado con éxito en Corea del Sur y está siendo desplegado de manera amplia en el sureste asiático.
- HSDPA (High Speed Downlink Packet Access) ha sido definido por 3GPP y proporciona mejoras en el enlace descendente sobre WCDMA R'99. Estas mejoras se traducen en un máximo de velocidad de transferencia (DL) de 14 Mbps en un canal de 5 MHz que, con la especificación HSUPA (High-Speed Uplink Packet Access) se extenderá al enlace de subida (UL).

En la figura 4.7 se puede apreciar la evolución que está teniendo y la prevista, de los estándares de 3GPP y 3GPP2.

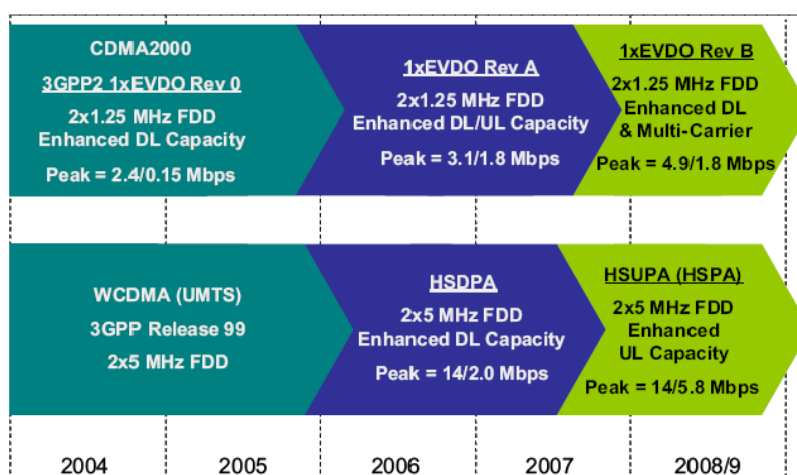


Figura 4.7: Evolución de los estándares 3G

### 4.3.2. Vídeo vigilancia

La transmisión inalámbrica de señales de vídeo puede realizarse de diversas maneras:

- Empleando transmisión analógica.
- Utilizando como soporte una red de comunicaciones basada en IP (vídeo IP)

#### 4.3.2.1. Vídeo IP

A la hora de dimensionar una red de comunicaciones para soportar transmisiones de vídeo IP hay que tener en cuenta diversos aspectos. El primero de ellos es que la señal de vídeo debe digitalizarse ajustando ciertos parámetros que definirán la calidad de la señal resultante y también las necesidades de ancho de banda de la red de comunicaciones:

- Resolución
- Número de tramas por segundo.
- Algoritmos de codificación.
- Flujo de datos.
- Ancho de banda requerido.

**Resolución** La resolución es el número de píxeles que contiene cada trama de imagen. En las cámaras empleadas para vídeo vigilancia se emplea la notación CIF (Common Intermediate Format) que se define con una resolución de 352x240. De esta forma Q-CIF se refiere a 176x120, 4-CIF a 704x480 y 16-CIF a 1408x960. El empleo de una resolución u otra tiene impacto en la calidad de la imagen resultante, en las necesidades de cómputo y en el ancho de banda requerido del sistema de comunicaciones.

**Número de tramas por segundo** El número de tramas por segundo (fps) define la capacidad de la señal de vídeo para representar el movimiento. Cuanto mayor sea este parámetro mejor calidad de vídeo se tendrá, pero más ancho de banda se necesitará. 24 fps es suficiente para imágenes en movimiento. Para aplicaciones de vídeo vigilancia puede ser suficiente con 10 fps.

**Algoritmos de codificación** Los algoritmos de codificación permiten reducir las necesidades de ancho de banda necesarios para la transmisión de la señal de vídeo. Los más habituales se indican la figura 4.8.

|                    | Characteristics                                                                                                                              |
|--------------------|----------------------------------------------------------------------------------------------------------------------------------------------|
| JPEG               | <ul style="list-style-type: none"> <li>• Still image technology</li> </ul>                                                                   |
| MPEG               | <ul style="list-style-type: none"> <li>• Allows streaming video to be “stopped” on one frame</li> <li>• Designed for motion video</li> </ul> |
| MPEG-4 H.264 (AVC) | <ul style="list-style-type: none"> <li>• Reduced bandwidth demand for image transfer</li> </ul>                                              |

Figura 4.8: Formato de codificación de imágenes

**Ancho de banda** Los parámetros anteriores definen los requerimientos de ancho de banda del canal de comunicaciones. La figura 4.9 resume estos requerimientos.

| Application        | Typical Screen Resolution (pixels) | Typical FPS (Frames per Second) | CODEC | Typical Upstream Bandwidth per Camera* (bits per second) | Image                                                                                                                |
|--------------------|------------------------------------|---------------------------------|-------|----------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------|
| Monitoring         | 176x120 QCIF                       | 30                              | MPEG4 | 150 kbps                                                 | <ul style="list-style-type: none"> <li>• Good Image</li> <li>• Smooth motion</li> <li>• Some text legible</li> </ul> |
|                    | 176x120 QCIF                       | 10                              | MPEG4 | 50 kbps                                                  | <ul style="list-style-type: none"> <li>• Cloudy Image</li> <li>• Step Motion</li> <li>• Some text legible</li> </ul> |
| Emergency Response | 352x240 CIF                        | 30                              | MPEG4 | 600 kbps                                                 | <ul style="list-style-type: none"> <li>• Clear Image</li> <li>• Full Motion</li> <li>• Clear Text</li> </ul>         |
|                    | 352x240 CIF                        | 5                               | MPEG4 | 100 kbps                                                 | <ul style="list-style-type: none"> <li>• Some blurring</li> <li>• Step motion</li> <li>• Text is legible</li> </ul>  |
| Investigations     | 640x480 4CIF                       | 30                              | MPEG4 | 11,200 kbps                                              | <ul style="list-style-type: none"> <li>• Sharp Image</li> <li>• Full motion</li> <li>• Text is legible</li> </ul>    |
|                    | 640x480 4CIF                       | 10                              | H.264 | 350 kbps                                                 | <ul style="list-style-type: none"> <li>• Sharp Image</li> <li>• Step motion</li> <li>• Text is legible</li> </ul>    |

\* Recommended Bandwidth Allocation set for 80% upstream and 20% upstream.

Figura 4.9: Requerimientos de ancho de banda para transmisión de vídeo

Además, el enlace de datos ascendente (de la cámara a la red) debe tener más ancho de banda que el enlace descendente (redes a la cámara). De ahí que para aplicaciones de vídeo vigilancia el ancho de banda asignado al canal de subida sea aproximadamente del 80 % y del 20 % al canal de bajada.

Por tanto, para la transmisión de vídeo entre tierra-tren, con resolución 1-CIF y considerando que se trata de un sistema que requiere una respuesta rápida ante posibles emergencias (30 fps), es necesario un ancho de banda de 600kbps, codificando las imágenes en MPEG4.

### **4.3.3. Sistemas inalámbricos para vídeo IP**

La transmisión de vídeo sobre redes IP inalámbricas es más flexible, rápida y económica a la hora de realizar montajes en exteriores. Su instalación y puesta a punto es fácil, dado que no hay que instalar cableado.

Con el uso de redes IP es posible monitorizar virtualmente cualquier lugar, siempre que esté conectado a Internet con el ancho de banda suficiente.

Los sistemas de transmisión analógicos compuestos normalmente de un transmisor, una cámara y un receptor, no pueden conectarse de manera inmediata a las redes IP y por tanto su uso está reducido a entornos cerrados o controlados. Además, debido a su rango limitado (sólo en aplicaciones militares se pueden emplear potencias suficientes para alcanzar coberturas superiores a algunos km) los sistemas analógicos sólo pueden aumentar su cobertura con el empleo de repetidores, lo que encarece enormemente este tipo de soluciones. La migración a sistemas de vídeo digital permite que las imágenes puedan convertirse en paquetes IP y, por tanto transportadas hasta cualquier lugar donde alcance la red de datos.

La mayoría de los sistemas comerciales de vídeo vigilancia disponibles en el mercado se apoyan en redes WiFi (802.11) dado su reducido coste y demostrada efectividad. Sin embargo, estas soluciones tienen una cobertura limitada entre 100 y 500 m, lo que normalmente las hace convenientes para aplicaciones de interior. Si bien este inconveniente se puede solventar con repetidores, se encarece el coste final.

Además el sistema de acceso al medio empleado en el estándar 802.11 (basado en contención) convierte en un reto implementar soluciones que proporcionen el ancho de banda y la calidad de servicio (QoS) necesarios para los sistemas de vídeo vigilancia. Por otro lado, los poco fiables sistemas de seguridad recogidos en el estándar 802.11 hace que este aspecto deba ser tenido muy en cuenta en aplicaciones donde la privacidad sea importante.

Para resolver los problemas de seguridad, algunos fabricantes ofrecen soluciones propietarias, con equipamientos de alto coste (debido fundamentalmente a la poca cantidad de unidades que pueden vender). Sin embargo la solución adecuada debe pasar por una nueva estandarización ya recogida por la IEEE en el estándar 802.16, en torno al cual ha surgido la denominación WiMAX.

#### **4.3.3.1. Transmisión analógica**

A la hora de determinar qué equipo es el ideal para realizar la transmisión de vídeo dentro de los requisitos del proyecto realizado, se deben evaluar diferentes aspectos que van a condicionar

la elección de un equipo u otro. Algunos de estos aspectos son:

- Frecuencia de funcionamiento  $\Rightarrow$  Se debe determinar cuál es la idónea dentro de las muchas posibilidades que existen. Es importante tener presente si la banda está licenciada o no. En nuestro caso, en todo momento se ha buscado bandas no licenciadas con objeto de disminuir los costes de explotación.
- Ámbito de aplicación del sistema de transmisión/recepción  $\Rightarrow$  Interior o Exterior.
- Potencia de señal  $\Rightarrow$  Se debe determinar en función de la antena de los equipos y de la distancia a cubrir.
- Distancia a cubrir  $\Rightarrow$  La distancia mínima a la que debe establecerse la comunicación entre el TX y el RX debe estar entorno a 1.5 km (distancia mínima de seguridad en el frenado de la locomotora).
- Objetos móviles/estáticos  $\Rightarrow$  El equipo transmisor estará ubicado en una estación base estática mientras que el receptor en la aplicación que nos atañe, irá alojado en una locomotora de un tren la cual podrá ir hasta velocidades de 140 km/h.

Las características de la aplicación a desarrollar, obliga al uso de un PC para procesar las imágenes captadas por una cámara de vídeo. Por ello, no es aconsejable emplear cámaras de vídeo wireless como [Consulting, 2005a], ya que su uso obligaría a establecer un enlace entre el PC y la cámara y luego otro que permita enviar desde el PC (equipo Transmisor) la imagen procesada al equipo Receptor.

En base a los aspectos anteriores, en el presente proyecto se realiza un barrido por los diferentes sistemas comerciales de radio preparados para la transmisión analógica de vídeo de forma inalámbrica. Este aspecto es muy importante, ya que si fuera a implementarse es necesario introducir una tarjeta conversora de vídeo dentro del PC para poder enviar la imagen en formato PAL/NTSC al equipo de transmisión. Este aspecto se puede ver en la figura 4.10. Otro aspecto importante de las soluciones mostradas, es la obligatoriedad de emplear los sistemas presentados en aplicaciones punto a punto. En caso de necesitar soluciones multipunto es necesario introducir multiplexores externos.

#### **4.3.3.2. Cámaras de vídeo vigilancia basadas en redes 3G**

Existen equipos de vídeo vigilancia que se apoyan en las redes 3G para transmitir vídeo, mediante el empleo de vídeo llamadas UMTS, por lo que no es necesario un servidor IP. La configuración de las mismas se realiza mediante mensajería SMS. En caso de cobertura adecuada, estas cámaras podrían transmitir señales de vídeo CIF a 10 fps (sobre 384 kbps), ligeramente insuficientes para la aplicación bajo estudio.

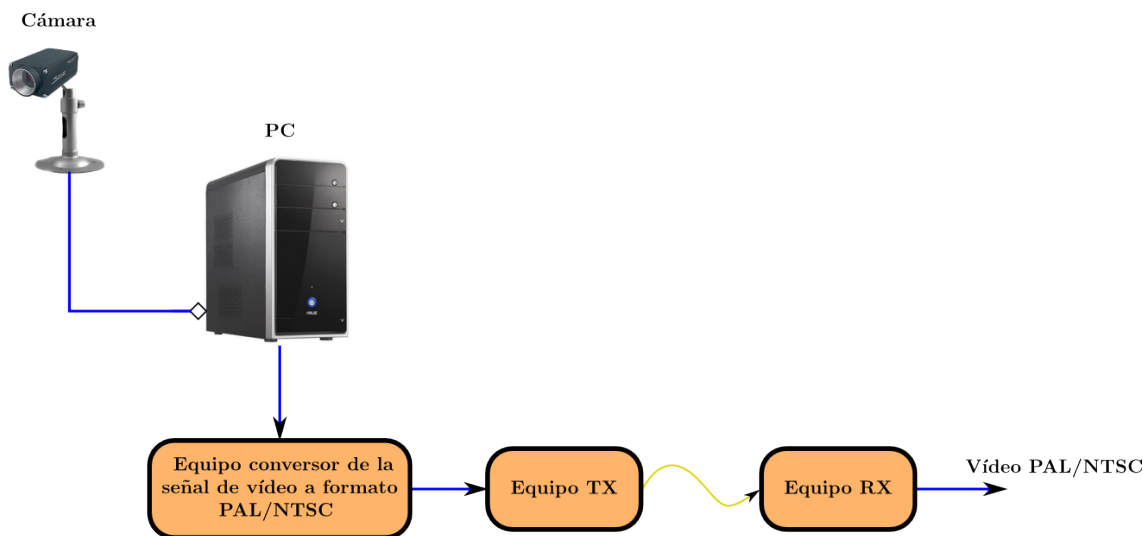


Figura 4.10: Diagrama de bloques del sistema de transmisión/recepción analógico

#### 4.3.4. Aspectos teóricos relacionados con la solución propuesta: WiMAX

WiMAX (Worldwide Interoperability for Microwave Access) es el nombre asociado a los estándares IEEE 802.16a/Rev d/e. El estándar 802.16 es también conocido como “Air Interface for Fixed Broadband Access Systems”.

Entre todos los miembros de la Alianza destaca Intel, que incluye en sus planes el desarrollo de nuevos chips WiMAX y la inversión de importantes sumas en torno a esta tecnología, lo que hace aumentar las expectativas sobre el desarrollo de este estándar. La incorporación de Intel, supone, pues, un antes y un después para WiMAX.

Los objetivos del WiMAX Forum se centran en:

- Conseguir la interoperabilidad entre los productos de diferentes fabricantes, de forma que se realicen procesos de certificación funcional más allá del cumplimiento del estándar. Esto conducirá a una mayor seguridad por parte de los usuarios a la hora de utilizar dispositivos procedentes de fabricantes diferentes.
- Reducir los costes. Si el estándar se consolida, muchos fabricantes optarán por soluciones homogéneas que garanticen la interoperatividad y mayores volúmenes de producción, gracias a la utilización de una tecnología base homogénea, lo que provocará una minimización de costes.
- Promover una marca que difunda en el mercado un estándar masivo en comunicaciones inalámbricas.
- Potenciar acciones de marketing que conviertan a WiMAX en el estándar para la banda ancha en ámbito metropolitano, a imagen de lo que hoy es WiFi para entornos de Área Local.

#### 4.3.4.1. Versiones del estándar

El objetivo del foro WiMAX es proporcionar soluciones para acceso inalámbrico en cualquier modalidad, esto es, fijo, nómada, portátil y móvil. Para ello se han definido dos versiones del estándar:

- 1.- **802.16-2004 WiMAX**  $\Rightarrow$  se basa en el estándar 802.16-2004, que agrupa diferentes versiones previas del estándar 802.16, y en el HiperMAN de ETSI (organismo de estandarización europeo). Emplea modulación OFDM y soporta acceso fijo y nómada, con o sin visión directa (LOS o NLOS). Las frecuencias usadas están en las bandas de 3.5 GHz y 5.8 GHz. Ya existen productos certificados para este estándar.
- 2.- **802.16e WiMAX**  $\Rightarrow$  está optimizado para emplear canales de radio móviles, soportando el tránsito (handoff). Utiliza SOFDMA (Scalable Orthogonal Frequency Division Multiplexing Access), una técnica de modulación multiportadora que emplea subcanalización (esta técnica se analizará posteriormente con más detalle). Los proveedores de servicios que desplieguen 802.16e podrán usar la red para proporcionar todo tipo de acceso (fijo, nómada, portátil o móvil). Las bandas de uso son las de 2.3 GHz y 2.5 GHz. La primera versión de este estándar fue aprobada a finales de 2006 y se espera disponer de productos certificados a lo largo de 2009.

Los productos basados en 802.16-2004 son menos complejos que los basados en 802.16e, pueden emplearse en un mayor rango de bandas de frecuencia libres (no licenciadas) y permiten un menor tiempo de puesta en el mercado y, en ciertos casos, mejores prestaciones de ancho de banda. Por otro lado, los productos basados en 802.16e soportan movilidad, tienen mejor cobertura en interiores, permiten a las operadoras gestión flexible del espectro y, gracias a su menor tamaño, pueden instalarse en un mayor rango de equipos.

En las tablas de la figura 4.11 se muestra un resumen de las capacidades de cada estándar.

| Definition      | Devices                                          | Locations/<br>Speed               | Handoffs      | 802.16-2004 | 802.16e |
|-----------------|--------------------------------------------------|-----------------------------------|---------------|-------------|---------|
| Fixed access    | Outdoor and indoor CPEs                          | Single/<br>Stationary             | No            | Yes         | Yes     |
| Nomadic access  | Indoor CPEs, PCMCIA cards                        | Multiple/<br>Stationary           | No            | Yes         | Yes     |
| Portability     | Laptop PCMCIA or mini cards                      | Multiple/<br>Walking speed        | Hard handoffs | No          | Yes     |
| Simple mobility | Laptop PCMCIA or mini cards, PDAs or smartphones | Multiple/<br>Low vehicular speed  | Hard handoffs | No          | Yes     |
| Full mobility   | Laptop PCMCIA or mini cards, PDAs or smartphones | Multiple/<br>High vehicular speed | Soft handoffs | No          | Yes     |

|                   | Fixed WiMAX<br>(IEEE 802.16-2004) | Mobile WiMAX<br>(IEEE 802.16e-2005)                      |
|-------------------|-----------------------------------|----------------------------------------------------------|
| Multiplexing      | OFDM                              | OFDMA                                                    |
| Duplexing mode    | TDD, FDD                          | TDD                                                      |
| Modulation        | BPSK, QPSK, 16-QAM, 64-QAM        | QPSK, 16-QAM, 64-QAM (optional uplink)                   |
| Channel bandwidth | 3.5, 7, 10 MHz                    | 5, 7, 10 MHz                                             |
| Frequency bands   | 2 GHz – 11 GHz                    | 2.3-2.4 GHz<br>2.5-2.7 GHz<br>3.3-3.4 GHz<br>3.4-3.6 GHz |

Figura 4.11: Tipos de acceso (izquierda) y principales características (derecha) en las redes WiMAX

#### 4.3.4.2. Principales Características

A continuación, se resumen las principales características, [WiMAX, 2006a]:

**Tasa de Transferencia** Por medio de un robusto esquema de modulación, el IEEE 802.16 entrega una alta tasa de transferencia con un alto nivel de eficiencia espectral, que es también tolerante a reflexiones de señal. La modulación dinámica adaptativa permite a la estación base negociar la tasa de transferencia por rangos. Por ejemplo, si la estación base no puede establecer un enlace robusto a un abonado distante usando el esquema de modulación de mayor orden, 64 QAM (Modulación por Amplitud en Cuadratura), el orden de modulación se reduce a 16 QAM o QPSK, que reduce la tasa de transferencia e incrementa el rango efectivo.

