



INSTITUTO POLITÉCNICO NACIONAL

CENTRO DE INVESTIGACIÓN EN COMPUTACIÓN

**DETECCIÓN TEMPRANA DE INCENDIOS MEDIANTE
FLUJOS DE VIDEO JPEG Y MPEG**

TESIS

QUE PARA OBTENER EL GRADO DE:

**MAESTRO EN CIENCIAS EN INGENIERÍA DE
CÓMPUTO**

PRESENTA:

GILBERTO GUSTAVO ROMERO HERNÁNDEZ

DIRECTORES: M. EN C. PABLO MANRÍQUE RAMÍREZ

DR. LUIS MARTÍN ROJAS CÁRDENAS



MÉXICO D. F. Diciembre de 2013



INSTITUTO POLITÉCNICO NACIONAL
SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

SI/14/06

ACTA DE REVISIÓN DE TESIS

En la Ciudad de México, D.F. siendo las 12:00 horas del día 18 del mes de junio de 2013 se reunieron los miembros de la Comisión Revisora de la Tesis, designada por el Colegio de Profesores de Estudios de Posgrado e Investigación del:

Centro de Investigación en Computación

para examinar la tesis titulada:

"Detección temprana de incendios mediante flujos de video JPEG y MPEG"

Presentada por el alumno:

ROMERO
Apellido paterno

HERNÁNDEZ
Apellido materno

GILBERTO GUSTAVO
Nombre(s)

Con registro:

A	9	7	0	1	3	6
---	---	---	---	---	---	---

aspirante de: **MAESTRÍA EN CIENCIAS EN INGENIERÍA DE CÓMPUTO CON OPCIÓN EN SISTEMAS DIGITALES**

Después de intercambiar opiniones los miembros de la Comisión manifestaron **APROBAR LA TESIS**, en virtud de que satisface los requisitos señalados por las disposiciones reglamentarias vigentes.

LA COMISIÓN REVISORA
Directores de Tesis

M. en C. Pablo Márquez Ramírez

Dr. Luis Martín Rojas Cárdenas

Dr. Sergio Suárez Guerra

Dr. Aleksiy Pogrebnyak

Dr. Luis Pastor Sánchez Fernández

Dr. José Giovanni Guzmán Lugo

PRESIDENTE DEL COLEGIO DE PROFESORES

Dr. Luis Alfonso Villa





INSTITUTO POLITÉCNICO NACIONAL
SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

CARTA CESIÓN DE DERECHOS

En la Ciudad de México D.F. el día 18 del mes Junio del año 2013, el (la) que suscribe Gilberto Gustavo Romero Hernández alumno (a) del Programa de Maestría en Ciencias en Ingeniería de Cómputo con Opción en Sistemas Digitales con número de registro A970136, adscrito al Centro de Investigación en Computación, manifiesta que es autor (a) intelectual del presente trabajo de Tesis bajo la dirección de M. en C. Pablo Manrique Ramírez y el Dr. Luis Martín Rojas Cárdenas y cede los derechos del trabajo intitulado Detección Temprana de incendios mediante flujos de video JPEG y MPEG, al Instituto Politécnico Nacional para su difusión, con fines académicos y de investigación.

Los usuarios de la información no deben reproducir el contenido textual, gráficas o datos del trabajo sin el permiso expreso del autor y/o director del trabajo. Este puede ser obtenido escribiendo a la siguiente dirección gilbertogromeroh@hotmail.com. Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.

Gilberto Gustavo Romero Hernández

Nombre y firma

Agradecimientos

Al Instituto Politécnico Nacional, por darme la oportunidad de ser parte de su comunidad.

Al Centro de Investigación en Computación, por haberme compartido sus conocimientos.

Al Dr. Luis Martín Rojas Cárdenas, por haberme invitado a participar en este proyecto de investigación y sobre todo por brindarme su tiempo, apoyo y valiosa ayuda incondicional.

A mi familia, por su paciencia y motivación para la culminación de este trabajo.

Resumen

En esta tesis se propone un algoritmo para efectuar la detección temprana de incendios a partir de las informaciones de la DCT (Discret Cosine Transform) contenidas en ciertos flujos de video comprimido, lo que permite obtener un método de detección de bajo costo computacional y directamente aplicable a sistemas de video-vigilancia convencionales IP, las cuales generan, entre otras, secuencias de video JPEG y H264. Se utiliza la “detección temprana” para la detección de humo grisáceo translucido, el cual es uno de los primeros síntomas del proceso combustión. Los resultados obtenidos muestran un alto grado de fiabilidad y un bajo costo computacional, lo que permiten procesar imágenes de 1920x1080 píxeles al ritmo que las genera la cámara.

Abstract

In this thesis an algorithm to make early detection of fires from the DCT (Discret Cosine Transform) information contained in certain compressed video streams is proposed, this allows to obtain a method of detection of low-cost computer and directly applicable to conventional IP conventional video surveillance systems, which generate, among other, JPEG and H264 video sequences. "Early detection" is used for the detection of greyish smoke translucent, which is one of the first symptoms of the combustion process. The results show a high degree of reliability and a low computational cost, which allows you to process images of 1920 x 1080 pixels at the pace that generated the camera.

Índice

Resumen.....	2
Abstract	3
Índice	4
Lista de figuras	6
Lista de tablas.....	8
Lista de Abreviaturas.....	9
1 Introducción	12
1.1 Motivación y Justificación	12
1.2 Identificación del Problema	13
1.3 Objetivo general.....	15
1.4 Objetivos particulares	15
1.5 Alcances y limitaciones	15
1.6 Organización del documento	16
2 Estado del Arte	17
2.1 El Preprocesamiento	18
2.2 Extracción de elementos útiles para el reconocimiento.....	19
2.3 La clasificación.....	20
2.4 El método de reconocimiento.....	20
2.5 Eliminación de resultados errátiles.....	21
2.6 Trabajos Relacionados.....	22
3 Marco Teórico	49
3.1 La transformada de coseno discreta	50
3.2 La norma Joint Picture Expert Group	54
3.3 El proceso de compresión y descompresión paso a paso.....	55
3.4 Impacto del humo en los coeficientes de la DCT	60
3.5 Cuantificando cambios en las propiedades de la imagen	66
4 Modelo propuesto para la detección de humo	78
4.1 Modelos con Mezclas Gaussianas.....	79
4.2 Creación del Modelo GMM para el reconocimiento del humo	82
4.2.1 Adquisición de muestra y extracción de EURIs	83

4.2.2 Extracción de características	86
4.2.3 Creación del modelo usando mixdist	89
4.3 El proceso de reconocimiento de humo	94
4.4 Eliminación de resultados errátiles	97
4.4.1 Acciones en el ámbito del bloque	98
4.4.2 Control de evento “humo” en un grupo de bloques.....	101
4.5 Verificación del movimiento de la masa de humo.....	104
5 Evaluación de Resultados.....	107
5.1. Comparación con otros trabajos.....	112
5.1.1 Tasa de detecciones correctas	113
5.1.2 Tasa de falsas detecciones	113
5.1.3 Cantidad de píxeles procesados por segundo.....	114
6 Conclusiones.....	116
6.1 Contribuciones	117
6.2 Trabajo futuro	118
7 Referencias	119

Lista de figuras

Figura 2.1 Esquema genérico de un método para la detección de incendios	17
Figura 2.2 Uso del Procesador cuando se decodifica video H.264.....	44
Figura 3.1 Clasificación de áreas indicativas en un bloque DCT.....	52
Figura 3.2 Orden de los coeficientes de la DCT.....	53
Figura 3.3 Representación de un códec JPG	56
Figura 3.4 Bloque de pixeles y su correspondiente bloque de coeficientes DCT.....	57
Figura 3.5 Tabla de cuantificación (a) y un bloque de coeficientes resultante del proceso de cuantificación (b).....	58
Figura 3.6 Coeficientes DCT descuantificados (a) e inversamente transformados por la DCT (b). 59	
Figura 3.7 Un bloque sin humo, cuadro punteado abajo izquierda (a), y el mismo bloque, ahora con humo en una imagen ulterior, cuadro punteado abajo derecha (b)	62
Figura 3.8 bloques de Coeficientes frecuenciales DCT sin y con humo	63
Figura 3.9 Imagen de referencia sobre la cual se efectúan modificaciones de brillo, contraste y definición.....	67
Figura 3.10 Cambios de brillo en los coeficientes DCT de un bloque	69
Figura 3.11 Efecto de los cambios de contraste en los coeficientes DCT	71
Figura 3.12 La cantidad de coeficientes AC cuyo valor difiere de cero, está ligado a la cantidad de información presente en el bloque.....	73
Figura 3.13 Pérdida de información provocada por la aplicación de filtros de desenfoco	75
Figura 4.1 Gráficas que entrega la herramienta de análisis de imágenes JPG.....	85
Figura 4.2 Distribución empleando el conjunto C.....	90
Figura 4.3 Distribución extraída del Conjunto U.....	91
Figura 4.4 Distribución extraída del conjunto E.....	91
Figura 4.5 Distribución extraída del conjunto H	91
Figura 4.6 Distribución extraída del conjunto Ω	91
Figura 4.7 Distribución extraída del conjunto T.....	91
Figura 4.8 Distribución extraída del conjunto Θ	92

Figura 4.9 Algoritmo que reconoce humo translúcido a partir del modelo propuesto en el capítulo anterior.....	96
Figura 4.10 Área de búsqueda circundante a un elemento considerado como ligado a la presencia de humo en la escena	103
Figura 4.11 Modelo de Mezclas Gaussianas para la detección de humo translúcido con fase de filtrado y eliminación de resultados erráticos	103
Figura 4.12 Siluetas de humo superpuestas que corresponden a una misma fuente de humo fotografiada a diferentes instantes.....	104
Figura 4.13 Disposición del método para verificar movimiento.....	106
Figura 5.1 Enfoque de la teoría de conjuntos para los posibles escenarios en la detección de humo	108
Figura 5.2 Interpretación de varios tipos de resultados de desempeño del algoritmo de acuerdo a su calidad.....	110
Figura 5.3 Tasa de detecciones correctas	113
Figura 5.4 Tasa de falsas detecciones	114
Figura 5.5 Cantidad de píxeles procesados por segundo.....	115

Lista de tablas

Tabla 2.1 Puntos esenciales de las propuestas para la detección de incendios	45
Tabla 3.1 Humo apareciendo en la escena	65
Tabla 3.2 Sombras apareciendo en la escena	65
Tabla 3.3 Cambios de iluminación en ciertas porciones	65
Tabla 3.4 Aparición de objetos opacos	66
Tabla 4.1 Gaussian Mixture Model λ_C para el conjunto C.....	92
Tabla 4.2 Gaussian Mixture Model λ_U para el conjunto U.....	93
Tabla 4.3 Gaussian Mixture Model λ_E para el conjunto E.....	93
Tabla 4.4 Gaussian Mixture Model λ_H para el conjunto H	93
Tabla 4.5 Gaussian Mixture Model λ_Ω para el conjunto Ω	93
Tabla 4.6 Gaussian Mixture Model λ_T para el conjunto T	93
Tabla 4.7 Gaussian Mixture Model λ_Θ para el conjunto Θ	93
Tabla 4.8 Humo desapareciendo de la escena.....	105
Tabla 5.1 Evaluación de resultados del algoritmo propuesto, usando medidas de evaluación global.....	111

Lista de Abreviaturas

$\pi^{-1}(S)$	Secuencias que el sistema propuesto detecta humo
ACC	Accuracy
CAMSHIFT	<i>“Continuously Adaptive Mean Shift”</i>
cd	<i>“Correct detections”</i> (detecciones correctas)
cr	<i>“Correct rejections”</i> (rechazos correctos)
DIVB	Detección de incendios mediante video en bruto (video sin compresión)
DIVC	Detección de incendios mediante video comprimido
DR	Detection Rate
DXVA	DirectX Video Acceleration
E	Secuencias detectadas por un observador humano
EM	<i>“Expectation Maximization”</i>
ERR	Error
EURI	Elementos útiles para el reconocimiento de Incendios
FAR	<i>“False Alarm Rate”</i>
FCM	<i>“Fuzzy C-Means”</i>
fd	<i>“False detections”</i>
FDR	<i>“Fire Detection Rate”</i>
FLC	Código de longitud fija
FN	<i>“False Negative Rate”</i>
FNR	<i>“False Negative Rate”</i>
FPR	<i>“False Positive Rate”</i>

GLCM	<i>“Gray Level Co-occurrences matrix”</i>
GMM	Modelo de mezclas Gaussianas
GMM	Gaussian Mixture Models
HD	Alta definición
HIS	Hue Intensity Saturation
HVS	Sistema Visual Humano
KLT	<i>“Karhunen-Loeve Transform”</i>
LBP	<i>“Local Binary Pattern”</i>
mcc	Correlación de Mathews
MCU	<i>“Minimum Coded Unit”</i>
md	<i>“missed detections”</i>
MHI	<i>“Motion History Image”</i>
MLE	Estimación de máxima Verosimilitud
MRD	<i>“Moving Region Decision”</i>
MV	Vectores de movimiento
NRLBP	<i>“Non-Redundant Local Binary Pattern”</i>
NRLMBP	<i>“Non-Redundant Local Motion Binary Pattern”</i>
PFK	<i>“Potential Fire Mask”</i>
PPM	<i>“Portable Pixel Mapping”</i>
ppv	<i>“positive predictive value”</i>
ROI	Regiones de interes
SVM	Máquinas de soporte vectorial
TN	Verdaderos negativos

TNR	True Negative Rates
TP	Verdaderos positivos
TPR	<i>“True Positive Rate”</i>
VLC	Código de longitud variable
X	Todas las secuencias usadas con humo y sin humo para evaluar al sistema

1 Introducción

1.1 Motivación y Justificación

El fuego provoca enormes pérdidas humanas, sufrimiento e incalculables daños materiales en todos los países del planeta. Contar con sistemas detectores de fuego en etapas tempranas de la combustión podría evitar en gran medida la pérdida material y sobre todo salvar vidas. Con esta intención, un número considerable de técnicas de detección de incendios han sido exploradas. Entre estas técnicas se puede citar: el muestreo de partículas, de la temperatura, o de la humedad relativa, las pruebas de la transparencia del aire, etc. Desafortunadamente, aunque estos sensores ofrecen una alta sensibilidad, su eficacia y fiabilidad es limitada ya que estos detectores solo pueden indicar la presencia de fuego cuando las partículas correspondientes llegan a ellos. Esto hace que este tipo de sensores sean poco recomendables en espacios abiertos. Frente a este inconveniente, la detección de incendios mediante video representa una solución alternativa con diversas ventajas tales como la posibilidad de detectar incendios a distancias significativamente mayores, en un lapso de tiempo menor y sobretodo en espacios abiertos. Además, la detección en video puede entregar informaciones que los sensores basados en partículas no pueden, tal como la ubicación de la fuente de combustión, su magnitud, etc. Finalmente, el uso de esta tecnología se ve motivado por la gran infraestructura de video-vigilancia [51] [53] ya presente en muchos edificios y otras instalaciones.

1.2 Identificación del Problema

La investigación en el campo del reconocimiento de patrones en video ha sido desarrollada desde hace años con resultados novedosos. En este campo, un área relativamente nueva es el reconocimiento de incendios, los resultados aquí obtenidos son alentadores pero aún queda mucho por hacer, en particular, la detección de incendios mediante video comprimido (DIVC) es un tema poco abordado. En efecto, la mayor parte de trabajos propuestos hasta ahora se centra en la detección de incendios mediante video en bruto (DIVB), también conocido como video sin compresión. A pesar de esta tendencia, existen diversas razones que motivan la creación de métodos DIVC:

a) La gran mayoría de sistemas para la video-vigilancia generan flujos de video comprimido que no pueden ser directamente procesados por los DIVB. Para que esto sea posible se requiere, en primer lugar, descomprimir el flujo de video. Esto implica un enorme costo computacional que puede comprometer la eficacia del método DIVB, recordar que detectar incendios en tiempo real es un requisito de vital importancia. Otro factor que puede comprometer la eficacia de los métodos DIVB, es el enorme volumen de datos generados por las cámaras de video-vigilancia actuales: alrededor de 30 imágenes por segundo con una resolución espacial de 1920x1080 píxeles.

b) El video comprimido contiene información que no está representada con píxeles sino con elementos de un nivel de abstracción superior tales como vectores de movimiento, información frecuencial e información de luminancia separada de la de color. Esta representación de la información permite realizar el reconocimiento de patrones de una manera más sencilla con respecto a los métodos DIVB. Es más, muchos de ellos [2][14][15][17][19][35][37][38][45][55][56] realizan la transformación de píxeles a una de

las representaciones que se cita aquí arriba. Sin embargo, estas representaciones solo se logran aplicando costosos métodos de transformación, tales como Fourier, wavelet u otros, para el caso de la información espacial. O bien métodos de predicción de movimiento para el análisis de la información temporal. Entonces, realizar un algoritmo de reconocimiento para trabajar directamente sobre flujos de video comprimido permite ahorrar recursos de cómputo, pero requiere tomar en cuenta las imperfecciones del video comprimido.

c) El video descomprimido presenta pérdidas de información y distorsiones de orden considerable no previstas por los métodos DIVB. Es bien sabido que los métodos de compresión de video utilizados por los sistemas de video-vigilancia son de tipo destructivo, lo que significa que el método de compresión, y en particular el procedimiento de cuantificación, elimina una parte de la información de video; en prioridad la información ubicada en las altas frecuencias. Esta información es la que determina la fineza de las texturas y los contornos en las imágenes. Frente a estas pérdidas, algunos métodos DIVB [2][14][15][17][19][35][37][38][45][55][56] están en problemas, ya que éstos fundan su funcionamiento en el análisis de las altas frecuencias. Entonces, si no existe esta información en un video descomprimido y se quiere aplicar en él un método DIVB, éste último debe ser reajustado o rediseñado, para tomar en cuenta esta degradación. Aquí es importante mencionar que en los trabajos antes citados, no se menciona la forma en que se adaptarán estos métodos DIVB a los sistemas de video-vigilancia que generan flujos de video comprimidos.

1.3 Objetivo general

Proponer un método para la detección temprana de incendios mediante video comprimido bajo la norma JPEG o MPEG, considerando la aparición de humo tenue grisáceo azulado en las primeras etapas de la combustión.

1.4 Objetivos particulares

1. No utilizar píxeles en la detección de humo, sino la información que se encuentre disponible en el flujo de video codificado. En este caso, tal información la constituye los coeficientes frecuenciales de la DCT.
2. Proponer un algoritmo de reconocimiento que presente un bajo costo computacional, de tal manera que éste pueda procesar imágenes de alta resolución (1980x1080 pixeles) al ritmo que son generados por la cámara de video IP (Internet Protocol); alrededor de 30 imágenes por segundo.
3. El algoritmo de detección debe presentar una buena precisión, al menos con un índice de 0.93 en la tasa de las detecciones correctas, en base a las medidas de evaluación global [22].

1.5 Alcances y limitaciones

Para desarrollar este estudio se emplea un marco controlado de experimentación, el cual no incluye, ni todas las clases de humo, ni todas las condiciones posibles, así, el marco considerado es el siguiente:

- a) Solo se considera humo translucido grisáceo-azulado
- b) Uso de cámaras fijas convencionales del tipo IP.

- c) Buena iluminación, solar, o equivalente. Desde que aparecen los primeros rayos del sol hasta su ocaso.
- d) Video “Full HD” codificado con JPEG baseline; 1920x1080 píxeles@30 fps.
- e) Tasa de compresión de 1.5 a 2.0 bits/píxel.
- f) Distancia entre la cámara y la fuente de combustión de hasta 100 metros.

1.6 Organización del documento

En lo que resta de este documento, el capítulo II presenta el estado del arte en el campo de la detección de incendios en video. La detección de incendios en video se puede efectuar, ya sea por detección de humo, o por detección de flama. Así, aunque la tesis está consagrada a la detección de humo, se analizan algunos trabajos relacionados con la detección de flama por considerarlos de interés. El marco teórico en el campo de la compresión de video y la idea básica de cómo abordar la detección de humo en video comprimido son expuestos en el capítulo III. El capítulo IV trata del desarrollo del algoritmo basado en los modelos de mezclas Gaussianas. Una sección dedicada a la eliminación de resultados errátiles es también presentada en este capítulo. Los resultados obtenidos son presentados y evaluados en el capítulo V. Finalmente, el capítulo VI aborda la conclusión y el trabajo futuro.

2 Estado del Arte

En los últimos años se ha desarrollado un buen número de propuestas destinadas a efectuar la detección de incendios en video, las cuales abordan el problema desde diferentes puntos de vista y aplicando técnicas diversas. La gran mayoría de estas propuestas, sin embargo, coinciden en el tipo de video sobre el cual operan, a saber: el video sin compresión. En efecto, a pesar de la conveniencia de trabajar directamente sobre video comprimido, el número de propuestas DIVC es realmente reducido.

En este capítulo se presenta un análisis de las técnicas propuestas en la literatura para efectuar la detección de incendios en video. El análisis porta un particular interés en dos aspectos esenciales: los procedimientos utilizados y el desempeño de la propuesta. Para este último se considera elementos objetivos, tanto el número de imágenes procesadas por segundo, como de la tasa de alarmas falsas o verdaderas. Es importante mencionar que en pocos trabajos se reportan estos dos últimos aspectos.

Para realizar este análisis, se utiliza como referencia el esquema presentado en la figura 2.1, el cual representa la organización de un método genérico para el reconocimiento de patrones, el cual se retoma para representar el caso particular de la detección de incendios. Los módulos que conforman este esquema representan cada uno de los principales elementos sobre los cuales reposa el método.

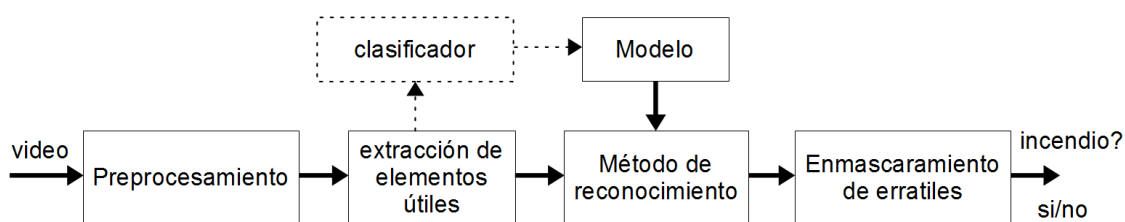


Figura 2.1 Esquema genérico de un método para la detección de incendios

2.1 El Preprocesamiento

La primera tarea efectuada por un método de reconocimiento de incendios consiste en convertir los datos del video recibido en una representación que convenga más a los intereses del módulo “extracción de elementos”. Los detalles de esta conversión dependen, por una parte, de la naturaleza del flujo de video entrante, y por la otra, de los objetivos perseguidos por el método de detección. Aquí es importante mencionar que no todos los métodos de reconocimiento de patrones incluyen una etapa de preprocesamiento. Para el caso del video en bruto, algunos métodos de preprocesamiento comúnmente utilizados son los métodos de transformación de dominio, tales como las transformadas DCT, Fourier, wavelet y otros. Los cambios en el espacio de representación de colores también son frecuentemente utilizados. El cambio del espacio de representación RGB (Red Green Blue) a YUV (Luminancia *Y* y diferencias de crominancias), HIS (*Hue Intensity Saturation*) y CIEL LAB son algunos ejemplos. Algunos métodos de detección de incendios también recurren a la eliminación de la información de color.

Para el caso del video comprimido, el preprocesamiento se realiza descomprimiendo parcialmente el flujo de video; solo se efectúa la primera fase de la descompresión JPEG, en particular la decodificación entrópica (Huffman) con el fin de obtener los coeficientes frecuenciales y los vectores de movimiento. Es importante mencionar que esta descompresión parcial es un procedimiento de bajo costo, si se compara con los métodos de conversión de dominio citados para el caso del video en bruto.

2.2 Extracción de elementos útiles para el reconocimiento

Un método de detección de incendios en video basa su funcionamiento en un modelo que describe la presencia de un incendio en términos de ciertas propiedades u objetos presentes en las imágenes al momento del incendio. Para estas propiedades y objetos se utiliza el término Elementos Útiles para el Reconocimiento de Incendios, o EURI. Estos elementos pueden ser extraídos, tanto en el dominio espacial, en el temporal, como en el frecuencial. En todo caso, los EURI deben ser aquellos que describan de la mejor manera posible la presencia de un incendio en la escena. Ahora, en el campo espacial es común extraer elementos relacionados con propiedades, tales como la textura, la colorimetría, el brillo y el contraste, u objetos tales como los contornos o aglomeraciones de píxeles que comparten propiedades similares. En el campo frecuencial, aparte de los elementos citados anteriormente, también es común extraer información ligada a los cambios de brillo, constraste, nitidez, etc. En efecto, los cambios de todas estas propiedades pueden ser analizadas con toda precisión analizando las bajas frecuencias, las intermedias y las altas frecuencias respectivamente.

Por otra parte, en el campo temporal se extrae la información de movimiento, tal como la frecuencia de parpadeo de las flamas, el desplazamiento de la masa de humo, e incluso la evolución caótica de éste. Además, los EURI también son utilizados por el módulo de reconocimiento para compararlos con un modelo ya creado que describe el comportamiento de un incendio en la escena. Aquí cabe aclarar que, en una primera fase, los EURI son utilizados para alimentar el módulo de clasificación; esta fase corresponde con una fase de aprendizaje del sistema de reconocimiento. Después, en una fase posterior, los EURI son empleados para reconocer el patrón buscado.

2.3 La clasificación

El proceso de clasificación se encarga de crear un modelo capaz de describir la ocurrencia de incendios en video gracias al análisis y caracterización de muestras relacionadas con el humo o las flamas. Estas muestras pueden corresponder con alguna de las múltiples propiedades ligadas a una imagen, tales como el brillo, el contraste, la nitidez, los colores, etc. O elementos de un nivel de abstracción superior como son la forma de los contornos, la textura de los objetos, el movimiento de los objetos en la escena, etc. Es importante mencionar, que en la mayor parte de propuestas reportadas, es el ser humano quien establece la relación entre una muestra y alguno de los síntomas de un incendio. Así, la precisión del método depende en parte del trabajo humano durante la selección de las muestras.

Por otra parte, existe una amplia gama de herramientas matemáticas, algunas más complejas que otras, que permiten la creación de modelos para el reconocimiento de patrones en video. Aquí cabe aclarar que la detección de un incendio es una tarea compleja que requiere, en la mayor parte de casos, de la intervención conjunta de varios modelos. Ahora, entre las técnicas para crear un modelo se puede encontrar desde reglas con umbrales obtenidos empíricamente, pasando por las clásicas herramientas estadísticas que entregan una distribución de probabilidad, hasta llegar a técnicas más complejas como las redes neuronales, las máquinas de soporte vectorial (SVM), las cadenas de Markov, etc.

2.4 El método de reconocimiento

Un método de reconocimiento de incendios en video basa su funcionamiento en un modelo que describe la presencia de un incendio en términos del “comportamiento” de ciertas propiedades u objetos presentes en las imágenes. Recordar que para estas

propiedades y objetos se ha establecido el término EURI. Tomando en cuenta esto, el método de reconocimiento recibe los EURI que le presenta el módulo de “extracción de elementos útiles” y prueba que éstos conformen el modelo. El procedimiento de prueba efectuado por un método de reconocimiento depende tanto de la naturaleza del EURI, como del método de clasificación empleado para generar el modelo. Si el modelo es un conjunto de reglas basadas en umbrales, entonces el EURI probablemente es un escalar y debe entrar en un conjunto de rangos para enunciar que éste conforma el modelo, entonces, se trata de un indicio de fuego.

En casos más sofisticados, como las redes neuronales, es común que estas presenten dos modos de operación. Una de aprendizaje y otra de reconocimiento. En el modo de aprendizaje, la red solo recibe EURI seleccionados – por estar ligados a la presencia de un incendio –. Con estos EURI la red forma el modelo que posteriormente será utilizado para el reconocimiento, el cual solo se puede efectuar cuando la red se pone en modo de reconocimiento.

2.5 Eliminación de resultados errátiles

La gran mayoría de métodos para el reconocimiento de incendios en video no son perfectos y tienden a dar un cierto número de resultados errátiles de forma imprevista. Para reducir este problema, algunos métodos utilizan mecanismos de enmascaramiento sobre los resultados finales; esta etapa es una operación de pos-procesamiento que busca eliminar resultados aislados, y por lo tanto, probables de ser incorrectos. Algunos de los métodos que se utilizan para eliminar los resultados errátiles son la lógica difusa, las operaciones de matemática morfológica, en particular la operación de apertura, la cual está compuesta de una dilatación y una erosión, acumulación de resultados, etc.

2.6 Trabajos Relacionados

Desde hace varios años en la literatura se puede encontrar múltiples propuestas para efectuar la detección de incendios, las cuales basan su operación en la búsqueda de información relacionada con las flamas, el humo, o ambos. Las técnicas para efectuar la búsqueda son muy variadas. En este capítulo, se presentan algunas de las propuestas más representativas en el campo de la detección de incendios.

En [3] se presenta un método para el reconocimiento de flamas que reposa en un modelo que describe las flamas en términos de la variabilidad de su área, de la rugosidad de sus contornos, de la aspereza de la textura y del grado de asimetría de la distribución del componente rojo. Esta propuesta trabaja sobre imágenes en bruto en formato RGB y la fase de preprocesamiento consiste en filtrar todo aquello cuyo color no corresponda con el color de una flama. Una máscara, llamada *Potential Fire Mask* (PFK), utiliza un modelo de color “flama” obtenido a partir de un estudio estadístico de las componentes de color RGB. En la fase de extracción de EURI se obtienen cuatro EURI. Uno que representa la variación del área de la flama de un cuadro a otro:

$$\Delta A_i = \frac{|A_i - A_{i-1}|}{A_i} \quad (2.1)$$

Otro representa la rugosidad de los contornos:

$$B_R = P_S / P_{CHs} \quad (2.2)$$

En donde P_S es el perímetro del área de la flama y P_{CHs} un índice de convexidad según el algoritmo *convexhull* [13]. Uno más evalúa la aspereza de la textura. Aquí se

argumenta que es difícil describir una textura, por lo tanto se recurre al uso de la desviación estándar σ de los píxeles que forman la flama. Finalmente, la asimetría de la distribución γ_R , con respecto al componente de color rojo, es obtenida utilizando la propuesta de [30]. Por otra parte, la fase de clasificación emplea un clasificador de Bayes con el siguiente vector:

$$d = \begin{bmatrix} \Delta A \\ B_R \\ \sigma \\ \gamma_R \end{bmatrix} \quad (2.3)$$

La evaluación del método se realizó utilizando videos con una resolución de 768×576 píxeles y una velocidad de presentación de 25 cuadros por segundo. El porcentaje de falsos positivos y falsos negativos es de aproximadamente 0.68% y 0.028% respectivamente.

En [4] se propone un modelo para detectar flamas y humo por su coloración. Un conjunto de reglas fundadas en el uso de umbrales determinan cuando una muestra es considerada como relacionada con una flama. Este método emplea las siguientes técnicas: para el preprocesamiento efectúa la conversión del espacio de representación RGB a $YCbCr$; los autores argumentan que la intensidad luminosa de las flamas se manifiesta de mejor manera en la componente Y . Ahora, los EURI son píxeles que verifican la siguiente relación:

$$Y(x, y) \geq Cr(x, y) \geq Cb(x, y) \quad (2.4)$$

La clasificación consiste en determinar que un píxel que verifica (4) realmente pertenece a una flama. Para esto se propone la medida $Pf(x, y)$. Esta medida expresa la

pertenencia al rango $[0,1]$ y es un mapeo de la observación establecida en (2.4) a una cantidad que describa la vecindad de un píxel con respecto a un flama. Para evaluar esta vecindad se utiliza lógica difusa utilizando la pertenencia de funciones triangulares y trapezoidales para las diferencias entre $Cr(x, y)$ y $Cb(x, y)$ pero también entre $Y(x, y)$ y $Cb(x, y)$. Está claro que los valores de $Pf(x, y)$ más cercanos a 1.0 son los que se toman en cuenta. La detección de humo resuelve en forma similar, sin embargo, aquí se considera que el humo, y en especial aquel que aparece en las primeras etapas de la combustión, presenta una coloración muy baja. Para detectar la presencia de humo se prefiere trabajar con el modelo de color HVS utilizando la siguiente condición:

$$S(x, y) \leq 0.1 \quad (2.5)$$

El método fue probado para un flujo de video con imágenes de tamaño 320×240 pixeles. La tasa de detección resultante fue de 99.0% y 4.5% de falsas alarmas.

El método presentado en [6] detecta humo gracias al análisis, tanto de los *Vectores de Movimiento* (MV), como de la información de color. Si bien los autores mencionan que la detección puede operar en video H.264 y MPEG II, en realidad no se trabaja con video comprimido. En efecto, el método aquí analizado considera que ciertos datos del proceso de codificación están disponibles a terceros, de tal manera que estos puedan ser reutilizados en tareas distintas a la codificación. Estos datos incluyen los MV y la información de color. Los primeros se encuentran disponibles en forma de los componentes MV_x^{MB} y MV_y^{MB} , mientras que los segundos son tomados directamente de las imágenes sin codificar bajo el formato RGB. La fase de preprocesamiento realiza una depuración y preparación de estos datos. Por parte de los MV, se elimina aquellos de pequeña talla y probablemente ligados al

ruido. Esto se logra aplicando la regla, llamada *Moving Region Decision* (MRD), que a continuación se presenta:

$$MRD^{MB} = \begin{cases} 1, & MV_{\Phi}^{MB} > \tau_1 \\ 0, & \text{otro} \end{cases} \quad (2.6)$$

Donde MV_{Φ}^{MB} es la magnitud del vector y τ_1 es un umbral determinado. En cuanto a la información de color, el preprocesamiento consiste en “diluir” el ruido en los componentes de color RGB de un bloque de $n \times n$ píxeles. Este proceso se efectúa en cada componente en forma independiente gracias a la siguiente expresión:

$$C_{\Omega}^{MB} = \frac{1}{s^2} \sum_{(x,y) \in MB} C_{\Omega}(x,y) \quad (2.7)$$

En donde $\Omega = \{R, G, B\}$ y C_{Ω}^{MB} es la suma de un componente de color en un bloque MB dividido por $s = n \times n$. La fase de extracción de EURI consiste en obtener el ángulo MV_{Ψ}^{MB} del vector MV_{Φ}^{MB} , ya que se considera que, en la mayoría de los casos, el humo presenta un movimiento ascendente. En cuanto a la información de color, considerando que el humo presenta una tonalidad grisácea, se buscarán elementos para determinar que las componentes de color presentan un valor próximo entre ellas. Así, en esta fase se obtiene el valor C_{Ω}^{MB} máximo y mínimo en el bloque. Ahora en la fase de clasificación no se emplearon técnicas sofisticadas. La clasificación es de carácter empírico

y consistió en determinar una serie de reglas con umbrales, las cuales deben ser verificadas antes de anunciar la presencia humo. La primer regla impone una diferencia máxima entre el valor máximo y mínimo de C_{Ω}^{MB} . La segunda regla exige que la intensidad I del bloque se deba encontrar en un rango de valores dado; no se indica cómo obtener I . Para la evaluación del método se usaron imágenes de 320×240 píxeles. Los resultados solo muestran algunas imágenes tomadas de [52] en donde se puede observar las marcas definidas por el método. No se presenta una evaluación exhaustiva del método en términos de la tasa de éxito, falsos positivos, falsos negativos, etc.

En [8] se propone un método de reconocimiento basado en la textura. Los datos iniciales son imágenes en bruto en formato RGB. La fase de preprocesamiento efectúa dos principales tareas. Primero, considerando que la textura depende en gran medida de la resolución de la imagen, todas las imágenes se normalizan a 320×240 píxeles. Segundo, se separan las regiones de primer plano de la imagen de fondo. El uso de un GMM adaptivo permite modelar el fondo de la imagen como una distribución compleja que resulta de la suma de k distribuciones Gaussianas:

$$p(x/\lambda) = \sum_{i=1}^k w_i g(x/\mu_i, \Sigma_i) \quad (2.8)$$

En donde x es un vector multidimensional, en este caso x representa un píxel con 3 componentes de color, w_i es el peso con el que cada distribución Gaussiana contribuye en la mezcla y $g(x | \mu_i, \Sigma_i)$ son los componentes de la mezcla; de hecho, cada uno de éstos es una distribución Gaussiana. Por otra parte, para crear la GMM se toman todos aquellos píxeles

que entre dos imágenes permanezcan estáticos. Como seguramente algunos píxeles se mueven, el GMM va cambiando a lo largo del tiempo. Cuando el modelo alcanza un cierto punto, éste es tomado como referencia durante el análisis de los píxeles de las imágenes subsecuentes para separar todos aquellos que no se encuentran en alguna de las k distribuciones. Así, se declara que estos píxeles pertenecen a las regiones de primer plano. Por otra parte, las imágenes normalizadas se dividen en bloques de 16×16 píxeles para aplicar en ellos los resultados estadísticos obtenidos. Si un bloque presenta más del 80% de píxeles de primer plano, entonces se considera como un bloque probablemente ligado al humo. La fase de extracción de EURI se practica sobre los bloques de primer plano y busca describir la energía, el contraste y la homogeneidad de cada bloque. Para obtener esta descripción se adoptan los algoritmos propuestos en [16] y que son llamados GLCM (*Gray Level Co-occurrences Matrix*).

$$energy = \sum_i \sum_j P^2 [i, j] \quad (2.9)$$

$$contrast = \sum_i \sum_j (i - j)^2 P[i, j] \quad (2.10)$$

$$homogeneity = \sum_i \sum_j \frac{P[i, j]}{1 + |i - j|} \quad (2.11)$$

La clasificación de los EURI se realiza con una red neuronal de tipo *Perceptron multicapa*. Este algoritmo ofrece una tasa verdaderas alarmas de 50% a 75% y solo el 5% de falsas alarmas. Además es capaz de procesar 25 imágenes por segundo, sin contar el proceso de normalización de imágenes a 320×240 píxeles, el cual es definitivamente muy costoso.

El trabajo presentado en [14] está destinado a la detección de humo en bosques. El método considera que la presencia de humo en la escena provoca la pérdida de información en las altas frecuencias. Para determinar la pérdida de información se compara una imagen G_i con respecto a una G_i' . La primera es una imagen que representa una escena cualquier, mientras que la segunda es la misma imagen a la cual se le ha aplicado un filtrado, el cual se intenta sea similar al que provoca la aparición de humo en la escena. Si en la comparación se encuentran regiones que presentan pocas diferencias, entonces se asume que estas regiones están ligadas a la presencia de humo. La fase de preprocesamiento de esta propuesta se desarrolla de la manera siguiente: Primero, las imágenes en formato RGB se normalizan a un tamaño de 704×560 píxeles usando una interpolación bicúbica. La imagen resultante se transforma en una imagen de tonalidades de gris G_i , sobre la cual se aplica una transformada de ondeletas estacionaria; esta ofrece un mejor desempeño para el filtrado. Enseguida, se aplica una transformada inversa pero solo considerando los niveles 3 y 4 obtenidos con la transformada directa. Esto elimina las bandas de información frecuencial más altas y da como resultado una imagen G_i' , la cual presenta una pérdida de información similar a la que provoca el humo. Los EURI de este método entonces son G_i y G_i' . La fase de clasificación establece, por una parte, los umbrales en los que se considera que la pérdida de información de G_i con respecto a G_i' está ligada a la presencia de humo, y por la otra, los umbrales de crecimiento del área del ROI para verificar que una secuencia de ROI corresponde a la presencia de humo en la escena. Es importante mencionar que la propuesta descrita en el artículo carece de especificación formal. En cuanto a los resultados experimentales, el método se probó en solo dos videos dando como resultado una tasa

detección del 100% y 10% de falsas alarmas. Además se estima que el tiempo de detección es de 15 segundos.

En [23] se presenta un método para la detección de flamas, el cual propone analizar tres elementos representativos: i) la coloración, ii) el crecimiento periódico del área y iii) la irregularidad de la forma. Es importante mencionar que si bien el método toma como referencia videos de cuatro diferentes tipos de fuentes de luz, a saber: fuego en papel, fuego en una vela, bulbo rojo y un bulbo incandescente, el método se centra en el caso de fuego en papel. Primero, con respecto a la coloración, la fase de preprocesamiento consiste en convertir las imágenes, del formato RGB al YUV; ambas representaciones son utilizadas en este método. La extracción de los EURIS se efectúa esencialmente a partir de las informaciones de luminancia Y y de la componente de color rojo. Primero se obtiene una descripción de las variaciones de la componente R (Rojo) con respecto a la luminancia Y , tal y como lo establece [33]. Esta descripción se establece como una ecuación cuadrática llamada R_r y como una derivada entre las variaciones de R_r y la luminancia Y . La fase de clasificación establece que el color de la flama puede ser descrito con la siguiente expresión:

$$\left\{ \begin{array}{l} Y_{mean} > 180 \\ R_{mean} \geq G_{mean} \geq B_{mean} \\ R_r > 0.33 \\ -0.007 < \frac{dR_r(x,y)}{dY(x,y)} < 0 \end{array} \right. \quad (2.12)$$

Se observa que la luminosidad de la flama Y_{mean} debe ser superior a 180 (valores de 8 bits). También, la proporción de la componente rojo R_{mean} debe superior a los otros

componentes. Además, la ecuación cuadrática $R_r > 0.33$ para valores de luminancia entre 150 y 240. Finalmente, la pendiente de la derivada debe conformar los valores especificados. En segundo lugar, en lo que respecta el crecimiento periódico del área de la flama. El análisis de las muestras de las diferentes fuentes de calor muestra que solo la combustión del papel presenta un área de flama que crece a medida que el fuego avanza. Esta se describe de la siguiente manera:

$$A_{pi} = \frac{A_{i+1} - A_i}{A_i}, \quad i = 1, 2, 3, 4 \quad (2.13)$$

$$A_{pi} > 10\%, \quad i = 1, 2, 3, 4 \quad (2.14)$$

La (2.13) permite calcular la diferencia entre dos regiones de flama en imágenes contiguas en el tiempo y (2.14) indica que este crecimiento debe ser superior al 10%. Por último, la regularidad de la flama es calculada con:

$$c = 4\% p_i S/L^2 \quad (2.15)$$

En donde c es un índice que indica que tal circular es la flama, S es el área de la flama y L es el perímetro. Durante regular sea S , más pequeña será c . El umbral aceptable establecido es $c > 2.56$. Los resultados de este método muestran una precisión del 94%, una tasa de falsos positivos del 2% y de falsos negativos del 4%.

Por otra parte, [18] propone un método que detecta humo en dos etapas, primero detecta los segmentos de imagen que presenta movimiento, después caracteriza la información espectral (colores), espacial y temporal de los segmentos asociados a movimiento. La primera fase, el preprocesamiento, la constituye una conversión del espacio

de colores RGB a HSI. Esta información es posteriormente utilizada para obtener una correlación espectral del humo. Por otra parte, la extracción de los EURI está representada por la información de movimiento que contiene la escena, la cual es posteriormente utilizada para determinar la forma irregular del humo y el movimiento representativo de la mansa de humo. Entonces, la extracción de la información de movimiento se obtiene por medio del algoritmo *Motion History Image* (MHI) [11], el cual entrega una secuencia de imágenes binarias que indican, en una tonalidad que tiende al blanco, en las regiones de movimiento.

$$H_{\tau}(x, y, t) = \begin{cases} \tau & \text{if } D(x, y, t) = 1 \\ \max(0, H_{\tau}(x, y, t - 1) - 1) & \text{otro} \end{cases} \quad (2.16)$$

La fase de clasificación ataca cuatro caracterizaciones, todas basadas, o en la información de color, o en la información de movimiento. Primeramente, un conjunto de muestras de humo y fugo en formato HSI permiten la obtención de un coeficiente de correlación del histograma de correlación Scorr. En segundo lugar, considerando que el humo es un fenómeno cuya forma es caótica, se procede a medir la irregularidad de la forma a través del tiempo haciendo uso de $\Omega = P_r / A_r$, en donde P_r representa el perímetro de la región en movimiento y A_r representa el área de la región en movimiento. Pero para incluir el tiempo en este análisis se seleccionó el uso de la lógica difusa en combinación con una función *level crossing rate* (tasa de cruzando el nivel cero). En tercer lugar, el algoritmo CAMSHIFT (Continuously Adaptive Mean Shift, algoritmo de segmentación de imágenes de color) es utilizado para seguir las regiones previamente declaradas como

ligadas a humo. Finalmente, el método fue probado en varias secuencias de video a 60 imágenes por segundo. Los resultados muestran una tasa de detección de 77.2%.

La propuesta presentada en [25] detecta humo primeramente separando los contornos en movimiento del fondo de la imagen y probando sobre estos últimos la coloración, el tamaño y la evolución de la forma a lo largo del tiempo. En la fase de preprocesamiento la imagen se convierte del espacio RGB a YUV. Los EURI son bloques de 2×2 que representan movimiento; éste se declara cuando la luminancia de un bloque de la imagen en curso sobrepasa, por Th unidades, la luminancia promedio de las imágenes previas. Este promedio es considerado como la luminancia del fondo. De esta manera se obtiene una lista de bloques que presentan movimiento y sobre la cual se aplican operaciones morfológicas de erosión y dilatación para eliminar los bloques aislados; estos últimos son potencialmente erróneos. La fase de clasificación se realiza caracterizando el área de los grupos de bloques, así como la desviación estándar y la media de la luminancia y las crominancias UV. No se dan estadísticas de la tasa de fiabilidad, ni de la velocidad del método.

[27] presenta una propuesta que detecta incendios gracias al uso de dos sub-métodos. Uno determina la presencia de flamas y el otro busca la presencia de humo. Además, la presencia de un incendio solo es declarada si los dos sub-métodos entregan un resultado positivo. En este método, la fase de preprocesamiento sigue dos pasos; primero se efectúa una conversión del espacio de representación de color RGB a YCbCr. Después se realiza una transformada de ondeletas sobre la componente Y ; las bandas frecuenciales son posteriormente utilizadas para determinar la pérdida de información. Los EURI extraídos por este método son, tanto las informaciones YCbCr, como la información frecuencial

entregada por las ondeletas. Por otra parte, la fase de clasificación la efectúa cada sub-método por separado. Para la detección de flamas se analiza la distribución de información del canal R y se verifica $C_r > C_b$. También se verifica que la inclinación de la flama entre dos cuadros no sobrepase los 30° . Los umbrales que verifican la presencia de flama son todos de carácter experimental. En cuanto a la detección de humo, ésta se basa en el análisis del histograma de tonalidades de gris. También se verifica la pérdida de la cantidad de información del fondo cuando el humo denso se propaga por la escena. Esta pérdida es evaluada obteniendo el coeficiente de correlación de la información entregada por las ondeletas. La detección es una fase que sigue las reglas y umbrales obtenidos en la fase de clasificación. Finalmente, una fase de eliminación de resultados errátiles es practicada. Esto se logra verificando que la unión de algunas regiones se vaya acumulando en un cierto número de cuadros. Las pruebas del algoritmo fueron realizadas con imágenes de reducido tamaño 320×240 y una baja velocidad de presentación, específicamente 5 imágenes por segundo. Además, no se entrega resultados objetivos de la tasa de fiabilidad.

En [28] se presenta un trabajo para detección de humo en espacios abiertos gracias a la detección de movimiento y luego a la caracterización del humo por su coloración. Una verificación posterior de la información de movimiento y frecuencial son caracterizadas para dar mayor certitud al método. La fase de preprocesamiento está constituida por una transformada wavelets que entrega información de orden frecuencial, la cual es utilizada en la fase final de verificación antes citada. Otro preprocesamiento que se realiza concierne la información de color RGB, la cual se normaliza tomando como referencia el canal B .

$$I'_b = I_b / \sqrt{I_G^2 + I_B^2 + I_R^2}, \quad b \in R, G, B \quad (2.17)$$

Los EURI de este método están representados por segmentos de imagen en movimiento, y en particular, sus datos de color normalizados. La localización de estos segmentos se logra combinando un filtro de Kalman y el algoritmo MHI, ver (2.16). Mientras que el filtro detecta movimiento a nivel píxel, el algoritmo MHI trata de ensamblar los segmentos dispersos que constituyen un mismo objeto. La fase de clasificación se realiza utilizando la información de color gracias al modelo de mezclas Gaussianas (GMM). Para que un píxel sea considerado como ligado al humo, éste debe cumplir la siguiente relación:

$$|I'_b - \mu_x| < 2.5\sigma_s \quad 2.18$$

En este punto, la fiabilidad el modelo de color para el humo se considera insuficiente, así se efectúan adicionalmente dos clasificaciones adicionales. Primero, la atenuación de la información en las altas frecuencias de los segmentos en movimiento con respecto al fondo de la escena. Esta atenuación es la división de la potencia de las altas frecuencias de cada segmento en movimiento entre la potencia de lo que se considera la imagen de fondo:

$$\rho = \frac{\sum w_l(n)}{\sum w_b(n)} \quad (2.19)$$

En segundo lugar se evalúa la transparencia de la masa de humo verificando la mezcla de colores del segmento en movimiento con respecto a los colores del fondo de la

escena. Finalmente, en este método aparece una fase de “eliminación de resultados erráticos”, la cual se basa en el uso de operaciones morfológicas. Este método se probó en secuencias de imágenes de 720×480 píxeles. No se menciona, ni el tamaño de las imágenes, ni el número de cuadros por segundo, ni el desempeño del método.

En [31] se propone un método para la detección de humo, el cual efectúa dos principales tareas. Primero, se busca los segmentos de imagen en movimiento. Después, verifica la textura utilizando un modelo obtenido gracias a una SVM. La fase de preprocesamiento opera sobre imágenes en tonalidades de gris y obtiene la diferencia acumulada de las imágenes. Un segundo paso efectúa la binarización de las imágenes sobre la cual se aplica una operación morfológica de cerradura. Esta última con el objeto de eliminar el ruido. Después, las regiones restantes se expanden utilizando los diámetros de Feret. El resultado es la creación de una máscara que se aplica sobre la imagen original en tonalidades de gris. La fase de extracción de EURI es la aplicación de esta máscara sobre las imágenes originales, con el fin de obtener las regiones de imagen potencialmente ligadas a la presencia de humo. La fase de clasificación toma como entrada estas regiones para alimentar una SVM, la cual entrega un modelo de la textura de estas regiones. Por supuesto, la SVM es alimentada con muestras definitivamente ligadas a la presencia de humo. Con el fin de aumentar la fiabilidad del método, se incluye una etapa de eliminación de resultados erráticos. Esta consiste en acumular los resultados positivos entregados por la SVM. La conclusión reporta una tasa de detección de 86.090% para video con 720×640 píxeles.

La propuesta presentada en [40] está diseñada para la detección de humo en bosques y está compuesto de dos principales partes. Primero, la detección de segmentos de imagen

en movimiento con respecto al fondo de la escena. Segundo, la caracterización de características estáticas y dinámicas de estos segmentos. La fase de pre-procesamiento consiste en convertir las imágenes del espacio de representación de color RGB a YUV. Los EURI que se buscan en este método están representados por segmentos de imagen en movimiento. Para esto se emplea la componente Y de las imágenes YUV. El movimiento se puede estimar efectuando la diferencia entre una imagen previa, la cual se considera como el fondo de la imagen – solo la primer vez –, y una imagen actual. El método continúa, solo si al menos un valores de $imDf(t, x, y)$ es diferentes de cero.

$$imDF(t, x, y) = |imBG(t, x, y) - imIN(t, x, y)| \quad (2.20)$$

Enseguida se binariza la imagen “diferencia” utilizando un umbral Th .

$$imBN(t, x, y) = \begin{cases} 1 & \text{if } imDF(t, x, y) \geq Th \\ 0 & \text{otro} \end{cases} \quad (2.21)$$

Después, para eliminar regiones aisladas de pequeño tamaño y probablemente ligadas al ruido, se aplican operaciones morfológicas de erosión y dilatación sobre la imagen $imBN(t, x, y)$. Un paso final de preprocesamiento se efectúa encerrando las áreas en movimiento en un polígono gracias al algoritmo *convexhull* [13], estas áreas forman lo que los autores llaman Regiones de Interés o ROI. Por otra parte, la fase de clasificación analiza y clasifica el color de los ROI por medio de las siguientes expresiones:

$$\alpha_{xy} \} = (R_{xy} + G_{xy} + B_{xy})/2 \quad (2.22)$$

$$c = \begin{cases} \text{gris} & |R_{xy} - \alpha_{xy}| \wedge |G_{xy} - \alpha_{xy}| \wedge |B_{xy} - \alpha_{xy}| < TH \\ \text{no gris} & \text{otro} \end{cases} \quad (2.23)$$

Un ROI es considerado como ligado al humo, si este presenta una tonalidad gris en al menos un 25%, según un umbral que corresponde a información de color RGB. La fase de clasificación es realizando una simple condición que declara la presencia de humo en un ROI si éste es de color gris, el área del ROI ha cambiado y mantiene un color gris por al menos 30 cuadros. La fase experimental se desarrolla sobre videos con imágenes de 640×480 píxeles a 30 cuadros por segundo capturadas bajo la norma JPEG. Esto significa que para procesar el video antes hay que descodificarlo, lo cual implica un importante costo computacional. Además, el algoritmo solo opera sobre imágenes separadas por 1 segundo entre sí. No se indica el desempeño del algoritmo en términos, ni de la fiabilidad, ni de su velocidad o costo computacional.

En [43] se propone un método para detectar humo, el cual se basa en un modelo de la textura de humo y un modelo que describe el movimiento inter imagen. Ambos modelos están basados en los *Local Binary Pattern* (LBP). La etapa de preprocesamiento consiste en separar los segmentos de imagen de primer plano del resto de la escena. El método estadístico GMM [39] es empleado aquí debido a su buen desempeño en este tipo de tareas. Las imágenes son posteriormente divididas en bloques de un mismo tamaño y solo se conservan los bloques que contienen mayoritariamente píxeles de primer plano. En la fase de obtención de los EURI intervienen los LBP. En términos generales, estos métodos

permiten describir la textura de regiones de imagen de radio R por medio de un código de P bits.

$$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{p-1} s(g_p - g_c)2^p \quad (2.24)$$

En donde g_c representa el valor de la tonalidad de gris del píxel central y $g_p (p=0, \dots, P-1)$ representa los valores de tonalidades de gris de P píxeles igualmente espaciados en un círculo de radio R ($R > 0$), $s(x)$ es una función que entrega 1 si x es positivo y 0 si éste es negativo. Por otra parte, dos variantes de éstos son propuestos para soportar el trabajo: una que describe la textura espacial, llamada *Non-Redundant Local Binary Pattern* (NRLBP), y otra que describe el movimiento de regiones entre imágenes. Este descriptor es llamado *Non-Redundant Local Motion Binary Pattern* (NRLMBP). La clasificación se efectúa alimentando una SVM con bloques muestra ligados a la presencia de humo en la escena. La eliminación de resultados errátiles se logra descartando todo aquel bloque que aparece aislado en un área de 5×5 bloques. Los resultados reportados indican, para bloques de 16×16 píxeles, una tasa de detección del 97.67% y falsas alarmas con un 1.165%. No se dan datos del número de cuadros por segundo, ni de la velocidad de procesamiento.