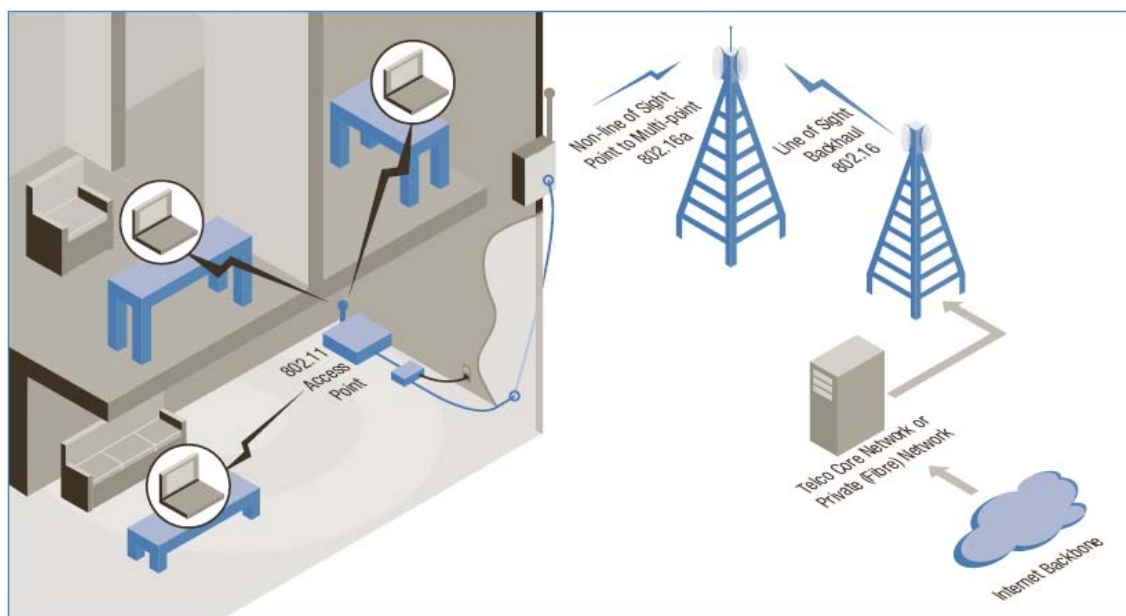


Figura 4.12: Estándar 802.16 y su integración con 802.11

**Escalabilidad** Para acomodar un planeamiento de celda fácil en el espectro de ambas bandas licenciada y no licenciada en todo el mundo, el 802.16 soporta canales de ancho de banda flexibles. Por ejemplo, si un operador tiene asignado 20 MHz de espectro, este operador puede dividirlo en 2 sectores de 10 MHz cada uno, o 4 sectores de 5 MHz cada uno. Focalizando potencia en sectores de pequeños incrementos, el operador puede incrementar el número de usuarios manteniendo una buena cobertura y tasa de transferencia. Para escalar aún más la cobertura, el operador puede reutilizar el mismo espectro en dos o más sectores creando islas propias entre las antenas de estaciones base.

**Cobertura** Junto con un esquema de modulación robusto y dinámico, el estándar IEEE 802.16 también soporta tecnologías que incrementan la cobertura, incluida la tecnología de malla (mesh) y las técnicas de “antena inteligente”. Mientras la tecnología de radio mejora y los



costos bajan, la habilidad de incrementar la cobertura y la tasa de transferencia usando múltiples antenas para crear “diversidad en transmisión y/o recepción” aumentará sensiblemente la cobertura en escenarios extremos.

**Calidad de Servicio** La capacidad de voz es extremadamente importante, especialmente en mercados internacionales no cubiertos por otros operadores. Por esta razón el estándar IEEE 802.16-2004 incluye características de calidad de servicio que permiten soportar servicios como voz y vídeo que requieren una red de baja latencia. Las características de “garantía” requeridas por el controlador de acceso al medio (MAC) del IEEE 802.16, permiten al operador brindar simultáneamente niveles de servicio Premium garantizados para negocios, tanto como niveles de servicio T1 y servicio a hogares, similares a niveles de servicio de cable, todos dentro de la misma área de servicio perteneciente a una estación base.

**Seguridad** Por lo que respecta a la seguridad, el estándar 802.16 realiza una autenticación con certificados x.509 usando DES en modo CBC, además de permitir cifrado de los datos mediante los algoritmos Triple DES (128 bits) y RSA (1024 bits).

El estándar x.509, se refiere a los formatos para certificados de claves públicas y algoritmos de validación de la ruta de los certificados. Este tipo de documentos es una parte esencial de la estructura de PKI y su función es obtener los datos de identificación del titular de la clave pública.

El uso de certificados de claves públicas permite a los usuarios tener la confianza de que las claves públicas que reciben para el cifrado de toda la información, por medio del algoritmo RSA, son de un titular legítimo y que por lo tanto al comenzar a enviar información cifrada, efectivamente la comunicación únicamente será entre usuarios auténticos de la red.

En una red inalámbrica es sumamente importante que la información que viaja a través de ésta se encuentre cifrada con un algoritmo suficientemente fuerte, debido a las condiciones del medio, ya que un usuario con la capacidad de conectarse a la red podría interceptar el tráfico de ésta y por lo tanto vulnerarla fácilmente, si es que no se tiene un algoritmo de cifrado lo suficientemente robusto.

La seguridad WiMAX soporta dos estándares de cifrado de alta calidad: 3DES y AES, este último considerado el mejor para esto.

El estándar define un proceso de seguridad enfocado directamente a la estación central. También se manejan algunas exigencias mínimas de cifrado para el tráfico y para la autenticación punto a punto, estándares que son adaptados de la especificación de interfaz del servicio de datos sobre cable y el protocolo de seguridad.

La figura 4.13 muestra el espectro de licencias disponibles.

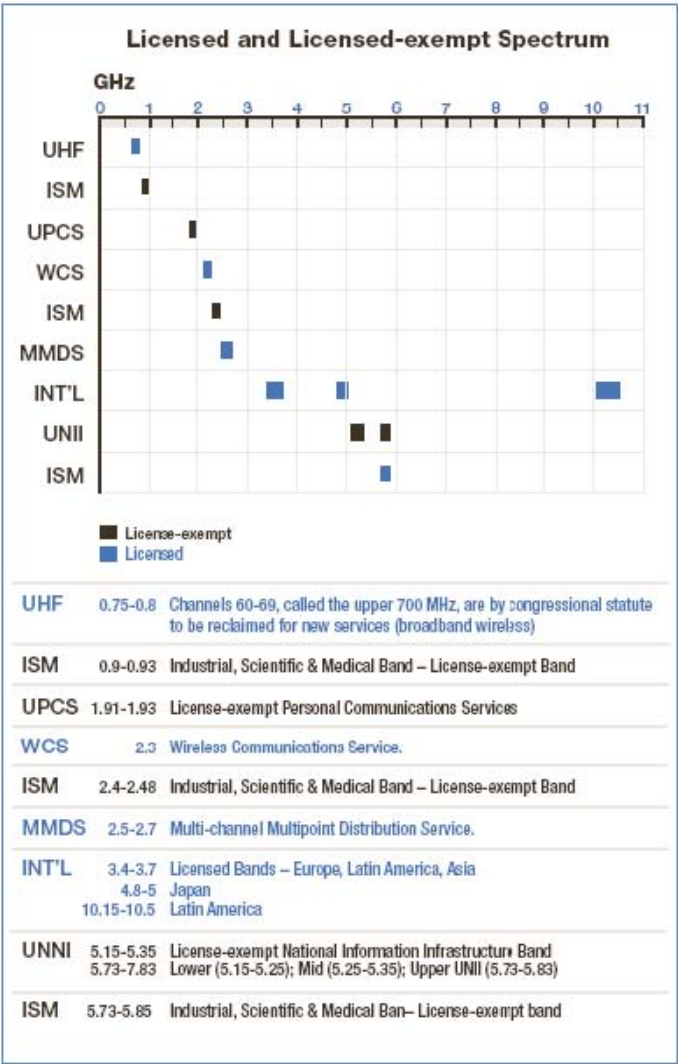


Figura 4.13: Espectro de licencias disponibles

La tabla de la figura 4.14 resume las características de WiMAX.

|                    | 802.16                             | 802.16a                                                              | 802.16e                |
|--------------------|------------------------------------|----------------------------------------------------------------------|------------------------|
| ESPECTRO           | 10 – 66 GHz                        | 2 – 11 GHz                                                           | < 6 GHz                |
| CONFIGURACION      | LOS                                | NLOS                                                                 | NLOS                   |
| VELOCIDAD DE DATOS | 34 a 134 Mbps<br>(Canal de 28 Mhz) | ≤ 70 o 100 Mbps<br>(Canal de 20 Mhz)                                 | Hasta 15 Mbps          |
| MODULACION         | QPSK, 16-QAM,<br>64-QAM            | 256 Subportadoras<br>OFDM usando<br>QPSK, 16-QAM,<br>64-QAM, 256-QAM | Igual que 802.16a      |
| MOVILIDAD          | Fijo                               | Fijo                                                                 | ≤ 113 Km/h             |
| ANCHO BANDA CANAL  | 20, 25, 28 Mhz                     | Seleccionable<br>1.25 a 20 Mhz                                       | 5 MHz<br>(Planificado) |
| RADIO CELDA TIPO   | 1.6 a 4.8 Km                       | 4.8 a 8 Km                                                           | 1.6 a 4.8 Km           |

Figura 4.14: Resumen de características WiMAX.

#### WiMAX frente a otras tecnologías inalámbricas de banda ancha.

**WiFi** El Fórum WiMAX presenta al 802.16 como una tecnología complementaria al estándar 802.11. En cierto modo esto es así, dado que WiFi proporciona accesos en ámbito de Red Local y, a medida que WiFi se desarrolla, WiMAX podría ser una solución en la creación de redes troncales para la interconexión de hotspots WiFi. Al mismo tiempo, WiMAX hace redundantes los esfuerzos de muchos entusiastas de la tecnología WiFi, que tratan de ampliar su alcance y mejorar otras características.

WiMAX tiene la oportunidad de ofrecer en comunicaciones inalámbricas características de alcance, soporte multimedia, fiabilidad y velocidad que WiFi nunca podrá lograr.

Otra diferencia del 802.16 frente al 802.11 es el uso de “time slots”, permitiendo una mayor eficiencia espectral. Se espera que gracias a la colaboración de fabricantes de productos de mercado masivo, se conseguirán economías de escala similares a las alcanzadas con dispositivos WiFi WLAN, lo que redundará en unos precios más competitivos para el equipamiento WiMAX que su alternativa “cableada”.

Los sistemas basados en la versión móvil del estándar (802.16e), que deberá estar en el mercado hacia finales del 2009, permitirán el despliegue de redes inalámbricas de largo alcance y tendrán un mayor potencial que los hotspots WiFi a la hora de proporcionar cobertura ubicua para competir con las redes celulares (móviles).

Bien sea de una forma aislada o bien en combinación con WiFi, WiMAX se presenta como una tecnología capaz de desafiar, en ciertos entornos, a las redes móviles de una manera realista.

**Tecnologías de Redes Móviles** La FCC (Federal Communications Commission, EEUU) está liberando progresivamente más espacio radioeléctrico para redes inalámbricas. Este tipo de movimientos amenazan el valor del activo de las licencias sobre el espectro de los operadores 3G licenciadas (en EEUU el 25 % del espectro concedido a los operadores móviles fue a coste cero al comienzo de los 90). De manera que se está trabajando en la apertura no licenciada de gran cantidad de espectro, destinado especialmente a nuevos entrantes, gracias al éxito de WiFi.

En ciertos ámbitos, WiMAX es una seria amenaza para la 3G, debido a sus capacidades de ancho de banda, alcance y capacidad para soportar voz con calidad de servicio. Estas características lo convierten en una alternativa tecnológica a las redes móviles, en un modo inimaginable para WiFi. Los operadores móviles están integrando WiFi en sus ofertas sin demasiado interés, tratando de controlar tanto el espectro licenciado, como los hotspots no licenciados.

Dado el lento progreso de 3G, especialmente en Europa, y el proceso racionalizado de desarrollo de WiMAX (poco usual), los operadores móviles están empezando a incluir en sus desarrollos estratégicos el análisis de amenazas y oportunidades de WiMAX.

**Solape MobileFi** En sus definiciones iniciales, WiMAX apareció como una tecnología pensada para el acceso en la última milla, mientras que el 802.20 trataba de concentrar esfuerzos para proporcionar una solución de banda ancha móvil IP. Incluso en la variante móvil 802.16e se veía al principio como WiMAX ha incrementado las prestaciones contempladas en sus especificaciones y expectativas iniciales cubriendo también el mundo móvil.

WiMAX ha usurpado gran parte del territorio Mobile-Fi. El esfuerzo inicial en el desarrollo del 802.20 fue realizado por algunas compañías pioneras en la cuarta generación de tecnologías IP inalámbricas que rechazaron inicialmente cualquier relación con el 802.16e, argumentando su deseo de desarrollar una solución pura IP.

**WiMAX móvil vs 3G** Las diferencias entre WiMAX y 3G y sus mejoras radican básicamente en el uso de OFMDA frente a WCDMA y de TDD frente a FDD. El empleo de diferentes opciones tecnológicas para modulación y multiplexación, se traducen en prestaciones claramente mejores en WiMAX.

En lo que a cobertura se refiere, ambas alternativas ofrecen prestaciones semejantes; véase figura 4.15.

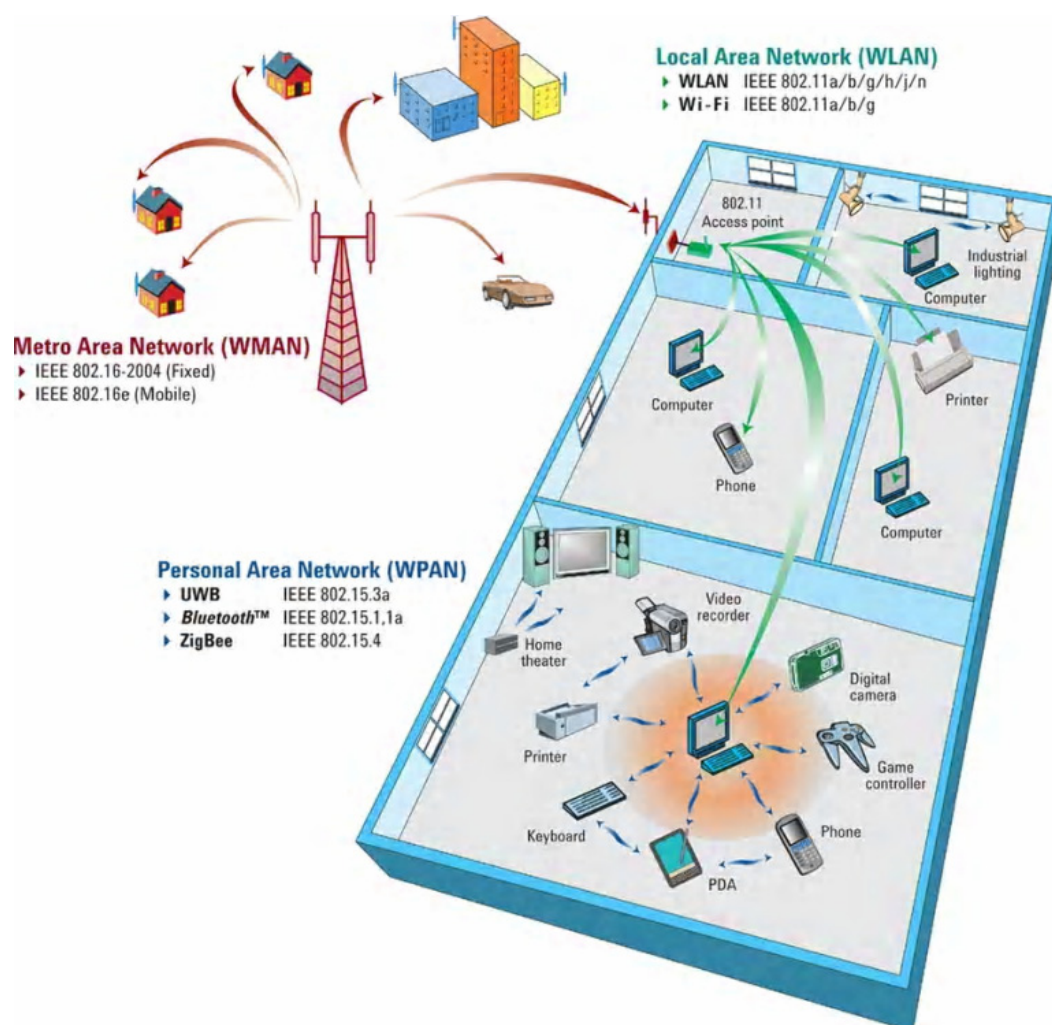


Figura 4.15: Comparativa en cobertura de las redes inalámbricas de banda ancha

Las ventajas de OFDMA son tan evidentes que 3GPP (encargado de la normalización de 3G y sus evoluciones) ha optado por esta alternativa en la especificación LTE (Long Term Evolution). Sin embargo, WiMAX cuenta ya con experiencia en OFDMA lo que supondrá una ventaja competitiva frente a las especificaciones 3GPP.

### 4.3.5. Solución probada: BreezeACCESS VL vídeo

Las soluciones anteriores presentan varios problemas, básicamente de disponibilidad de equipos, de coste y fundamentalmente de estandarización. Se trata de soluciones propietarias que implican renunciar a parte de los requerimientos iniciales del proyecto. Por ello se optó por probar una solución basada en estándar WiMAX, que no haga necesarios los añadidos propietarios a la solución. Dentro de las empresas implicadas en el desarrollo de WiMAX destaca Alvarion, tanto por la disponibilidad de soluciones como por el precio asequible de los equipos necesarios.

El producto adquirido para pruebas fue BreezeACCESS VL vídeo, la plataforma inalámbrica de banda ancha de Alvarion en la frecuencia no licenciada de 5 GHz, que forma parte de la familia BreezeACCESS, la plataforma de banda ancha inalámbrica más desplegada en el mundo.

Características tales como enlace fuera de la línea de visión (NLOS), alcance extendido, alta capacidad en todos los tamaños de paquete, cifrado y calidad del servicio (QoS) de extremo a extremo para aplicaciones donde el tiempo es crítico, son la clave de su éxito en los despliegues, a escala mundial.

Las características más importantes del equipo son:

- Funcionalidad de puente  $\Rightarrow$  configuración simple y rápida instalación de soporte VLAN 802.1Q con modos troncal, acceso, y 802.ad híbrido y QinQ.
- QoS  $\Rightarrow$  QoS de extremo a extremo con MAP utilizando priorización de paquetes.
- Refuerzo de SLA  $\Rightarrow$  soporta las velocidades de información comprometidas (CIR) y las velocidades de información máximas.
- (MIR) por usuario, por dirección; priorización de paquetes con clasificación de gama de puertos IP TOS, VLAN, DiffServ y UDP/TCP, y una degradación elegante en el caso de congestión.

Opciones de Seguridad y Filtrado:

- Opciones de cifrado AES 128 y WEP 128  $\Rightarrow$  y el nuevo modo de cifrado FIPS-197, certificado de acuerdo con los Estándares de Procesamiento de Información Federales (Federal Information Processing Standards), lista de acceso/denegación que permite conectarse solamente a las CPE autorizadas.
- Control de acceso con protocolo de dirección IP y filtrado basado en MAC  $\Rightarrow$  ofreciendo un mejor control capaz de limitar el número autorizado de direcciones IP, posibilitando una fuente adicional de beneficios, o para evitar que las transmisiones locales invadan el enlace inalámbrico.

Flexibilidad y Modularidad:

- Topología flexible que permite configuraciones autónomas o basadas en chasis para soluciones modulares y capaces de crecer en escala permitiendo el “pague según crezca”.
- Desplegable en múltiples sectores usando diversas selecciones de antenas.
- Opciones de alimentación CA y CC.
- Soporta velocidades de CPE de 3, 6 y 54 Mbps con opciones de antena incorporada o externa.
- Ancho de banda de CPE actualizado sobre el aire.

La Solución “Espectro Completo”™:

- Cubre toda la banda de 5 GHz y se integra fácilmente con las bandas de 900 MHz, 2.4 GHz, 3.5 y 4.9 GHz de BreezeACCESS usando la misma infraestructura y gama de tecnologías.
- Soporta la concurrencia de LOS, NLOS y multi-frecuencias con velocidades de abonado de 3 a 54 Mbps.
- Permite a los operadores personalizar redes para diversos segmentos del mercado, para lograr el máximo beneficio por celda.