El método propuesto en [46] detecta fuego gracias al uso de redes neuronales basadas en códigos genéticos, para los cuales los autores reportan una mejor precisión con respecto a las soluciones clásicas. La fase de preprocesamiento tiene como objetivo obtener porciones de imagen que presenten movimiento, las cuales se obtienen siguiendo dos pasos: Primero se detectan las regiones en movimiento utilizando información de gris y un método

de media aproximada que permite distinguir el fondo de la imagen, en principio estático, de las regiones de primer plano que se consideran estar en movimiento. En segundo lugar, tomando en cuenta que en la escena pueden existir diversos objetos en movimiento que no son, ni fuego, ni del color del fuego, se procede a eliminar aquellos cuyo color no corresponde con el fuego. El conocido método *Fuzzy C-Means* (FCM) es aquí utilizado en las regiones declaradas con movimiento pero utilizando ahora la información de color, la cual ha sido transformada del espacio RGB al CIE LAB. Según los autores de este documento, este espacio de representación es más preciso para tareas de visión. Un aspecto importante para el Fuzzy C-Means, es el número inicial de aglomeraciones de elementos, en este caso, píxeles. Este trabajo se basa en uno previo [41], que determina este número. La parte de extracción de EURI, entrega tres características de la representación frecuencial de las regiones que corresponde con la coloración del humo. Esta representación se obtiene dividiendo las regiones seleccionadas en bloques de 16×16 y aplicando sobre ellas la transformada por ondeletas discreta. Las tres características son: la media, la desviación estándar y la energía de diferentes subbandas. Así se obtiene un vector de 10 elementos. La fase de clasificación se efectúa con una red neuronal basada en códigos genéticos. El método de reconocimiento fue aplicado a secuencias de imágenes de 320×240 píxeles. Los resultados experimentales muestran una precisión del 94.86% y un error del 2.58%. No se da información sobre la velocidad o complejidad de esta propuesta.

En [15] se aborda un método de detección de humo en videos JPEG. Si bien el método recibe imágenes JPEG, el método en realidad opera sobre imágenes en bruto. En efecto, la fase de preprocesamiento descomprime las imágenes, de las cuales solo se utiliza la componente de color verde. Después se aplican, por una parte, la DCT, y por la otra las

ondeletas. Después de una somera evaluación, se concluye que las ondeletas presentan un mejor desempeño. Los EURI se obtienen a partir de los coeficientes de tres niveles sub-bandas y son de seis diferentes tipos: la media aritmética, la media geométrica, la desviación estándar, la simetría de la distribución (*skewness*), la *kurtosis* y la entropía. La herramienta de clasificación utilizada es una máquina de soporte vectorial. No se dan detalles del proceso de clasificación. Una evaluación del desempeño de este método de reconocimiento es presentada con ayuda del procedimiento de evaluación descrito en [22]. Los resultados indican una precisión del 88.75%.

La propuesta presentada en [7] detecta humo por medio de la tonalidad de este y de su comportamiento dinámico. La fase de preprocesamiento efectúa la conversión de las imágenes originalmente representadas en el espacio RGB al espacio HSI. Después, la fase de extracción de EURIS obtiene las regiones que presentan algún tipo de movimiento. Esto se logra efectuando la diferencia entre imágenes. La fase de clasificación determina las condiciones en las cuales estas regiones están ligadas al humo. Dos condiciones son verificadas: la presencia de coloración grisácea y las propiedades dinámicas de las regiones grisáceas. Estas dos últimas propiedades son el desorden y la tasa de crecimiento. La clasificación de la tonalidad del humo se obtiene por medio de un estudio estadístico y concluye que en el espacio RGB se debe cumplir la condición $R \pm \alpha = G \pm \alpha = B \pm \alpha$, en donde α va de 15 a 20 y en el espacio HSI la condición $L_1 \leq I \leq L_2$ y $D_1 \leq I \leq D_2$, en donde $D_1 = 80$ y $D_2 = 150$ mientras que $L_1=150$ y $L_2=220$. En cuanto a la clasificación del desorden, este se determina con la expresión siguiente:

$$d = \begin{cases} \text{verdad} & SEP/STP \geq STD \\ \text{falso} & \text{otro} \end{cases} \quad (2.25)$$

En donde SEP es la suma de circunferencias asociadas a cada región de humo potencial, STP es el número de píxeles en la región es un umbral de desorden. Ahora, la tasa de crecimiento del área del humo es evaluado como el promedio de la siguiente expresión:

$$\overline{\Delta A_{dt}} = \frac{1}{n} \sum_{i=1}^n \frac{dA}{dt} \quad (2.26)$$

Los resultados experimentales no presentan la tasa de detección, ni las dimensiones de las imágenes utilizadas, ni el desempeño en términos computacionales.

En [45] se presenta un método para la detección de humo, el cual está basado en la detección de la pérdida frecuencial y la atenuación de la información de color. El método utiliza imágenes 160×120 píxeles en formato RGB. La fase de preprocesamiento realiza tres acciones diferentes: i) transforma la imagen RGB en representación YUV, ii) extrae a partir de Y las regiones en movimiento utilizando el algoritmo descrito en [9] y iii) transforma por ondeletas la información Y. Los EURI que extrae son: a) la energía acumulada en los bloques R_i de tamaño $K1 \times K2$ que provienen de una subimagen obtenida por medio de una transformada de ondeletas:

$$e_{(l1,l2)} = \sum_{(x,y) \in R_i} w_n(x + l1K1, y + l2K2) \quad (2.27)$$

En donde x e y representan la esquina superior izquierda del bloque R_i y w_n corresponde con un coeficiente frecuencial de una subimagen. Por otra parte, considerando que el humo afecta considerablemente la nitidez de los contornos, se busca los R_i que contengan contornos para evaluar la pérdida de energía $e(l1, l2)$ en el bloque R_i a instantes

diferentes. Para la detección de los contornos se recurre al cálculo llamado “local extrema” [5] , la cual es una forma de detectar contornos gracias a la presencia de “picos” de información en un conjunto de coeficientes frecuenciales, en este caso, R_i ; b) la intensidad de la información de color presente en las componentes U y V es otro EURI que ofrece información importante; c) el parpadeo de la luminancia Y en los bordes de la masa de humo. La fase de clasificación se basa en reglas y umbrales experimentales para el caso de la pérdida frecuencial y para la atenuación de la información de crominancia. El caso del parpadeo del entorno de la masa de humo se funda en el uso de un modelo de Markov de tres estados. Una alarma solo se lanza cuando estos tres elementos asociados al humo están presentes. Por otra parte, un cierto número de falsas alarmas se emitía en escenas nocturnas con presencia de faros de automóviles, las cuales fueron eliminadas por medio de análisis del entorno. En particular se observa que una fuente luminosa en la noche no provoca pérdida frecuencial. El método se probó con videos de 160×120 píxeles y el tiempo de procesamiento por trama es de 5 ms. El método detectó humo en todos los casos y emitió cero falsas alarmas.

Partiendo de este estudio del estado del conocimiento se pueden efectuar las siguientes observaciones:

1. La mayor parte de métodos propuestos se funda en el uso de más de un modelo para detectar, ya sea humo o flamas.

2. A pesar que en muchos trabajos se menciona el uso de sistemas de video-vigilancia, estos trabajos consideran poco las características y capacidades de éstos. Mientras que los sistemas de video-vigilancia convencionales generan video comprimido,

en muchos casos de alta resolución, la gran mayoría de trabajos propuestos solo operan con video en bruto o no comprimido con imágenes de pequeña talla.

3. La gran mayoría de métodos de detección de incendios no pueden ser directamente aplicados a los sistemas de video-vigilancia convencionales. Por una parte, los sistemas de video-vigilancia convencionales emplean intensivamente algoritmos de compresión de video, los cuales basan su desempeño en la destrucción controlada de información, tales como el submuestreo de la información de color y el proceso de cuantificación. Por la otra, muchos de los métodos de detección no consideran la existencia pérdidas de información. Algunos no tolerarían la pérdida de información, ya que basan su operación en un análisis fino de las texturas [3] [8] [43] o en el seguimiento de píxeles entre tramas para segmentar las regiones en movimiento [18] [25] [28] [40]. Otros, en cambio, asocian la pérdida de información con la presencia de humo [2] [14] [15][17] [19] [35] [37] [38] [45] [55] [56] en la escena. Por otra parte, la gran mayoría de los métodos utiliza la información de color para detectar incendios, sin embargo ellos no toman en cuenta que la gran mayoría de sistemas de video-vigilancia submuestra esta información. La pregunta que aquí se realiza es: ¿Cómo afecta esta pérdida de información a todos estos métodos? Hacer cohabitar los métodos propuestos con las capacidades de los sistemas de video-vigilancia actuales requiere una revisión de muchos de los métodos de detección propuestos en la bibliografía.

4. En el contexto de sistemas convencionales para la video-vigilancia, muchos de los métodos de detección analizados difícilmente pueden operar en tiempo real debido a que los flujos de video se encuentran comprimidos. Recordar que la gran mayoría de métodos de detección de incendios trabajan con píxeles, y por lo tanto, antes de comenzar su trabajo, ellos deben decodificar el video. Este primer paso, que constituye una fase de preprocesamiento, es de un costo elevado. En [26] se presenta un estudio que muestra el uso del CPU durante el proceso de decodificación de diferentes tipos de video. Aquí abajo se puede observar el desempeño de diferentes sistemas basados en procesadores AMD Athlon de 64 bits frente a la descompresión de video H.264 de alta definición, 1920×1080 píxeles y 25 cuadros por segundo. Este estudio incluye la carga del CPU para efectuar el proceso de presentación en pantalla y si el sistema, Windows, está dotado con DirectX Video Acceleration (DXVA).

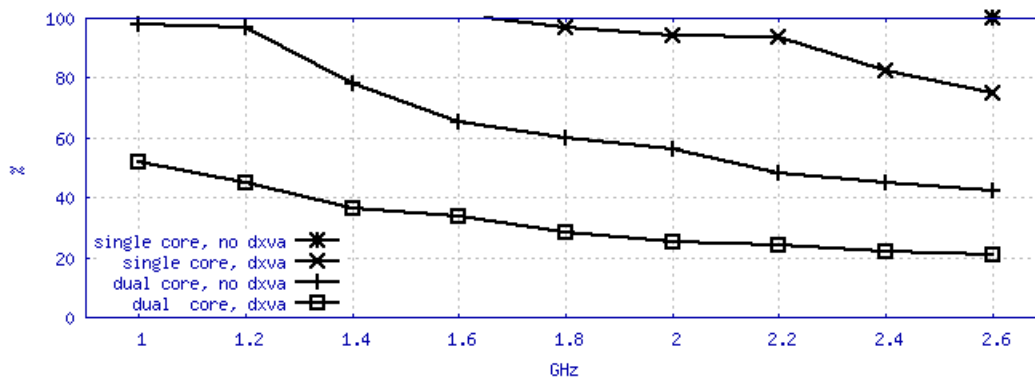


Figura 2.2 Uso del Procesador cuando se decodifica video H.264

Se observa que para un sistema con un solo núcleo y sin DXVA, la decodificación de video de alta resolución consume el 100% de un procesador a 2.6 Ghz. Solo en procesadores potentes el procesador puede efectuar tareas adicionales. Entonces, un método

que detecta incendios en video no comprimido debe tomar ciertas consideraciones cuando se quiera aplicar a sistemas de video vigilancia convencionales actuales.

5. Los algoritmos de detección de humo y flama superan el 90% de éxito. Contar con algoritmos que lleguen al 100% de éxito parece improbable en este momento.

Finalmente una tabla comparativa conteniendo información de todos estos trabajos es presentada en la tabla 2.1. Hay que aclarar que en la columna llamada desempeño, se utilizó las siglas que aparecen en los artículos correspondientes. Para la publicación [3], FP significa False Positive Rate y FN False Negative Rate. En [4] FDR significa Fire Detection Rate y FAR False Alarm Rate. En [8], TP corresponde con True Positive Rate, mientras que FP con False Positive Rate. En [14], mientras que DR significa Detection Rate, FAR significa False Alarm Rate. En [15] y [20], ACC significa accuracy. En [18], DR significa Detection Rate. En [23], ACC corresponde con accuracy, FP con False Positive y FN con False Negative. En [43], DR, FAR y MR significan Detection Rate, False Alarm Rate y Missed Detection Rate correspondientemente. En [46] DR abrevia la palabra *Detection Rate* y ERR, *error* de detección. Por otra parte, hay que notar que no todos los trabajos presentan sus resultados con una evaluación de desempeño y otros tampoco ofrecen una idea de la velocidad de sus algoritmos.

Tabla 2.1 Puntos esenciales de las propuestas para la detección de incendios

Detecta	Pre-procesamiento	Extracción EURI	Clasificador	supresión de resultados Errátiles	Desempeño	Cantidad de datos procesados
flama [3]	- extrae los píxeles de la flama por su color rojo	- variación área flama - rugosidad contornos - textura flama - "skewness" de la distribución del rojo.	Bayes		FP: 0.68% FN: 0.028%	- 768×576 píxeles > 25 ips

Detecta	Pre-procesamiento	Extracción EURI	Clasificador	supresión de resultados Errátiles	Desempeño	Cantidad de datos procesados
flama y humo [4]	- conversión RGB a YCbCr (para flama) - conversión RGB a HSV (para humo)	- píxeles con ciertos Y, Cb y Cr (flama) - saturación menor a un umbral (flama)	Lógica difusa		- FDR: 99.0% - FAR: 4.5%	- 320×240 píxeles
Humo [6]	- filtra los vectores de movimiento ligados al ruido. - diluye el ruido de la imagen RGB.	- ángulo vector de movimiento - componente color máximo y mínimo en bloque $n \times n$ píxeles	- conjunto de reglas con umbrales experimentales		- sin datos	- 320×240 píxeles
Humo [8]	- normaliza imagen RGB a 320x240 píxeles - extrae regiones de primer plano. - divide la imagen en bloques 16x16 píxeles.	- obtención de energía, contraste y homogeneidad en bloques [34].	- Red Neuronal Back-Propagation.		TP > 50 % FP < 5%	- 320×240 píxeles - 25 ips - detection delay: 1 seg
Humo [14]	-normaliza la imagen RGB a 704x560 píxeles. - Conversión RGB a Y - transformada por ondeletas	- Información de los niveles 3 y 4 de las ondeletas.	- conjunto de reglas con umbrales experimentales		DR: 55% (media) FAR: 3.75% (media)	- 704×560 píxeles - detection delay: 15 seg.
Humo [15]	-Decodificación JPEG.. -Transforma por ondeletas la componente G.	- media aritmética, media geométrica, desviación estándar y la simetría de la distribución (skewness), la kurtosis y la entropía.	- Máquina de Soporte Vectorial.		ACC: 88.75% (Wavelet) ACC:76.16 (DCT). Sensitivity: 0.9 Specificity: 0.89.	- detection delay: 0.98 seg.
flama y humo [18]	- conversión RGB a HSI - extrae regiones en movimiento con MHI [36].	- coeficiente de correlación de la distribución de color para humo y flama. - índice de irregularidad de área de la flama	- lógica difusa para el color e índice de irregularidad. - camshift en información de movimiento.		- DR: 77.2%	- 320×240 píxeles - 11 a 60 ips.
Humo [19]	- conversión RGB a YUV - extrae regiones en movimiento de Y [46] - Transforma componente Y por ondeletas	- Energía frecuencial por bloque - información color U y V	- conjunto de reglas con umbrales experimentales para pérdida frecuencial y atenuación de colores - modelo de Markov aleatorio para detección por	- análisis del contorno.	- Sin datos	- 160×120 píxeles - detection delay: 5 ms.

Detecta	Pre-procesamiento	Extracción EURI	Clasificador	supresión de resultados Errátiles	Desempeño	Cantidad de datos procesados
			parpadeo.			
Humo [20]	- obtención de coeficientes de la DCT por decodificación entrópica de imágenes JPEG. [21][22]	- Coeficientes DC y DC inter-transformado de Y, Cr y Cb [21][22] - Persistencia de la "mancha" de humo [21][22]	- conjunto de reglas con umbrales experimentales	- Operaciones morfológicas: erosión y dilatación.	ACC: 0.97 Sensitivity: 0.96 Specifity: 0.98	-1920×1080 píxeles
Flama [23]	- conversión RGB a YUV	- media de R, G, B e Y de un píxel. - obtención de la curva Y contra R. - obtención del área de una flama. - crecimiento del área de la flama - irregularidad del área de la flama.	- conjunto de reglas con umbrales experimentales		- ACC: 94% - FP: 2% - FN: 4%	- sin datos
Humo [25]	- conversión RGB a YUV - extrae regiones de primer plano - fragmenta las regiones en bloques 2×2 píxeles	- bloques 2×2 píxeles que presentan movimiento. - media y desviación estándar de Y, U y V.	- conjunto de reglas con umbrales experimentales	- Operación morfológica: erosión y dilatación	- sin datos	- sin datos
Flama y humo [27]	- conversión RGB a YCbCr. - transforma componente Y por ondeletas.	- Para flama, distribución de Y cuando $C_b < C_r$ - inclinación de la flama - Para humo, distribución de Y. - Coeficiente de correlación de ciertas subbandas Y.	- conjunto de reglas con umbrales experimentales	- adaptación del sistema a la luminosidad del fondo de la escena.	- sin datos	- 320×240 píxeles - 5 ips
Humo [28]	- normalización de la información del canal B (Blue). - Transforma B por Ondeletas	- regiones en movimiento con MHI [36] y filtro de Kalman. - información frecuencial de B.	- GMM de canal B normalizado. - modelo de atenuación de las altas frecuencias.	- análisis de las variaciones de energía. - análisis de la mezcla de coeficientes de color.	- Sin datos	- 720×480 píxeles
Humo [31]	-conversión de RGB a Y -segmentación de regiones considerando el movimiento; diámetro de Feret.	- textura de regiones en movimiento	- Máquina de Soporte Vectorial	-acumulación de resultados de la SVM.	Cross Validation Rate:86.09%	- 720×640 píxeles
humo [40]	- Descompresión JPEG - Conversión de	- Regiones con movimiento en color RGB.	- conjunto de reglas con umbrales	- operaciones morfológicas: erosión y	- Sin datos	- 640×480 píxeles

ESTADO DEL ARTE

Detecta	Pre-procesamiento	Extracción EURI	Clasificador	supresión de resultados Errátiles	Desempeño	Cantidad de datos procesados
	RGB a YUV. - extrae regiones en movimiento por diferencia de tramas Y.		experimentales	dilatación.		- 1 ips
Humo [43]	- segmenta la imagen en bloques de 16x16 píxeles - elimina los bloques que no son de primer plano usando un GMM.	- códigos LBP que describen la textura de los bloques	- Máquina de Soporte Vectorial.	- erosión morfológica	DR: 97.69% FAR: 1.165% MR: 1.622%	- 320×240 píxeles
Flama [46]	- Conversión de RGB a CIE LAB. - extrae regiones con movimiento. - conserva regiones con color "flama" usando Fuzzy C-means. - divide regiones en bloques de 16x16 píxeles. - transforma los bloques por ondeletas.	- Media, desviación estándar y energía de ciertas subbandas frecuenciales.	- Red Neuronal basada en códigos genéticos.		DR: 94.86% ERR: 2.58%.	- 320×240 píxeles.

3 Marco Teórico

Una trama JPEG, no contiene píxeles, sino información de orden frecuencial organizada en bloques de 8x8 elementos y comprimida, en la gran mayoría de los casos, con el método de Huffman. La información contenida en estos bloques es generada a partir de los píxeles de la imagen por medio de una transformación bidimensional de coseno discreta (DCT por sus siglas en inglés). La información de la imagen, estando representada en el dominio de la frecuencia, facilita las tareas relacionadas con el procesamiento de señales, tales como el filtrado, la detección de ciertos eventos, la caracterización de propiedades, etc. Para el caso de la detección de humo grisáceo claro, el cual es nuestro objetivo, su ocurrencia en la escena se manifiesta de distintas maneras. Entre otros, los coeficientes frecuenciales que representan el color sufren una reducción en su magnitud, igual que los coeficientes que representan el contraste de la imagen. En cambio, el coeficiente que representa el brillo se incrementa. El conjunto de coeficientes que representa, las texturas, los contornos y la información más fina, no solo se atenúa sino que también se pierde. En este capítulo se presenta brevemente la transformada DCT para luego pasar a analizar la norma JPEG. Por último se presenta una primera caracterización empírica del impacto del humo en los coeficientes frecuenciales de la DCT. Adicionalmente, se presenta la caracterización de otros eventos similares que hay que evitar confundir con humo. En particular, la aparición de sombras y los cambios de iluminación.

3.1 La transformada de coseno discreta

La transformada DCT es la piedra angular de diversas normas para la compresión de imágenes y video tales como MPEG, H.261, H.263 y H.264. Su uso está motivada por las propiedades que ella posee, entre ellas, las más comúnmente citadas para la compresión de imágenes son [24]:

a) La descorrelación. Esta propiedad remueve la redundancia entre píxeles vecinos. En efecto, los coeficientes frecuenciales obtenidos a partir de un proceso de transformación están descorrelacionados y son en cierta forma independientes.

b) La compactación de la energía. Esta propiedad concentra la información en un número mínimo de coeficientes. Esto permite a un proceso de cuantificación descartar los coeficientes con relativa pequeña amplitud sin que esto provoque distorsiones significativas en la imagen reconstruida. La DCT ofrece excelentes concentración de la energía en imágenes altamente correlacionadas.

Además, aunque en este campo existen otras transformadas consideradas como próximas al ideal, tal como la transformada *Karhunen-Loeve* (KLT), se prefiere el uso de la DCT en un marco aplicativo debido a su bajo costo computacional, el cual se refleja en implementaciones más económicas y rápidas. Este último aspecto es uno de los principales requerimientos de la norma JPEG [50].

Por otra parte, la transformada DCT representa una imagen como una suma de senoidales que presentan diversas magnitudes y frecuencias. Para una porción de imagen f , los coeficientes frecuenciales DCT de la imagen transformada F son obtenidos gracias a la

(3.1). En esta ecuación, f es una imagen de $M \times N$ píxeles y $f(x, y)$ es la intensidad del píxel en el renglón x y la columna y de la imagen.

En cuanto a $F(u, v)$, éste es un coeficiente DCT en el renglón u y la columna v de la matriz de coeficientes DCT

$$F(u, v) = \frac{1}{4} C(u) C(v) \left[\sum_{x=0}^7 \sum_{y=0}^7 f(x, y) * \cos \frac{(2x+1)u\pi}{16} \cos \frac{(2y+1)v\pi}{16} \right] \quad (3.1)$$

donde:

$$C(u), C(v) = 1/\sqrt{2} \quad \text{Para } u, v = 0;$$

$$C(u), C(v) = 1 \quad \text{otro}$$

La imagen es reconstruida aplicando una operación DCT inversa de acuerdo a la siguiente ecuación:

$$f(x, y) = \frac{1}{4} \left[\sum_{u=0}^7 \sum_{v=0}^7 C(u) C(v) F(u, v) * \cos \frac{(2x+1)u\pi}{16} \cos \frac{(2y+1)v\pi}{16} \right] \quad (3.2)$$

Se observa que el tamaño de los bloques sobre los cuales opera la transformada DCT es de 8×8 elementos. En diversas normas de compresión de imágenes y video, se ha estimado que este tamaño de bloque es el que ofrece la mejor relación entre el costo computacional y utilidad de la transformada. Entonces, en JPEG, para aplicarla a una imagen de tamaño mayor a 8×8 píxeles, ésta se debe fragmentar en bloques de 8×8 que no se traslapen entre sí y solo después puede ser aplicada la transformada. Por otra parte, en los bloques DCT resultantes se puede distinguir diversos tipos de información [44], a saber: información de brillo, la cual está asociada al coeficiente DC, información de baja

frecuencia relacionada con la textura, información de alta frecuencia que describe la presencia de contornos, ver figura 3.1.

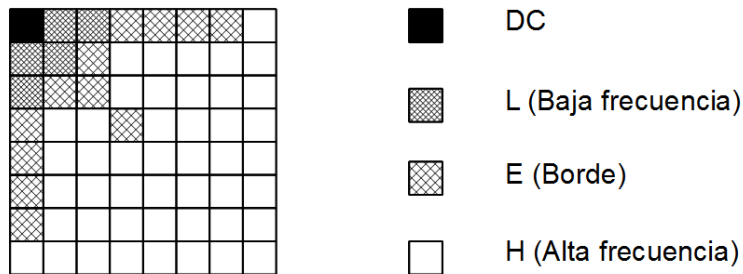


Figura 3.1 Clasificación de áreas indicativas en un bloque DCT

Esta característica es útil tanto para los algoritmos de compresión, como para los algoritmos de detección de eventos y clasificación de propiedades de la imagen. Por último, en un campo aplicativo, el bloque de coeficientes frecuenciales de la DCT puede interpretarse en forma lineal siguiendo un orden en ZIG-ZAG como el que se muestra en figura 3.2. Además, el primer coeficiente recibe el nombre de DC (Corriente Directa) mientras que para el resto de los coeficientes se utiliza el término AC (Corriente Alterna). El término DC denomina un elemento que no representa una oscilación, mientras que AC denota una cierta frecuencia de oscilación.

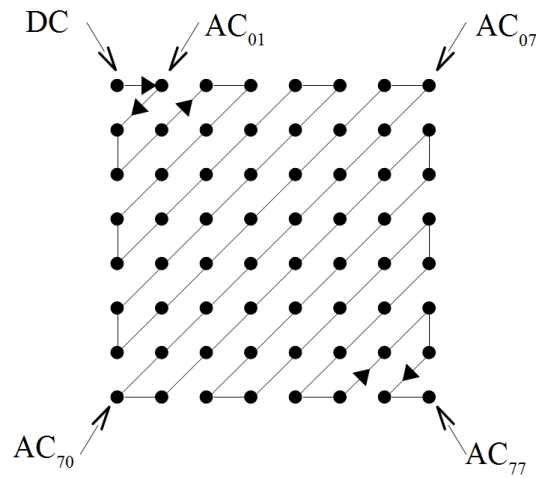


Figura 3.2 Orden de los coeficientes de la DCT

En el marco de esta tesis, por conveniencia, se define el bloque de la DCT y sus componentes de la siguiente manera:

Sea un componente de luminancia o crominancia de una imagen JPEG, el cual puede representarse como una matriz I de $m \times n$ bloques.

$$I = \left\{ \begin{array}{cccc} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & & & A_{2n} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ A_{m1} & A_{m2} & \dots & A_{mn} \end{array} \right\} \quad (3.3)$$

con $i=1,2,3,\dots,m$ y $j = 1,2,3,\dots,n$

Para el arreglo I , cada bloque A_{ij} representa una sub-matriz con 8×8 elementos de tipo coeficiente DCT, los cuales están etiquetados con sub-índices que siguen un orden en ZIG-ZAG, de tal forma que cada bloque puede ser representado como:

$$A_{ij} = \begin{pmatrix} a_0 \rightarrow a_1 & a_5 \rightarrow a_6 & a_{14} \rightarrow a_{15} & a_{27} \rightarrow a_{28} \\ a_2 \rightarrow a_4 & a_7 \rightarrow a_{13} & a_{16} \rightarrow a_{26} & a_{29} \rightarrow a_{42} \\ a_3 \rightarrow a_8 & a_{12} \rightarrow a_{17} & a_{25} \rightarrow a_{30} & a_{41} \rightarrow a_{43} \\ a_9 \rightarrow a_{11} & a_{18} \rightarrow a_{24} & a_{31} \rightarrow a_{40} & a_{44} \rightarrow a_{53} \\ a_{10} \rightarrow a_{19} & a_{23} \rightarrow a_{32} & a_{39} \rightarrow a_{45} & a_{52} \rightarrow a_{54} \\ a_{20} \rightarrow a_{22} & a_{33} \rightarrow a_{38} & a_{46} \rightarrow a_{51} & a_{55} \rightarrow a_{60} \\ a_{21} \rightarrow a_{34} & a_{37} \rightarrow a_{47} & a_{50} \rightarrow a_{56} & a_{59} \rightarrow a_{61} \\ a_{35} \rightarrow a_{36} & a_{48} \rightarrow a_{49} & a_{57} \rightarrow a_{58} & a_{62} \rightarrow a_{63} \end{pmatrix} = [a_r]_{ij} \quad (3.4)$$

con $r = 0, 1, 2, 3, \dots, 63$

Así, un arreglo I que contiene información, o de luminancia, o de crominancia, puede ser considerada como teniendo la forma: $I = [[a_r]]_{ij}$ con $i = 1, 2, 3, \dots, m; j = 1, 2, 3, \dots, n$ y $r = 0, 1, 2, \dots, 63$, donde el subíndice r sigue la progresión en ZIG-ZAG antes indicada.

3.2 La norma Joint Picture Expert Group

La norma de compresión de imágenes JPEG (*Joint Photographic Experts Group*) es la primera norma internacional para la compresión de imágenes de tonos continuos, tanto a color, como en blanco y negro. Para cubrir las necesidades de compresión de imágenes de la época, JPEG integra dos métodos básicos de compresión. Uno basado en la DCT, catalogado como “destructivo” y el otro, fundado en métodos predictivos, que no introduce pérdidas de información; cada uno de estos métodos incluye diferentes modos de operación. En todo caso, los métodos basados en la DCT permiten obtener la mejor relación entre la tasa de compresión y la calidad psicofísica de las imágenes comprimidas. En el marco de este trabajo de tesis solo se aborda el modo de compresión conocido como

“Baseline”, ya que es el más ampliamente utilizado. En efecto, este modo de compresión se encuentra disponible en la gran mayoría de cámaras fotográficas, en teléfonos celulares, y en sistemas de video vigilancia. Para estos últimos, las normas más ampliamente disponibles en los equipos modernos son el video comprimido con JPEG y H.264. El primero persiste en estos sistemas por la calidad de imagen que éste ofrece, mientras que el segundo es preferido por su alto desempeño en términos de tasa de compresión.

3.3 El proceso de compresión y descompresión paso a paso

Para explicar el funcionamiento de un codec (*coding-decoding*) JPEG, se toma, por su simplicidad, el caso de una imagen en tonalidades de gris. El caso de la compresión de imágenes a color varía poco, solo es cuestión de codificar por separado cada uno de los componentes de color y tomar en cuenta que cierta parte de la información puede estar submuestreada. La norma JPEG integra una infraestructura logística para considerar estos aspectos.

La codificación de imágenes JPEG comienza con la fragmentación de la imagen de entrada en bloques de 8×8 píxeles, los cuales se presentan al proceso de transformación DCT, ver punto a) en la figura 3.3. La transformada hace su trabajo descorrelacionando la información redundante y concentrando el grueso de la información en los primeros coeficientes del bloque DCT de 8×8 , ver punto b) figura 3.3.

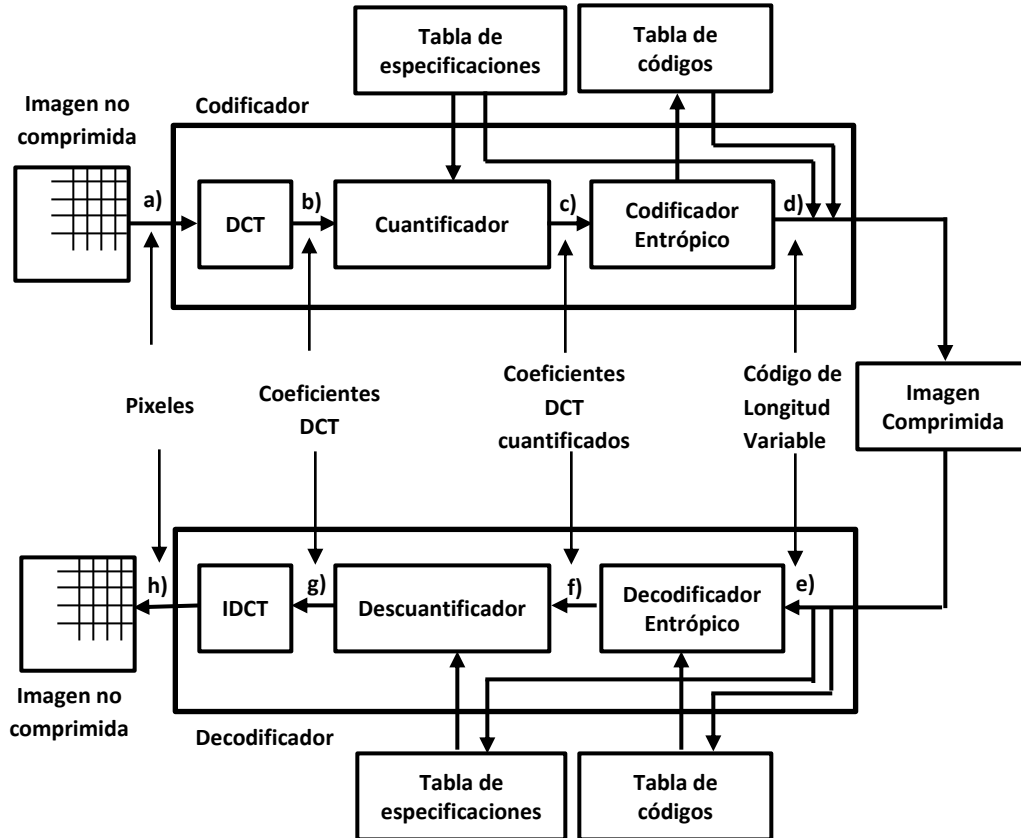


Figura 3.3 Representación de un códec JPG

Para observar más de cerca el trabajo que desarrolla una transformada DCT, se ha presentado con gráficas de barras, en la figura 3.4, un bloque de píxeles y su correspondiente transformación DCT. Observar que los coeficientes AC que circundan el coeficiente DC presentan magnitudes superiores al resto de los coeficientes; esta aglomeración de magnitudes se le denomina concentración de la información y es una de las propiedades que motivaron el uso de la DCT en la norma JPEG.

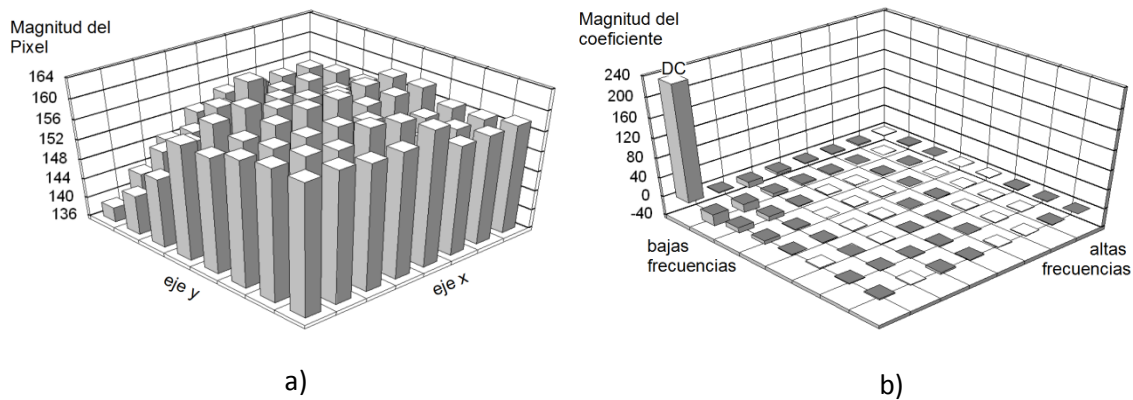


Figura 3.4 Bloque de pixeles y su correspondiente bloque de coeficientes DCT

Es importante mencionar que el proceso de transformación no efectúa por si solo la compresión de datos. En efecto, el número de datos de 8 bits que entra a la transformada es el mismo número que sale, ver figura 3.4. La acción de comprimir comienza con la eliminación controlada de información, la cual es efectuada por el proceso de cuantificación, el cual se define como la división de cada coeficiente $F(u, v)$ del bloque por su correspondiente cuantificador $Q(u, v)$; estos últimos se encuentran contenidos en una tabla de cuantificación Q de 8×8 elementos.

$$F^Q(u, v) = \left[\frac{F(u, v)}{Q(u, v)} \right] \quad (3.5)$$

donde $[]$ denota la operación de redondeo al número entero.

La figura 3.5a, se presenta la tabla de cuantificación Q que han sido aplicada al bloque de coeficientes frecuenciales presentado en la figura 3.4b. El bloque cuantificado F^Q resultante es presentado en la figura 3.5b. Con respecto al esquema general de compresión y descompresión, figura 3.3, el bloque F^Q se encuentra en el punto c).

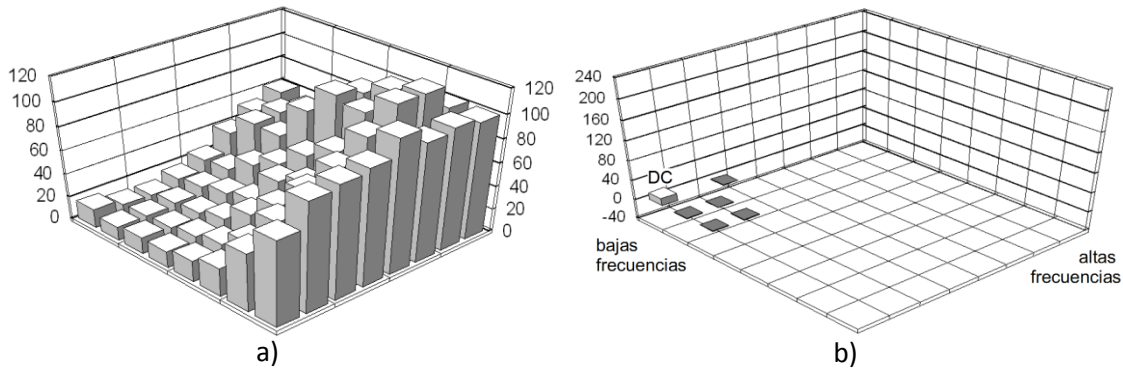


Figura 3.5 Tabla de cuantificación (a) y un bloque de coeficientes resultante del proceso de cuantificación (b)

El trabajo de cuantificación puede ser más o menos agresivo, todo depende de la relación entre la tasa de compresión deseada y la degradación de las imágenes que el usuario final desea obtener. De hecho, el comité desarrollador de la norma JPEG impuso como requerimiento contar con un codificador parametrizable capaz de responder a esta necesidad. Normalmente, para que un codificador pueda ofrecer una tasa de compresión variable, el codificador debe aplicar una tabla de cuantificación que corresponda a la tasa de compresión que se desea obtener. Contar con múltiples tablas de cuantificación es una solución, de hecho la norma [50] ofrece un par de ellas las cuales se pueden ajustar en forma empírica a las necesidades requeridas por el usuario. Por otra parte, los bloques de coeficientes cuantificados deben pasar aun por dos métodos de compresión, específicamente, por el método diferencial Runlength y por el método de codificación entrópica de Huffman, el cual permite presentar los elementos del conjunto de datos, con códigos de longitud variable (VLC), lo que permite obtener una compresión considerable dependiendo de la estadística del conjunto de datos. Existe un método alternativo, poco utilizado, llamado Aritmético que realiza una función similar. Es importante mencionar que tanto el método diferencial como el entrópico son de carácter no destructivo; una explicación detallada de éstos sobrepasa los objetivos de este documento. Finalmente los

datos comprimidos son dispuestos en un flujo binario a sintaxis normalizada. Ver figura 3.3, punto d). En cuanto al proceso de decodificación, cada uno de los procesos descritos aquí arriba posee su propio proceso inverso. La descompresión de Huffman transforma los coeficientes de la DCT representados con VLC en *Códigos de Longitud Fija* (FLC), ver figura 3.3, punto f). La descuantificación se logra con la siguiente expresión:

$$F^{Q'}(u, v) = F^Q(u, v) * Q(u, v) \quad (3.6)$$

Los bloques descuantificados se encuentran en el punto g) de la figura 3.3. Y la transformada inversa (IDCT) descrita en la sección anterior, permite reconstruir la imagen original, figura 3.3, punto h).

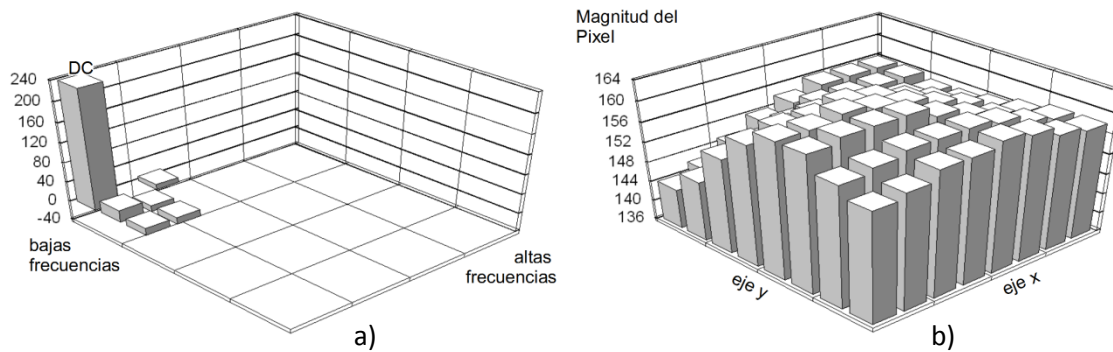


Figura 3.6 Coeficientes DCT descuantificados (a) e inversamente transformados por la DCT (b)

Finalmente, se observa que el bloque original, figura 3.4a y el bloque reconstruido por la DCT durante la decodificación JPEG son muy similares, ver figura 3.6b. Para el ejemplo que se utiliza en esta sección, si bien en estas gráficas de barras se puede notar una pequeña diferencia, cuando las intensidades de los píxeles se muestran en una pantalla o en una impresión, es difícil encontrar la diferencia.

3.4 Impacto del humo en los coeficientes de la DCT

El humo que aparece en las primeras etapas de la combustión [18] es percibido por el sistema visual humano como un fluido a densidad fluctuante y translucidez variable, cuyo color se aproxima al blanco con una ligera tonalidad azulada. Desde la perspectiva del sistema visual humano (HVS), este tipo de humo provoca cambios en diversas propiedades de la escena, tal como el brillo, el contraste, la nitidez y la crominancia. En el marco de esta tesis se considera que el brillo es la intensidad luminosa o luminancia de la imagen, mientras que el contraste es la diferencia notable entre las regiones más claras y más oscuras de la imagen. Para la crominancia, se asume que ésta es la información de color que irradia un objeto sin considerar el brillo. Ésta integra dos componentes el matiz y la saturación. Es importante mencionar que la información completa de la imagen está dada por la crominancia y el brillo. En cuanto a la definición, ésta es la claridad y precisión con la que se presentan los detalles finos de la imagen, tales como los contornos y las texturas.

En esta sección se presenta un estudio que busca comprender la naturaleza del impacto que el humo provoca en los coeficientes de la DCT de una imagen comprimida bajo la norma JPEG. Es importante mencionar que, si bien el estudio presentado tiene un carácter empírico, éste representa un elemento útil para comprender el problema de reconocer humo en video JPEG.

Para desarrollar este estudio se emplea un marco controlado de experimentación, el cual no incluye, ni todas las clases de humo, ni todas las condiciones posibles, así, el marco considerado es el siguiente:

- a) Solo se considera humo translucido grisáceo-azulado.
- b) Uso de cámaras fijas.
- c) Buena iluminación, solar, o equivalente. Desde que aparecen los primeros rayos del sol hasta su ocaso.
- d) Video “Full HD” codificado con JPEG baseline; 1920x1080 píxeles@30 fps.
- e) Tasa de compresión de 1.5 a 2.0 bits/píxel.
- f) Distancia entre la cámara y la fuente de combustión de hasta 100 metros.

El procedimiento seguido en el presente estudio consiste, primero, en adquirir diversas secuencias de video mostrando escenas con humo. La adquisición se realizó gracias a una cámara IP AXIS Q1755 en modo captura de video JPEG. En segundo lugar, de estas secuencias de video se seleccionan diversos pares de imágenes que presentan como el humo va cubriendo la escena. Las imágenes que componen este par son identificadas con I e I' , en donde I es tomada a un instante t_i , mientras que I' a un instante $t_i + \Delta$. Se considera que un retraso Δ de aproximadamente 500 ms es suficiente para observar con claridad el desplazamiento del humo a través de la escena a una distancia de observación de hasta 100 m; para distancias mayores Δ debe ser mayor. En tercer lugar, para estudiar el efecto que provoca el humo en los coeficientes de la DCT, se debe trabajar en el ámbito del bloque. Así, se seleccionan aquellos bloques A_{ij} que no presenten humo y sus correspondientes A'_{ij} con humo; los primeros pertenecen a la imagen I , mientras que los segundos a la imagen I' . Además, se utiliza el término “correspondencia” para indicar una correspondencia espacial entre el bloque A_{ij} y el bloque A'_{ij} con respecto a la imagen I e I' .

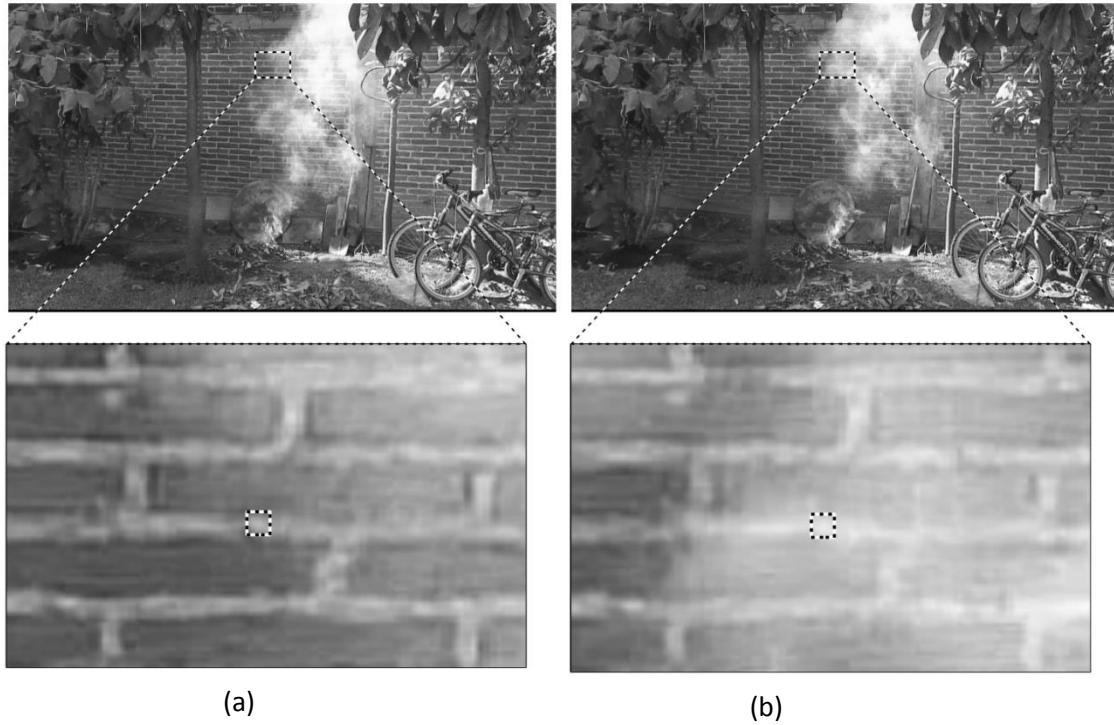


Figura 3.7 Un bloque sin humo, cuadro punteado abajo izquierda (a), y el mismo bloque, ahora con humo en una imagen ulterior, cuadro punteado abajo derecha (b)

Para observar de una mejor manera el efecto provocado por la aparición de humo en los coeficientes DCT del bloque A'_{ij} , se grafican los coeficiente frecuenciales de los bloques involucrados (figura 3.8). Es importante notar que un bloque de coeficientes DCT normalmente presenta una forma matricial, sin embargo, en esta figura se muestra en forma lineal gracias a la interpretación llamada “de ZIG-ZAG order” (3.4) [50].

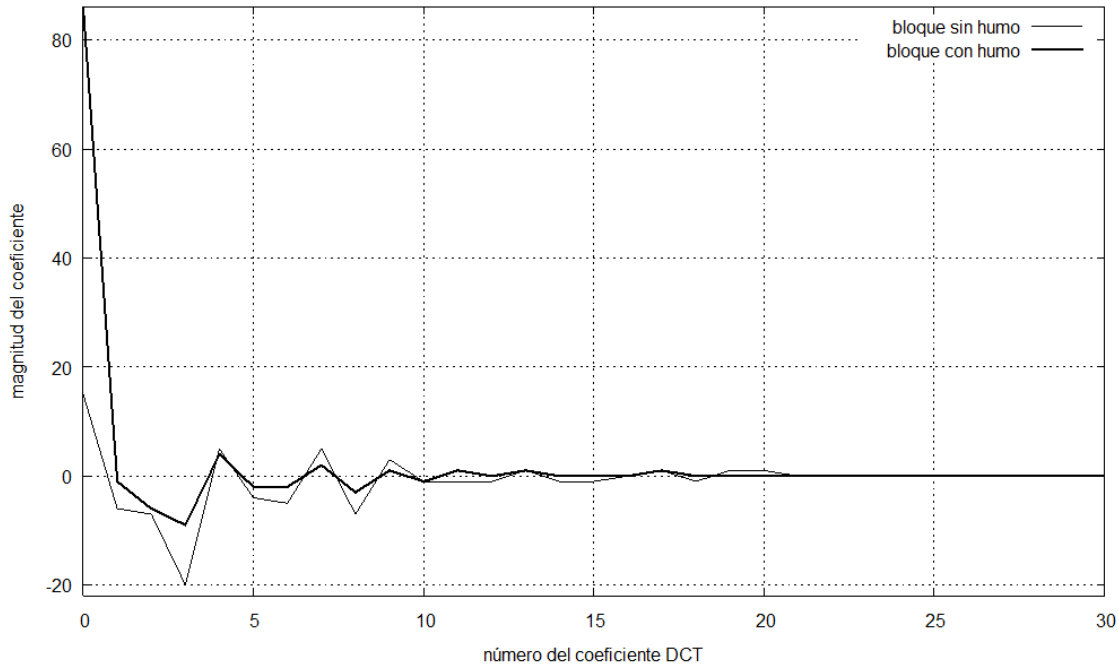


Figura 3.8 bloques de Coeficientes frecuenciales DCT sin y con humo

En la figura 3.8, el bloque libre de humo se presenta con una línea delgada, mientras que el bloque con humo es representado con una línea gruesa. Observar que el coeficiente 0, también llamado DC, presenta un incremento de su valor, el cual pasa de 15 a 86. Desde la perspectiva del HVS, este incremento es percibido como un incremento del brillo. Es importante mencionar que un incremento del valor de DC no puede denunciar, por sí solo, la presencia de humo. Sin embargo, este cambio representa un primer indicio hacia la detección de humo. Además, la prueba del incremento de brillo evita practicar un análisis más profundo a toda la imagen, lo cual resultaría innecesario y sobretodo costoso en términos computacionales.

En lo que respecta al contraste, desde la perspectiva del HVS, la presencia de humo provoca un decremento del contraste en las porciones de imagen afectadas. El cual se

observa, en términos de la DCT, como un decremento de la magnitud de los coeficientes AC. En este caso es importante observar que, mientras que la magnitud del coeficiente DC se incrementa, la magnitud de los coeficientes AC disminuye. En forma inversa, la desaparición del humo provoca una amplificación de los coeficientes AC y un decremento del coeficiente DC. En realidad estos valores no se amplifican, ni se decrementan, solo regresan a su estado original. Es importante recordar que los cambios de contraste aquí enunciados están ligados al humo translucido, por lo tanto un bloque que cambia de contraste de una imagen a otra, debe conservar su significado visual, es decir, los mismos contornos, las mismas texturas y los mismos colores aunque estos estén degradados.

En cuanto al impacto en la nitidez de la imagen, ésta es percibida por el HVS como una degradación de los contornos y texturas, el cual se manifiesta, en términos de los coeficientes de la DCT, como la pérdida de la información de más alta frecuencia. Aquí se considera la pérdida de un coeficiente DCT como aquel cuya magnitud se reduce a cero. En la figura 3.8 se observa la desaparición de los coeficientes de alta frecuencia, en particular los coeficientes 12, 14, 15, 18 19 y 20.

Tomando en cuenta estas consideraciones, se realiza una primera caracterización de comportamiento frecuencial del humo y de otros eventos similares al humo, tales como las sombras, los cambios de iluminación y la aparición de objetos opacos en la escena. El interés de caracterizar estos otros eventos reside en el hecho que éstos pueden ser confundidos con humo, entonces es importante caracterizarlos para poder distinguirlos. Esta caracterización es de orden subjetivo y fue realizada gracias a una herramienta interactiva que despliega dos imágenes JPG con un retraso inter imagen ajustable. Esta forma de presentar las imágenes permite distinguir fácilmente los cambios de una imagen a

otra. Es importante mencionar que las imágenes presentadas fueron tomadas con una diferencia de 500 ms. La herramienta también permite marcar un bloque de interés y desplegar, en otra ventana, los coeficientes de la DCT. Las conclusiones a las que llegaron con este análisis subjetivo son las siguientes:

Tabla 3.1 Humo apareciendo en la escena

Propiedad	Impacto en los coeficientes de la DCT
Brillo	Amplificación del coeficiente DC
Contraste	Atenuación de los coeficientes AC
Definición	Pérdida de coeficientes AC puede ocurrir en cualquier elemento, sin embargo es más probable en las altas frecuencias.
Similitud	Alta

Tabla 3.2 Sombras apareciendo en la escena

Propiedad	Impacto en los coeficientes de la DCT
Brillo	Atenuación de los coeficientes AC
Contraste	Atenuación de los coeficientes AC
Definición	En las altas frecuencias, un pequeño número de modificaciones. Igual número de pérdidas y ganancias.
Similitud	Alta

Tabla 3.3 Cambios de iluminación en ciertas porciones

Parámetro	Impacto en los coeficientes de la DCT
Brillo	Amplificación del coeficiente DC
Contraste	Amplificación de los coeficientes AC
Definición	Ganancia de coeficientes en las altas frecuencias
Similitud	Alta

Tabla 3.4 Aparición de objetos opacos

Parámetro	Impacto en los coeficientes de la DCT
Brillo	Amplificación/Atenuación del coeficiente DC
Contraste	Amplificación/Atenuación de los coeficientes AC
Definición	Ganancia/Pérdida de los coeficientes AC en las altas frecuencias. Todo depende de la textura tanto de la escena como del fondo.
Similitud	Baja o Nula

Observar que el evento “aparición de humo” está ligado a un comportamiento frecuencial bien definido y que difiere de eventos similares. En la siguiente sección se presenta una propuesta para cuantificar los cambios que sufren algunas las propiedades de una imagen cuando se presenta la aparición de humo, en particular los cambios de brillo, de contraste y la pérdida de nitidez en imágenes que han sido comprimidas bajo la norma JPEG.