Robustez y Confiabilidad:

- Modulación adaptativa con 8 esquemas de velocidades y una suave transición entre velocidades en respuesta a las condiciones del enlace, facilitando la robustez del enlace, establecido a la máxima velocidad por abonado posible.
- Control automático de potencia de transmisión (ATPC)  $\Rightarrow$  la unidad de acceso mide automáticamente y ajusta la potencia de transmisión de la unidad del abonado, permitiendo una instalación más fácil y optimizando el funcionamiento de la red.

Soporta varias opciones de redundancia:

- Corrección de Error Adelantada incorporada y retransmisión corrigiendo bits dañados o perdidos.
- Opción de equipo totalmente apto para exterior con OPS-AC-HD.

#### 4.3.5.1. Componentes del Sistema

La solución BreezeACCESS VL (figura 4.16) consta de una estación base y unidades de abonados en sitios del cliente (CPE). Las estaciones base están disponibles ya sea como elementos modulares o como unidades de micro celda autónomas. Las CPE están disponibles en diversos modelos para los diferentes anchos de banda y las configuraciones de usuario simple o múltiple.



Figura 4.16: Elementos de la solución BreezeACCESS VL

**Unidades de Acceso (AU)** Instaladas en el sitio de la estación base, cada AU incluye una unidad interna y una externa. La interna se conecta con la red mediante una interfaz estándar Ethernet 10/100 BaseT (RJ-45), y la unidad externa se conecta con la unidad interna mediante un cable CAT-5.

La estación base modular de estante (BS-SHVL) de chasis universal de 19' 3U alojando hasta 6 módulos AU. En un chasis BS-SH-VL pueden usarse dos módulos de fuente de alimentación (ya sea CC o CA) para una operación libre de fallos. El conjunto AU-D-BS incluye una unidad interna basada en chasis, una unidad externa montada sobre mástil y antenas de sector.

El conjunto de la micro estación base autónoma (AU-D-SA) incluye una pequeña unidad interna, una unidad externa montada sobre mástil y una antena de sector.

Con la estación base pueden utilizarse diversas antenas: 360, 120, 60 y 90 grados.

**Unidades de Abonado (SU)** La unidad de abonado (SU) permite al cliente la conexión con la estación base. Puede soportar un usuario único o múltiples usuarios. Las SU proveen una plataforma eficiente para Internet e Intranet de alta velocidad siempre conectado, VoIP, VPN y otros servicios. Cada SU se conecta con la red mediante una interfaz estándar Ethernet 10/100 BaseT (RJ-45), y se conecta con la unidad interna mediante un cable CAT-5. Cada conjunto de SU incluye una unidad interna con un único puerto de datos, cable CAT-5 interior-exterior, unidad externa de montaje sobre mástil y antena integrada en la mayoría de los casos. Existe una serie de módulos que se pueden añadir a la unidad de abonado, incluyendo el gateway de red que ofrece a los abonados residenciales, SOHO y SME una gama flexible de servicios de red alámbricos e inalámbricos, y el gateway de voz que ofrece el suministro eficaz de voz y datos.

Existen varios modelos de CPE disponibles (ff - banda de frecuencias).

- El SU-A-ff-3-1D-VL soporta una velocidad bruta de hasta 3 Mbps para un usuario único,



incluye antena integrada.

- El SU-A-ff-6-BD-VL soporta una velocidad bruta de hasta 6 Mbps para usuarios múltiples, incluye antena integrada.
- EL SU-A-ff-54-BD-VL soporta una velocidad bruta de hasta 54 Mbps para usuarios múltiples, incluye antena integrada.
- El SU-E-ff-54-BD-VL soporta una velocidad bruta de hasta 54 Mbps para usuarios múltiples, no incluye antena.

El listado completo de especificaciones del equipo se puede obtener en [Solution, 2005].

#### 4.3.5.2. Pruebas realizadas de la alternativa adoptada

**Resumen de resultados** Las pruebas realizadas con el equipo adquirido han demostrado que proporciona las características requeridas en la aplicación:

- 1.- Se trata de equipos basados en el estándar WiMAX.
- 2.- La banda de frecuencias empleada es no licenciada.
- 3.- Proporciona un ancho de banda útil de 3 Mbps, ampliable hasta 54 Mbps.
- 4.- Con velocidades de desplazamiento de 170 km/h se han recibido sin ningún problema las imágenes de vídeo.
- 5.- En enlaces LOS la distancia cubierta por el enlace supera los 7 km.
- 6.- En enlaces con obstáculos próximos a la antena (camiones en circulación) se pierde la conectividad, pero en cuanto desaparece el obstáculo se recupera.
- 7.- El coste de los equipos es comedido.

Aunque se podrían completar las pruebas con otras en entornos ferroviarios, se puede asegurar que la tecnología existente es suficiente para resolver el problema planteado en este proyecto.

**Conclusiones** En esta sección se han descrito, atendiendo a las especificaciones establecidas en el proyecto AVIZOR, las alternativas tecnológicas disponibles, se ha elegido una para realizar pruebas y se han mostrado los resultados obtenidos.

Dentro del abanico de tecnologías inalámbricas disponibles, las soluciones comerciales disponibles se reducen a cuatro:

- 1.- Sistemas analógicos de transmisión de vídeo.
- 2.- Redes inalámbricas con tecnología propietaria basada en los estándares 802.11x pero no interoperables.

- 3.- Redes inalámbricas basadas en el estándar 802.16 con mayor interoperatividad, aunque no total.
- 4.- Redes de banda ancha basadas en tecnología celular (HSDPA y HSUPA).

La primera alternativa es la más probada y barata, sin embargo sólo permiten la transmisión de vídeo dentro de cierto alcance y, al no basar su funcionamiento en redes IP, son soluciones de recorrido limitado, pues su funcionalidad no puede extenderse. Son una buena alternativa si sólo se transmite vídeo sin más pretensiones.

Las otras soluciones permiten añadir nuevas funcionalidades al sistema, pues la red inalámbrica montada puede emplearse para transmitir cualquier tipo de datos.

Sin embargo la adopción de una solución u otra depende de varios factores:

- 1.- La disponibilidad de equipos en el mercado, a ser posible de distintos fabricante e interoperables entre sí.
- 2.- El cumplimiento de las normativas aplicables y el precio de los equipos.
- 3.- El uso de estándares abiertos sin modificaciones propietarias que impiden la sustitución por otros equipos equivalentes.
- 4.- El empleo de bandas de frecuencia no licenciadas.
- 5.- Soporte para el ancho de banda necesario en vídeo vigilancia y para el reparto asimétrico del mismo, dado que en esta aplicación el ancho de banda debe asignarse en un 80 % al enlace de subida.

A la vista de estos condicionantes, parece descartable por el momento el empleo de enlaces basados en redes celulares, por problemas de cobertura y de ancho de banda de disponible. Sin embargo con el despliegue de las soluciones HSPA será factible el empleo de esta solución, siempre que la cobertura alcance a los pasos a nivel necesarios. De los planes de las compañías parece que esto no ocurrirá hasta dentro de varios años (sobre 2012).

El empleo de enlaces basados en modificaciones de estándares WiFi es una opción ya disponible en el mercado y probada en condiciones similares. Sin embargo presentan varios inconvenientes: son soluciones propietarias y requieren de una infraestructura significativa (puntos de acceso cada 500 m y enlaces entre ellos, bien mediante cableado o mediante troncales inalámbricas).

La solución probada presenta como ventaja el apoyarse en el estándar WiMAX que incorpora soluciones estandarizadas a los problemas que el caso de WiFi se resuelven de manera propietaria. El coste de la solución no es muy elevado y ha mostrado prestaciones suficientes para la aplicación.

## 4.4. Sistema de transmisión de vídeo en tiempo real al equipo receptor

### 4.4.1. Especificación de la interfaz externa

Este apartado tiene como objetivo definir las interfaces externas del sistema AVIZOR, a montar en una unidad de tren. Se definirá la interfaz SW de los equipos del sistema AVIZOR-Sistema Embarcado.

Se define el DVR como un sistema grabador digital de vídeo ubicado en el sistema embarcado. La central DVR se conecta con el equipo Gestor de Comunicaciones por medio de una línea Ethernet para obtener vídeo de las cámaras remotas (cámaras ubicadas en los puntos conflictivos de la vía, por ejemplo en pasos a nivel) y alarmas para la gestión de las grabaciones.

#### 4.4.1.1. Protocolo de comunicación con el equipo gestor de comunicaciones

El equipo DVR recibirá el streaming de vídeo y datos de alarmas y estados a través de la conexión Ethernet con el Punto de Acceso (AP). A continuación se detalla cada nivel OSI de esta comunicación, figura 4.17.



Figura 4.17: Pila de protocolos OSI

**Nivel Físico** El estándar de transmisión en este enlace será 100Base-TX, siendo la velocidad binaria máxima alcanzable de 100 Mbps. La interfaz física del conmutador de la red Ethernet soporta la negociación automática en 10Base-TX y 100Base-TX.

**Nivel de Enlace de Datos** Los mensajes en la red deben soportar tramas Ethernet de 64 bytes de tamaño mínimo y 1518 bytes de tamaño máximo.

**Nivel de Red** Se utilizará IPv4. El equipo AP tendrá una dirección IP fija, el resto de los equipos de la central DVR pueden tener IP dinámicas. El equipo AP debe soportar los protocolos ARP e ICMP.

**Nivel de Transporte** El puesto de Tierra del paso a nivel, definido como PT, debe implementar dos protocolos:

1. UDP para el transporte de mensajes de control desde el PT, pasando por el punto de acceso (AP) como bridge, hacia la central DVR (mensajes de alarma y de estado). Se utilizará un mismo número de puerto UDP para mensajes de control en todos los PT. El valor por defecto del puerto será 10000.
2. RTP sobre UDP para el streaming de vídeo real desde el PT a la central DVR. Se contempla la posibilidad de que el PT indique el puerto UDP por el que se envíe el vídeo, existiendo no obstante un valor por defecto (10001).

### **Nivel de Sesión**

- **Asignación de IP**  $\Rightarrow$  la central DVR usará direcciones IP dinámicas, y se comunicará mediante DHCP con el AP u otro servidor DHCP para su asignación.
- **Mensajes de Control**  $\Rightarrow$  mediante el sistema de “polling” (interrogación síncrona de estado), la central DVR detectará la existencia de un PT accesible desde la red del tren. La central DVR enviará un paquete (sin datos de aplicación), para que el PT detecte la dirección IP y el puerto de la central, y comience a enviar mensajes de control a la central DVR.
- **Streaming de Vídeo**  $\Rightarrow$  se usará el control de sesión que ofrece RTP, pero no se usa el canal de retorno RTCP ni RTSP. El cliente RTP que reside en la central DVR simplemente establece la sesión y recibe los streams sin dar información sobre la calidad de dicha sesión.

### Nivel de Presentación

**Mensajes de Control** Se configurará un puerto UDP para datos de alarmas y estados, que serán codificados en base a XML, un estándar abierto y fácilmente legible e interpretable debido a que es textual.

- **Formato de la Trama**  $\Rightarrow$  los datos de la aplicación serán representados por etiquetas XML y empaquetados en una trama binaria para prevenir su corrupción, tal y como se observa en la tabla 4.1.

| Identificación                                                                | Longitud                                                                                                                      | Datos                                                  | CRC                                                                   |
|-------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------|-----------------------------------------------------------------------|
| 1 byte                                                                        | 2 bytes                                                                                                                       | "N" bytes                                              | 2 bytes                                                               |
| Identificación del comienzo de la trama. Todos los bits menos uno están fijos | Longitud total del mensaje (incluyendo todos los campos de la trama, es decir, "N"+5 bytes), los valores válidos son: 0-65535 | Datos de aplicación con/sin compresión en notación XML | Checksum de los campos anteriores (Identificación + Longitud + Datos) |

Tabla 4.1: Formato de la trama

- **Identificación de la trama**  $\Rightarrow$  el byte de identificación está compuesto por los bits mostrados en la tabla 4.2.

| Bit 7 | Bit 6 | Bit 5 | Bit 4 | Bit 3 | Bit 2 | Bit 1 | Bit 0 |
|-------|-------|-------|-------|-------|-------|-------|-------|
| 1     | 0     | 0     | 1     | 0     | X     | 0     | 1     |

Tabla 4.2: Identificación de la trama

El *bit 2* indica si los datos incluidos en la trama están comprimidos con LZW (= *true*) o no (= *false*); en hexadecimal serían los valores 0x91 y 0x95.

- **Algoritmo del Checksum**  $\Rightarrow$  el algoritmo para calcular el CRC será el estándar CRC-16-CCITT con el polinomio generador:

$$x^{16} + x^{12} + x^5 + 1$$

Este CRC se usa en protocolos como: X.25, V.41, CDMA, Bluetooth, PPP, IrDA.

**Streaming de Vídeo** Se prevé la recepción de video-streaming procedente de cuatro cámaras diferentes. Cada una de estas cámaras tendrá asignada una dirección IP diferente.

Se hará uso de tráfico Unicast.

El vídeo recibido desde el equipo PT debe estar comprimido con el Codec MPEG-4 Simple Profile (SP) con la resolución CIF y 12,5 fps. El elementary stream resultante estará empaquetado en tramas RTP con una cabecera mínima de 12 bytes.

En la tabla 4.3 se muestra la estructura de paquete RTP.

| Bits | 0-1                      | 2 | 3 | 4-7  | 8 | 9-15                | 16-31               |
|------|--------------------------|---|---|------|---|---------------------|---------------------|
| 0    | Versión                  | 0 | 0 | 0000 | 0 | Formato del Payload | Número de Secuencia |
| 32   | Timestamp                |   |   |      |   |                     |                     |
| 64   | Identificador del Stream |   |   |      |   |                     |                     |
| 96+  | Datos del Vídeo          |   |   |      |   |                     |                     |

Tabla 4.3: Estructura del paquete RTP.

**Nivel de Aplicación** Este nivel afecta solamente a los mensajes de control que manda el PT a la central DVR. El streaming de vídeo por RTP no incluye ninguna definición en este nivel, se usa un streaming RTP estándar y simple sin extensiones propietarios.

- **Mensajes de Control**  $\Rightarrow$  en este nivel se especifica una notación XML que se puede consultar y validar. Se define, en este protocolo, un único tipo de mensaje:
  - **Mensaje de Estado y Alarma**  $\Rightarrow$  la central DVR pedirá periódicamente mensajes del Puesto de Tierra para reflejar el estado / alarmas de los pasos controlados por el equipo de tierra. La periodicidad de estos pedidos será de medio segundo (pudiendo ser configurable dentro de la central DVR).

En la tabla 4.4 se especifican los parámetros del mensaje; los mensajes de estado contienen los datos en caso de que todo el sistema vaya bien.

|              |                                                                                  |                      |                                         |                                                 |                                              |
|--------------|----------------------------------------------------------------------------------|----------------------|-----------------------------------------|-------------------------------------------------|----------------------------------------------|
| <status>     | <ok>                                                                             |                      |                                         |                                                 |                                              |
| <version>    | ...                                                                              | </version>           | Contiene la versión del software del PT |                                                 |                                              |
| <crossing>   | Información sobre el paso a nivel                                                |                      |                                         |                                                 |                                              |
|              | <id>                                                                             | ...                  | </id>                                   | Identificador del paso a nivel                  |                                              |
|              | <protection>                                                                     | on, off              | </protection>                           | Barreras de paso cerradas “on” o abiertas “off” |                                              |
| </crossing>  |                                                                                  |                      |                                         |                                                 |                                              |
| <cameras>    | Lista de elementos de tipo cámara. Dichos elementos serán de la siguiente forma: |                      |                                         |                                                 |                                              |
|              | <camera>                                                                         | 1...n                |                                         |                                                 |                                              |
|              |                                                                                  | <id>                 | ...                                     | </id>                                           | Identificador de la cámara                   |
|              |                                                                                  | <state>              | on/off                                  | </state>                                        | Estado de la cámara (funcionando o no)       |
|              |                                                                                  | <port>               | ...                                     | </port>                                         | Puerto UDP para vídeo                        |
|              | </camera>                                                                        |                      |                                         |                                                 |                                              |
|              | ...                                                                              |                      |                                         |                                                 |                                              |
| </cameras>   |                                                                                  |                      |                                         |                                                 |                                              |
| <intrusion>  | Informa de intrusión en el paso a nivel identificado                             |                      |                                         |                                                 |                                              |
|              |                                                                                  | <state>              | on/off                                  | </state>                                        | Indica si hay una intrusión en este instante |
|              |                                                                                  | <level>              | normal, serious                         | </level>                                        | Nivel de la intrusión                        |
| </intrusion> |                                                                                  |                      |                                         |                                                 |                                              |
| <datetime>   | Informa de la fecha y hora del Puesto de Tierra (en formato UTC)                 |                      |                                         |                                                 |                                              |
|              | <date>                                                                           | Información de fecha |                                         |                                                 |                                              |
|              |                                                                                  | <year>               | ...                                     | </year>                                         | Año                                          |
|              |                                                                                  | <month>              | ...                                     | </month>                                        | Mes                                          |
|              |                                                                                  | <day>                | ...                                     | </day>                                          | Día                                          |
|              | </date>                                                                          |                      |                                         |                                                 |                                              |
|              | <time>                                                                           | Información horaria  |                                         |                                                 |                                              |
|              |                                                                                  | <hour>               | ...                                     | </hour>                                         | Hora                                         |
|              |                                                                                  | <minute>             | ...                                     | </minute>                                       | Minuto                                       |
|              |                                                                                  | <second>             | ...                                     | </second>                                       | Segundo                                      |
|              |                                                                                  | <milisec>            | ...                                     | </milisec>                                      | Milisegundo                                  |
|              | </time>                                                                          |                      |                                         |                                                 |                                              |
| </datetime>  |                                                                                  |                      |                                         |                                                 |                                              |
| </ok>        |                                                                                  |                      |                                         |                                                 |                                              |

Tabla 4.4: Parámetros del mensaje de control

#### 4.4.1.2. Notas

Hay definido una lista de cámaras, aunque en el sistema actual, el puesto de paso a nivel solamente manda un vídeo de una cámara (el identificador de la cámara en el mensaje siempre será “1”). El mensaje contempla ya un posible escenario futuro donde un paso a nivel mandará varios flujos de vídeo al tren. Cada cámara enviará el stream de vídeo por un puerto UDP diferente, el cual se especifica en un campo extra destinado a tal efecto. En el caso de que el PT envíe vídeo desde una sola cámara se define un puerto UDP por defecto (10001).

Los parámetros de fecha y hora se ignorará en el escenario actual, pero podría ser útil en un futuro para sincronizar las grabaciones en el tren con el reloj del puesto de paso a nivel. Se definen por ahora dos niveles de gravedad de la intrusión, aunque posiblemente siempre será “normal” por no haber algorítmica de distinción en el puesto de paso a nivel.





## Capítulo 5

# Implementación del sistema en tiempo real

### 5.1. Introducción

Basándose en la teoría vista hasta el momento, se ha implementado un sistema, ubicado en entornos exteriores, capaz de detectar obstáculos en tiempo real y enviar la información relevante, mediante un radioenlace WiMAX, al módulo receptor ubicado en el tren; el cual se aproxima al escenario en cuestión.

La metodología seguida pasa por utilizar una arquitectura cliente-servidor, donde la solución software/hardware propuesta con anterioridad, es tratada como un servidor que proporciona imágenes “compuestas” al cliente, siendo este último el encargado de visualizar la información al conductor del tren. Las imágenes a color de la escena son modificadas de manera que se muestre la ubicación y el área ocupada por los posibles obstáculos. Además, se transmiten mensajes de alarma en formato XML donde se describe, más detalladamente, el resto de información sustraída de la escena; véase 4.4.1.

Este capítulo se centra en describir cual ha sido el esquema seguido para implementar el servidor del módulo de tierra, ubicado en las zonas de riesgo potencial para la circulación ferroviaria. La idea pasa por diseñar un esquema de programación basado en lenguaje C y concurrente, optimizando así su tiempo de cómputo, y permitiendo su viabilidad para aplicaciones en tiempo real.