3.5 Cuantificando cambios en las propiedades de la imagen

La tesis que aquí se presenta busca detectar humo en imágenes JPEG utilizando la única información directamente disponible, es decir, los coeficientes de la DCT. Es esto lo que motiva comprender el efecto que el evento “humo” provoca en los coeficientes de la DCT. Para lograr esto se efectuaron una serie de experimentos que consistieron en emular la aparición de humo sobre una imagen no comprimida (mejor conocida como raw image) con el fin de observar, una vez comprimida bajo la norma JPEG, los efectos provocados en los coeficientes de la DCT. Sabiendo que la aparición de humo provoca, entre otros, cambios de brillo, de contraste y de definición, se emularon estos cambios en forma independiente sobre la imagen de prueba. Para efectuar estos cambios se utilizó como imagen de prueba la conocida imagen “goldhill” (figura 3.9) y el software para edición de imágenes GIMP en su versión 2.6.11.



Figura 3.9 Imagen de referencia sobre la cual se efectúan modificaciones de brillo, contraste y definición

Estudio sobre los cambios de brillo

La aparición de humo translucido incrementa el brillo de la porción de imagen en donde éste se presenta; ver sección 3.4. En forma inversa, la desaparición de humo se manifiesta como un decremento del brillo. La evaluación de los cambios de brillo se pueden efectuar analizando el coeficiente $F(0, 0)$ o DC. Según la (3.1), este coeficiente representa, en forma proporcional, el brillo promedio del bloque de píxeles $f(x, y)$:

cuando:

$$C(u), C(v) = 1/\sqrt{2} \quad \text{Para } u, v = 0;$$

entonces:

$$F(0,0) = \frac{1}{4} \frac{1}{\sqrt{2}} \frac{1}{\sqrt{2}} \left[\sum_{x=0}^7 \sum_{y=0}^7 f(x, y) * \cos \frac{(2x+1)0\pi}{16} \cos \frac{(2y+1)0\pi}{16} \right]$$

$$F(0,0) = \frac{1}{8} \left[\sum_{x=0}^7 \sum_{y=0}^7 f(x,y) \right] \quad (3.7)$$

Observar que el coeficiente DC de un bloque de coeficientes DCT representa una parte proporcional del brillo del bloque. Por otra parte, una entidad deseando analizar los cambios de brillo en imágenes JPEG no tiene que calcular $F(0, 0)$ porque éste valor ya se encuentra en la imagen JPEG y ha sido calculado por el dispositivo que generó la imagen. Hay que mencionar que $F(0, 0)$ también será identificado con a_0 , tal como se especificó en la (3.4).

Para comprender el impacto de los cambios de brillo en los coeficientes de la DCT, se realizó el experimento que consiste en: 1) modificar el brillo de una imagen en el dominio espacial, específicamente la imagen goldhill, la cual no presenta, ni pérdidas, ni compresión alguna. En efecto, esta imagen se encuentra codificada bajo el formato PPM (Portable Pixel Mapping). 2) la imagen se comprime haciendo uso de la norma JPEG. Aquí el objetivo es transformar la información espacial al dominio de la DCT; dominio en donde se desarrolla nuestro estudio. Para efectuar la compresión se recurrió a la utilidad software cjpeg [20] utilizando la calidad por defecto, es decir, 75%. 3) Se extraen y comparan los coeficientes DCT de un bloque A_{ij} que no presenta cambios de contraste con su correspondiente A'_{ij} que si los presenta. Recordar que el primer bloque pertenece a una imagen I , mientras que el segundo a una imagen I' . La selección y extracción de los bloques se efectuó gracias a una herramienta software creada en el marco de este proyecto, la cual implementa parte del proceso de decodificación JPEG, específicamente, el proceso de decodificación de Huffman.

Además, esta herramienta integra, gracias a la librería GTK+ 2.0, la selección del bloque por medio del ratón y un sistema de ventanas para presentar resultados gráficos, ver figura 4.1. De esta manera, los coeficientes del bloque seleccionado son extraídos y almacenados en un archivo.

Con el fin de obtener una idea más general del impacto de los cambios de brillo en un bloque, se aplican diversas intensidades de brillo en la imagen de referencia. Las intensidades disponibles en el GIMP van desde -127 a +128 unidades, para las cuales no se indica su naturaleza. Para abarcar toda la escala se aplicaron los siguientes cambios de brillo: -127, -96, -64, -32, 0, 32, 64, 96 y 127. Los resultados de este estudio son mostrados en la figura 3.10.

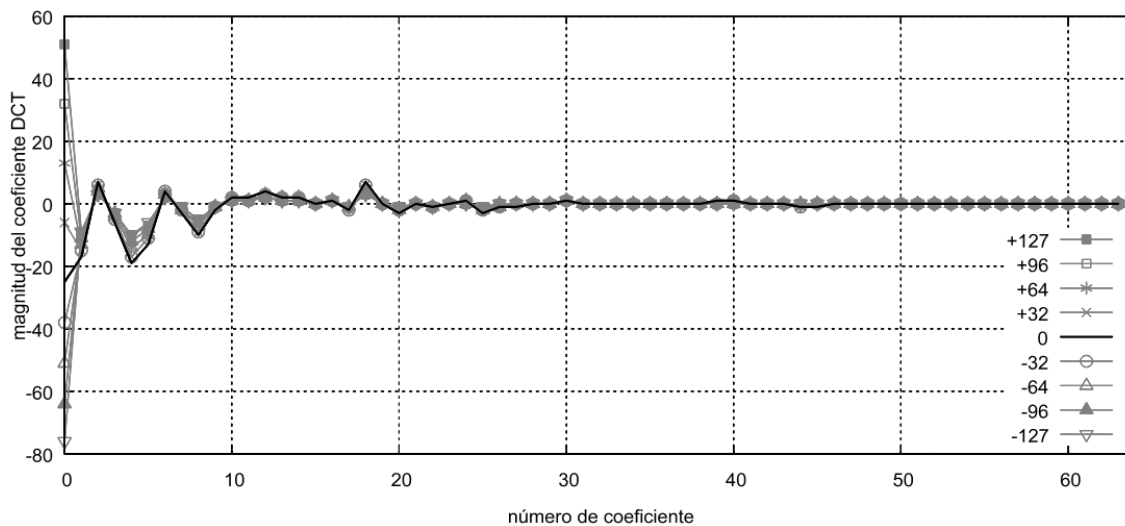


Figura 3.10 Cambios de brillo en los coeficientes DCT de un bloque

En la figura 3.10, mientras que la imagen de referencia y sin cambios de brillo se representa con una línea negra gruesa etiquetada con el valor “0”, los incrementos o

decrementos de brillo se representan con líneas grises más delgadas. Observar que los cambios de brillo se manifiestan esencialmente en el coeficiente DC. Por otra parte, tomando como referencia la representación establecida en (3.4), se puede decir que un cambio de brillo en un bloque entre dos imágenes puede ser descrita como sigue:

$$b(A_{ij}, A'_{ij}) = a'_0 - a_0 \quad (3.8)$$

En donde a_0 es el coeficiente DC de un bloque A_{ij} cualquiera en una cierta imagen I , mientras que a'_0 es el coeficiente DC del bloque A'_{ij} correspondiente en una imagen ulterior I' . De esta manera, cuando b entrega un valor positivo ocurre un incremento de brillo, mientras que un valor negativo indica un decremento del brillo. De esta forma, en principio, un valor cero corresponde a un bloque sin cambios de brillo.

Estudio sobre los cambios de contraste

La aparición de humo se manifiesta en la escena como una reducción del contraste. Para comprender como utilizar la información disponible en una imagen JPEG, se efectúa un estudio similar al que se desarrolló para el caso del brillo. El procedimiento también utiliza la herramienta GIMP, con la cual se provoca cambios de contraste en la imagen, en el dominio espacial, dentro del rango -127 a 127. Aunque no se indica la naturaleza de las unidades, se observa que un valor de -127 elimina todo contraste en la imagen llevando la a convertirse en un cuadro gris sin ningún tipo de textura o bordes más que el código de tonalidad 128 decimal, el cual corresponde con el punto medio en una escala de grises representada con 8 bits. En sentido inverso, un valor 127 enfatiza los contornos y texturas llevándolos a una representación cercana al blanco y negro puro.

En la figura 3.11 se presentan los coeficientes DCT de un bloque A_{ij} al cual se han aplicado cambios de contraste; este bloque pertenece a la imagen de referencia goldhill (figura 3.9). Regresando a la figura 3.11, mientras que la línea negra etiquetada con “0” representa la imagen de referencia, las líneas grises representan los cambios de contraste. Se observa que en el dominio de la DCT un aumento de contraste se manifiesta como una amplificación de la magnitud en cada coeficiente AC. En forma inversa, una reducción del contraste se manifiesta como una atenuación; esta última está ligada a la aparición de humo en la escena. Es importante notar que el coeficiente DC se ve poco afectado, por lo cual se considera que los cambios de contraste se concentran en los coeficientes AC.

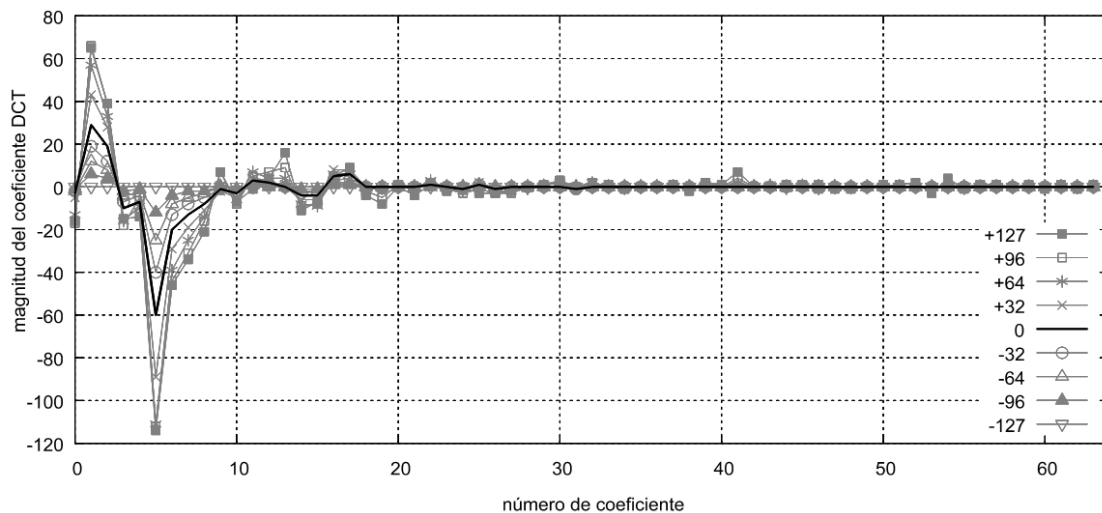


Figura 3.11 Efecto de los cambios de contraste en los coeficientes DCT

Tomando en cuenta lo anterior, se propone evaluar la reducción del contraste entre dos imágenes como la suma de la diferencia del coeficiente a_r con respecto a su correspondiente a_r' . Hay que notar que solo se consideran los coeficientes AC y que el

resultado se hace relativo a la “cantidad de información” que se estima contiene el bloque.

La siguiente expresión indica esto en forma más precisa:

$$c(A_{ij}, A'_{ij}) = \frac{1}{u(A_{ij}, A'_{ij})} \sum_{r=1}^{63} c_r \quad (3.9)$$

donde:

$$c_r = \begin{cases} |a_r| - |a'_r|, & \text{signo}(a_r) = \text{signo}(a'_r) \wedge a_r \neq 0 \wedge a'_r \neq 0 \wedge |a_r| > |a'_r| \\ 0, & \text{en otro caso} \end{cases}$$

En efecto, en un misma imagen no todos los bloques presentan la misma cantidad de información, ver figura 3.12, lo cual hace necesario hacer relativo el cálculo del contraste.

La expresión que se propone para estimar la cantidad de información $u(0)$ (3.10) involucra la pareja de bloques (A_{ij}, A'_{ij}) , y de estos solo se consideran los coeficientes AC de la siguiente manera:

$$u(A_{ij}, A'_{ij}) = \sum_{r=1}^{63} u_r \quad (3.10)$$

$$\text{donde: } u_r = \begin{cases} 1, & a_r \neq 0 \vee a'_r \neq 0 \\ 0, & \text{en otro caso} \end{cases}$$

Por otra parte, el elemento u también permite reaccionar al hecho de que, para bloques con poca información, no es conveniente intentar reconocer cambios de contraste, pérdida de información, u otro. En efecto, un bloque que representa una superficie lisa presenta pocos, o ningún coeficiente AC, en cambio, un bloque que presenta contornos

finos y una textura compleja contendrá una riqueza frecuencial mayor, lo cual permite tener mayor precisión en la detección de un evento particular. La figura 3.12 presenta los coeficientes frecuenciales de la DCT de diferentes bloques pertenecientes a la misma imagen. Comenzando de arriba hacia abajo, la primer gráfica corresponde con un bloque de píxeles con textura casi lisa y sin contornos. En los bloques que siguen, se va incrementando la complejidad de la textura y la fineza de los contorno, hasta llegar al último bloque, el cual presenta, en el dominio espacial, la mayor cantidad de detalles y por lo tanto más coeficientes de la DCT.

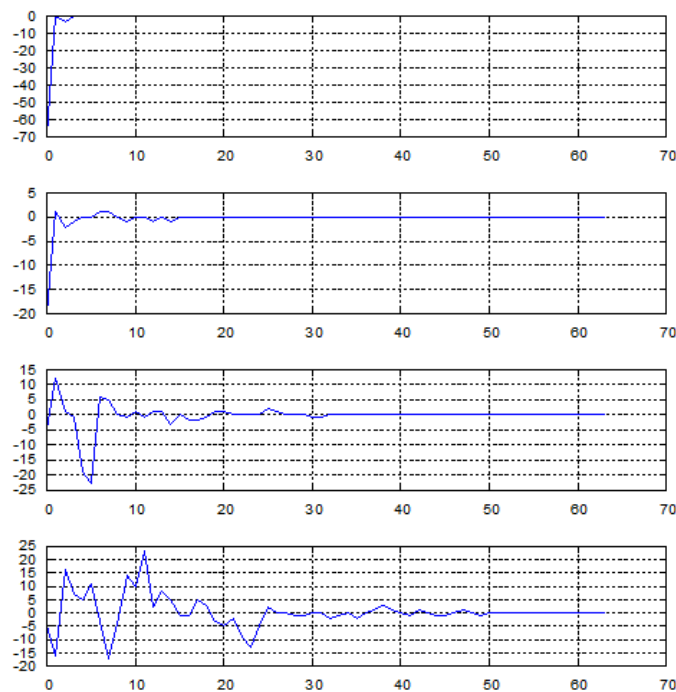


Figura 3.12 La cantidad de coeficientes AC cuyo valor difiere de cero, está ligado a la cantidad de información presente en el bloque

Regresando a la (3.9), observar que la suma de diferencias está condicionada de tal manera que solo se considera aquel par de coeficientes a_r y a_r' , que estén ligados a una atenuación y no a un cambio de signo o una pérdida. Se descartan los cambios de signo porque tal evento no es consecuencia de un cambio de contraste, ver figura 3.11. En cuanto a la pérdida de un coeficiente AC, éstas se consideran ligadas a la pérdida de definición. Evento que debe ser evaluado de otra manera.

Estudio sobre los cambios de definición.

En cuanto a la definición, mientras que la aparición de humo translucido en la escena provoca una degradación visual en los contornos y las texturas, el humo denso destruye completamente la información de la escena. Solo el primer tipo de humo es considerado en esta tesis, del cual es de interés su propiedad para filtrar la información en las altas frecuencias de la imagen [2] [14] [15] [17] [19] [35] [37] [38] [45] [55] [56]. Para comprender y saber cuantificar el impacto de este filtro pasa-bajas, se emula en el dominio espacial y en la medida de lo posible, esta pérdida frecuencial. Para esto se recurre de nuevo a la herramienta GIMP, pero esta vez haciendo uso de un filtro Gaussiano de desenfoque con un radio de 1, 2, 3, 4, 5 y 6 píxeles en ambos ejes. Es claro que no existe ninguna relación entre este filtro y el filtrado que el humo provoca, sin embargo esta prueba representa una forma de comprender como afectan los filtros pasa-bajas a los coeficientes de la DCT. En la figura 3.13 se presenta el impacto del filtro antes mencionado aplicando diferentes intensidades de desenfoque.

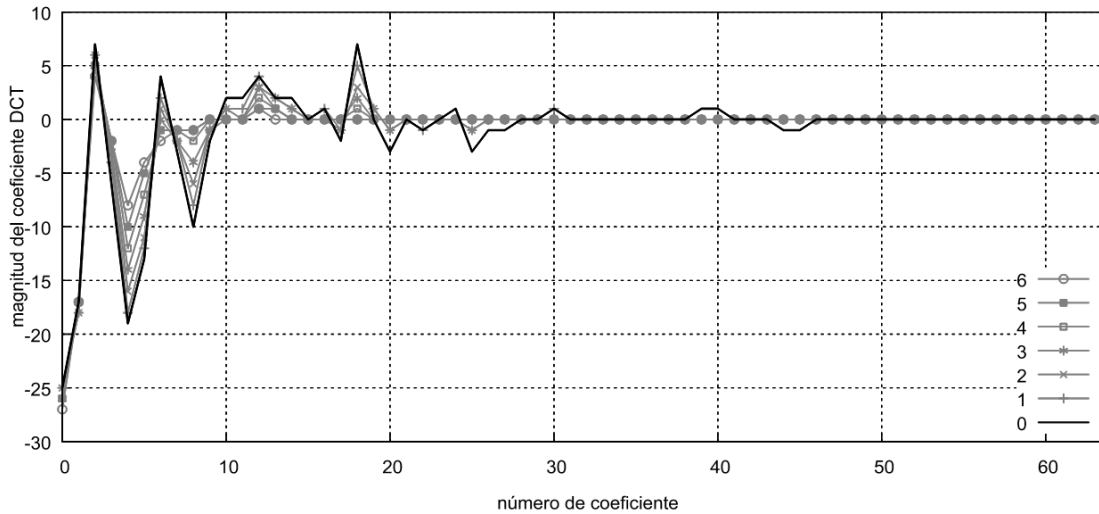


Figura 3.13 Pérdida de información provocada por la aplicación de filtros de desenfoco

Se observa que, tanto el coeficiente DC, como aquellos que lo circundan, son los menos afectados, ver figura 3.13. Entre estos, la magnitud de los coeficientes a_1 , a_2 , a_3 , a_4 y a_5 no llegan a cero a pesar de la aplicación de diferentes magnitudes de filtro (líneas en gris). Ahora, considerando que una pérdida se presenta cuando un coeficiente AC pasa de $a_r \neq 0$ a $a'_r = 0$, la pérdida de información puede ser evaluada a través de la siguiente expresión:

$$q(A_{ij}, A'_{ij}) = \frac{1}{u(A_{ij}, A'_{ij})} \sum_{r=1}^{63} q_r \quad (3.11)$$

$$\text{donde } q_r = \begin{cases} 1, & a_r \neq 0 \wedge a'_r = 0 \\ 0, & \text{en otro caso} \end{cases}$$

En donde la suma de las pérdidas también se hace relativa a la cantidad de información que se estima contiene el bloque.

Verificación de la similitud entre bloques

La evaluación de las diferentes propiedades analizadas anteriormente solo son validas si y solo si se ha verificado que el par de bloques involucrado en un cálculo corresponde a una misma porción de escena. En efecto, las expresiones para calcular los cambios de brillo, de contraste y de definición que se han propuesto pueden entregar un mismo resultado para una escena que presenta humo, como para una escena que no tiene nada que ver con él. De aquí la necesidad indispensable de verificar la similitud entre los bloques.

En el marco de este documento se utiliza el término “similitud” y no igualdad porque igualdad es una comparación cuyo resultado es binario. En cambio, una prueba de similitud puede entregar una gran variedad de resultados. En este punto, es importante mencionar que las múltiples propuestas presentadas en la literatura [2] [14] [15] [17][19] [35] [37] [38] [45] [55] [56] basadas en la evaluación de la pérdida frecuencial no verifican la similitud. En esta tesis se propone, un método para estimar el grado de similitud entre dos bloques, el cual se basa en la observación siguiente: a pesar de los cambios de brillo, contraste y definición aplicados a una imagen, los signos de los coeficientes AC no cambian, ver figura 3.10, figura 3.11, figura 3.13.

Es justo mencionar que durante los cambios de contraste algunos coeficientes pueden llegar a cero, en particular para aquellos que presentan una magnitud es débil; este caso también debe ser considerado.

El método propuesto se basa esencialmente en contar el número e de coeficientes AC que cambian de signo y en el número h de coeficientes que no sufren modificación

alguna; mientras que e estima que tan diferentes son los bloques, h determina que tan parecidos son. Las expresiones correspondientes se incluyen a continuación:

$$e(A_{ij}, A'_{ij}) = \frac{1}{u(A_{ij}, A'_{ij})} \sum_{r=1}^{63} e_r \quad (3.12)$$

$$\text{donde } e_r = \begin{cases} 1, & a_r \neq 0 \wedge a'_r \neq 0 \wedge \text{signo}(a_r) \neq \text{signo}(a'_r) \\ 0, & \text{en otro caso} \end{cases}$$

$$h(A_{ij}, A'_{ij}) = \frac{1}{u(A_{ij}, A'_{ij})} \sum_{r=1}^{63} h_r \quad (3.13)$$

$$\text{donde } h_r = \begin{cases} 1, & a_r \neq 0 \wedge a'_r \neq 0 \wedge \text{signo}(a_r) = \text{signo}(a'_r) \\ 0, & \text{en otro caso} \end{cases}$$

4 Modelo propuesto para la detección de humo

Para modelar el impacto que el humo provoca en las diferentes propiedades de la imagen, se fijaron dos principales prioridades: la velocidad y la precisión del algoritmo. Si bien en la literatura se reportan diversos sistemas con una precisión aceptable, su velocidad de operación no es suficiente para enfrentar flujos de video provenientes de los sistemas de video-vigilancia actuales. El tamaño de las imágenes, pero sobre todo que el video se encuentra comprimido, son elementos con los que estos sistemas no pueden enfrentar. En la tabla I, se observa que el tamaño de las imágenes consideradas por estas propuestas es muy pequeño comparado con el generado por los sistemas actuales de video vigilancia, los cuales pueden generar imágenes de hasta 1920×1080 píxeles. Además, de todos estos trabajos, ninguno trabaja en el dominio del video comprimido. Pero en la misma tabla, aun, estos sistemas solo procesan video no comprimido, lo que implica que para aplicar estas propuestas a un sistema de video-vigilancia actual, el video debe ser descomprimido antes de todo. De esta manera, una computadora que corre alguna de las propuestas de la tabla I, debe no solo lidiar con los algoritmos de reconocimiento, sino también con el costoso proceso de decodificación de video. En una máquina este proceso puede consumir casi en su totalidad el tiempo de CPU, ver figura 2.2.

En este trabajo se propone evitar el proceso de descompresión de video tomando como datos a procesar la única información directamente disponible en las imágenes comprimidas, en este caso, los coeficientes de la DCT. Además, esta información debe ser caracterizada y reconocida por métodos que sean simples y rápidos. Un primer reflejo conduce hacia la caracterización de las propiedades de la imagen, frente al humo, por

medio de distribuciones Gaussianas. Sin sorpresas, estas distribuciones presentan una forma compleja que no puede caracterizada con estos métodos. Sin embargo, existe una variante de estas distribuciones, la cual se conoce bajo el nombre de Gaussian Mixture Models (GMM), que permite caracterizar distribuciones complejas en las cuales intervienen más de una variable aleatoria. Así, en este capítulo se presenta un algoritmo para el reconocimiento de humo translucido grisáceo basado en los GMM. Su uso está motivado por la simpleza que él presenta al momento de reconocer una muestra con respecto al modelo previamente obtenido. En este capítulo se hace una breve presentación de modelo de mezclas Gaussianas, después se aborda el uso del GMM en el reconocimiento de humo, para finalmente presentar el modelo que permite reconocer el evento deseado.