Se ha de tener en cuenta que la arquitectura hardware dispuesta (PC con 4 núcleos) permite asociar los distintos procesos a CPU's diferentes de manera que se realice un procesamiento distribuido. Así, se consigue repartir la carga computacional entre los diferentes procesadores; optimizando los tiempos de cómputo y minimizando la sobrecarga de los procesadores.

## 5.2. Servidor del módulo de tierra

Se pueden distinguir, a grandes rasgos, 6 subsistemas dentro del servidor del módulo de tierra. Éstos se caracterizan por poseer una funcionalidad diferenciada dentro del servidor y por ejecutarse independientemente en diferentes hilos.

1. **Subsistema de sincronización:** se encarga de dar paso a los hilos de captura, envío de vídeo y envío de alarma de manera que cumplan sus restricciones temporales.
2. **Subsistema servidor de imágenes:** es el encargado de garantizar el envío de imágenes compuestas al “frame ratio” especificado en el proyecto AVIZOR (12’5 fps). Se trata de imágenes “compuestas” ya que se modifican para localizar, mediante un recuadro, los obstáculos detectados.
3. **Subsistema servidor de mensajes:** se encarga de enviar mensajes UDP en formato “XML”. Estos mensajes contienen información acerca del número de obstáculos detectados, identificador del paso a nivel, estado de las cámaras, marca de tiempo, etc..
4. **Subsistema de monitorización:** es el encargado de monitorizar el estado del sistema y ejecutar las medidas correctoras oportunas en caso de fallo.
5. **Subsistema de captura:** se encarga de realizar la captura de las cámaras Gigabit-Ethernet mediante un trigger generado de manera software.
6. **Subsistema de procesamiento:** soporta la mayor parte de la carga computacional. Es el encargado de obtener un “streaming” de vídeo donde queden localizados y dimensionados los obstáculos detectados dentro de la región de interés.

En la figura 5.1 se presenta el esquema de los procesos concurrentes necesarios para la implementación del sistema, indicando los procesadores a los que son asignados.

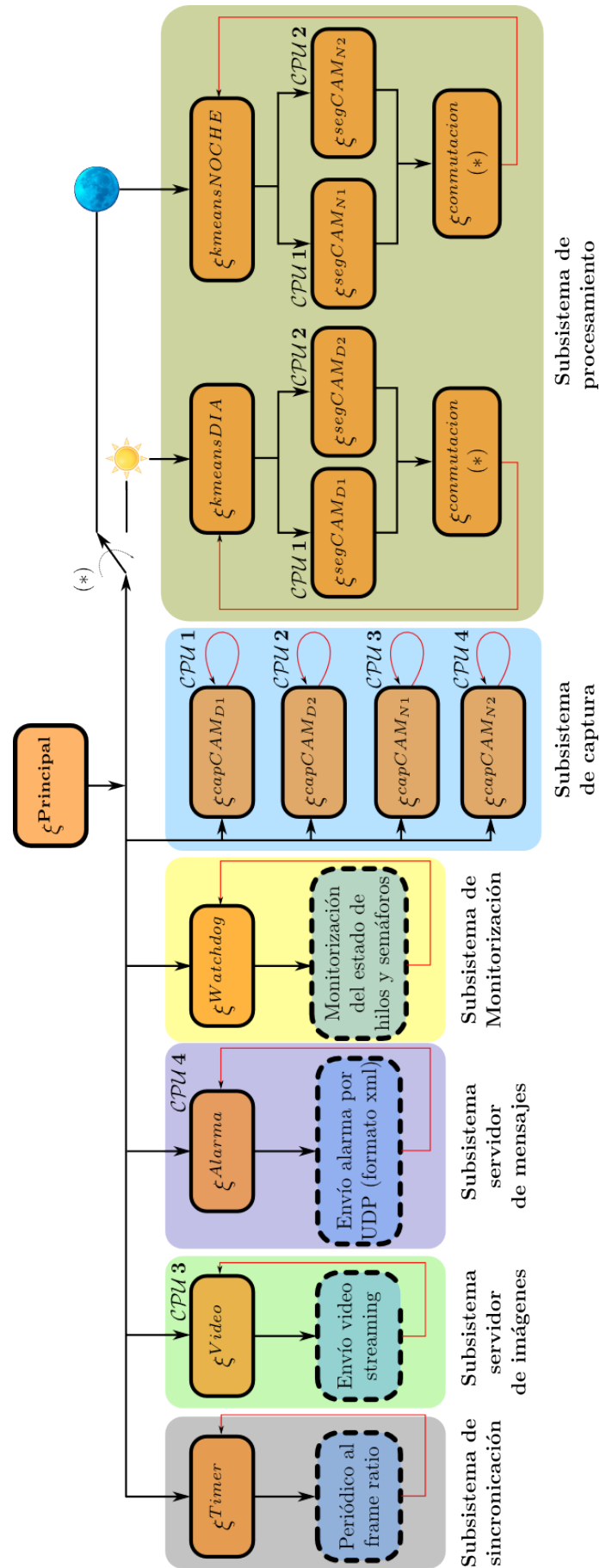


Figura 5.1: Esquema de los hilos empleados para el funcionamiento del sistema; procesado concurrente

donde:

- $\xi^{Timer}$ : hilo encargado de asegurar el frame ratio requerido. Genera señales periódicas y da paso a los semáforos de vídeo, alarma y captura de las cámaras pertinentes.
  - $\xi^{Video}$ : hilo encargado del envío de video-streaming MPEG, mediante protocolo RTP, al sistema receptor ubicado en la cabina del tren. Se implementa como un servidor de imágenes al que se conecta un tren que se aproxima a la región monitorizada.
  - $\xi^{Alarma}$ : hilo encargado del envío de alarmas al sistema receptor mediante protocolo UDP. Se implementa como un servidor UDP que proporciona mensajes en formato “XML” al tren que se aproxima a la región de interés.
  - $\xi^{Watchdog}$ : hilo encargado de monitorizar de manera periódica el estado de los hilos, semáforos y demás anomalías del correcto funcionamiento del sistema; ejecutando algún tipo de medida correctiva si es necesario.
  - $\xi^{kmeansDIA}$ : hilo encargado de la algoritmia de segmentación e identificación de objetos en modo diurno (en caso de fusionar varios algoritmos, se implementarán de manera paralela a este).
    - $\xi^{segCAM_{Dx}}$ : hilo encargado de la segmentación de las cámara de color “x” (las imágenes de las dos cámaras son segmentadas de manera síncrona y concurrente).
    - $\xi^{conmutacion}$ : hilo encargado de equilibrar los niveles de color de la captura de ambas cámaras y con capacidad de decisión para la conmutación DIA/NOCHE.
  - $\xi^{capCAM_{Dx}}$ : hilo encargado de la captura de imágenes provenientes de la cámara de color “x” a la tasa especificada en hilo del timer.
  - $\xi^{kmeansNOCHE}$ : hilo encargado de la algoritmia de segmentación e identificación de objetos en modo nocturno.
    - $\xi^{segCAM_{Nx}}$ : hilo encargado de la segmentación de las cámara monocromáticas “x” (las imágenes de las dos cámaras son segmentadas de manera síncrona y concurrente).
    - $\xi^{conmutacion}$ : hilo encargado de equilibrar los niveles de color de la captura de ambas cámaras y con capacidad de decisión para la conmutación NOCHE/DIA.
  - $\xi^{capCAM_{Nx}}$ : hilo encargado de la captura de imágenes provenientes de la cámara monocromáticas “x” a la tasa especificada en hilo del timer.
- (\*) La conmutación NOCHE  $\leftrightarrow$  DIA provoca el cambio de la adquisición de las cámaras monocromáticas a color o viceversa, en función del sentido del cambio.

A continuación se detalla el proceso de implementación llevado a cabo en cada uno de los subsistemas.

### 5.2.1. Subsistema de sincronización

Este subsistema está compuesto por un solo proceso, el cual se lanza al comenzar la ejecución de la solución propuesta. Se implementa como un proceso periódico e independiente cuya implementación se define a continuación:

1. Se lanza el proceso “timer”, como un hilo independiente, al comienzo del cuerpo principal de la solución software. A dicho proceso se le asocia un semáforo (“sem\_timer”), con valor inicial la unidad.
2. Se inicializa, en dicho proceso, una señal periódica (SIGALRM: “Alarm clock”, definida en “signal.h”) que ejecuta una rutina de manera periódica; al “frame ratio” que se ha de garantizar.
3. Se duerme el proceso mediante una primitiva **P(sem\_timer)** (*wait*) de manera que se reduzca el consumo de CPU a la ejecución periódica de la rutina del timer.
4. Rutina del timer: es la encargada de reiniciar el contador de la señal periódica del timer y dar paso a los procesos de captura de ambas cámaras (subsistema de captura), de envío de video-streaming (subsistema servidor de imágenes) y de envío de mensajes (subsistema servidor de mensajes). Para ello, se ejecuta una primitiva **V** (*signal*) sobre los semáforos que sincronizan dichos procesos.

Está, por tanto, planteado como un único hilo de manera que minimice el consumo de CPU.

### 5.2.2. Subsistema servidor de imágenes

Se plantea como un único proceso encargado de garantizar el envío de imágenes como un flujo de vídeo codificado mediante MPEG-4. A dicho proceso se le asocia un semáforo (“sem\_video”), inicializado con valor nulo al inicio del proceso. De esta manera, se consigue que sea el subsistema de sincronización el encargado de dar paso al envío de imágenes a una tasa definida por el “frame ratio”. En el siguiente párrafo se indican los pasos seguidos para su implementación.

1. Se lanza el proceso a continuación del subsistema de sincronización, permitiendo así el envío de imágenes de la escena desde el primer momento.
2. Se ejecuta la primitiva **P(sem\_video)** al comienzo del proceso, de manera que su ejecución esté sincronizada por el subsistema anterior.
3. A continuación, se da paso a la rutina de etiquetado de los obstáculos detectados, componiendo la imagen final que se envía al módulo receptor ubicado en la cabina del tren.
4. Por último, se codifican y envían las imágenes según el estándar definido en MPEG-4 mediante el protocolo RTP (“Real-time Transport Protocol”). El puerto por defecto por el cual se envía el “streaming” de vídeo es el “10001”.

### 5.2.3. Subsistema servidor de mensajes

Este subsistema tiene como objetivo el envío permanente de mensajes de control al módulo receptor, cuyo formato se especificó en el apartado 4.4.1. Para ello se emplea el protocolo UDP. Se utilizará un mismo número de puerto para mensajes de control en todos los PT. El valor por defecto del puerto será “10000”.

La implementación del servidor de mensajes “XML” se realiza de manera equivalente a la del subsistema servidor de imágenes; en un único proceso y sincronizado mediante un semáforo (“sem\_msg”) inicializado con valor nulo. Los pasos se indican a continuación:

1. El proceso es lanzado de manera secuencial a continuación del proceso servidor de imágenes.
2. Se ejecuta la primitiva **P(sem\_msg)** al comienzo del proceso. Con esto se consigue que su ejecución sea síncrona a la tasa deseada.
3. Se espera a que un cliente se conecte al servidor; momento en el cual comienza el envío de mensajes de control con la información relevante de la región de monitorizada.

### 5.2.4. Subsistema de monitorización

Este hilo se implementa ante la necesidad de disponer de un proceso independiente que sea capaz de supervisar el estado de los diferentes hilos. Es el encargado de realizar las medidas correctoras oportunas, de manera que el sistema pueda continuar su funcionamiento ante el fallo repentino de alguno de los procesos.

Su implementación, discurre entre los siguientes pasos:

1. Se crea el proceso como un hilo independiente en el cuerpo principal de la solución software (se le asocia un semáforo (“sem\_monit”) con valor inicial la unidad).
2. Al igual que en el subsistema de sincronización, se inicializa una señal periódica (SIG-PROF: “Profiling timer expired”, definida en “signal.h”) que ejecuta la rutina de monitorización periódicamente a una frecuencia definida a priori.
3. Se “duerme” el proceso mediante una primitiva **P(sem\_monit)** (*wait*) de manera que se reduzca el consumo de CPU a la ejecución periódica de la rutina de monitorización.
4. Los parámetros monitorizados, así como las medidas correctoras, se muestran en la tabla 5.1.

| Parámetro Monitorizado        | Posible Error                      | Acción Correctora                     |
|-------------------------------|------------------------------------|---------------------------------------|
| Actividad hilo timer          | Hilo Inactivo                      | Relanzar proceso                      |
| Actividad hilo vídeo          | Hilo Inactivo                      | Relanzar proceso                      |
| Actividad hilo mensajes       | Hilo Inactivo                      | Relanzar proceso                      |
| Actividad hilos de captura    | Hilo Inactivo                      | Relanzar proceso                      |
| Actividad hilo procesamiento  | Hilo Inactivo                      | Relanzar proceso                      |
| Estado semáforo vídeo         | Número de hilos esperando excesivo | Reiniciar semáforo y relanzar proceso |
| Estado semáforo alarma        | Número de hilos esperando excesivo | Reiniciar semáforo y relanzar proceso |
| Estado semáforo captura       | Número de hilos esperando excesivo | Reiniciar semáforo y relanzar proceso |
| Estado semáforo procesamiento | Número de hilos esperando excesivo | Reiniciar semáforo y relanzar proceso |

Tabla 5.1: Parámetros supervisados en el subsistema de monitorización y medidas correctoras aplicadas

### 5.2.5. Subsistema de captura

Este subsistema es el encargado de capturar las imágenes de las cámaras Basler Scout (Gigabit Ethernet) y almacenarlas en variables accesibles por todos los subsistemas; de esta manera, cada subsistema puede proveerse de imágenes bajo demanda y en cualquier instante.

Se crean cuatro hilos, uno para cada cámara, de manera que cada par de cámaras, a color o monocromo, comparten el mismo proceso. Por tanto, se dispone de dos procesos, uno para la captura de las cámaras a color y otro para la captura de las cámaras monocromáticas. Las imágenes son capturadas con tamaño 4-CIF y en el espacio de color YUV, por lo que es necesario convertirlas al espacio de color RGB y redimensionarlas a tamaño CIF.

El método de adquisición se realiza mediante un “trigger” software síncrono. La implementación del subsistema de captura se indica a continuación:

1. El proceso se lanza de manera secuencial al proceso servidor de mensajes, y previo al subsistema de procesamiento. Se ha de tener en cuenta que se debe haber inicializado, con valor nulo, cuatro semáforos asociados a la captura de cada una de las cámaras ( $sem\_cam_i$ ).
2. Se ejecuta la primitiva  $P(sem\_cam_i)$  al comienzo del proceso. Con esto se consigue que su ejecución sea síncrona y se garantice la tasa indicada en la sección de especificaciones 2.2.1 de 12'5 fps.
3. Se almacena la imagen capturada en un array accesible por todos los subsistemas, al cual accederán de manera asíncrona y bajo demanda.

### 5.2.6. Subsistema de procesamiento

Este último subsistema es el encargado de realizar todo el procesamiento algorítmico de la propuesta. Los procesos creados en su ejecución se orientan tomando dos enfoques; por un lado se persigue mantener el esquema de sincronismo planteado hasta el momento, y por otro lado se plantea como una necesidad para reducir los tiempos de cómputo y distribuir la carga computacional a varios procesadores.

Inicialmente se crean 2 procesos, relativos a la ejecución del sistema bajo condiciones de luminosidad diurna y nocturna (reforzada mediante iluminación infrarroja), respectivamente. A ambos procesos padre, se les asocia sendos semáforos definidos como “sem\_día” y “sem\_noche”, respectivamente. Dentro de ambos procesos, y de manera dual, son creados varios hilos, los cuales se citan a continuación:

1. **Hilos de segmentación**  $\Rightarrow$  se trata de los hilos encargados de realizar el proceso de segmentación de fondo. Se crean dos hilos, cada uno asociado a una de las cámaras en funcionamiento, cuyo objetivo es obtener a su salida la imagen segmentada; los obstáculos aparecen etiquetados como zonas blancas sobre fondo negro.  
Existe un punto de espera entre ambos procesos; hasta que ambos no concluyan su ejecución, no se evoluciona en el bucle. Esto es debido a que la siguiente rutina a ejecutar, relativa a la eliminación de sombras, necesita de ambas imágenes a su entrada.  
Por tanto, ambos hilos son creados y eliminados en cada iteración.
2. **Hilo de equilibrado y decisión de conmutación**  $\Rightarrow$  dos son las funcionalidades que recaen sobre este hilo:
  - a) Es el encargado de **monitorizar** el valor medio de las imágenes vistas por las dos cámaras en funcionamiento y **equilibrar** la imagen vista por ambas en torno a un valor especificado a priori. El ajuste se realiza actuando sobre los parámetros internos de la cámara (ganancia y tiempo de exposición) y se ejecuta en aquella cuya desviación, respecto del valor medio, supere un determinado umbral. El proceso concluirá cuando las dos cámaras posean un valor medio, dentro de la región de interés, situado entorno a la constante fijada con anterioridad.
  - b) La **decisión de conmutación** se implementa bajo la siguiente premisa:  
*Siempre y cuando las cámaras de color puedan situarse en torno al valor medio definido como una constante, se ejecutará el sistema en modo diurno. En caso contrario, cuando la modificación de los parámetros de las cámaras no permita mantener las imágenes capturadas en torno a la media, se deberá conmutar a modo nocturno. La vuelta al modo día se realizará en el momento en el que las cámaras a color vuelvan a ser capaces de capturar imágenes con media en torno al umbral.*

Si, durante la ejecución de este hilo, se decide que el sistema ha de conmutar de día a noche, o viceversa, se detiene el hilo padre, en el modo de ejecución en el que se encuentre (día o noche), mediante la primitiva P(sem\_dia/noche), y se ejecuta el hilo con el modo de funcionamiento oportuno.



# Capítulo 6

## Pruebas y resultados

En este capítulo se realiza una descripción de las pruebas realizadas para validar los métodos propuestos y analizar el comportamiento de los mismos ante diferentes situaciones. Por tanto, los resultados mostrados se dividen en dos grandes bloques; detección de obstáculos y servidor de imágenes y mensajes de control. El objetivo es plasmar, de manera visual, los resultados obtenidos para cada uno de los procesos que componen dichos bloques funcionales.

Las imágenes que se presentan a continuación han sido capturadas bajo las condiciones impuestas por los entornos exteriores. Para el desarrollo de las pruebas se han ubicado las cámaras en la terraza de la última planta del Departamento de Electrónica de la Universidad de Alcalá. Su situación, bajo una malla de hormigón y rodeado de columnas y paredes, hace que este escenario se encuentre sometido constantemente a circunstancias de las consideradas problemáticas: evolución del modelo de fondo, proyección de sombras de objetos no presentes en la escena, cambios bruscos de iluminación, etc.. De esta manera, los resultados arrojados pueden considerarse cercanos a los del peor caso en una situación real.

De ahora en adelante, se asume el mismo modelo de fondo para todas las pruebas realizadas. Dicho modelo queda plasmado en la figura 6.1.



Figura 6.1: Escenario empleado como fondo para las pruebas realizadas

## 6.1. Detección de obstáculos

La detección de obstáculos es la consecuencia lógica de la concatenación y superposición de sus procesos formantes. Es en esta sección donde se muestran los resultados obtenidos en cada uno de ellos, observando cómo influyen en la detección final.

### 6.1.1. Segmentación de fondo mediante el algoritmo propuesto

#### 6.1.1.1. Segmentación mediante el algoritmo “KAM”

En las imágenes mostradas en la figura 6.2 se indican un conjunto de imágenes capturadas por las cámaras y su correspondiente segmentación de fondo tras haber aplicado el algoritmo “KAM” (K-medias Adaptativo con distancia de Mahalanobis).



Figura 6.2: Resultados obtenidos en la segmentación de fondo en un instante previo a la actualización.

Como se puede observar en las imágenes, la presencia de las sombras proyectadas por los objetos es evidente. Este efecto es debido a que los resultados mostrados pertenecen a una etapa previa a la eliminación de sombras y la actualización del modelo de fondo. Aunque el objetivo primordial del sistema es el aviso y localización de obstáculos, y no la precisión en la detección, se puede percibir como los obstáculos detectados aparecen segmentados del fondo con bastante exactitud.

### 6.1.1.2. Eliminación de sombras

La evaluación del grado de discriminación de sombras se trata de una prueba importante en el presente capítulo. Debe recordarse que dotar al sistema de una detección de objetos invariante a la iluminación ha supuesto una de las tareas más laboriosas. El conjunto de imágenes presentadas a continuación muestra cual ha sido el proceso realizado hasta llegar a una detección en la que las sombras quedan totalmente suprimidas. Por tanto, se representan en la figura 6.3 los pasos especificados en el capítulo 3.5.3.

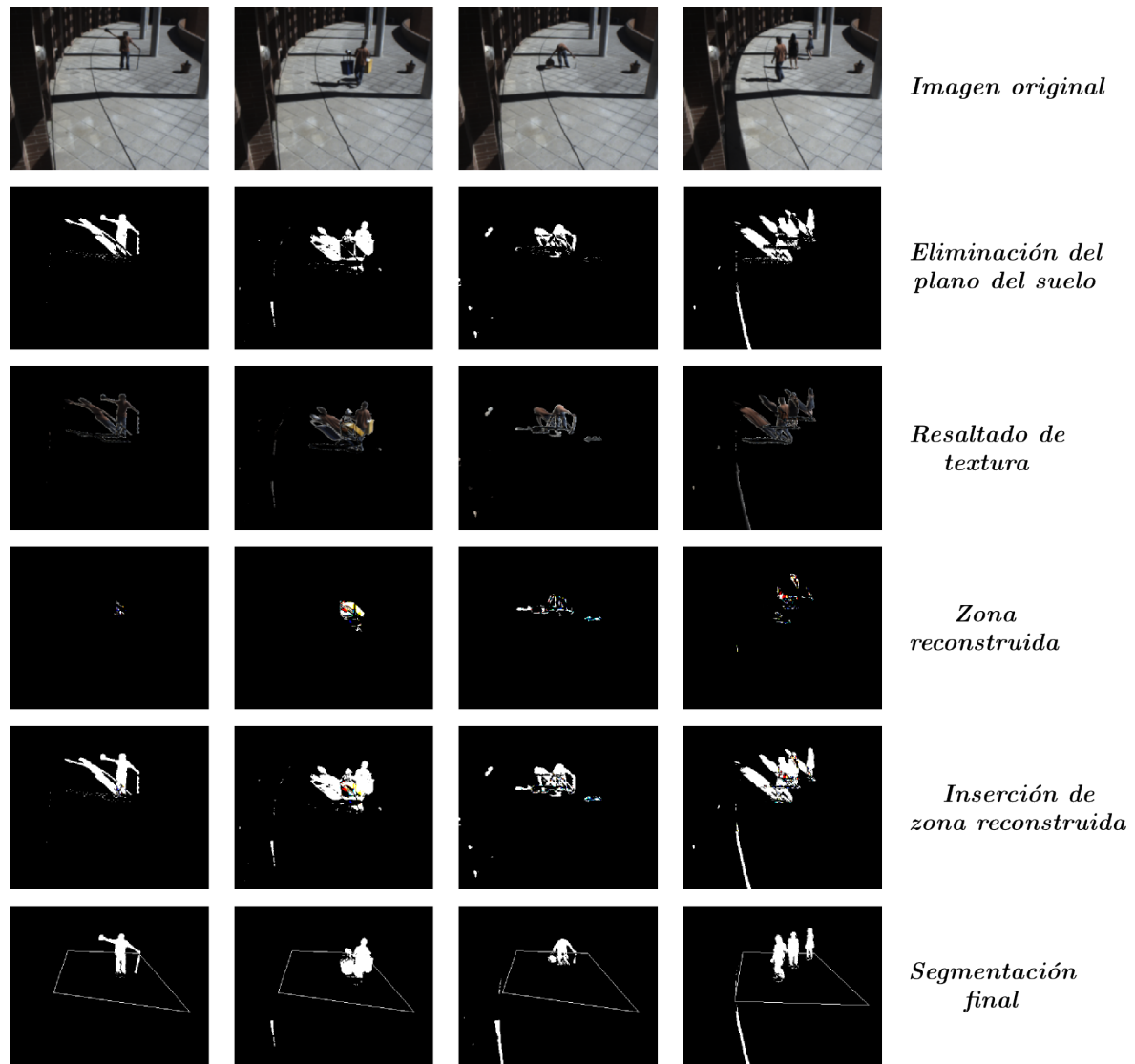


Figura 6.3: Separación del proceso de eliminación de sombras en cinco pasos

Se ha de asumir un pequeño margen de error en aquellas circunstancias en las que las imágenes observadas por ambas cámaras no presentan suficiente diferencia de “textura” como para poder reconstruir la zona de obstáculo eliminada con máxima exactitud. Este es el coste de haber implementado un mecanismo sencillo y rápido para recuperar las zonas de objeto indebidamente suprimidas.

### 6.1.1.3. Actualización del modelo de fondo. Mapa de disparidad

En este punto se comprueba rápidamente la consecuencia directa de la actualización de fondo. Se representan, en 4 columnas, los resultados obtenidos en un proceso típico de actualización del modelo estadístico de fondo; figura 6.4.

- En la primera columna se muestra una serie de imágenes capturadas por una de las cámaras.
- En la segunda columna se representa cual ha sido el resultado de la segmentación en una etapa previa a la eliminación de sombras y la actualización de fondo.
- La tercera columna muestra el mapa de actualización o disparidad obtenido. Un píxel en blanco implica su incorporación al modelo de fondo, mientras que un píxel en negro implica que ese punto no pertenece a lo que se supone como modelo de fondo.
- Por último, se indica en la última columna cual ha sido la segmentación obtenida tras la etapa de actualización y un recuadro blanco con la región de interés especificada.

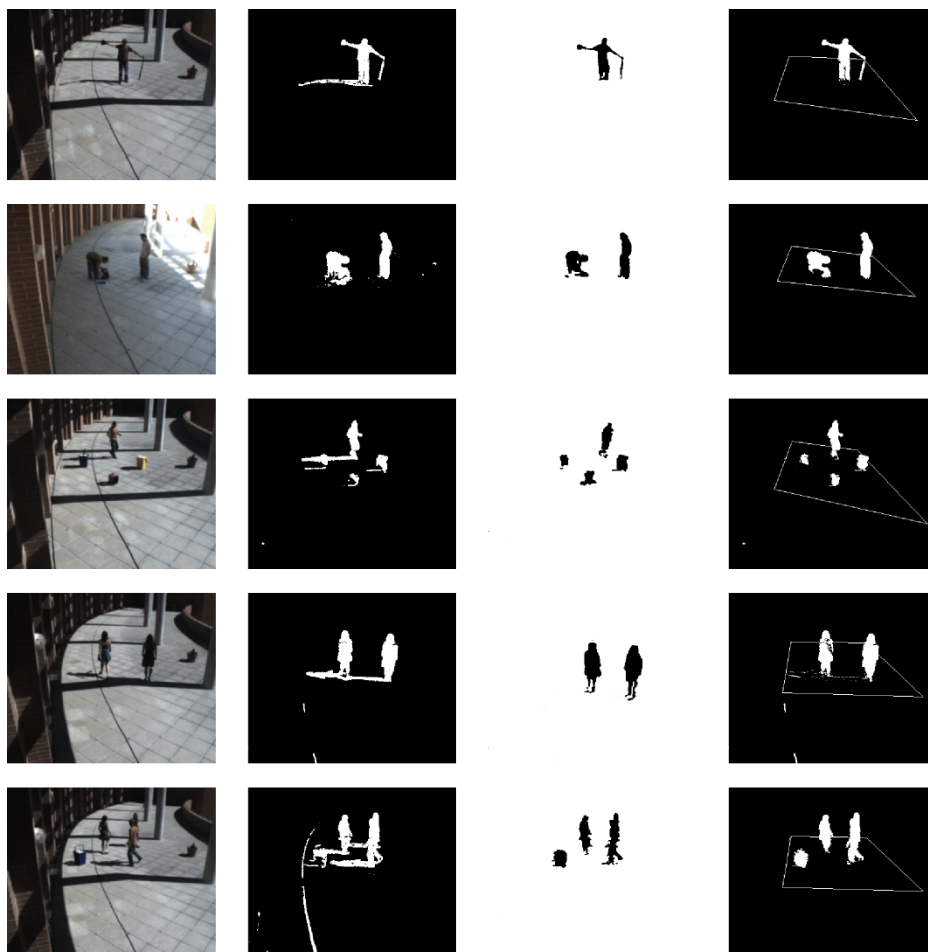


Figura 6.4: Ejemplos del proceso de actualización del modelo de fondo

Es importante señalar cómo las sombras y los cambios de iluminación que acontezcan serán incorporados al modelo de fondo. De esta manera, si otro objeto proyectase sombra en la misma zona, ésta ya no será segmentada como objeto en la etapa previa a la actualización. Este efecto se plasma en la figura 6.5, en la que se expone cómo, en un instante “ $t$ ” previo a la actualización, la sombra es segmentada, mientras que en el instante “ $t+1$ ”, posterior a la actualización, ésta ya es incorporada al fondo y no es segmentada.

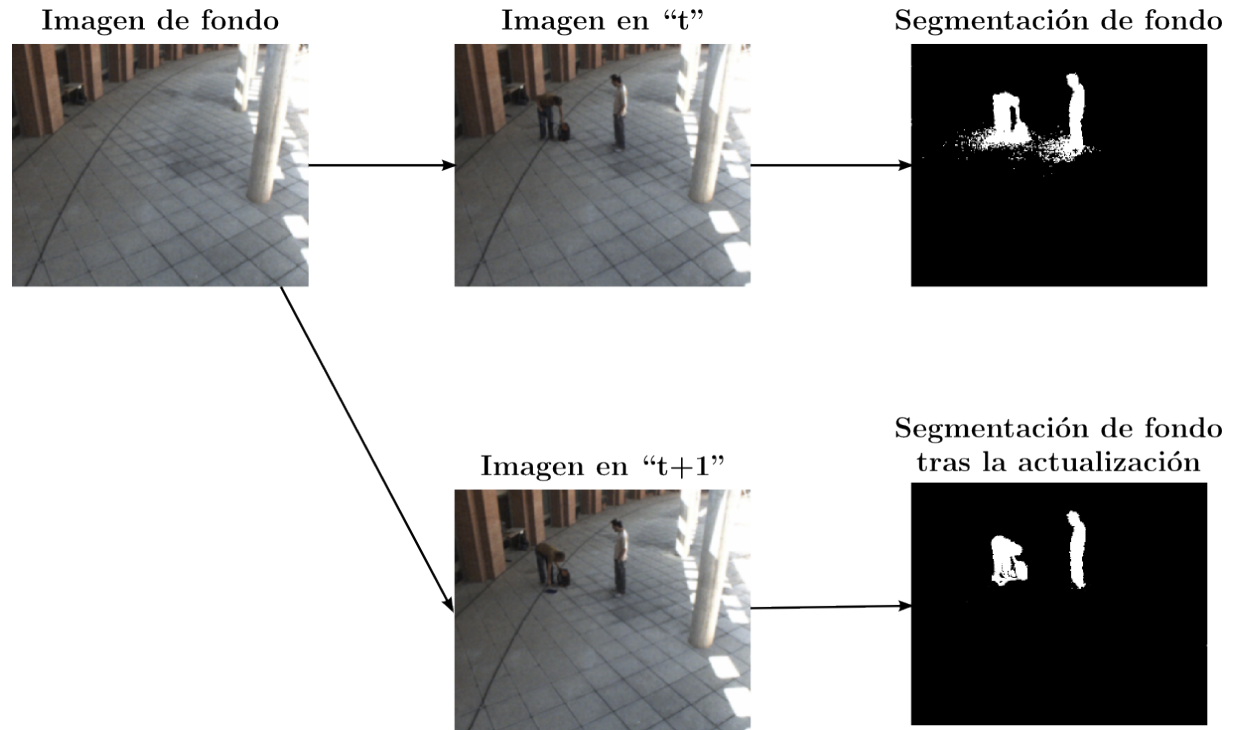


Figura 6.5: Resultado obtenido en un instante posterior a la actualización de fondo

#### 6.1.1.4. Segmentación mediante el algoritmo completo

Para finalizar con el apartado del algoritmo propuesto, se muestra el resultado obtenido para la segmentación de fondo final en distintos escenarios. En vista de los resultados de la figura 6.6, se entiende el proceso como exitoso.

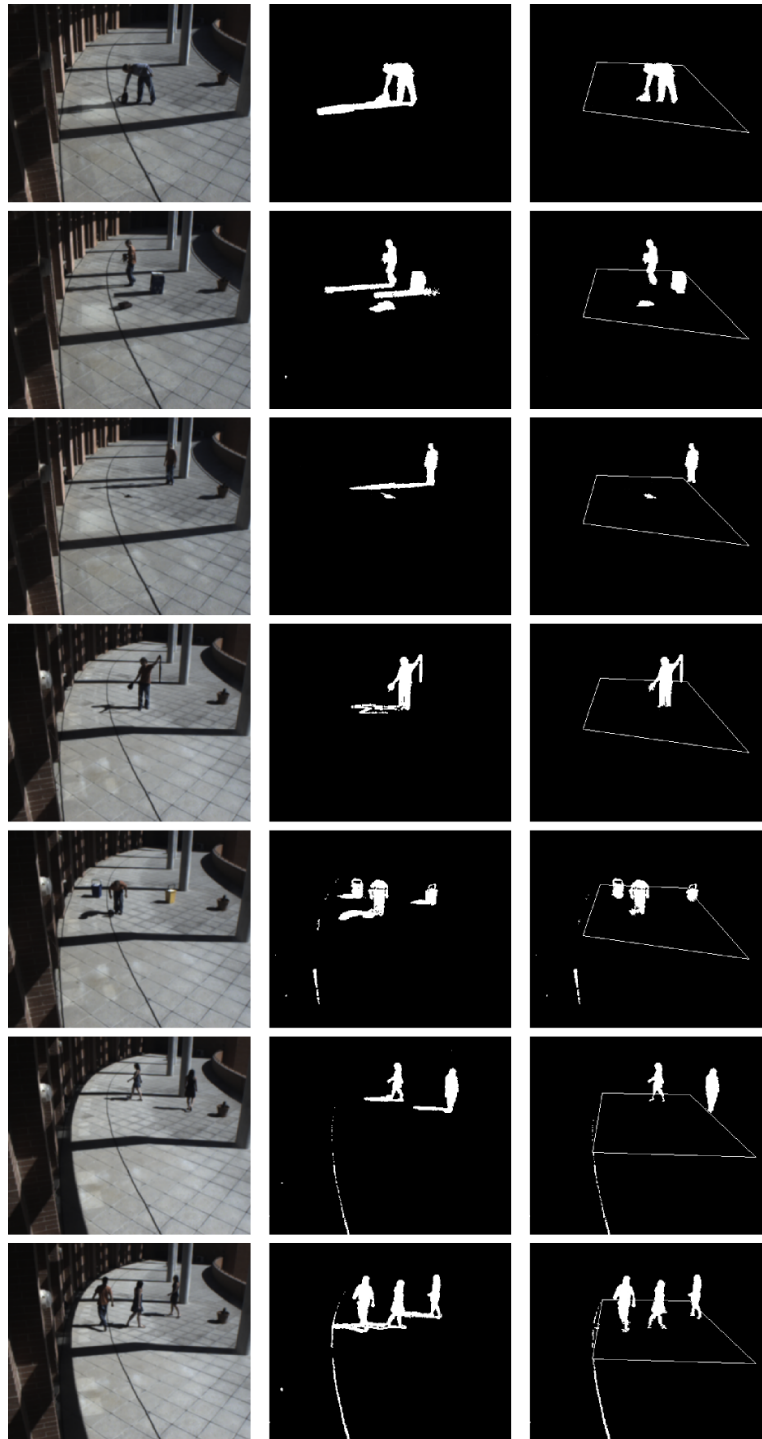


Figura 6.6: Ejemplos de las imágenes segmentadas tras la aplicación de una iteración del algoritmo completo



### 6.1.2. Segmentación en el espacio invariante a la iluminación

A continuación, se muestra en la figura 6.7 cómo, la segmentación en un espacio invariante a la iluminación, se trata de una alternativa válida para la detección de obstáculos en este tipo de entorno. Se ha de tener en consideración que, aunque en todas las imágenes aparecen sombras proyectadas, éstas no se identifican como obstáculos dentro de la segmentación.

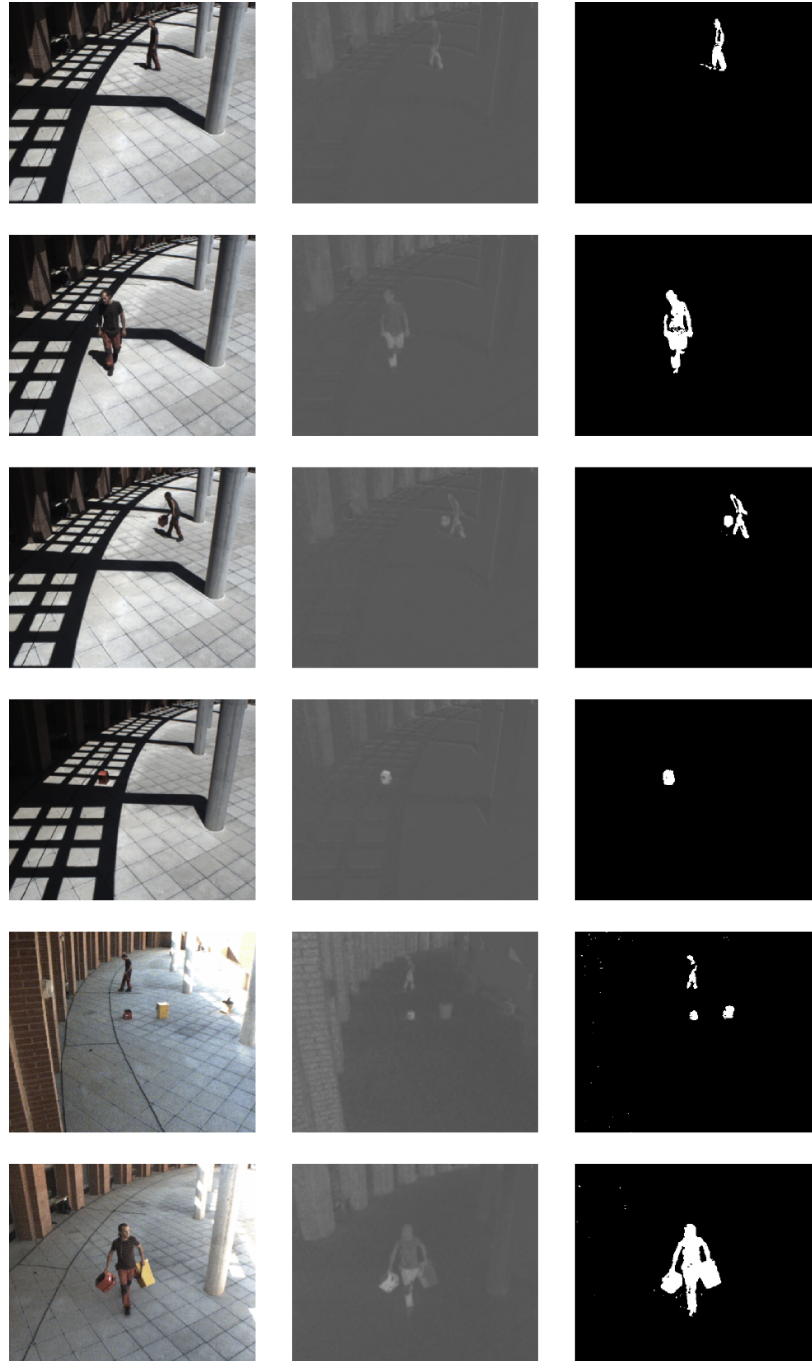


Figura 6.7: Ejemplos de las imágenes segmentadas tras la aplicación de una iteración del algoritmo completo en el espacio invariante a la iluminación

### 6.1.3. Absorción de la problemática en entornos exteriores

En esta sección se muestran los resultados obtenidos ante algunos de los casos más desfavorables presentes en entornos exteriores. Se pretende así reflejar el correcto funcionamiento del sistema ante este tipo de situaciones.

#### 6.1.3.1. Sombras proyectadas

Se presentan, en tres columnas dentro de la figura 6.8; la imagen capturada por una de las cámaras; la segmentación en una etapa previa a la eliminación de sombras y la segmentación obtenida tras una ejecución del algoritmo completo.