4.1 Modelos con Mezclas Gaussianas

Los Modelos con Mezclas Gaussianas son una composición de diferentes distribuciones estadísticas. En este tipo de distribuciones, cada una de las componentes presenta una función de densidad de probabilidad diferente. Ejemplos de mezcla de distribuciones son: la distribución de frecuencia del tamaño en un grupo de animales en donde se mezclan especímenes de diferentes edades. La medida de la distribución de algún síntoma en un grupo de pacientes que padecen enfermedades diversas también es una mezcla de distribuciones. En todos estos casos, se verifica que una mezcla finita de distribuciones tiene un número finito de componentes. Un GMM es una suma ponderada de M componentes:

$$p(\vec{x}/\lambda) = \sum_{i=1}^M p_i b_i(\vec{x}) \quad (4.1)$$

donde \vec{x} es un vector aleatorio de dimensión D, $b_i(\vec{x})$, $i = 1, \dots, M$, son los componentes de densidad y p_i , $i = 1, \dots, M$, son los pesos de las mezclas. Cada componente de densidad es una función Gaussiana de forma siguiente:

$$b_i(\vec{x}) = \frac{1}{(2\pi)^{D/2} |\Sigma_i|^{1/2}} \exp \left\{ -\frac{1}{2} (\vec{x} - \vec{\mu}_i)' \Sigma_i^{-1} (\vec{x} - \vec{\mu}_i) \right\} \quad (4.2)$$

Aquí, $\vec{\mu}_i$ es el vector que representa la media y Σ_i es la matriz de covarianza. Además, la mezcla de pesos debe satisfacer la condición $\sum_{i=1}^M p_i = 1$. La mezcla de densidades es completamente parametrizada por los vectores de media, las matrices de covarianza y los pesos provenientes de todas las densidades componentes. Estos parámetros son colectivamente representados por:

$$\lambda = \{p_i, \mu_i, \Sigma_i\} \quad i = 1, \dots, M \quad (4.3)$$

En el marco de este trabajo, se plantea obtener un λ para cada una de las propiedades de la imagen que se ven afectadas por la aparición de humo en la escena. Existen diversas técnicas que permiten estimar los elementos de λ , en donde el más popular y mejor establecido es la estimación ML o máxima verosimilitud. Estas técnicas han sido exitosamente aplicadas en diversos trabajos de investigación [28] que persiguen objetivos similares a los nuestros. La estimación ML puede ser obtenido iterativamente utilizando un caso particular del algoritmo Expectation-Maximization (EM). La idea básica del algoritmo EM es arrancar la búsqueda de los elementos p_i , μ_i y Σ_i con un modelo λ inicial para estimar un nuevo modelo λ tal que $p(x|\vec{\lambda}) \geq p(x|\lambda)$. El nuevo modelo se convierte

entonces en el modelo inicial en la siguiente iteración y el proceso se repite hasta que algún umbral de convergencia es alcanzado. En cada iteración, se utilizan las siguientes fórmulas de re-estimación:

Mezclas ponderada:

$$\bar{p}_i = \frac{1}{T} \sum_{t=1}^T p(i|\vec{x}_t, \lambda) \quad (4.4)$$

Medias:

$$\vec{\mu}_i = \frac{\sum_{t=1}^T p(i|\vec{x}_t, \lambda) \vec{x}_t}{\sum_{t=1}^T p(i|\vec{x}_t, \lambda)} \quad (4.5)$$

Varianzas:

$$\bar{\sigma}_i^2 = \frac{\sum_{t=1}^T p(i|\vec{x}_t, \lambda) x_t^2}{\sum_{t=1}^T p(i|\vec{x}_t, \lambda)} - \bar{\mu}_i^2 \quad (4.6)$$

donde σ_i^2 , x_i^2 y μ_i^2 son elementos arbitrarios de los vectores $\vec{\sigma}_i^2$, \vec{x}_i^2 y $\vec{\mu}_i^2$

respectivamente. Por otra parte, la probabilidad a posteriori está dada por:

$$p(i|\vec{x}_t, \lambda) = \frac{p_i b_i(\vec{x}_t)}{\sum_{k=1}^M p_k b_k(\vec{x}_t)} \quad (4.7)$$

Hay dos factores críticos en el manejo del algoritmo EM: la selección del orden T de la mezcla y la iniciación de los parámetros del modelo. En efecto, no hay buenos medios teóricos para guiar uno u otro de estos factores. En [27] se recomienda que para llegar a un modelo que describa en forma fiel una mezcla Gaussiana, hay que intentar con varios conjuntos de parámetros de configuración antes de converger en un modelo preciso.

Por otra parte, la programación del algoritmo EM no es una tarea sencilla y para los objetivos buscados en este proyecto es innecesario desarrollar tal programación. Así, se utiliza la herramienta llamada R-environment [27], el cual es presentado como un software gráfico que integra múltiples métodos matemáticos para el análisis de datos y la elaboración de gráficas. La herramienta R-environment integra un lenguaje derivado de otro lenguaje llamado S que fue desarrollado en los laboratorios Bell. Entre los métodos que éste ofrece, se aplica la librería mixdist [27] debido a que ésta implementa el método de estimación de máxima verosimilitud (MLE), el cual permite estimar los pesos p_i de cada componente de la mezcla de distribuciones, así como los parámetros μ_i y σ_i de cada distribución. Estos componentes pueden ser expresados en términos de diferentes tipos de distribución, a saber: normal finita, log-normal, gamma, Weibull, binomial, Poisson y otras. Entre estas distribuciones, se utilizan las distribuciones normales. Adicionalmente, mixdist integra una herramienta para graficar de manera interactiva, la cual permite ir probando diferentes componentes de distribución inicial, los cuales ayudan a la herramienta a converger en un resultado final adecuado.

4.2 Creación del Modelo GMM para el reconocimiento del humo

El modelo que aquí se propone está esencialmente expresado en términos de los GMM. El primer paso para crear este modelo consiste en adquirir muestras de información que estén ligadas a la aparición de humo en la escena; una muestra es un par de bloques A_{ij} y A'_{ij} . Después, utilizando las expresiones del capítulo anterior, y un buen número de muestras, se calculan los cambios de brillo, de contraste, de la definición y la similitud entre bloques. Los resultados obtenidos se procesan independientemente con el fin de obtener una distribución para los cambios de brillo, otra para el contraste, otra para la

definición, etc. Todos los cálculos a partir de la obtención de estas distribuciones son efectuadas con la herramienta R-environment, y en particular, con la librería *mixdist*, la cual permite generar un modelo GMM. Los parámetros que este proceso arroja constituyen el modelo del comportamiento del humo.

Este modelo se llevó a un plano aplicativo mostrando una buena caracterización de la aparición de humo en la escena, sin embargo, éste presentó un cierto número de falsos positivos, los cuales se anulan con un método que denomino eliminación de resultados errátiles. Además de eliminar los falsos positivos, este método enmascara resultados positivos correctos. Para recuperar estos resultados positivos se reaplica el un modelo laxo a los bloques que circundan una muestra declarada como positiva. La aplicación de estos dos métodos, corresponden a las operaciones morfológicas llamadas erosión y la dilatación. A continuación se presenta con más detalle los diversos pasos que permiten obtener un modelo para el reconocimiento de humo.

4.2.1 Adquisición de muestra y extracción de EURIs

La creación del modelo comienza con la adquisición de muestras, las cuales son seleccionadas por un observador humano. Es útil comentar que este es un proceso subjetivo que se ve influenciado por factores tales como el criterio, la edad, la profesión, etc. En el marco de esta tesis de maestría, la adquisición de muestras fue realizada en su mayor parte por estudiantes, de licenciatura a maestría, así como por un par de profesores. Para las personas que participaron en esta tarea, la consigna fue: “encontrar los bloques sin humo que sean cubiertos, en un segundo tiempo, por humo. Además, éste debe ser completamente translucido. De esta manera, se desechan los bloques que presenten al mismo tiempo humo

translucido y opaco. El procedimiento para adquirir las muestras se desarrolló de la siguiente manera:

Primero, se toman en video diversos escenarios que presenten humo. Esta adquisición se realizó con la ayuda la cámara IP AXIS Q1755, en modo compresión JPEG, con una resolución de 1920×1080 píxeles.

En segundo lugar, de estas secuencias de video se escogen pares de imágenes separadas en el tiempo, aproximadamente 500 ms; este tiempo es suficiente para dejar que el humo se desplace y permita observar porciones de escena, sin y con humo, a una distancia aproximada de hasta 100 m.

En tercer lugar, con la ayuda de una herramienta de software especialmente creada para este proyecto (ver sección 3.5), las imágenes se presentan en forma alternada para facilitar al observador la aparición de humo. La herramienta se invoca con el siguiente comando en línea:

```
$ analiza_diferencia <file1> <file2> <delay>
```

En donde `file1` y `file2` corresponden al archivo que contiene las imágenes a desplegar, mientras que `delay` es el retraso de presentación en milisegundos entre las imágenes especificadas. Por otra parte, cuando se selecciona un bloque y se decide lanzar la ventana de representación gráfica de los coeficientes DCT de los bloques involucrados, se presenta otra ventana que muestra las amplificaciones, atenuaciones, las pérdidas, las ganancias, los cambios de signo y los coeficientes que aparecen intactos. Esta información gráfica, ver figura 4.1, facilitan al observador humano la decisión de conservar o eliminar los bloques muestra; la herramienta guarda en un archivo los datos que se desean conservar.

Además la decisión de conservar una muestra debe verificar las observaciones establecidas en la tabla 3.1.

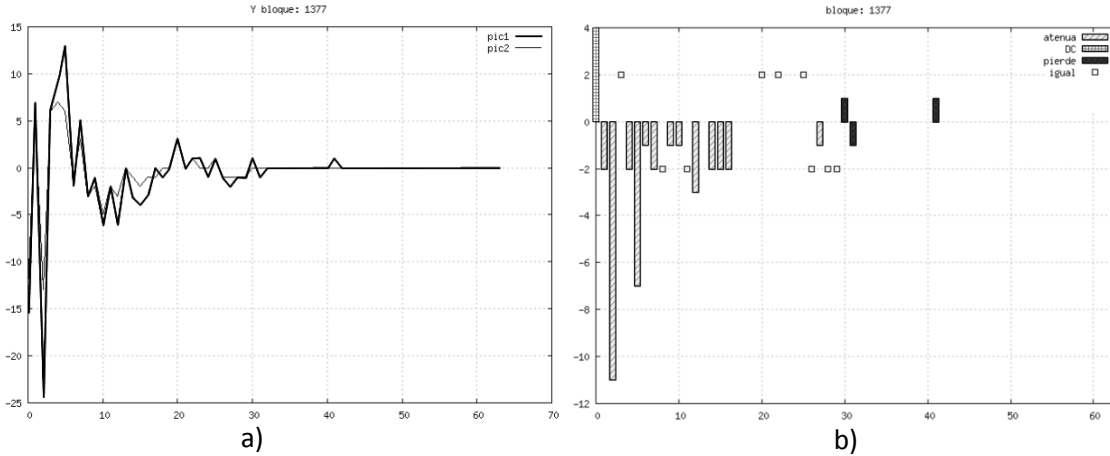


Figura 4.1 Gráficas que entrega la herramienta de análisis de imágenes JPG

En la figura 4.1a se presentan, en forma lineal, los coeficientes frecuenciales de la DCT, en donde la línea *pic1* corresponde a un bloque sin humo, mientras que *pic2* representa un bloque con humo. Se observa, que un aumento del brillo se manifiesta como un incremento en el coeficiente DC. Este cambio se representa en la figura 4.1b, bajo la forma de una barra de valor 4; en realidad el coeficiente DC pasa de -16 a -14. Con respecto a los cambios de contraste, se observa que ciertos coeficientes AC sufren una atenuación. Hay que notar que algunos coeficientes permanecen intactos; un cuadrado blanco los representa. Por último, la pérdida de coeficientes frecuenciales se puede distinguir perfectamente; estas son representadas con las barras negras.

De esta manera, la herramienta facilitó la adquisición de aproximadamente 8000 muestras sobre una docena de videos, todos relacionados con la aparición de humo en la escena. Cada muestra está constituida por los datos de dos bloques de coeficientes de la DCT. El primero no presenta humo, el segundo sí.

4.2.2 Extracción de características

Las características a considerar en el modelo de humo que se propone son los cambios contraste, de definición y de similitud. Un número de otras características, derivado de estas que se acaban de enunciar, es presentado más adelante. Entonces, para obtener las características enunciadas aquí arriba se utiliza (3.9), para los cambios de contraste, la (3.11) para los cambios de definición y las (3.12) y (3.13) para estimar la similitud entre bloques. Es importante mencionar que los cambios de brillo no son modelados con ayuda de los GMM. Al tratarse de un solo coeficiente, este solo se emplea en el algoritmo de reconocimiento para conocer si un cambio de brillo amerita pasar a un análisis más exhaustivo. También es importante mencionar que, aunque se utilizan imágenes en color, el algoritmo propuesto solo toma en cuenta la componente de luminancia Y. La razón es que las cámaras de video-vigilancia normalmente codifican el video con el esquema de submuestreo 4:2:0, el cual reduce la información de color al 25%. Se considera entonces, que esta cantidad de información puede resultar insuficiente para efectuar un trabajo de reconocimiento.

Por otra parte, la aparición de humo en términos frecuenciales no siempre se presenta en la forma “ideal” descrita en la tabla 3.1. Por ejemplo, para el caso del contraste, además de la atenuación en ciertos coeficientes AC se puede presentar una amplificación. La cantidad de amplificaciones a tolerar durante el proceso de reconocimiento de humo es una respuesta que se puede resolver con un método estadístico, específicamente con el uso del GMM. Con este fin se establece la relación entre la cantidad de atenuaciones y amplificaciones, se propone la siguiente expresión:

$$\omega(A_{ij}, A'_{ij}) = c(A_{ij}, A'_{ij}) - d(A_{ij}, A'_{ij}) \quad (4.8)$$

En donde d representa una estimación de la amplificación experimentada por los coeficientes AC y es obtenida por medio de la siguiente ecuación:

$$d(A_{ij}, A'_{ij}) = \frac{1}{u(A_{ij}, A'_{ij})} \sum_{r=1}^{63} d_r \quad (4.9)$$

donde:

$$d_r = \begin{cases} |a_r| - |a'_r|, & \text{signo}(a_r) = \text{signo}(a'_r) \wedge a_r \neq 0 \wedge a'_r \neq 0 \wedge |a_r| < |a'_r| \\ 0, & \text{en otro caso} \end{cases}$$

Observar que en el caso ideal $\omega(A_{ij}, A'_{ij}) = c(A_{ij}, A'_{ij})$; este valor se va reduciendo a medida que más amplificaciones van apareciendo. Si $\omega = 0$, la cantidad de amplificaciones iguala a las atenuaciones, y si $\omega < 0$ entonces hay más amplificaciones que atenuaciones.

En lo que respecta a los cambios de nitidez en la imagen, cuando ésta se expresa en términos de la pérdida de coeficientes frecuenciales, el caso ideal es aquel en donde solo hay pérdidas y cero ganancias (ver tabla 3.2). Desafortunadamente este caso no siempre se presenta. En efecto, en la gran mayoría de los casos, las pérdidas van acompañadas de ganancia de coeficientes. La cantidad de pérdida con respecto a la ganancia es una característica propia a la aparición de humo. Para confrontar la cantidad de pérdidas contra las ganancias, y así caracterizar las cantidades que corresponden con la aparición de humo, se propone la siguiente expresión:

$$\tau(A_{ij}, A'_{ij}) = q(A_{ij}, A'_{ij}) - g(A_{ij}, A'_{ij}) \quad (4.10)$$

En donde g es la ganancia de coeficientes frecuenciales entre dos imágenes.

$$g(A_{ij}, A'_{ij}) = \frac{1}{u(A_{ij}, A'_{ij})} \sum_{r=1}^{63} g_r \quad (4.11)$$

$$\text{donde } g_r = \begin{cases} 1, & a_r = 0 \wedge a'_r \neq 0 \\ 0, & \text{en otro caso} \end{cases}$$

Hay que notar que el caso ideal, descrito en la tabla 3.2, se presenta cuando no hay ganancias, es decir $\tau(A_{ij}, A'_{ij}) = q(A_{ij}, A'_{ij})$; además, a medida que g va creciendo, q se va reduciendo. Una vez más, para caracterizar el rango de valores τ que corresponden con la aparición de humo se emplea un método estadístico basado en el uso de GMM.

Por otra parte, existe una relación entre los cambios de contraste y los cambios de definición. Observando la figura 3.11 se encuentra que los cambios de contraste puede provocar la pérdida de los coeficientes AC que presentan la magnitud más baja. Así, con el fin de establecer una relación entre los cambios de contraste ω y de definición τ , se propone la siguiente expresión:

$$\theta(A_{ij}, A'_{ij}) = \frac{\omega(A_{ij}, A'_{ij})}{\tau(A_{ij}, A'_{ij})} \quad (4.12)$$

El objetivo de esta expresión, la cual se incluye en el modelo, es no dejar los cambios de contraste y de definición como dos eventos completamente independientes. Se considera que la aparición de humo en la escena se presenta alrededor de un determinado ω y un determinado τ los cuales se pueden evaluar por medio de θ .

4.2.3 Creación del modelo usando mixdist

El modelo propuesto para describir el comportamiento de humo en video está compuesto por siete diferentes GMM que describen los cambios de siete diferentes propiedades de un A_{ij} con respecto a A'_{ij} , los cuales se obtienen gracias a las ecuaciones que se enlistan a continuación:

- 1) La atenuación del contraste c (3.9).
- 2) La estimación de la cantidad de información u de (3.10).
- 3) El número de coeficientes e que sufren cambios de signo (3.12).
- 4) El número de coeficientes h que permanecen sin cambios (3.13).
- 5) El índice ω que describe la atenuación de contraste con respecto a su amplificación (4.8).
- 6) Los cambios de definición τ expresados como la diferencia entre la pérdida y la ganancia de coeficientes (4.10).
- 7) El índice θ que expresa los cambios de contraste con respecto a los cambios de definición (4.12).

Para crear cada GMM se utilizaron alrededor de 8000 bloques “muestra”, los cuales fueron adquiridos por medio de la herramienta descrita anteriormente. En seguida, a cada muestra se aplicó cada una de las ecuaciones enlistadas aquí arriba de tal manera que a partir de cada par de bloques A_{ij} , A'_{ij} , se obtienen siete diferentes datos, los cuales se agrupan de acuerdo a su tipo para formar los siete diferentes conjuntos de datos, a saber: el

MODELO PROPUESTO PARA LA DETECCIÓN DE HUMO

conjunto C contiene la colección de valores c , Q contiene los valores entregados por q , U es el conjunto de valores estimados por u , E es el conjunto de valores entregados por e , H contiene los valores producidos por h , Ω es el conjunto de valores entregados por ω , el cual es la confrontación de los incrementos y decrementos de contraste en un par de bloques, T que representa el conjunto de valores entregados por τ y cuyo valor corresponde a la pérdida de coeficientes frecuenciales con respecto a su ganancia. Por último, Θ reúne los valores entregados por θ , el cual estima la relación entre los cambios de contraste y de nitidez. Después, los datos de los conjuntos son analizados para obtener las correspondientes distribuciones gracias al uso del software R-environment y en particular de la librería *mixdist*. El resultado de este primer análisis es presentado en las figura 4.2 a figura 4.8. Observar que una curva “principal” de la distribución compleja sigue el contorno del histograma. Esta curva es la estimación de la mezcla de distribuciones propuesto por la herramienta a partir de los componentes “normales” iniciales solicitados por *mixdist*. Estos componentes corresponden con las distribuciones normales trasladadas de menor tamaño y que en suma determinan la distribución compleja o curva.

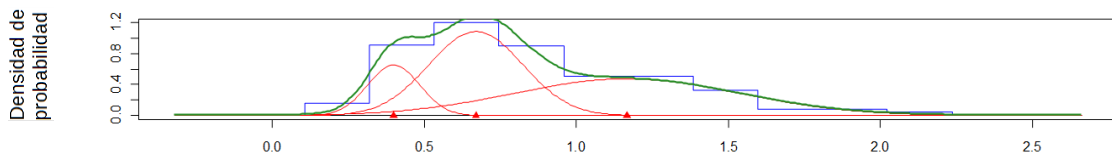


Figura 4.2 Distribución empleando el conjunto C

MODELO PROPUESTO PARA LA DETECCIÓN DE HUMO

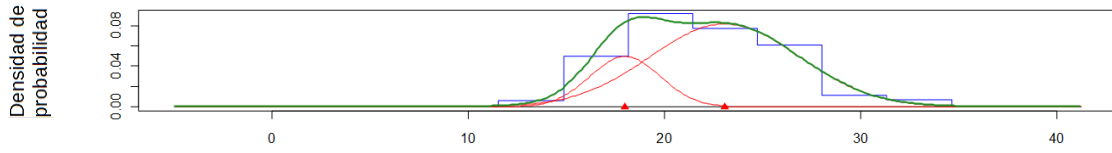


Figura 4.3 Distribución extraída del Conjunto U

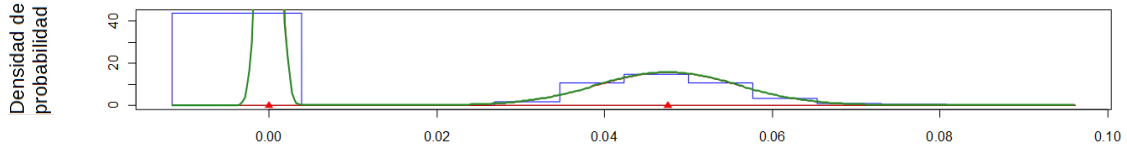


Figura 4.4 Distribución extraída del conjunto E

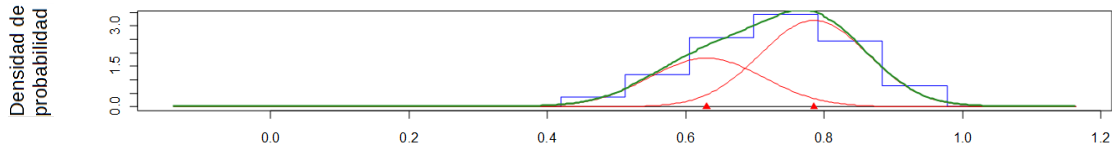


Figura 4.5 Distribución extraída del conjunto H

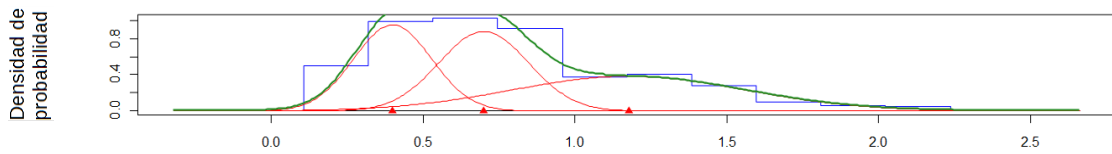


Figura 4.6 Distribución extraída del conjunto Ω

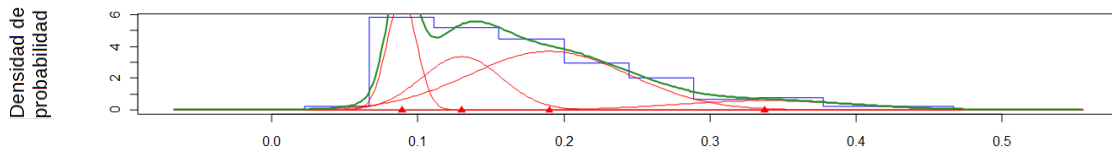


Figura 4.7 Distribución extraída del conjunto T

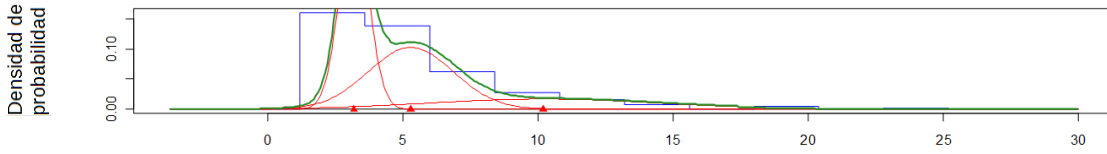


Figura 4.8 Distribución extraída del conjunto Θ

Es importante resaltar que llegar a una mezcla de distribuciones validada por mixdist requiere de múltiples pruebas, de los parámetros iniciales, una distribución que se parezca al histograma de distribuciones. La causa de esto es que no hay un procedimiento establecido para mixdist que indique cómo obtener una mezcla de distribuciones; de otra forma el procedimiento sería completamente automático. De esta manera se debe intentar con diferentes parámetros iniciales $(\lambda = \{p_i, \mu_i, \Sigma_i\} \quad i = 1, \dots, \eta)$ siempre intentando conformar, cada vez más, el contorno del histograma. Una vez que las distribuciones, y sus respectivos parámetros, son validados como correctos, mixdist entrega un cierto número de distribuciones descritas por medio de $(\lambda = \{p_i, \mu_i, \Sigma_i\} \quad i = 1, \dots, M)$, los cuales son los más próximos a la distribución representada por el histograma. Observar que los parámetros iniciales y que integran un número η de componentes puede diferir del número M de componentes que finalmente entrega mixdist. Para las distribuciones presentadas en las figura 4.2 a figura 4.8, mixdist entregó los siguientes resultados a los cuales se identifican con $\lambda_C, \lambda_U, \lambda_E, \lambda_H, \lambda_\Omega, \lambda_T$ y λ_Θ .

Tabla 4.1 Gaussian Mixture Model λ_C para el conjunto C

i	p_i	μ_i	σ_i
1	0.1397	0.400	0.08559
2	0.4190	0.670	0.15436
3	0.4412	1.167	0.36956

MODELO PROPUESTO PARA LA DETECCIÓN DE HUMO

Tabla 4.2 Gaussian Mixture Model λ_U para el conjunto U

i	pi	μ_i	σ_i
1	0.2230	18.000	1.794
2	0.7770	23.080	3.801

Tabla 4.3 Gaussian Mixture Model λ_E para el conjunto E

i	pi	μ_i	σ_i
1	0.6689	0.00000	0.0009988
2	0.3311	0.04755	0.0084900

Tabla 4.4 Gaussian Mixture Model λ_H para el conjunto H

i	pi	μ_i	σ_i
1	0.372	0.6300	0.08236
2	0.628	0.7854	0.07822

Tabla 4.5 Gaussian Mixture Model λ_Ω para el conjunto Ω

i	pi	μ_i	σ_i
1	0.3056	0.400	0.1284
2	0.3250	0.700	0.1480
3	0.3694	1.178	0.3833

Tabla 4.6 Gaussian Mixture Model λ_T para el conjunto T

i	pi	μ_i	σ_i
1	0.17748	0.0890	0.01004
2	0.22932	0.1300	0.02727
3	0.50289	0.1900	0.05442
4	0.09031	0.3376	0.05974

Tabla 4.7 Gaussian Mixture Model λ_Θ para el conjunto Θ

i	pi	μ_i	σ_i
1	0.4199	3.20	0.543
2	0.4045	5.30	1.577
3	0.1755	10.19	4.153

4.3 El proceso de reconocimiento de humo

Para estimar la presencia de humo en la escena con estos modelos, se confrontan los valores c, e, u, h, ω, τ y θ extraídos del par bloques A_{ij} y A'_{ij} , con los modelos $\lambda_C, \lambda_U, \lambda_E, \lambda_H, \lambda_\Omega, \lambda_T$ y λ_Θ . La prueba consiste en determinar la probabilidad total p , ver (4.1) para cada parámetro. De esta manera, se considera que un bloque de la escena presenta humo con respecto a una probabilidad de referencia P_x sí se verifica la siguiente regla:

$$s(A_{ij}, A'_{ij}) = \begin{cases} SI, & \begin{aligned} & p(c(A_{ij}, A'_{ij}) / \lambda_C) \geq P_x \wedge \\ & p(u(A_{ij}, A'_{ij}) / \lambda_U) \geq P_x \wedge \\ & p(e(A_{ij}, A'_{ij}) / \lambda_E) \geq P_x \wedge \\ & p(h(A_{ij}, A'_{ij}) / \lambda_H) \geq P_x \wedge \\ & p(\omega(A_{ij}, A'_{ij}) / \lambda_\Omega) \geq P_x \wedge \\ & p(\tau(A_{ij}, A'_{ij}) / \lambda_T) \geq P_x \wedge \\ & p(\theta(A_{ij}, A'_{ij}) / \lambda_\Theta) \geq P_x \end{aligned} \\ NO, & \text{al menos una no se cumple} \end{cases} \quad (4.13)$$

El proceso de reconocimiento aquí presentado fue programado según el diagrama de flujo mostrado en la figura 4.9. Primeramente, el programa se alimenta con un par de imágenes JPEG sacadas del flujo producido por una cámara AXIS Q1755 [51]. Las imágenes son tomadas del flujo con una separación de aproximadamente 500 ms; se considera que este tiempo es suficiente para observar claramente el desplazamiento del humo en la escena bajo las condiciones establecidas en la sección 3.4. Después, el programa comienza preguntando si hay más bloques en las imágenes I e I' . Si es el caso, se toma un bloque de cada una de las imágenes; tomar un bloque implica lanzar el proceso de descompresión de Huffman para recuperar un MCU (*Minimum Coded Unit*) [50]. Para un

submuestreo 4:2:0, tal como el que se utiliza en las cámaras AXIS Q1755, un MCU integra cuatro bloques, dos que contienen la información de luminancia Y y dos más con la información de crominancia Cr y Cb. En este trabajo de tesis solo se utiliza la información de luminancia por las siguientes razones: En esta tesis solo busca detectar humo translucido a partir de la información de tonalidades de gris, y en segundo lugar, porque la información de color generada por la gran mayoría de cámaras de video-vigilancia es de una resolución considerablemente inferior con respecto a la luminancia. En efecto, el submuestreo 4:2:0 destruye $\frac{3}{4}$ partes de la información de cada componente de color.

A continuación, el algoritmo de detección solo se aplica si existe un cambio de brillo $b(A_{ij}, A'_{ij})$, entre 3 y 43 unidades.

Además, dentro de una trama JPEG los MCU están acomodados en renglones cuyo orden va de izquierda a derecha y de arriba hacia abajo. Así, el primer MCU en ser decodificado, vía Huffman, es aquel que corresponde con la esquina superior izquierda de la imagen. En cuanto al último bloque, éste corresponde con la esquina inferior derecha. Si bien existen muy diversas formas de acomodar los MCU en una trama JPEG, las tramas que generan las cámaras AXIS se forman de esta manera. En cuanto a las funciones que realizan una descompresión parcial de una imagen JPEG y que extraen los coeficientes frecuenciales del MCU, éstas fueron creadas en nuestro laboratorio en el marco de un trabajo previo [36]. De esta manera se recuperan los bloques de luminancia A_{ij} y A'_{ij} con respecto a las imágenes I e I' . Además, este proceso perdura hasta que no hay más bloques en las tramas JPEG. En cuanto a la información de crominancia, esta es descartada por las siguientes razones: Primeramente, esta tesis solo busca detectar humo translucido a partir de la información de tonalidades de gris, y en segundo lugar, porque la información de

color generada por la gran mayoría de cámaras de video-vigilancia es de una resolución baja con respecto a la luminancia. Las cámaras que se utilizaron en el marco de esta tesis efectúan un submuestreo 4:2:0.

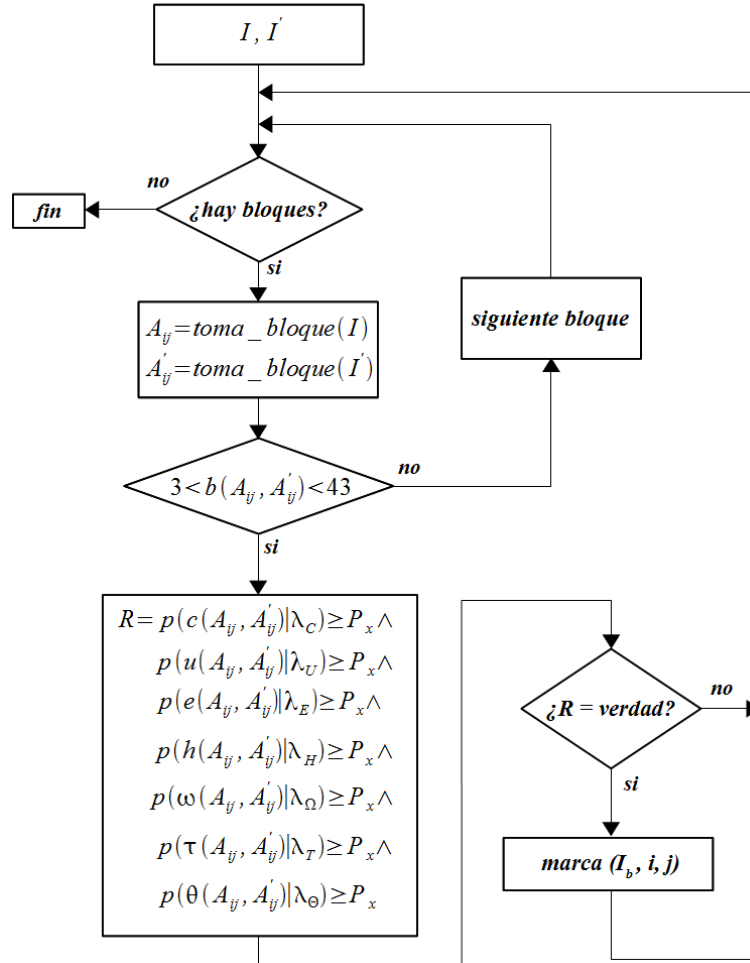


Figura 4.9 Algoritmo que reconoce humo translúcido a partir del modelo propuesto en el capítulo anterior

Por otra parte, sobre los bloques A_{ij} y A'_{ij} , se aplican las ecuaciones c, e, u, h, ω, τ y θ , las cuales se confrontan individualmente su modelo GMM correspondiente. Observar

que la aplicación de las ecuaciones antes citadas solo se efectúa si los cambios de brillo (3.8) se encuentran entre 3 y 43 unidades; haciendo notar que el algoritmo se comporta mejor cuando se restringe el cálculo a estos umbrales. Una vez pasando esta primer condición, un bloque es declarado como “ligado a la aparición de humo”, si todos los valores c , e , u , h , ω , τ y θ conforman su correspondiente modelo λ con al menos una probabilidad P_x ; ver (4.13).

Para saber si alguno de los valores antes citados conforma su correspondiente modelo λ , los valores c , e , u , h , ω , τ y θ se considera como la variable aleatoria x con la cual se calcula su probabilidad con respecto a cada una de las distribuciones normales que componen el GMM correspondiente. La suma de estas diferentes probabilidades será la probabilidad a confrontar con P_x . Por otra parte, para calcular la probabilidad de cada valor con respecto a su correspondiente distribución, se recurrió a la librería propuesta por [1] y conocida bajo el nombre de “*Lower tail quantile for standard normal distribution function*”, la cual permite obtener una aproximación de la función de distribución normal estándar cumulativa inversa. Por ejemplo, dado P_x , esta librería regresa una aproximación del punto X que satisface $P_x = P_r \{ Z \leq X \}$ donde Z es una variable aleatoria proveniente de una distribución normal estándar.

4.4 Eliminación de resultados errátiles

Si bien el método presentado en la sección anterior es capaz de detectar humo, éste presenta algunos falsos positivos. En esta sección se presentan dos propuestas para atacar el problema. La primera emprende acciones en el ámbito del bloque, mientras que la segunda controla que el evento “humo” se realice sobre un grupo de bloques en forma simultánea.

4.4.1 Acciones en el ámbito del bloque

Estas acciones buscan controlar las condiciones bajo las cuales el método propuesto falla. Dicho de otra manera, estas acciones tienen como objetivo evitar aplicar el método en bloques cuyo contenido se considera adverso. Entre otras, se considera inadecuado aplicar el método en bloques que no incluyen movimiento, en bloques que pierden demasiada información con la aparición de humo y en bloques demasiado claros u oscuros.

a) *Bloques sin eventos visuales y el ruido.* Aunque el ruido está presente en todas las imágenes de un flujo de video, el ruido solo representa problemas en las porciones de imagen que carecen de eventos visuales, los cuales corresponden con todo aquel cambio en la imagen que provoca que la atención del observador humano se dirija hacia el cambio. Los cambios de luminancia, el movimiento de sombras, los desplazamientos de objetos sólidos o translucidos, etc. son considerados como eventos visuales. Entonces, el objetivo de esta acción es determinar cuándo un par de bloques (A_{ij}, A'_{ij}) no presenta movimiento, sabiendo que el ruido implícito en toda imagen puede ser interpretado como movimiento. Así es necesario definir un umbral (o conjunto de reglas, u otro medio) que permita determinar la “quietud” en el par de bloques (A_{ij}, A'_{ij}) . El método que se propone es un conjunto de reglas obtenidas por medios estadísticos basadas en un estudio realizado sobre aproximadamente 100 bloques provenientes de diferentes escenas, que no presentan ningún tipo de evento visual y que sin embargo fueron erróneamente marcados como presentando humo. Más específicamente, se considera que una porción de escena está en quietud si los cambios de los coeficientes frecuenciales del bloque A_{ij} con respecto al bloque A'_{ij} son pequeños, tanto en número de coeficientes como en magnitud. De manera formal, las siguientes

expresiones definen un conjunto de condiciones bajo las cuales se considera un bloque en “quietud”. Primero, se define la (4.14) para contabilizar el número de coeficientes que, de un bloque a otro, sufren un cambio de magnitud igual a η .

$$\psi(A_{ij}, A'_{ij}, \eta) = \sum_{r=1}^{63} \psi_r \quad (4.14)$$

$$\text{donde } \psi_r = \begin{cases} 1, & |a_r - a'_r| = \eta \\ 0, & \text{de otro modo} \end{cases}$$

En segundo lugar, un conjunto de reglas establece como pueden presentarse estos cambios.

$$q(A_{ij}, A'_{ij}) = \left\{ \begin{array}{l} SI, \psi(A_{ij}, A'_{ij}, 2) < 4 \wedge \psi(A_{ij}, A'_{ij}, 3) = 0 \wedge \psi(A_{ij}, A'_{ij}, 4) = 0 \\ SI, \psi(A_{ij}, A'_{ij}, 2) < 3 \wedge \psi(A_{ij}, A'_{ij}, 3) < 2 \wedge \psi(A_{ij}, A'_{ij}, 4) = 0 \\ SI, \psi(A_{ij}, A'_{ij}, 2) < 3 \wedge \psi(A_{ij}, A'_{ij}, 3) = 0 \wedge \psi(A_{ij}, A'_{ij}, 4) < 2 \\ SI, \psi(A_{ij}, A'_{ij}, 2) = 0 \wedge \psi(A_{ij}, A'_{ij}, 3) < 3 \wedge \psi(A_{ij}, A'_{ij}, 4) = 0 \\ SI, \psi(A_{ij}, A'_{ij}, 2) = 0 \wedge \psi(A_{ij}, A'_{ij}, 3) < 2 \wedge \psi(A_{ij}, A'_{ij}, 4) < 2 \\ NO, otro \end{array} \right. \quad (4.15)$$

La primera regla de la (4.15) establece que hay “quietud” cuando se presenta menos de cuatro coeficientes con un cambio de magnitud de dos unidades. En siguiente regla, se establece que hay “quietud” cuando menos de tres coeficientes presentan un cambio de dos

unidades y un cambio de tres unidades, etc. Lo que implica que si existe quietud no se aplica la propuesta del método de detección.

b) *La pérdida excesiva de información en un solo bloque.* Se sabe que la aparición de humo en la escena provoca la pérdida de coeficientes frecuenciales, en particular aquellos ubicados en las altas frecuencias o aquellos cuya magnitud es pequeña. El modelo que se propone no distingue los casos en los que la pérdida se concentra en un solo coeficiente frecuencial y lo cual se ha observado no está ligado a la presencia de humo en la escena. La siguiente expresión permite determinar esto:

$$l(A_{ij}, A'_{ij}) = \begin{cases} SI, & |a_r| > 3 \wedge a'_r = 0 \text{ para } r = 1, \dots, 63 \\ NO, & \text{de otro modo} \end{cases} \quad (4.16)$$

El umbral con valor tres fue obtenido por medios estadísticos. Entonces cuando se detecta que un coeficiente frecuencial AC presenta una pérdida superior a tres unidades, indicando que sí hay pérdida excesiva de información y sobre este bloque no se aplica el método de detección propuesto.

c) *Bloques demasiado claros u oscuros.* Otro error en la detección de humo ocurre cuando el brillo del bloque a evaluar se aproxima al blanco o al negro. La causa de esto es que las muestras adquiridas para obtener el modelo propuesto provienen de escenas con buena iluminación, ver sección 3.4. Entonces, para descartar los bloques cuyo brillo no fue considerado en el modelo, se establece un rango de brillo en el que el modelo se comporta satisfactoriamente. Este se puede expresar de la manera siguiente:

$$o(A_{ij}, A'_{ij}) = \begin{cases} NO, & -60 < a_0 < 60 \\ SI, & \text{otro} \end{cases} \quad (4.17)$$

Hay que notar también, que esta condición no es redundante con la primera condición del algoritmo presentado en la figura 4.9, ya que en esta figura la condición considera los cambios de brillo entre imágenes y no los límites aceptables del brillo. Los umbrales -60 y 60 en a_0 también fueron obtenidos estadísticamente, lo que indica que si se encuentra en este rango no es un bloque muy oscuro o muy claro y sobre este bloque sí se aplica el método de detección.

Tomando en cuenta estas tres verificaciones, si el bloque a confrontar con el modelo propuesto no pasa una sola verificación, entonces simplemente la búsqueda de humo se detiene para el par de bloques en curso. El bloque en cuestión es entonces declarado como “sin presencia de humo” y se pasa a otro bloque.

4.4.2 Control de evento “humo” en un grupo de bloques

El proceso de verificación presentado en la sección anterior reduce hasta un cierto punto los errores en la detección de humo. Para dar más certeza al algoritmo propuesto se plantea eliminar los bloques que se encuentren total o parcialmente aislados. Este planteamiento está motivado por la siguiente observación: el humo se presenta la mayor parte del tiempo sobre aglomeraciones de bloques más o menos conexas. Ahora, para efectuar esta eliminación, en la literatura existen diversas herramientas pertenecientes al campo de las operaciones morfológicas. Entre ellas, la operación de “erosión” es la más adecuada a los intereses. Se propone una variación de la operación convencional, en la cual se considera que existe una matriz binaria $M=I_b$ de tamaño $m \times n$, ver (3.3) cuyos elementos m_{ij} reportarían el estado del par de bloques (A_{ij}, A'_{ij}) pertenecientes al par de imágenes (I, I') , a saber: marcado o no marcado. Solo los bloques marcados indican la presencia de humo en la escena. Además, gracias a que entre los elementos de M y el par

(I, I') existe una correspondencia 1:1, los bloques marcados en M permiten obtener la posición de los bloques de (I, I') ligados a la presencia de humo. La expresión que define la operación de erosión es la siguiente:

$$M' = M \otimes W \quad (4.18)$$

En donde \otimes es el operador de erosión y W es un elemento estructural cuadrado de tamaño 5×5 (figura 4.10). El algoritmo de erosión propuesto en este documento se efectúa de la siguiente manera: la búsqueda de bloques aislados se realiza sobre el arreglo M probando solo los elementos m_{ij} que presenten un estado "marcado". Si en el área W alrededor de m_{ij} se encuentran al menos otros tres elementos marcados, estén o no conexos, se reafirma el estado "marcado" de m_{ij} . En caso contrario m_{ij} pasa a un estado "no marcado". Tomando en cuenta que el elemento estructural W tiene un tamaño de 5×5 , el entorno de m_{ij} durante una búsqueda va desde $(i - 2, j - 2)$ a $(i + 2, j + 2)$. Además, la búsqueda no se debe desarrollar fuera de los límites de la imagen, así ésta debe respetar las siguientes condiciones: $0 < i \pm 2 < m$ y $0 < j \pm 2 < n$. El procedimiento de erosión que aquí se presenta elimina en forma satisfactoria los bloques erróneamente marcados como ligados a la presencia de humo.

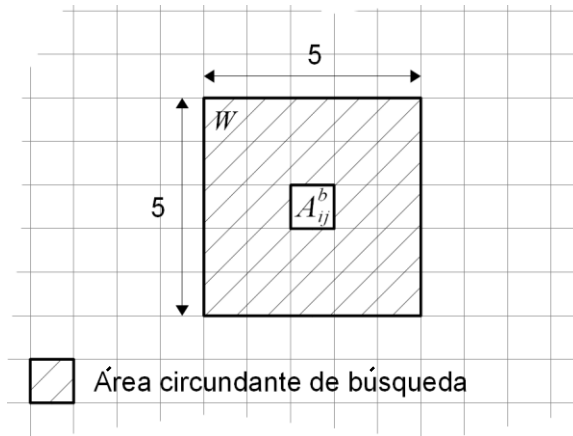


Figura 4.10 Área de búsqueda circundante a un elemento considerado como ligado a la presencia de humo en la escena

Tomando en cuenta las propuestas presentadas en esta sección, el algoritmo de detección de humo basado en un GMM, ver figura 4.9, adquiere una fase pre-procesamiento a la que se denominada “filtrado” y que corresponde con la propuesta presentada en la sección 4.4.1. En forma conceptual, esta fase recibe un par de imágenes (I, I') sobre las cuales se filtran los bloques que se consideran no aptos para aplicar el algoritmo de detección de humo. De esta forma, este filtrado entrega un par de imágenes filtradas (I_f, I'_f) sobre las cuales se aplica el algoritmo de detección basado en el GMM. Una segunda fase de eliminación de resultados errátiles se integra al final del algoritmo, el cual corresponde con el procedimiento presentado en la sección 4.4.2, ver figura 4.11, y que tiene como objetivo eliminar los bloques asilados.

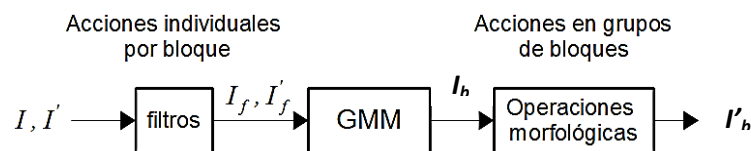


Figura 4.11 Modelo de Mezclas Gaussianas para la detección de humo translúcido con fase de filtrado y eliminación de resultados errátiles

Finalmente al método que integra la detección de humo basada en el GMM, los filtros y la operación morfológica erosión se le llama $\Phi(I, I')$ y entrega su resultado en una matriz $M' = I'_b$.

4.5 Verificación del movimiento de la masa de humo

La verificación del movimiento de la masa de humo es un elemento más que permite reforzar la certeza del método de detección. En este punto es importante considerar que el humo es una masa cuya forma evoluciona y se desplaza en forma irregular debido a diversos factores tales como su densidad, las corrientes de aire, la forma del entorno en donde se ubica la fuente de combustión, etc. pero que en cualquier caso va cubriendo la escena en algunas partes y las va descubriendo en otras. La figura 4.12 muestra un par de siluetas provenientes de la misma fuente de combustión pero adquiridas a diferentes instantes de tiempo; estas corresponden con las fotos presentadas en la figura 3.7.

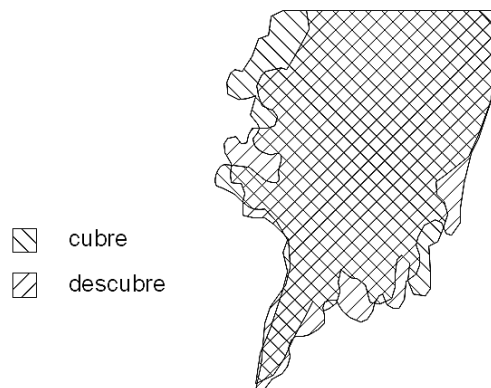


Figura 4.12 Siluetas de humo superpuestas que corresponden a una misma fuente de humo fotografiada a diferentes instantes

Con respecto a las siluetas que se muestran, a estas se le ha puesto un patrón diferente para que, una vez superpuestas, se pueda distinguir las porciones que no intersectan entre sí y que muestran las áreas que se van siendo cubiertas y descubiertas por el humo.

MODELO PROPUESTO PARA LA DETECCIÓN DE HUMO

En esta sección se propone adicionar al método de la detección de humo presentado en las secciones anteriores, la verificación del movimiento de masa de humo. Las condiciones a verificar son precisamente que la escena de humo presente, tanto regiones que se van cubriendo por el humo, como otras que se van descubriendo. Aquí es importante mencionar que el método propuesto en las secciones anteriores solo permite detectar el humo cuando va cubriendo la escena. Sin embargo, se observa que el efecto de cubrir la escena con humo es perfectamente inverso al efecto de descubrirla. Esto se corrobora analizando centenas de muestras; no se requirió un estudio tan profundo como en el caso del evento cubrir con humo (al cual también se llama aparición de humo). En la tabla 4.8 se observa el comportamiento de los coeficientes frecuenciales durante el evento desaparición de humo.

Tabla 4.8 Humo desapareciendo de la escena

Propiedad	Impacto en los coeficientes de la DCT
Brillo	Atenuación del coeficiente DC
Contraste	Amplificación de los coeficientes AC
Definición	Ganancia de coeficientes AC puede ocurrir en cualquier elemento, sin embargo es más probable en las altas frecuencias.
Similitud	Alta

Se observa que en la tabla, el impacto de la desaparición de humo en la escena es completamente inverso al descrito en la tabla 3.1 con la excepción de la similitud entre bloques. Esta última debe ser de un alto grado en ambos casos ya que la similitud entre bloques es uno de los pilares del método propuesto.

MODELO PROPUESTO PARA LA DETECCIÓN DE HUMO

Gracias al comportamiento de los coeficientes de la DCT descrito en la tabla 4.8, la detección del evento “humo descubre región” se puede obtener gracias al método $\Phi(I, I')$ con la única condición de invertir los argumentos de la siguiente manera: $\Phi(I', I)$. El resultado de la detección se repetiría también en la matriz I'_b . Ahora, para distinguir entre aparición y desaparición de humo se utilizó la siguiente disposición de los métodos con respecto a un tercer método para verificar el movimiento de humo.

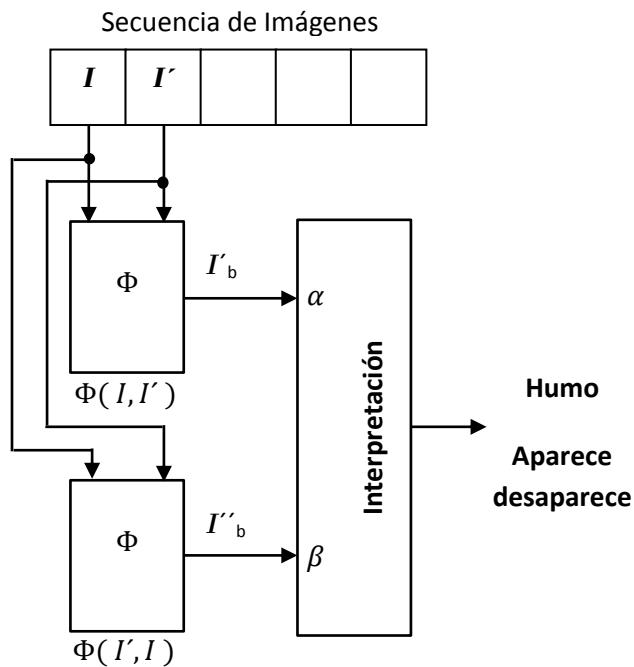


Figura 4.13 Disposición del método para verificar movimiento

Por otra parte, en un cuadro de video no siempre ocurre que el humo cubre y descubre porciones de escena, de esta manera se da un margen de tiempo para que esto suceda. Específicamente, si en un número de cuadros equivalente a 3 segundos aparece la condición de aparición y desaparición de humo, entonces se considera que el humo se encuentra en movimiento.

5 Evaluación de Resultados

El algoritmo propuesto fue evaluado usando el criterio propuesto por [22]. En este algoritmo se considera que se pueden presentar cuatro posibles escenarios: i) la detección verdadera o detección correcta, ii) detección falsa o falsa alarma, iii) las no detecciones falsas u omisión de la detección y iv) las no detecciones verdaderas o rechazos correctos.

La tarea del algoritmo es clasificar los videos de entrada en dos grupos, con humo presente (detección) o sin humo presente (no detección). Posteriormente se compara la clasificación obtenida del algoritmo contra una clasificación manual por un observador humano. De lo anterior se define a X como el conjunto de todas las secuencias usadas con humo y sin humo para evaluar al sistema, $\pi^{-1}(S)$ es un subconjunto de X , el cual sólo contiene aquellas secuencias en las cuales el sistema propuesto detecta humo, y finalmente E es un subconjunto de X que contiene aquellas secuencias que fueron clasificadas como humo por un observador humano. De lo anterior se tiene que TP (*verdaderos positivos*) es el conjunto de aquellas secuencias con humo que indican una *detección correcta*.

$$TP = \pi^{-1}(S) \cap E \quad (5.1)$$

La situación en la que una secuencia no se encuentra presente en $\pi^{-1}(S)$ ni en E , es llamada *rechazo correcto*. Un conjunto TN (*verdaderos negativos*) contiene aquellas secuencias con humo que han sido rechazadas correctamente.

$$TN = \pi^{-1}(S)^C \cap E^C \quad (5.2)$$

En el caso en que una secuencia está presente en $\pi^{-1}(S)$ y no está presente en E será considerada como una *falsa alarma*. FP (*falsos positivos*), el cual es el conjunto de las secuencias detectadas con humo de forma errónea.

$$FP = \pi^{-1}(S) \cap E^c \quad (5.3)$$

Por último, la situación en la que la secuencia está presente en E y no está presente en $\pi^{-1}(S)$, es llamada *omisión de la detección*. El conjunto FN (*falsos negativos*) contiene todas las secuencias con humo que fueron omitidas.

$$FN = \pi^{-1}(S)^c \cap E \quad (5.4)$$

En la figura 5.1 se muestra lo antes descrito.

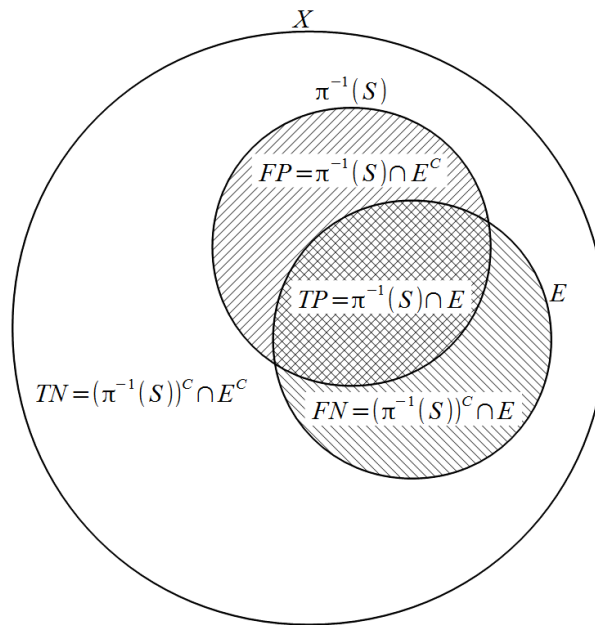


Figura 5.1 Enfoque de la teoría de conjuntos para los posibles escenarios en la detección de humo

Una vez que se han definido cada uno de los posibles escenarios, se especifican varias medidas claramente aplicables para establecer una *medida de evaluación global* del algoritmo propuesto; es decir, una medida de la correcta detección de las secuencias evaluadas, siendo la *sensibilidad* y la *especificidad* las más importantes.