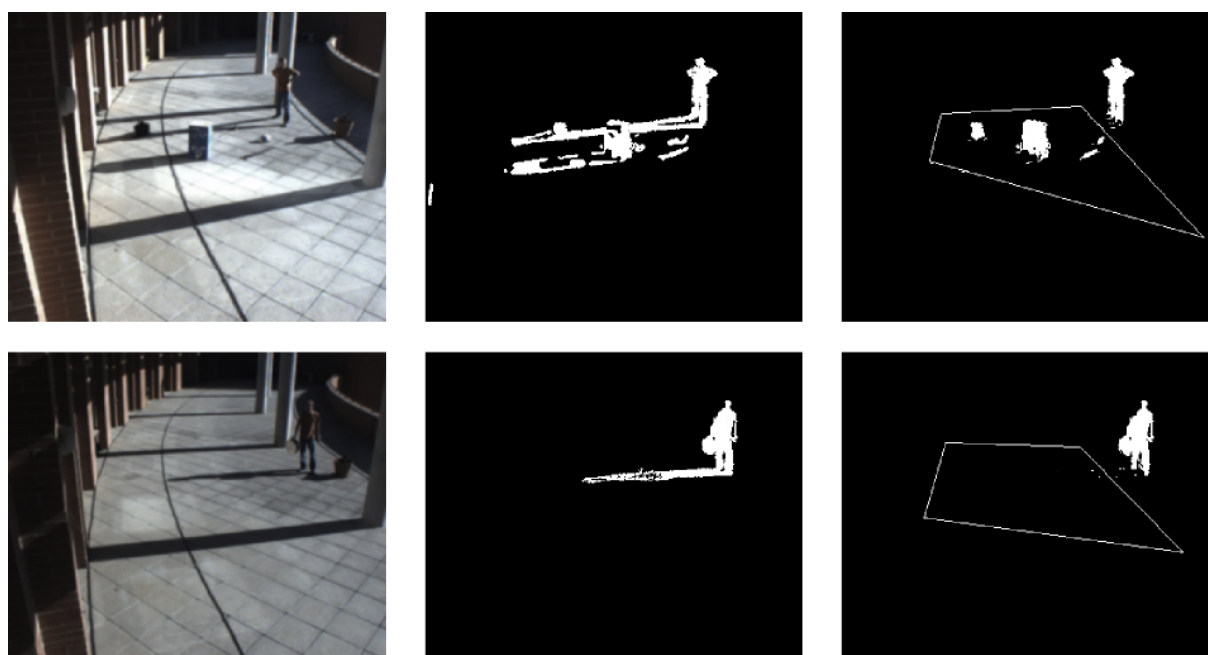


Figura 6.8: Ejemplos de eliminación de sombras proyectadas por objetos que no se encuentran presentes en el interior de la región de interés

De ahora en adelante y sirviendo de anticipo para la sección 6.2.1, se muestran las imágenes originales, compuestas para su envío al sistema receptor embarcado en el tren. En éstas, se representa; la región considerada como obstáculo, en un recuadro; la región de interés, en un cuadrilátero rojo; y un indicador circular, el cual si es de color rojo indica la existencia de obstáculo y en color verde la situación en la que no existe obstáculo dentro del área de interés monitorizada.



### 6.1.3.2. Cambio del modelo de fondo; evolución de sombras proyectadas

Un claro ejemplo del cambio del modelo de fondo se manifiesta en la evolución de la sombra proyectada por objetos fijos cercanos a la zona monitorizada. En las imágenes de la figura 6.9 se muestra la manera en la que dicho efecto es absorbido por el sistema propuesto

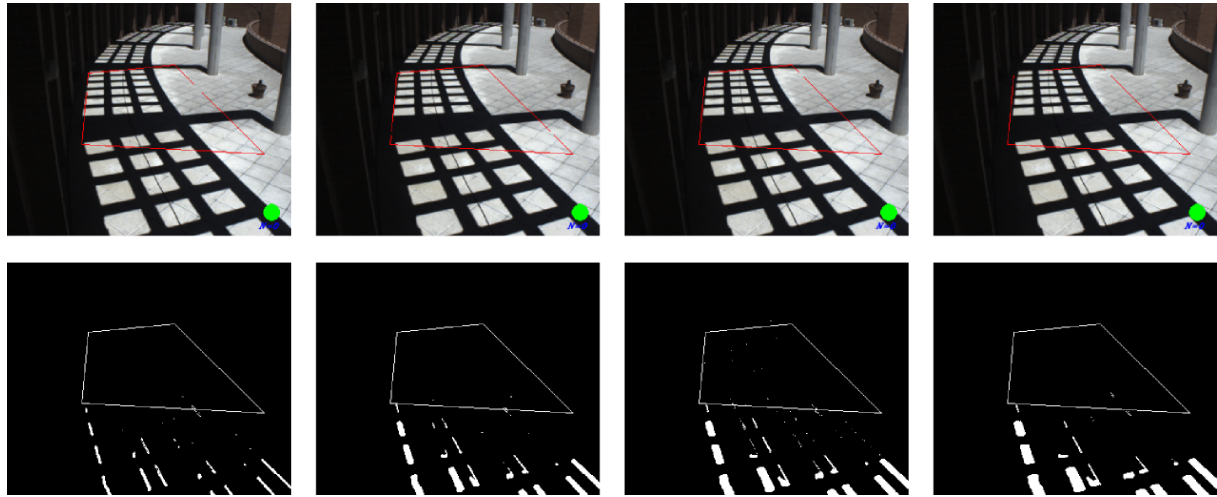


Figura 6.9: Resultado de la segmentación ante cambios en el modelo de fondo (evolución de sombras)

y la manera en la que la detección de obstáculos sigue funcionando, figura 6.10.

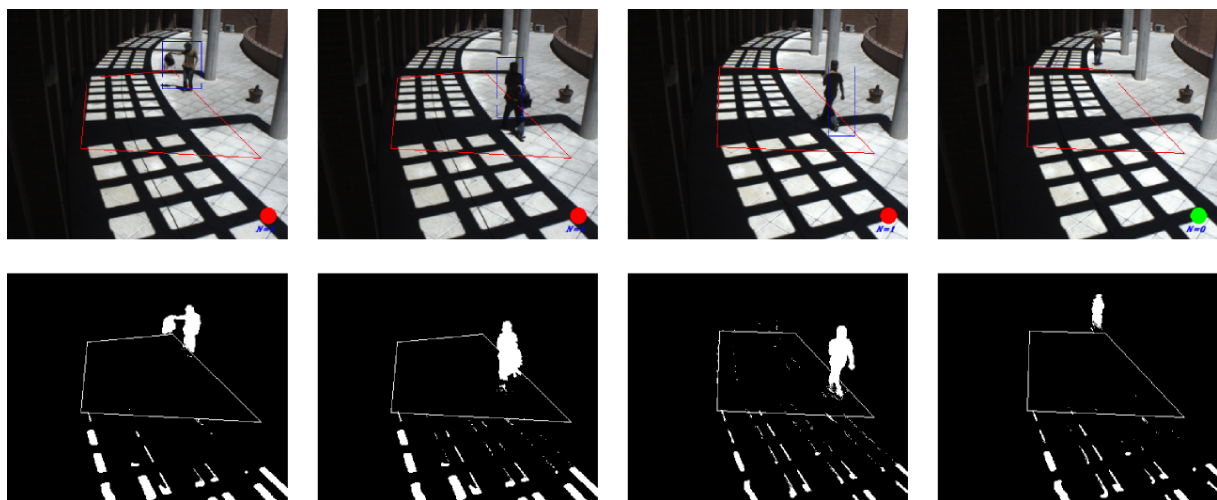


Figura 6.10: Resultado de la detección de obstáculos mientras el fondo varía

### 6.1.3.3. Cambios de iluminación

En la figura 6.11 se muestra la manera en la que la segmentación se mantiene funcionando correctamente en el área de interés, a pesar de la existencia de cambios bruscos de iluminación.

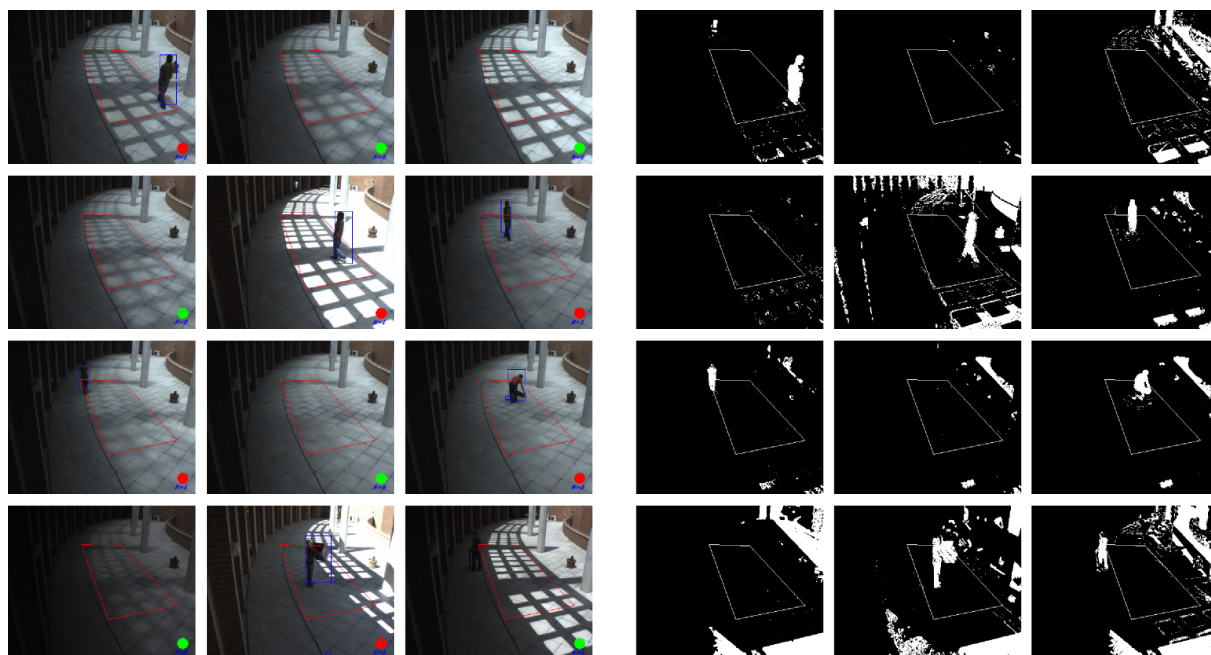


Figura 6.11: Ejemplos de detección ante cambios de iluminación

### 6.1.4. Fusión de alternativas de segmentación

Como se ha indicado, la fusión con diferentes alternativas de segmentación de fondo proporciona mayor fiabilidad en la detección, reduciendo, así, el número de falsos negativos; éstos, representan las situaciones más desfavorables y son objeto de minimización.

En las imágenes presentadas se emplean 3 alternativas de segmentación, cuyos resultados de detección se enmarcan en rectángulos de distinto color:

- En azul, el algoritmo “KAM” propuesto.
- En verde, el algoritmo “KAM” en el espacio invariante a la iluminación.
- En naranja, un tercer algoritmo de segmentación de movimiento basado en la segmentación por zonas.

A continuación, se representan algunas de las posibles situaciones en el resultado de la fusión de alternativas:

1. Los tres algoritmos indican la existencia de los 2 obstáculos, figura 6.12.

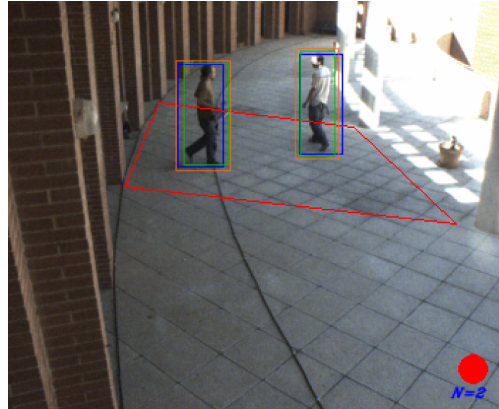


Figura 6.12: Los tres obstáculos son detectados por las distintas alternativas de segmentación

2. Excepto en el caso de la papelerita amarilla, en la que dos alternativas indican su existencia, todos los algoritmos detectan correctamente la presencia de objetos dentro de la región de interés, figura 6.13.

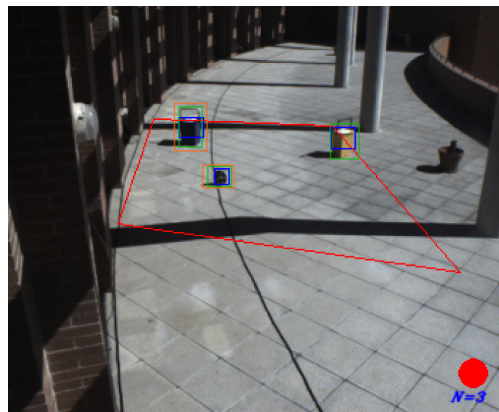


Figura 6.13: La papelerita amarilla es detectado por dos de las alternativas. Se mantiene la detección de los 3 obstáculos

3. Aunque la papelerita azul solo es detectada por uno de los algoritmos, se informará al tren de la existencia de los cuatro obstáculos, figura 6.14.

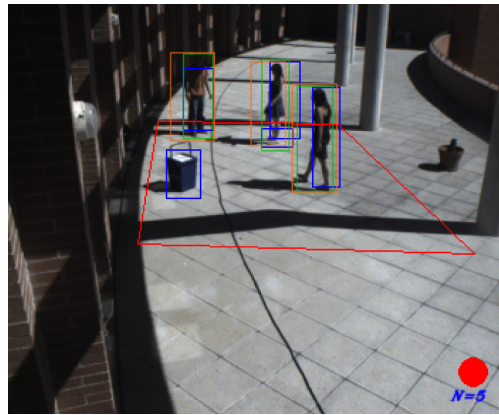


Figura 6.14: La papelerera azul solo es detectada por uno de los algoritmos. Se mantiene la identificación de los 4 obstáculos

### 6.1.5. Conmutación día $\Leftrightarrow$ noche

Por último, se muestra, en la figura 6.15, el comportamiento del sistema en el momento de la conmutación a modo nocturno bajo tres situaciones diferentes; con y sin obstáculos dentro de la región de interés. En la última fila se muestra la imagen preparada para su envío al sistema receptor.

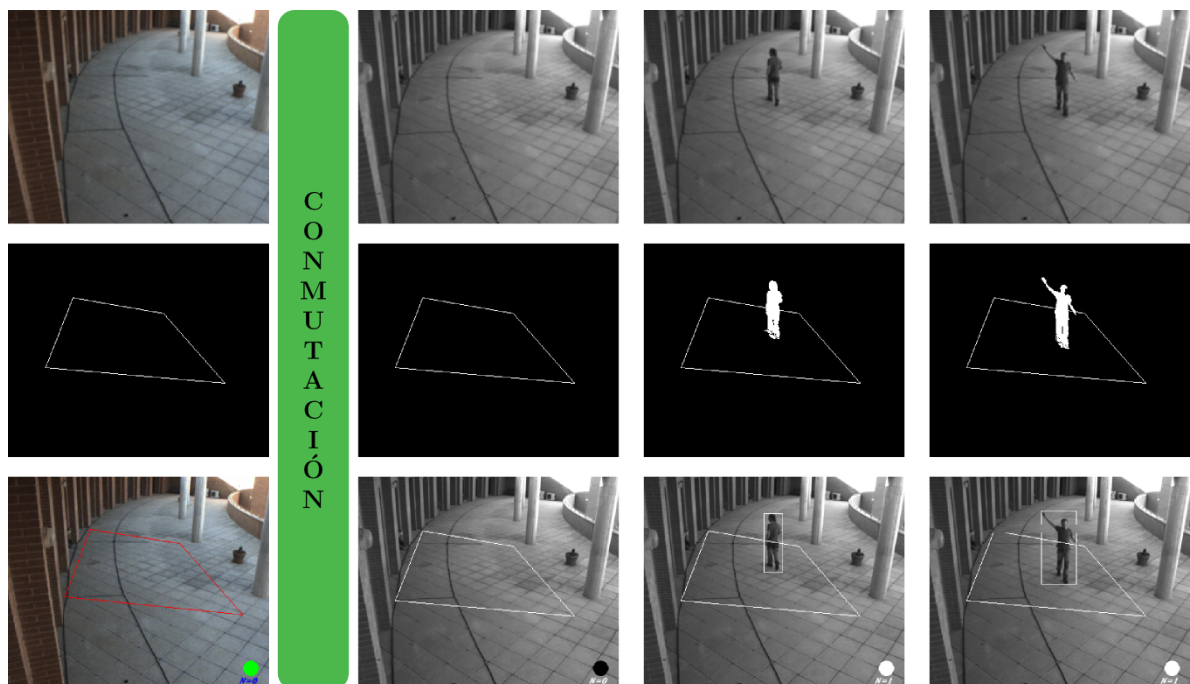


Figura 6.15: Resultados obtenidos tras la conmutación del sistema a modo nocturno

### 6.1.6. Tiempos de cómputo

Las tablas 6.1 y 6.2 muestran un resumen de los tiempos de cómputo más representativos de la implementación software del sistema de detección de obstáculos; tanto para el modo de funcionamiento diurno como para el nocturno, respectivamente.

■ **MODO DIURNO (imágenes a color):**

| PROCESO                                                | TIEMPO DE CÁLCULO(ms) |
|--------------------------------------------------------|-----------------------|
| Inicialización del modelo de fondo mediante K-medias   | 587                   |
| Segmentación mediante distancia de Mahalanobis         | 76.8                  |
| Eliminación de sombras                                 | 22.64                 |
| Obtención del mapa de disparidad para la actualización | 14.53                 |
| Actualización del modelo de fondo                      | 480                   |
| <b>Total en una iteración</b>                          | <b>99.4</b>           |

Tabla 6.1: Tiempos de conmutación más relevantes en la implementación del sistema de detección de obstáculos en ejecución diurna

■ **MODO NOCTURNO (imágenes monocromáticas):**

| PROCESO                                                | TIEMPO DE CÁLCULO(ms) |
|--------------------------------------------------------|-----------------------|
| Inicialización del modelo de fondo mediante K-medias   | 314                   |
| Segmentación mediante distancia de Mahalanobis         | 25.15                 |
| Eliminación de sombras                                 | 10.06                 |
| Obtención del mapa de disparidad para la actualización | 11.36                 |
| Actualización del modelo de fondo                      | 286.24                |
| <b>Total en una iteración</b>                          | <b>35.21</b>          |

Tabla 6.2: Tiempos de conmutación más relevantes en la implementación del sistema de detección de obstáculos en ejecución nocturna



## 6.2. Servidor de imágenes y mensajes de control

En esta sección se muestran los resultados más relevantes que reflejan el funcionamiento del sistema como servidor de imágenes y mensajes de control.

### 6.2.1. Composición de la imagen transmitida y envío de mensajes de control

Se representa, en la figura 6.16, ejemplos de un conjunto de imágenes enviadas al módulo receptor, así como el mensaje de control correspondiente a la última imagen (tabla 6.3), cuyo formato quedó definido en la tabla 4.3.