La sensibilidad del esquema propuesto evalúa la calidad observada de acuerdo a una detección correcta *cd* (*correct detections*). Esto podría definirse como la tasa de positivos verdaderos *TPR* (*True Positive Rate*):

$$cd = TPR = \frac{|TP|}{|E|} = \frac{|TP|}{|TP| + |FN|} \quad (5.5)$$

Donde $|\cdot|$ denota el número de elementos en el conjunto. De forma similar la calidad del esquema con respecto a detecciones falsas *fd* (*false detections*), se puede definir como la tasa de falsos positivos *FPR* (*False Positive Rate*):

$$fd = FPR = \frac{|FP|}{|FP| + |TN|} \quad (5.6)$$

Por otra parte, la calidad del sistema con respecto a los correctos rechazos *cr* (*correct rejections*), puede ser definida como la tasa de especificidad o tasa de negativos verdaderos *TNR* (*True Negative Rates*).

$$cr = TNR = \frac{|TN|}{|FP| + |TN|} = 1 - FPR \quad (5.7)$$

La calidad del sistema con respecto a la omisión de la detección *md* (*missed detections*), puede ser definida como la tasa de falsos negativos *FNR* (*False Negtive Rate*).

$$md = FNR = \frac{|FN|}{|TP| + |FN|} = 1 - TPR \quad (5.8)$$

Todas las medidas toman valores dentro del intervalo $[0,1]$. Para obtener resultados óptimos, cd y cr deberán ser, en la medida de lo posible, cercanos a 1. Mientras que fd y md deben ser lo más pequeño posible. En la figura 5.2 se muestran algunos casos especiales, si cd y cr son igual a 1, entonces fd y md son igual a 0. Así, el resultado del desempeño del algoritmo será declarado como un *sistema ideal*, por lo menos para las secuencias de prueba. Esto significa que todas las secuencias han sido correctamente clasificadas como humo o no humo. El desempeño del algoritmo es considerado como *sistema sin fallo*, si $cd = 1$ (o $md = 0$). Un buen desempeño del algoritmo no contiene errores y matemáticamente implica $E \subset \pi^{-1}(S)$. El desempeño del algoritmo es considerado como *malo* si $cd = 0$ (o $md = 1$), lo que implica $\pi^{-1}(S) \cap E = \emptyset$. El último caso del desempeño del algoritmo es el *peor*. Esto sucede cuando $cd = cr = 0$ ($fd = md = 1$), lo cual implica que $\pi^{-1}(S) = E^C$.

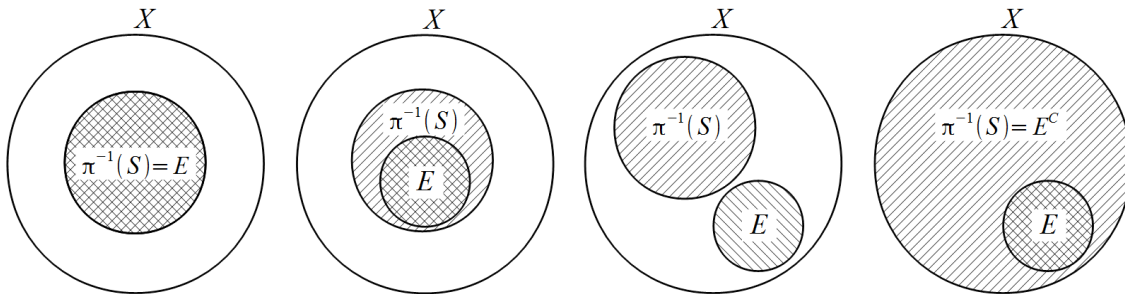


Figura 5.2 Interpretación de varios tipos de resultados de desempeño del algoritmo de acuerdo a su calidad

Algunas otras medidas, bien conocidas en la teoría de detección de señales y análisis de errores, también son adecuadas para evaluar el desempeño del algoritmo. La *precisión acc* (*accuracy*) se define como el grado de proximidad de las mediciones de una cantidad determinada a su valor (verdadero) real. En términos del desempeño del sistema,

también podría ser definida como una relación entre la correcta detección de secuencias y el número total de secuencias.

$$acc = \frac{|TP| + |TN|}{|X|} = \frac{|TP| + |TN|}{|TP| + |TN| + |FP| + |FN|} \quad (5.9)$$

El valor predictivo positivo o valor de precisión, reproducibilidad o *repetibilidad ppv* (*positive predictive value*), corresponde al grado en el que las mediciones repetidas en condiciones similares muestran los mismos resultados.

$$ppv = \frac{|TP|}{|TP| + |FP|} \quad (5.10)$$

La correlación de Mathews *mcc* es una medida de calidad para el problema de una clasificación binaria. Esta toma en cuenta, tanto la detección de los falsos y verdaderos, como los positivos y negativos. Además se considera generalmente como una medida equilibrada, se puede usar incluso aunque las clases son de tamaños muy diferentes.

$$mcc = \frac{|TP||TN| - |FP||FN|}{\sqrt{(|TP| + |FP|)(|TP| + |FN|)(|TN| + |FP|)(|TN| + |FN|)}} \quad (5.11)$$

La *medida de evaluación global* de los resultados obtenidos sobre el esquema propuesto, usando los criterios de [22] ya explicados, se muestran en la tabla 5.1.

Tabla 5.1 Evaluación de resultados del algoritmo propuesto, usando medidas de evaluación global

TP	TN	FP	FN	cd	fd	cr	md	acc	ppv	mcc
1720	991	18	93	0.9487	0.0178	0.9821	0.0513	0.9606	0.9896	0.9172

Para obtener estos resultados se analizó una decena de videos con alrededor 4000 imágenes cada uno. Para estos videos, aproximadamente la mitad de la secuencia no presenta humo y la otra mitad sí. Es importante notar que el mejor desempeño de un método de detección se puede medir de diferentes maneras. Una de ellas es observando el parámetro cd , también llamado TPR. El valor ideal es 1.0. Este parámetro se complementa con fd , el cual idealmente deber valer 0.0. El método de detección que se propone muestra así un buen desempeño ya que estos valores están muy próximos de los ideales.

5.1. Comparación con otros trabajos

La evaluación de resultados debe incluir al menos dos elementos básicos. El primero es el uso de un método objetivo de evaluación, como el que se acaba de presentar. Además, es preferible que el método de evaluación sea de uso común en la comunidad que se dedica a la detección de humo. Y en segundo lugar, la evaluación arrojada por tal método debe ser comparada con trabajos que persiguen objetivos similares. Los aspectos del desempeño que se emplearan para comparar la propuesta con aquellas que se pueden encontrar en la literatura, serán los siguientes: *tasa de detecciones correctas* (cd), *tasa de detecciones incorrectas* (fd) y *la cantidad de píxeles procesados por segundo*.

El estudio comparativo que se presenta a continuación es tomado del análisis del estado del arte presentado en el capítulo 2 y que se presenta resumido en la tabla 2.1. Hay notar que no todas las propuestas analizadas ofrecen una evaluación objetiva del desempeño de sus métodos, ni en términos de la tasada de detecciones correctas u otro similar, ni en términos de la cantidad de información que pueden procesar por segundo.

5.1.1 Tasa de detecciones correctas

Esta comparación solo toma los valores cd (5.5) disponibles en las publicaciones que se han analizado, o bien aquellos valores para cd que pueden ser deducidos utilizando las (5.7) y (5.8). Se observa que las referencias [3] y [4] son aquellas que presentan el mejor desempeño ya que son las que más se aproximan al valor ideal de 1.0. El trabajo propuesto, a la cual se denomina [pro] en este documento, no está lejos de este ideal.

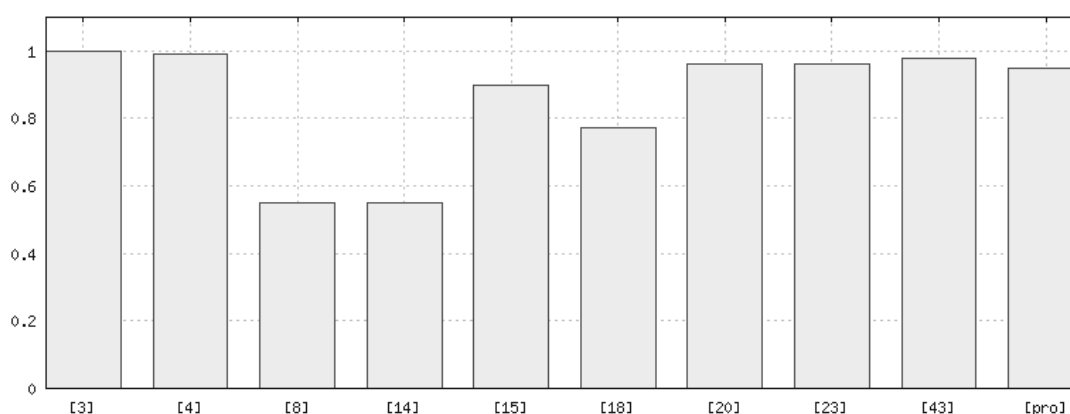


Figura 5.3 Tasa de detecciones correctas

5.1.2 Tasa de falsas detecciones

Esta comparación solo toma los valores fd (5.6) disponibles en las publicaciones que se han analizado, o bien aquellos valores para fd que pueden ser deducidos utilizando las (5.7) y (5.8). Se observa que las referencias [3] y [4] son aquellas que presentan el mejor desempeño ya que son las que más se aproximan al valor ideal de 0.0. Una vez más el trabajo propuesto, asociada a la etiqueta [pro], alcanza un valor próximo a 0.0.

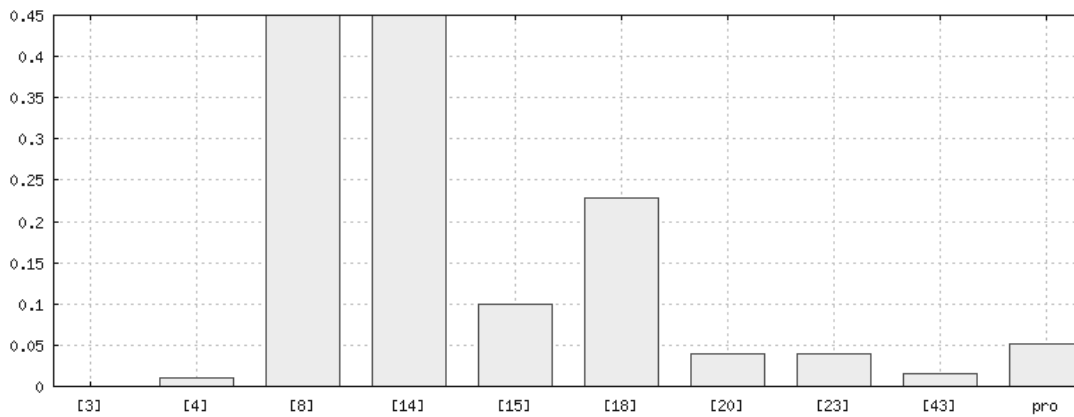


Figura 5.4 Tasa de falsas detecciones

5.1.3 Cantidad de píxeles procesados por segundo

Tomando en cuenta que los trabajos que reportan resultados de velocidad de operación, utilizan tamaños de imagen variados y procesan un número de imágenes por segundo muy diverso, en esta sección se intenta normalizar estos resultados a una representación que sea comparable una con otra. Esta normalización consiste en medir cuantos píxeles por segundo son capaces de procesar cada uno de los métodos propuestos incluyendo el nuestro. De esta manera se observa que el trabajo propuesto supera, en velocidad a todos los métodos propuestos. Es importante mencionar que en este estudio no se toma en cuenta la velocidad del CPU o algún otro parámetro físico del sistema de cómputo en donde fueron realizadas las pruebas, ya que son pocos los trabajos que reportan las condiciones bajo las cuales se efectuaron las medidas presentadas. Ver tabla 2.1, columna “cantidad de datos procesados”

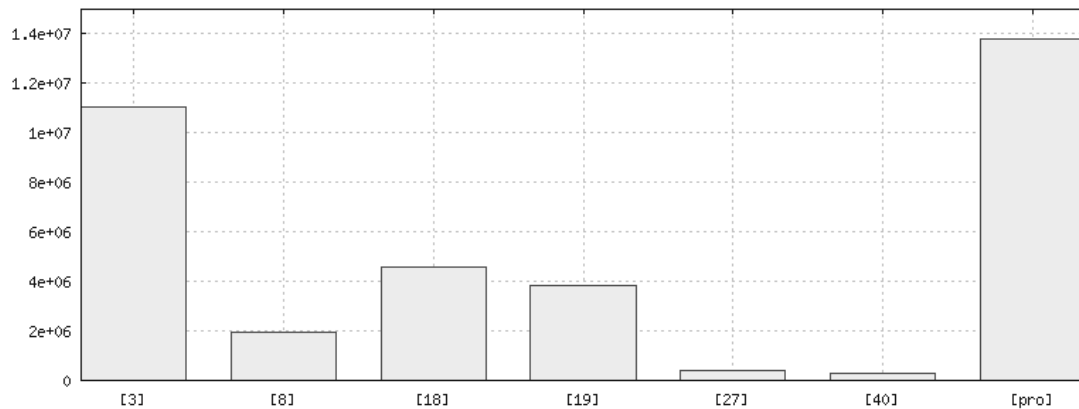


Figura 5.5 Cantidad de pixeles procesados por segundo

La evaluación de los resultados muestra que con el sistema propuesto es posible detectar a presencia de humo, más rápido que otros algoritmos reportados previamente además de economizar el CPU de forma importante ya que lo propuesto no demanda la decodificación total del video, proceso que puede consumir una parte importante de los recursos de cálculo de un sistema, ver figura 2.2.

Por otra parte, llevar a la práctica los métodos considerados en la figura 5.5 es en realidad más costoso que lo que se presenta, lo cual no ocurre con lo propuesto ya que no requiere efectuar la decodificación total del video sino solo una pequeña parte, específicamente la codificación entrópica.

Es importante mencionar que el tiempo de procesamiento del algoritmo es de aproximadamente 150 ms, dependiendo de la complejidad de las imágenes, y si le suma los 500 ms que existe entre imagen e imagen da como resultado 650 ms para la detección del humo.

6 Conclusiones

En la presente tesis, se propuso un algoritmo para el reconocimiento de humo translucido grisáceo azulado que presenta un bajo costo computacional, de tal manera que el método pueda procesar en directo el video proveniente de cámaras de video vigilancias IP. El objetivo fue alcanzado ya que el algoritmo propuesto puede procesar en promedio un par de imágenes de alta resolución en alrededor 100 ms con una computadora de tipo Laptop dotada de un procesador 5i de cuatro núcleos corriendo a 1.7 GHz. Es importante mencionar que los programas que implementan el algoritmo propuesto no están optimizados en lo absoluto, por lo cual es posible alcanzar un desempeño significativamente mayor.

El algoritmo propuesto no utiliza píxeles en la detección de humo, sino la información que se encuentre disponible en el flujo de video codificado. En este caso, el objetivo fue alcanzado ya que el video no se descomprimió y solo se utilizaron los coeficientes frecuenciales de la DCT para afectar el proceso de detección.

También, el algoritmo desarrollado utiliza imágenes de alta definición a un ritmo de presentación de 30 imágenes por segundo. El algoritmo de detección presenta una buena precisión. Las pruebas presentadas en el capítulo anterior muestran que el algoritmo presenta un buen desempeño con respecto al único trabajo similar que aborda imágenes de alta definición. La detección de humo mediante video solo se efectúa en distancias cercanas no mayores a 100 metros. Las pruebas desarrolladas demuestran que es posible detectar humo translucido grisáceo a tales distancias.

6.1 Contribuciones

En este documento se presentó un algoritmo para la detección temprana de incendios en video que ha sido comprimido alrededor de la transformada DCT (*Discret Cosine Transform*), tal y como aquel comprimido bajo las normas JPEG y MPEG I/II. Se utiliza el término “detección temprana” porque el algoritmo propuesto se enfoca en la detección de humo grisáceo claro translucido, el cual comúnmente aparece en las primeras etapas de la combustión[18]. Este trabajo de tesis hace las siguientes contribuciones:

1. Se propone un método que detecta humo mediante video codificado bajo las normas JPEG o MPEG utilizando las informaciones directamente disponibles en el flujo binario, específicamente, los coeficientes frecuenciales de la DCT. De esta manera, al no efectuar una costosa descompresión total de flujo de video, el método es de bajo costo computacional; inferior a la gran mayoría de métodos propuestos hasta la fecha.

2. Se crea un modelo que describe la presencia de humo en la escena, el cual se apoya de los *Gaussian Mixture Models (GMM)*. El modelo propuesto describe los cambios que sufren diversas propiedades de la imagen con la aparición del humo, entre otras, los cambios de brillo, de contraste y de nitidez. Los cambios de estas propiedades se obtienen comparando los bloques de coeficientes DCT de una imagen I con respecto a los bloques de una imagen I' ; estas imágenes están separadas por un cierto período de tiempo.

3. Se proponen dos métodos para medir los cambios que se pueden presentar a través de tiempo en las propiedades de una imagen, a saber: el contraste y la definición. Estos métodos operan en el ámbito de un bloque. Es importante mencionar que el método de detección de humo también se verifica los cambios de brillo, sin embargo, el método para medirlo es tan obvio que no se puede citar como una contribución.

4. Se propone un método para verificar que los bloques empleados para medir los cambios antes descritos presenten la misma información visual. En efecto, no es válido declarar cambios de brillo, de contraste, de definición u otro si antes no se ha verificado que las porciones de imagen en curso de prueba correspondan a la misma parte de la escena. Esta verificación está motivada por el hecho que un objeto solido (por ejemplo una persona) atravesando la escena puede provocar los mismo cambios de contraste, brillo, etc. que el humo. Es importante mencionar que de los diversos trabajos que miden la pérdida de definición (o pérdida de información en las altas frecuencias) debido a la aparición de humo, ninguno efectúa una verificación similar.

5. Este modelo presenta una buena precisión, la cual se incrementa con la aplicación de un método de erosión, similar a la conocida erosión morfológica que elimina los bloques aislados y probablemente no ligados a la presencia de humo. El modelo se complementa con otro que describe la forma en la que la mancha de humo debe moverse a través del tiempo. Esto ayuda a reducir considerablemente el número de falsos positivos.

6.2 Trabajo futuro

Mejorar el desempeño de la propuesta en términos de la tasa de detecciones correctas, de la tasa de falsas detecciones y mejorar la velocidad, lo que implica optimizar el algoritmo. Por otra parte, es interesante tomar en cuenta las informaciones adicionales a la DCT que ofrecen los videos MPG o H.264/AVC, tales como los vectores de movimiento o la información de color. Esto con el fin de mejorar la certeza del desempeño de detección. Es también interesante intentar otros caminos diferentes al empleado en esta tesis de maestría, tal y como el que se propuso en [32] y que es complementario a este trabajo.

7 Referencias

- [1] Acklam P. J., "Lower tail quantile for standard normal distribution function", www.math.uio.no/~jacklam, 2002.[55]
- [2] Aird et al., "Fire detection algorithm", United States Patent: 6956485, October 18, 2005.
- [3] Borges P. V., Izquierdo E., "A probabilistic Approach for Vision-Based Fire Deteccion in Videos", IEEE Transactions on Circuits and Systems for Technology, vol. 20, No. 5, may 2010.
- [4] Celik T., Ozkaramanli H., Demirel H., "Fire and Smoke Detection without Sensors: Image Processing Based Approach", EURASIP 2007.
- [5] Cetin A. E. and Ansari R., "Signal recovery from wavelet transform maxima," IEEE Trans. On Signal Processing, vol. 42, pp. 194-196, 1994.[47]
- [6] Changwoo H., Gwanggil J. and Jechang J, "Vision-Based Smoke Detection Algorithm for Early Fire Recognition in Digital Video Recording System", Signal-Image Technology and Internet-Based Systems (SITIS), 2011.
- [7] Chen T., Yin Y., Huang S., and Yen Y. "The smoke detection for early fire-alarming system based on video processing," in Proc. The International Intelligent Information Hiding and Multimedia Signal, pp. 247-430, 2006.[33][17]
- [8] Chunyu Y., Yongming Z., Jun F., Jinjun W., "Texture analysis of smoke for real-time fire detection", International Workshop on Computer Science and Engineering, 2009.
- [9] Collings R. T., Lipton A. J., and Kanade T. , "A system for video surveillance and monitoring," in 8th Int. Topical Meeting on Robotics and Remote Systems. 1999, American Nuclear Society. [46].

- [10] Davis, B.J; Nawab, S.H. The Relationship of Transform Coefficients for Differing Transform and/or Differing Subblock Sizes. IEEE Trans. Sign. Process. 2004, 52, 1458-1461.[22]
- [11] Davis I. W. and Bobick A. F., “The cognition of human movement using temporal templates”, IEEE Transaction on Pattern Analysis and Machine Intelligence, 23 (3):257-267, 2001.[36]
- [12] Gale T., Main I. and the GTK team, “GTK+2.0 Tutorial”, <http://developer.gnome.org/gtk-tutorial/stable/> [51].
- [13] Gonzalez R. C. and Woods R. E. , Digital Image Processing. Reading, MA: Addison-Wesley, 1992.[31]
- [14] Gonzalez-Gonzalez R. et als. "Wavelet-based smoke detection in outdoor video sequences", Circuits and Systems (MWSCAS), 2010.
- [15] Gubbi J., Marusic S. , Palaniswami M., Smoke detection in video using wavelets and support vector machines”, Fire Safety Journal, 44 (2009) pág. 11 10-1115. [16]
- [16] Haralick R. M. , Shanmugam K., and Dinstein I. , “Textural Features for Image Classification”, IEEE Transaction on Systems, Man, and Cybernetics, 1973, pp. 610-621.[34]
- [17] Haynes, "Image analysis for smoke detection, United States Patent: 7804522, September 28, 2010.
- [18] Ho C., Kuo T, “Real-time video-based fire smoke detection system”, Advanced Intelligent Mechatronics 2009.
- [19] Hou, "Multi-wavelength video image fire detecting system", United States Patent: 7786877, August 31, 2010.

- [20] Independent JPEG Group, cjpeg software, march 1998.[56]
- [21] Jianmin, J.; Guocan, F. The Spatial Relationship of DCT Coefficients between a Block and its Sub-Blocks. IEEE Trans. Sign. Process. 2002, 50, 1160-1169.[21]
- [22] Jakovcevic, T.; Setic L.; Stipanicev D. Krstinic D.; Wildfire Smoke-Detection Algorithms Evaluation. In Proceedings of International Conference on Forest Fire Research, Coimbra, Portugal, 15–18 November 2010.
- [23] Jianning C., Chengdong W., Yunzhou Z., “Early Fire Disaster Features Research and Detection Method Design in Video Images”, Symposium on Photonics and Optoelectronics (SOPO), 2011.
- [24] Khayam S. A., Discret Cosin Transform (DCT): Theory and Application, Department of Electrical & Computer Engineering, Michigan University, 2003[49]
- [25] Kim D. K.and Wang Y., "Smoke detection in video", World Congress on Computer Science and Information Engineering, 2009, San Francisco, USA.
- [26] Kirill Kochetkov, "Playing HD Content on AMD-based Machines", <http://ixbtlabs.com/articles2/video/hdvideodisk1-page1.html> [43]
- [27] Lai C., Yang J., "Advanced real time fire detection in video surveillance system", ISCAS, 2008.
- [28] Ma L., Kaihua Wu K., Zhu L., Fire Smoke Detection in Video Images Using Kalman Filter and Gaussian Mixture Color Model, AICI, 2010.
- [29] Macdonald P. , J. Du, “Package 'mixdist', Finit Distribution Models, version 0.5-4”, Licence GPL, CRAN, 2012.[48]
- [30] Manolakis D. G., Ingle V. K., and Kogon S. M., Statistical and Adaptive Signal Processing. New York: McGraw-Hill, 2000.[32]

- [31] Maruta H., Nakamura A., Kurokawa F., "A new approach for smoke detection with texture analysis and support vector machine", ISIE, 2010.
- [32] Millán-García L., Sanchez-Perez G., Nakano M., Toscano-Medina K., Perez-Meana H. and Rojas-Cardenas L., An Early Fire Detection Algorithm Using IP Cameras, Sensors Journal, Vol. 12, 2012, pp. 5670-5686.[20]
- [33] Qichang D., Shuen L. "Research on Application of Cloud Model in Fire Alarm System", Proceedings of the Second International Symposium on Test Automation & Instrumentation, vol 1:411-414, 2008.[35]
- [34] Redhat Inc., cygwin, <http://www.cygwin.com/> [53]
- [35] Rizzoti et al., "Process and device for detectin fires based on image analisis", United States Patent: 6937743, August 30, 2005.
- [36] Rojas-Cardenas L, Millan-Garca L., Rojas-Cardenas G. : Low Cost Visual Event Detection over JPEG Flows. CONIELECOMP 2006.[54]
- [37] Sadok et al., "Distinguishing between fire and non-fire conditions using cameras", United States Patent: 7245315, July 17, 2007.
- [38] Sadok et al., "Detecting Fire using cameras", United States Patent: 7256818, August 14, 2007.
- [39] Stuffer C. and Grimson W.E.L., "Adaptive background mixture models for realtime tracking," in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1999, pp. 246-252.[37]
- [40] Surit S., Chatwiriya W., "Forest Fire Smoke Detection in Video Based on Digital Image Processing Approach with Static and Dynamic Characteristic Analysis", International Conference on Computers, Networks, Systems and Industrial Engineering.

- [41] Tan K.S., and Isa N. A. M. , “Color image segmentation using histogram thresholding – fuzzy c-means hybrid approach,” *pattern Recognitions*, vol. 44, pp. 1 – 15 july 2010.[30]
- [42] Theodoridis S.and Koutroumbas K., *Pattern Recognition*. New York: Academic, 2006.[44]
- [43] Tian H., Li W., Ogunbona Ph., D. T. Nguyen, C. Zhan, “Smoke Detection in Video Using Non-Redundant Local Binary Pattern-Based Features”, *IEEE Workshop on Multimedia Signal Processing 2011*.
- [44] Tong H.H.Y.and Venetsanopoulos A.N. , "A Perceptual Model for JPEG Applications Based on Block Classification, Texture Masking, and Luminance Masking", *International Conference on Image Processing (ICIP 98)*, 1998.[50]
- [45] Toreyin B. U. , Dedeoglu Y., and Cetin A. E., “Contour Based Smoke Detection in Video Using Wavelets”, *14th European Signal Processing Conference*, 2006.[19]
- [46] Truong T., Yongmin Kim Y., Jongmyon Kim J., "Fire Detection in Video Using Genetic-Based Neural Networks", *ICISA*, 2011.
- [47] Ugur Töreyn B., Yigithan Dedeoglu, A. Çetin E., “Contour Based Smoke Detection in Video Using Wavelets”, *EUSIPCO 2006*, Septiembre 2006.[41]
- [48] Ugur Töreyn B., Yigithan D., Çetin E., “Wavelet Based Real-Time Smoke Detection in Video”, *EUSIPCO 2005*, Septiembre 2005. [40]
- [49] Vasyukov V., Kalennikova E., "An adaptive procedure of smoke and background discrimination in the early fire detection video system", *IFOST*, 2011.
- [50] Wallace, G. *The JPEG Still Picture Compression Standard*, *Communications of the ACM*, vol.34, No.4, pp. 30-44,1991.[44]
- [51] www.axis.com [38]

- [52] www.ee.bilkent.edu.tr/~signal/VisiFire/Demo/SampleClips.html [45]
- [53] www.trendnet.com[39]
- [54] Xu Z., Xu J., “Automatic Fire Smoke Detection Based on Image Visual Features”, 2007 International Conference on Computational Intelligence and Security Workshops, 2007. [18]
- [55] Zakrzewski et al. "Method for detection and recognition of fog presence within an aircraft compartment using video images", United States Patent: 7505604, March 17, 2009
- [56] Zakrzewski et al. "Method for distinguishing a smoke condition from a dust condition using video images", United States Patent: 7729510, June 1, 2010
- [57] Zhengguang Xu, Jialin Xu, “Automatic Fire Smoke Detection Based on Image Visual Features”, CISW 2007, Diciembre 2007.[42]