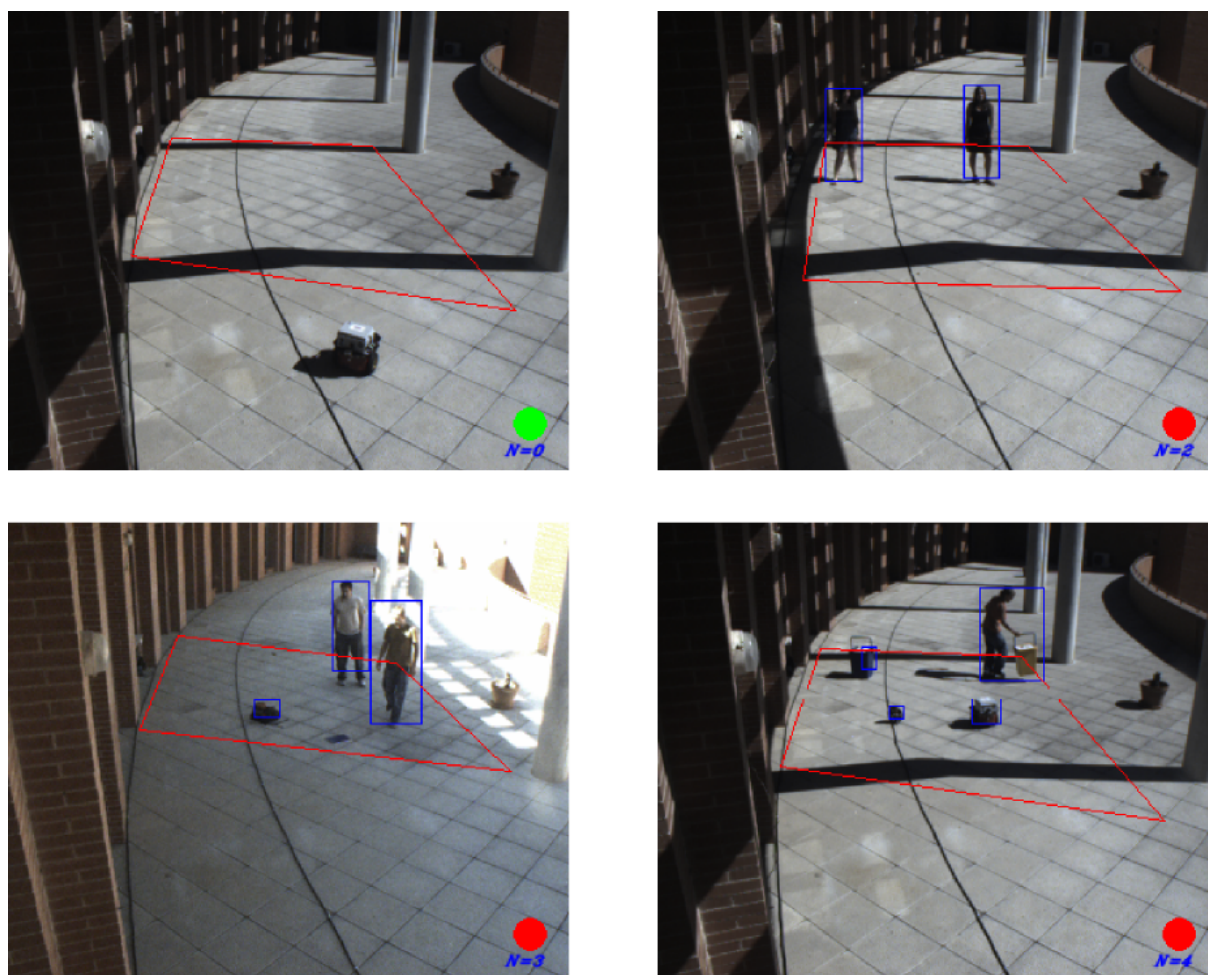


Figura 6.16: Imágenes compuestas para su envío al módulo receptor ubicado en el tren.  
Ejemplos para  $N=0, 2, 3$  y 4 obstáculos detectados

|              |              |            |               |            |  |
|--------------|--------------|------------|---------------|------------|--|
| <status>     | <ok>         |            |               |            |  |
| <version>    | VERSIÓN 1.0  | </version> |               |            |  |
| <crossing>   |              |            |               |            |  |
|              | <id>         | I          | </id>         |            |  |
|              | <protection> | ON         | </protection> |            |  |
| </crossing>  |              |            |               |            |  |
| <cameras>    |              |            |               |            |  |
|              | <camera>     |            |               |            |  |
|              |              | <id>       | I             | </id>      |  |
|              |              | <state>    | ON            | </state>   |  |
|              |              | <port>     | 10001         | </port>    |  |
|              | </camera>    |            |               |            |  |
| </cameras>   |              |            |               |            |  |
| <intrusion>  |              |            |               |            |  |
|              |              | <state>    | ON            | </state>   |  |
|              |              | <level>    | NORMAL        | </level>   |  |
| </intrusion> |              |            |               |            |  |
| <datetime>   |              |            |               |            |  |
|              | <date>       |            |               |            |  |
|              |              | <year>     | 2009          | </year>    |  |
|              |              | <month>    | 6             | </month>   |  |
|              |              | <day>      | 25            | </day>     |  |
|              | </date>      |            |               |            |  |
|              | <time>       |            |               |            |  |
|              |              | <hour>     | 18            | </hour>    |  |
|              |              | <minute>   | 24            | </minute>  |  |
|              |              | <second>   | 50            | </second>  |  |
|              |              | <milisec>  | 46            | </milisec> |  |
|              | </time>      |            |               |            |  |
| </datetime>  |              |            |               |            |  |
| </ok>        |              |            |               |            |  |

Tabla 6.3: Ejemplo de mensaje de control enviado al módulo del tren para la última imagen de la figura 6.16

### 6.2.2. Reproducción del video-streaming

Para poder verificar el correcto funcionamiento de la transmisión del video-streaming al módulo receptor se ha empleado el reproductor de vídeo VLC, bajo licencia GPL. La figura 6.17 muestra una serie de capturas de la recepción del video-streaming en un PC remoto conectado a la red.

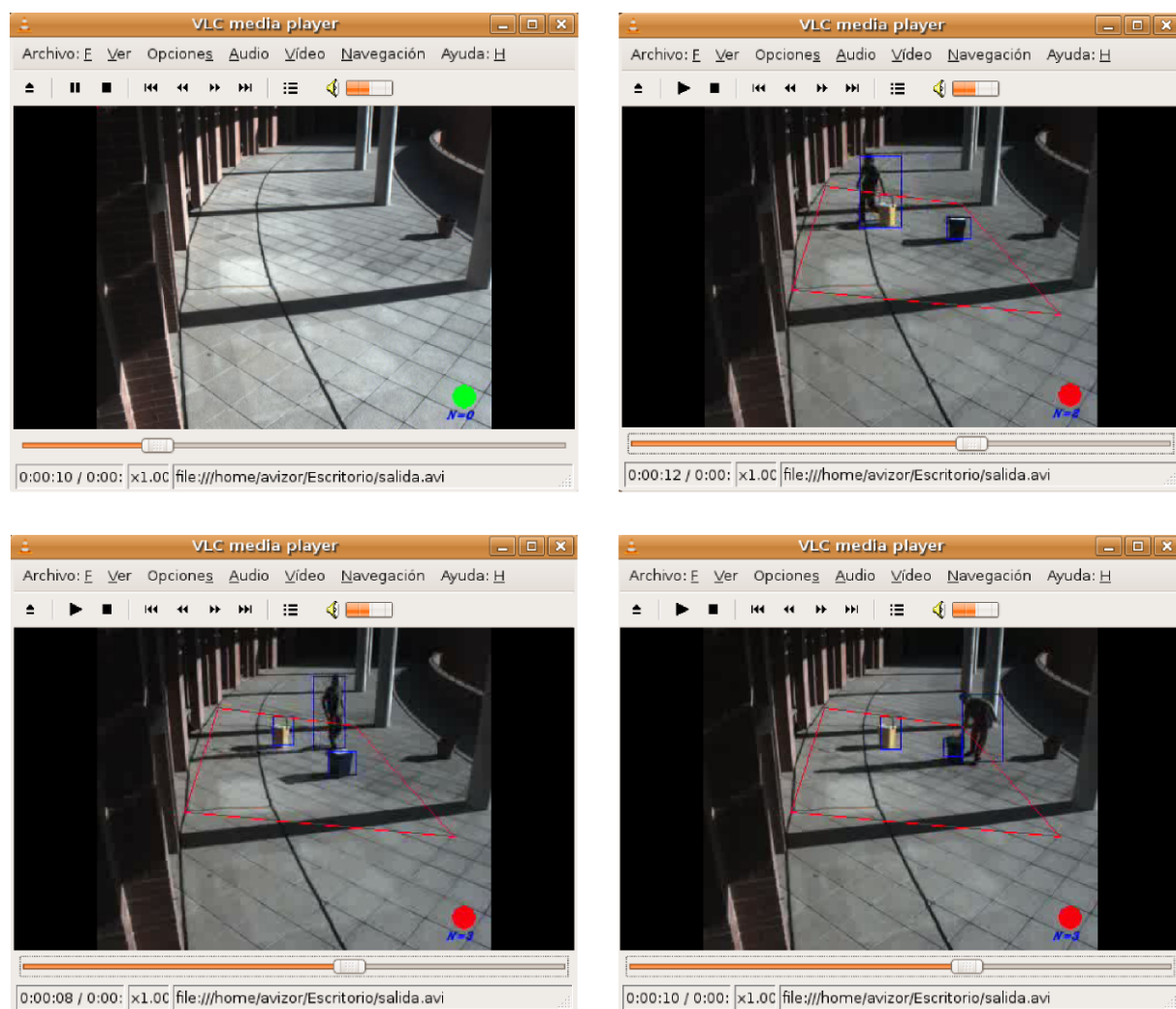


Figura 6.17: Capturas realizadas en la verificación del correcto funcionamiento de la transmisión de vídeo empleando el reproductor VLC



### 6.2.3. Análisis del ancho de banda transmitido

En la tabla 6.4 se muestran las pruebas realizadas del ancho de banda transmitido como una relación entre la tasa de bits configurada en el servidor de imágenes y la recibida en el sistema receptor.

| TASA DE BITS DE SALIDA (kbps) | TASA DE BITS DE ENTRADA (kbps) |
|-------------------------------|--------------------------------|
| 40                            | 39.628                         |
| 80                            | 78.95                          |
| 150                           | 148.74                         |
| 300                           | 279.15                         |
| 450                           | 437.2                          |
| 600                           | 576.53                         |

Tabla 6.4: Análisis del ancho de banda

### 6.2.4. Tiempos de cómputo

La tabla 6.5 muestra un resumen de los tiempos de cómputo más representativos de la implementación software del sistema servidor de imágenes y mensajes de control.

| PROCESO                                           | TIEMPO DE CÁMPUTO(ms) |
|---------------------------------------------------|-----------------------|
| Composición de la imagen para su transmisión      | 4.19                  |
| Codificación MPEG-4 de la imagen                  | 1.54                  |
| <b>Total</b>                                      | <b>5.73</b>           |
| Composición del mensaje de control en formato XML | 1.03                  |
| Envío del mensaje de control                      | 2.8                   |
| <b>Total</b>                                      | <b>3.83</b>           |

Tabla 6.5: Tiempos de cómputo más relevantes en la implementación del servidor de imágenes y mensajes de control



# Capítulo 7

## Conclusiones

Una vez concluido el desarrollo del presente proyecto final de carrera, se consideran cumplidos los objetivos propuestos en su inicio. Por tanto, se proporciona un sistema de visión artificial capaz de monitorizar zonas de riesgo potencial para la circulación ferroviaria. Este sistema se considera apto para la detección de obstáculos en entornos exteriores y el envío al tren de la información relevante acerca del estado del escenario en cuestión.

Se desprenden varias conclusiones del estudio y ejecución del sistema propuesto. Éstas se exponen a continuación:

- En primer lugar, se considera de especial relevancia el reto planteado a la hora de diseñar un sistema capaz de detectar obstáculos en entornos exteriores. Este tipo de escenarios presentan una casuística tan amplia que resulta imposible predecir las diferentes situaciones a las que se tendrá que hacer frente el sistema de detección. En las pruebas realizadas para comprobar la viabilidad de los algoritmos, se ha podido verificar cómo, en las condiciones más adversas, se revelaban constantemente carencias en los primeros diseños. De esta manera, se destaca la evolución sufrida por el sistema planteado en un inicio, cuyo objetivo ha sido reflejar el mayor número de casos que se puedan presentar en este tipo de entornos.
- Como propuesta inicial, se planteó un sistema de segmentación basado en el modelado de fondo haciendo uso del conocido algoritmo k-medias, al cual se le realizaron una serie de pequeñas modificaciones. En dicha propuesta, el modelo de fondo se forma en una etapa inicial de entrenamiento y permanece estático durante el resto de ejecución del sistema. Como se ha explicado en capítulos anteriores, esta estrategia se presentó como válida en aquellos entornos donde las condiciones del fondo estuvieran controladas y acotadas, considerándose una buena alternativa de segmentación para “Espacios Inteligentes”; lugar donde se realizaron las primeras pruebas. Su ejecución en entornos exteriores demostró, rápidamente, la necesidad de un sistema que se encargase de la actualización “on-line” del modelo de fondo.
- Los efectos provocados por la iluminación se especifican como los principales factores limitantes para el correcto funcionamiento de la segmentación de fondo en entornos exteriores. Gracias a las pruebas realizadas, se observa que es en la proyección de sombras

y los cambios bruscos de iluminación donde residen las mayores trabas para una correcta detección. El proceso que trata de absorber esta problemática, parte de las premisas de disponer dos cámaras, cuya configuración permita ver la misma escena desde ángulos diferentes, y de capturar imágenes cuyo valor, para el mismo píxel que referencia el mismo punto del espacio 3D visto por ambas cámaras, sea el mismo. Para conseguirlo se prepara una rutina que modifica dos de los parámetros internos de las cámaras (ganancia y tiempo de exposición) de manera que el valor medio de ambas imágenes se sitúe en torno a un umbral (con una cierta tolerancia). Debido a las desigualdades presentes entre la intensidad de los píxeles, se producen pequeños errores en la reconstrucción de las imágenes segmentadas al haber eliminado el plano de suelo. A pesar de ello, si se ajustan los umbrales empleados, teniendo en consideración este hecho, los resultados obtenidos se acercan, en gran medida, a los deseados, minimizándose dichos errores.

- En una etapa previa a la actualización de fondo y la eliminación de sombras, gracias al análisis píxel a píxel realizado en el algoritmo de segmentación propuesto se obtienen resultados que muestran gran precisión en la detección. Los obstáculos son detectados en función de su diferencia al modelo de fondo obtenido; por tanto, como el sistema solo recoge información luminosa del entorno, existe la posibilidad de no detectar objetos constituidos del mismo material o intensidad de color que el fondo. De esta manera, si el fondo es de cemento, puede que un bloque constituido de ese mismo material no sea detectado por el sistema. Este error es asumido ya que los sensores empleados no proporcionan más información que la luminosa.
- El proceso de estimación de la matriz de homografía es una etapa crucial para el correcto funcionamiento del sistema de detección. La fuerte dependencia con dicha matriz se hace patente en las rutinas de reconstrucción de la imagen segmentada y en la actualización del modelo de fondo, más concretamente en la creación del mapa de disparidad. Errores en el cálculo de la matriz provoca la incorrecta superposición de las imágenes y la acumulación de errores en la sucesiva ejecución de las rutinas anteriormente citadas.  
Se ha tener en cuenta que, dado que la homografía se optimiza para la región de interés especificada en el proceso de inicialización asistido, fuera de dicha zona la matriz de homografía tiende a perder precisión. Este efecto repercute en el sistema de eliminación de sombras y de actualización del modelo de fondo, de manera que, con el paso del tiempo y la evolución del modelo de fondo (p.ej. con el cambio de iluminación a lo largo de un día), se tiende a un modelo en el que la superficie exterior a la región de interés se interpreta como parte detectada. No obstante, este resultado no afecta a la correcta detección de obstáculos, ya que el funcionamiento del sistema se ha de centrar en la zona monitorizada.
- De igual manera que en el punto anterior, es importante darse cuenta de la necesidad de que las cámaras, en su colocación, dispongan de un punto de equilibrio estable, ya que el desplazamiento de alguna de ellas, una vez se haya obtenido la matriz de homografía, induciría, del mismo modo, errores en el solapamiento entre imágenes.
- Para dotar de procesamiento en tiempo real al sistema, se plantea un esquema concurrente en el que varios hilos, encargados de funcionalidades diferentes, se ejecutan de manera paralela. Para ello fue necesario adquirir un PC con las características más altas del mercado, dando una idea de lo difícil que resulta realizar tareas que para las personas no

suponen esfuerzo alguno. El PC adquirido dispone de 4 núcleos, necesarios para poder separar los procesos y asignarlos a procesadores diferentes; distribuyendo así el consumo de CPU y minimizando el tiempo de cómputo.

- La inclusión del sistema de envío de video-streaming y el de envío de mensajes de control XML, hace que se pueda englobar la solución final como un módulo servidor que proporciona imágenes y mensajes a los sistemas embarcados en los trenes. Estos se conectan, como clientes, al módulo servidor, momento a partir del cual se comienza el envío de información.
- Dado que el sistema debe estar en funcionamiento durante largos periodos de tiempo, debe estar preparado para afrontar la mayor cantidad de eventualidades posibles. Así, se ha diseñado un proceso encargado de supervisar el estado de los diferentes hilos y semáforos y realizar las medidas correctoras oportunas que mantengan el sistema activo.
- Con el mismo objetivo que el punto anterior, se implementa la conmutación día↔noche. El paso al modo nocturno implica el cambio de la captura de las cámaras a color a las cámaras monocromáticas, por presentar mayor sensibilidad al infrarrojo, y el encendido de los focos de iluminación infrarroja. El criterio para decidir si se está bajo iluminación diurna o nocturna pasa por la observación de las cámaras a color. De esta manera, se decide que la conmutación a modo nocturno venga dada por la incapacidad de las cámaras a color de capturar las imágenes con una intensidad luminosa que se sitúe alrededor de un umbral. De manera análoga, cuando las cámaras a color vuelvan a poder capturar imágenes en torno a dicho umbral, se volverá al modo diurno. Por tanto, en días nublados, en los que la iluminación es escasa, el sistema conmutará a modo nocturno mucho más rápido que en situaciones normales.
- Para poder realizar pruebas, que se asemejen a los escenarios donde finalmente se ubicará el sistema, se ha instalado el esquema hardware propuesto, véase Manual de Usuario 8, en un entorno exterior ubicado en la terraza de la última planta del Departamento de Electrónica de la Universidad de Alcalá. Por su ubicación bajo una malla de hormigón y rodeado de columnas y paredes, este escenario se encuentra sometido constantemente a circunstancias de las consideradas problemáticas: evolución del modelo de fondo, proyección de sombras de objetos no presentes en la escena, cambios bruscos de iluminación, etc.. Por tanto, el hecho de haber pasado con éxito las pruebas en este tipo de escenarios, es indicativo del correcto funcionamiento del sistema según las exigencias presentadas en el proyecto AVIZOR.



# **Parte III**

## **Manual de Usuario**





# Capítulo 8

## Manual de Usuario

En esta sección se resumen los pasos necesarios para la puesta en marcha del sistema. Se indicarán los aspectos necesarios relativos a la compilación e inicialización de la solución software, así como a la preparación y colocación de los elementos hardware.

### 8.1. Compilación de la solución software

Para la compilación del sistema se requiere tener instalados una serie de paquetes que, en función de la distribución linux empleada, estarán ya instalados:

*“libpthread-dev, ffmpeg, libavcodec-dev, libavutils-dev, libavformat-dev, libk8055.0.2, Pylon 2.0.3”*

además de tener debidamente instaladas las librerías de visión *“OpenCV”*.

Se proporciona un archivo *“Makefile”* encargado de compilar y linkar, de manera automática, todos los archivos fuente y librerías necesarias. Por tanto, la intervención externa, a la hora de compilar la solución y generar el ejecutable, reside en la instalación correcta de los paquetes y librerías anteriormente citados.

Tecleando la instrucción *“make”* en la línea de comandos, dentro del directorio donde se encuentran todos los archivos fuente y cabecera, se generará el archivo ejecutable; el cual recibe el nombre de *“avizor”*.

### 8.2. Inicialización de la solución software

Una vez que se ha realizado correctamente la compilación, y generado el ejecutable. Su puesta en funcionamiento se describe en los siguientes pasos:

1. Bajo el sistema operativo linux, se lanza un terminal; ubicándose en el directorio donde se encuentren todos los archivos fuente y cabecera.

2. Se procede a la puesta en marcha del archivo ejecutable escribiendo la sentencia “./avizor”
3. Comienza el proceso de inicialización asistido 3.5.1, en el que es necesario la intervención por parte de un operario. Se muestran las imágenes capturadas por cada una de las 4 cámaras.
4. El proceso se ejecuta dos a dos; primero con las imágenes provenientes de las cámaras de color y, después, con las monocromáticas. En cada una de las imágenes, se deben indicar 4 puntos, de manera que formen un cuadrilátero, cuyo área indica la región de interés monitorizada. Las 4 marcas deben referenciar los mismos puntos, teniendo especial atención a su pertenencia al plano del suelo. Este paso se ilustra en la figura 3.12 del capítulo 3 de la memoria.
5. Una vez marcados los puntos, se ejecuta, de manera automática, el algoritmo de registro de imágenes para la estimación de la matriz de homografía. Se debe observar cómo se solapan las imágenes de ambas cámaras en la zona perteneciente al plano del suelo.

A partir de este momento, y tras la configuración automática de los parámetros que equilibran las imágenes observadas por cada par de cámaras, comienza la ejecución del sistema implementado.

Para comprobar que el funcionamiento es el correcto, se muestran 2 ventanas con información relativa de:

1. Las sucesivas imágenes binarias, resultado del proceso de segmentación, para ambas cámaras.
2. El flujo de vídeo que se transmite al modulo receptor ubicado en la cabina del tren. Las imágenes enviadas, según los requisitos del proyecto AVIZOR, son directamente las capturadas por una de las cámaras, pero compuestas de manera que se pueda observar la información relevante para el conductor del tren. En ellas se puede observar:
  - La región de interés dentro de la imagen.
  - Número de obstáculos identificados.
  - En caso de detectar algún obstáculo, se presenta en la imagen:
    - Un indicador visual de obstáculo detectado.
    - El área de ocupación de los obstáculos detectados en forma rectangular.

En la figura 8.1 se muestra un ejemplo de la imagen compuesta.

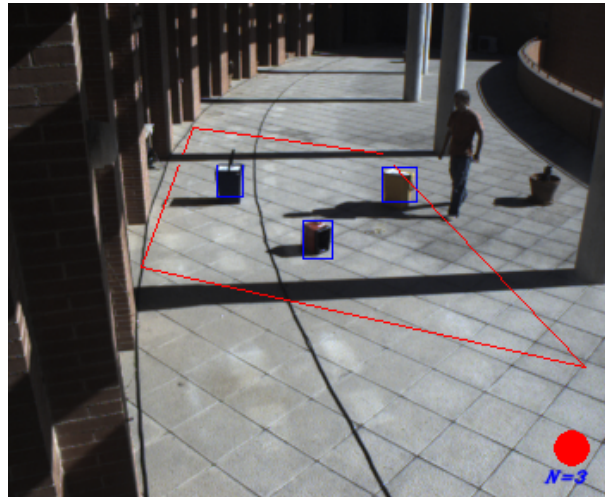


Figura 8.1: Ejemplo de imagen transmitida al tren.

Uno de los requisitos especificados en el pliego de condiciones 9.2 consiste en la asignación de una dirección IP estática de dominio público. Para comprobar que la transmisión de vídeo es correcta se lanza el reproductor de vídeo “VLC”, observando la correcta visualización de la imagen compuesta para su envío al tren, 8.2.



Figura 8.2: Ejemplo de la correcta visualización del vídeo transmitido.

Se proporciona un archivo de configuración, definido como “*sdp.file*”, necesario para la correcta recepción del video-streaming por VLC. Para la conexión con el servidor de vídeo implementado en la solución software, se debe ejecutar el reproductor VLC pasando como parámetro el archivo de configuración proporcionado, escribiendo en un terminal la sentencia “*vlc /ubicación/sdp.file*”.

### 8.3. Preparación y colocación de los elementos hardware

El correcto funcionamiento del sistema desarrollado, pasa por la preparación y colocación de los elementos hardware especificados en el pliego de condiciones 9.1. La disposición de dichos elementos responde al diagrama 8.3.

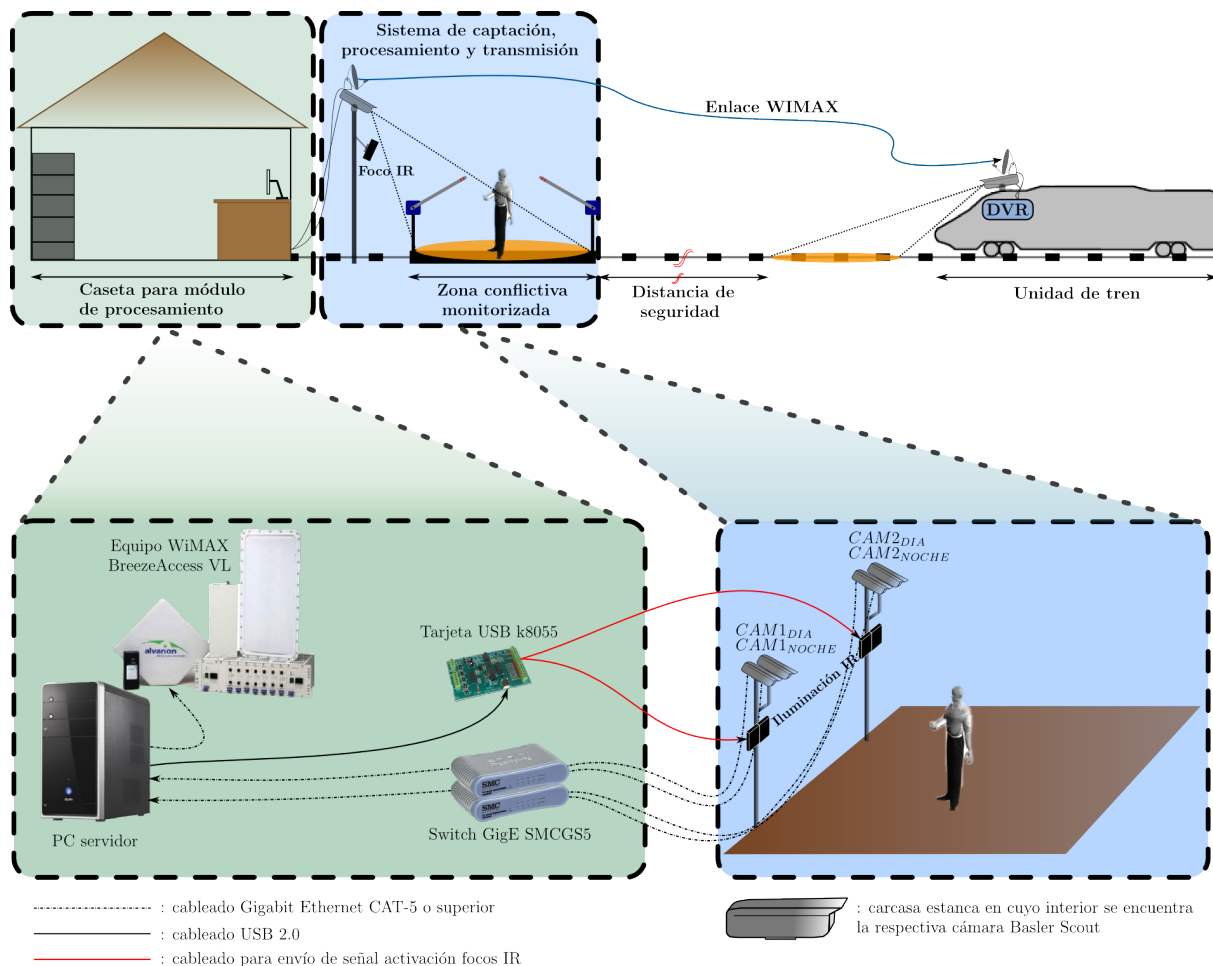


Figura 8.3: Esquema para la colocación y preparación de los elementos hardware.

## **Parte IV**

### **Pliego de condiciones**



# Capítulo 9

## Pliego de condiciones

Para el correcto desarrollo del sistema de detección propuesto se han visto implicados tanto recursos hardware como software. En el presente capítulo, se muestra cuales son los requisitos necesarios para la realización del sistema en completo en su versión final.

### 9.1. Hardware

- Ordenador de sobremesa con las siguientes características mínimas:
  - Procesador  $\Rightarrow$  Intel(R) Core(TM)2 Quad CPU Q9550 @ 2.83GHz
  - Número de núcleos  $\Rightarrow$  4
  - Memoria RAM  $\Rightarrow$  4 GB
  - Caché  $\Rightarrow$  6144 KB
  - Disco duro  $\Rightarrow$  465 GiB (500 GB)
  - Cuatro interfaces USB
  - Dos interfaces de red Gigabit Ethernet
- Dos switch Gigabit Ethernet SMCGS5 10/100/1000
- Cableado CAT-5 o superior
- Cuatro estructuras de soporte para cámaras
- Cuatro carcasas estancas de exterior para cámaras de vigilancia
- Dos cámaras de color GigE Basler Scout scA780-54gc
- Dos cámaras monocromáticas GigE Basler Scout scA780-54gm
- Cuatro ópticas 6 mm 1:1.4
- Cuatro focos de iluminación infrarroja
- Tarjeta interfaz USB de experimentación Velleman K8055
- Equipo de comunicaciones WiMAX BreezeACCESS(R) VL video (Alvarion)

## 9.2. Software

Se deberá tener debidamente instalado y configurado el siguiente software:

- Sistema operativo GNU/Linux Ubuntu 8.04 (LTS) con kernel 2.6.24 o superior
- Librerías OpenCV
- API Pylon 2.0 (Basler)
- Resto de librerías y paquetes, indicados en el manual de usuario 8.1
- Reproductor multimedia “VLC”

Por último, indicar la necesidad de disponer de una ***dirección IP estática de dominio público*** para establecer la conexión con el módulo ubicado en la cabina del tren.

Cabe destacar que, a excepción del API proporcionado al adquirir las cámaras “*Basler Scout*”, se trabaja con software libre distribuido bajo la licencia GPL.



# **Parte V**

## **Presupuesto**



# Capítulo 10

## Presupuesto

En este capítulo se se realiza una estimación del importe necesario para la ejecución del proyecto. Para ello se dividen los costes en función de su origen.

### 10.1. Costes de ejecución material

Se incluyen tres elementos en los costes de ejecución material:

1. Costes de equipamiento hardware
2. Costes de software
3. Costes por tiempo empleado

#### 10.1.1. Costes de equipamiento hardware

| CONCEPTO                                                                                      | PRECIO UNITARIO | CANTIDAD | SUBTOTAL        |
|-----------------------------------------------------------------------------------------------|-----------------|----------|-----------------|
| Ordenador sobremesa Intel Core2 Quad Q9550 @ 2.83GHz                                          | 1.300 €         | 2        | 2.600 €         |
| Cámara digital Gigabit Ethernet Basler Scout + carcasa IP65 + óptica + iluminación infrarroja | 1.500 €         | 4        | 6.000 €         |
| Equipo de comunicaciones WiMAX BreezeACCESS VL video                                          | 7.000 €         | 1        | 7.000 €         |
| Tarjeta interfaz USB Velleman K8055                                                           | 60 €            | 1        | 60 €            |
| Switch Gibabit Ethernet SM-CGS5 10/100/1000                                                   | 40 €            | 2        | 80 €            |
| Bienes fungibles                                                                              | 2.000 €         | 1        | 2.000 €         |
| <b>Subtotal</b>                                                                               |                 |          | <b>17.740 €</b> |

### 10.1.2. Costes de software

| CONCEPTO                                                                  | PRECIO UNITARIO | CANTIDAD | SUBTOTAL   |
|---------------------------------------------------------------------------|-----------------|----------|------------|
| Sistema operativo: Ubuntu 8.04                                            | 0 €             | 1        | 0 €        |
| Captura de imágenes: API Pylon 2.0 (incluido en el precio de las cámaras) | 0 €             | 1        | 0 €        |
| Entorno de programación: Gedit                                            | 0 €             | 1        | 0 €        |
| Elaboración de documentación: Latex                                       | 0 €             | 1        | 0 €        |
| Edición de gráficos vectoriales: Inkscape                                 | 0 €             | 1        | 0 €        |
| <b>Subtotal</b>                                                           |                 |          | <b>0 €</b> |

### 10.1.3. Costes por tiempo empleado

| FUNCIÓN         | PRECIO UNITARIO | Nº HORAS | SUBTOTAL        |
|-----------------|-----------------|----------|-----------------|
| Ingeniería      | 50 €            | 1.440    | 72.000 €        |
| Mecanografía    | 35 €            | 240      | 8.400 €         |
| <b>Subtotal</b> |                 |          | <b>80.400 €</b> |

El número de horas de ingeniería corresponde a un trabajo de seis horas diarias durante un año. El mecanografiado se ha estimado como seis horas diarias durante dos meses.

### 10.1.4. Coste total del presupuesto de ejecución material

| CONCEPTO                        | SUBTOTAL        |
|---------------------------------|-----------------|
| Costes de equipamiento hardware | 17.740 €        |
| Costes de software              | 0 €             |
| Costes por tiempo empleado      | 80.400 €        |
| <b>Subtotal</b>                 | <b>98.140 €</b> |

## 10.2. Gastos generales y beneficio industrial

Los gastos generales y beneficio industrial son los gastos obligados que se derivan de la utilización de las instalaciones de trabajo más el beneficio industrial. Se estima un porcentaje del 16 % sobre el coste de ejecución material.

| CONCEPTO                                | SUBTOTAL          |
|-----------------------------------------|-------------------|
| Gastos generales y beneficio industrial | 15.702,4 €        |
| <b>Subtotal</b>                         | <b>15.702,4 €</b> |

## 10.3. Importe total del presupuesto

| CONCEPTO                                          | SUBTOTAL            |
|---------------------------------------------------|---------------------|
| Coste total del presupuesto de ejecución material | 98.140 €            |
| Gastos generales y beneficio industrial           | 15.702,4 €          |
| <b>Subtotal</b>                                   | <b>113.842,4 €</b>  |
| <b>IVA 16 %</b>                                   | <b>18.214,78 €</b>  |
| <b>TOTAL IVA INCLUIDO</b>                         | <b>132.057,18 €</b> |

El importe total del proyecto suma la cantidad de:

*“Ciento Treinta y Dos mil Cincuenta y Siete euros con Dieciocho céntimos”*

Alcalá de Henares a 31 de Julio de 2009

Fdo: David Jiménez Cabello  
Ingeniero de Telecomunicación



# **Parte VI**

## **Bibliografía**





# Bibliografía

- [REO, 2004] (2004). REOST Project Description. <http://spt.dibe.unige.it/ISIP/reost/project-description.html>.
- [Allili et al., 2007a] Allili, M., Bouguila, N., and Ziou, D. (2007a). A robust video foreground segmentation by using generalized gaussian mixture modeling. In *Computer and Robot Vision, 2007. CRV'07. Fourth Canadian Conference on*, pages 503–509.
- [Allili et al., 2007b] Allili, M., Bouguila, N., and Ziou, D. (2007b). Finite generalized Gaussian mixture modeling and applications to image and video foreground segmentation. In *Computer and Robot Vision, 2007. CRV'07. Fourth Canadian Conference on*, pages 183–190.
- [Baker and Matthews, 2004] Baker, S. and Matthews, I. (2004). Lucas-kanade 20 years on: A unifying framework. *International Journal of Computer Vision*, 56(3):221–255.
- [Brown, 1965] Brown, E. (1965). Modern optics. *Modern Optics*.
- [Chinrungrueng and Sequin, 1995] Chinrungrueng, C. and Sequin, C. (1995). Optimal adaptive k-means algorithm with dynamic adjustment of learning rate. *IEEE Transactions on Neural Networks*, 6(1):157–169.
- [Consulting, 2005a] Consulting, S. (2005a). Can WiMAX address Your Applications. Technical report, WiMAX Forum.
- [Consulting, 2005b] Consulting, S. (2005b). Fixed, nomadic, portable and mobile applications for 802.16-2004 and 802.16 e WiMAX networks. Technical report, WiMAX Forum.
- [Cucchiara et al., 2003] Cucchiara, R., Grana, C., Piccardi, M., and Prati, A. (2003). Detecting moving objects, ghosts, and shadows in video streams. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1337–1342.
- [Finlayson et al., 2002] Finlayson, G., Hordley, S., and Drew, M. (2002). Removing shadows from images. *Lecture Notes in Computer Science*, pages 823–836.
- [Goodman and Myers, 2005] Goodman, D. and Myers, R. (2005). 3G cellular standards and patents. In *Wireless Networks, Communications and Mobile Computing, 2005 International Conference on*, volume 1.
- [Hartigan and Wong, 1979] Hartigan, J. and Wong, M. (1979). A k-means clustering algorithm. *JR Stat. Soc., Ser. C*, 28:100–108.

- [Hartley and Zisserman, 2003] Hartley, R. and Zisserman, A. (2003). *Multiple view geometry in computer vision*. Cambridge Univ Pr.
- [Mahalanobis, 1936] Mahalanobis, P. (1936). On the generalised distance in statistics. Proceedings National Institute of Science. India.
- [Michael Tilleman, 1998] Michael Tilleman, K. K. (1998). Off the line-of-sight laser radar. *Applied Optics*, 37:565–572.
- [Mockel et al., 2003] Mockel, S., Scherer, F., Schuster, P., Bildverarbeitungssysteme, S., and Wiesbaden, G. (2003). Multi-sensor obstacle detection on railway tracks. In *IEEE Intelligent Vehicles Symposium, 2003. Proceedings*, pages 42–46.
- [Okutani et al., 2000] Okutani, T., Nakamura, N., Shimazoe, T., and Shimada, M. (2000). Safety of Level Crossings on the Integrated Crossing Control System Using Electronic Controller. *IEIC Technical Report (Institute of Electronics, Information and Communication Engineers)*, 100(512):39–46.
- [Ota, 2004] Ota, M. (2004). Obstacle Detection System with Stereo Cameras for Level Crossings. *Railway Technology Avalanche*, 5:32.
- [P. Bon, 2004] P. Bon, C. C. (2004). Overview of National and European projects European Commission. Fifth Framework programme SAMNET thematic Network.
- [Pajares and De la Cruz, 2003] Pajares, G. and De la Cruz, J. (2003). Visión por computador. Imágenes digitales y aplicaciones. *Alfaomega Grupo Editor*.
- [Pappas, 1992] Pappas, T. (1992). An adaptive clustering algorithm for image segmentation. *IEEE Transactions on Signal Processing*, 40(4):901–914.
- [Quintas et al., 1996] Quintas, M. et al. (1996). Visión artificial.
- [Ruder et al., 2003] Ruder, M., Mohler, N., and Ahmed, F. (2003). An obstacle detection system for automated trains. In *IEEE Intelligent Vehicles Symposium, 2003. Proceedings*, pages 180–185.
- [Safety, 2003] Safety, H. L. G. R. (2003). Second Report of the Working Group on Level Crossings. Safety at Level Crossings. Technical report, European Commission DG Energy and Transport.
- [Solution, 2005] Solution, A. W. (2005). Documentation: BreezeACCESS VL.
- [Transportation, 2005] Transportation, G. E. (2005). Documentation: General Electric.
- [Ukai, 2004] Ukai, M. (2004). *Obstacle detection with a sequence of ultra telephoto camera images*. Computers in Railways IX, 2004 WIT Press.
- [Vazquez et al., 2004] Vazquez, J., Mazo, M., Lazaro, J., Luna, C., Urena, J., Garcia, J., Cabello, J., and Hierrezuelo, L. (2004). Detection of moving objects in railway using vision. In *2004 IEEE Intelligent Vehicles Symposium*, pages 872–875.
- [Velez et al., 2003] Velez, J., Moreno, A., Sánchez, Á., and Esteban, J. (2003). Visión por computador. *Universidad Rey Juan Carlos*.

- [Wang et al., 2006] Wang, Y., Loe, K., and Wu, J. (2006). A dynamic conditional random field model for foreground and shadow segmentation. *IEEE transactions on pattern analysis and machine intelligence*, pages 279–289.
- [Wang et al., 2005] Wang, Y., Tan, T., Loe, K., and Wu, J. (2005). A probabilistic approach for foreground and shadow segmentation in monocular image sequences. *Pattern Recognition*, 38(11):1937–1946.
- [WiMAX, 2006a] WiMAX, M. (2006a). Mobile WiMAX Usage Scenarios. In *WiMAX Forum*, February.
- [WiMAX, 2006b] WiMAX, M. (2006b). Part II: Competitive Analysis. In *WiMAX Forum*, February.
- [WiMAX-Part, 2006] WiMAX-Part, M. (2006). Part I: A technical overview and performance evaluation. In *WiMAX Forum*.

