

Universidad Politécnica de Valencia

Departamento de Comunicaciones



Tesis Doctoral

**Técnicas de análisis de secuencias de vídeo.
Aplicación a la restauración de películas antiguas**

Presentada por:

Valery Naranjo Ornedo

Dirigida por:

Dr. Antonio Albiol Colomer

Valencia, 2002.

A Luis y a Fran

“La mera formulación de un problema suele ser más esencial que su solución, la cual puede ser una simple cuestión de habilidad matemática o experimental. Plantear nuevas preguntas, nuevas posibilidades, contemplar viejos problemas desde un nuevo ángulo, exige imaginación creativa y marca adelantos reales en la ciencia.”

Albert Einstein

Agradecimientos

Es muy difícil mostrar mi agradecimiento, con unas simples palabras, a todas aquellas personas que han hecho que haya llegado hasta aquí, aún así, no quería dejar pasar la oportunidad de intentarlo.

En primer lugar quiero mostrar mi agradecimiento a Antonio Albiol, que ha sido no sólo mi director de tesis, sino también mi amigo, y mi maestro en todo lo que sé de procesado de señal.

A mi familia y amigos por estar ahí siempre que los necesito, sin esperar nada a cambio, y sobre todo, por tener fe en mí. A Luis, mi marido, que siempre me apoya y me ayuda en todo, y hace que todos los esfuerzos tengan sentido.

A mis compañeros del Departamento de Comunicaciones que me han echado una mano en esta empresa: a José Manuel, por su paciencia, sus consejos y su ayuda desinteresada e inestimable; a Luis Vergara por tantas dudas de tratamiento de señal resueltas, a María y Ángel por sus observaciones y revisiones, a Paco y Pablo por sus consejos lingüísticos, y a Juan Carlos por sus consejos burocráticos. Además, no puedo olvidar a todas la personas que han trabajado en el Grupo de Tratamiento de Imagen, aportando su trabajo, directa o indirectamente, al resultado final de esta tesis: Alberto Albiol, Josep, David, Esther, Inma, Pepa, Álvaro, Toñi, Manolo, Rafa, Guillermo, Jesús Molina y especialmente Jesús Angulo. A todos ellos, también gracias por el apoyo y el cariño que siempre me han demostrado.

Finalmente, gracias a Canal 9 Televisión Valenciana, a Domingo Santander y a Daniel Boluda por el material audio-visual cedido, que ha sido de gran utilidad en el desarrollo de este trabajo.

Resumen

La presente tesis propone una serie de técnicas de análisis de secuencias de vídeo con el fin de aplicarlas a la restauración de películas antiguas. Dicha restauración consiste en eliminar, manteniendo la máxima cantidad posible de información original en buen estado, todos aquellos defectos que, debidos al paso del tiempo, la mala conservación de las películas, y los errores mecánicos, hacen que la calidad de la proyección quede degradada notablemente.

De todos los defectos presentes en una película antigua, en esta tesis nos hemos centrado en el tratamiento de algunos de ellos, aportando soluciones novedosas a los problemas que nos han surgido. Los aspectos tratados han sido:

- Segmentación de las secuencias degradadas en planos, como paso previo a la restauración. Se han presentado una serie de métodos, basados en diferencias de luminosidad entre imágenes de la secuencia comprimida, que consiguen detectar las transiciones abruptas o cortes con una elevada probabilidad de detección y baja probabilidad de falsa alarma, incluso en el caso de imágenes de baja calidad, que es el que nos ocupa. También se presentan métodos para la detección de transiciones graduales, como fundidos y cortinillas.
- Corrección de las variaciones de intensidad entre fotogramas del mismo plano. Para resolver este defecto, típico de las películas en blanco y negro, se proponen dos métodos: uno basado en la estimación de parámetros para modelar la media y la varianza del plano, y otro basado en la ecualización de histograma de las diferentes imágenes, para conseguir que dicho histograma no varíe rápidamente a lo largo del plano.
- Reducción del ruido de grano dentro de cada plano. Se propone un filtrado FIR muy eficiente computacionalmente que reduce notablemente el ruido estacionario, basándose en la respuesta espacio-temporal del sistema visual humano, y consiguiéndose un buen factor de reducción, sin introducir una degradación visible.
- Detección e interpolación de defectos incorrelados en el tiempo (manchas). Se presentan tres métodos que se basan en operadores morfológicos y no realizan estimación de movimiento. Los métodos propuestos se utilizan tanto en la fase de detección como en la de interpolación. Para restaurar este tipo de defectos se han explotado las características que presentan los mismos: alta variación temporal, alto contraste y baja varianza espacial.

Abstract

In this PhD thesis, different methods to analyse video sequences are proposed, with the aim of applying them to the restoration of old movies. This restoration consists of removing those artifacts that, due to the material degradation along time, the bad conservation of the movies, and mechanical errors, lead to an important degradation of film quality. The restoration process must protect the correct original information as much as possible. We have focused in the treatment of some of these artifacts, contributing with novel solutions to the problems that have arisen us. The studied subjects have been:

- Segmentation in shots of degraded sequences has been taken as a starting point of the restoration process. We have presented several methods, based on brightness differences between images that belong to the compressed sequence. These methods are able to detect the abrupt transitions or cuts with a high detection probability and a low false alarm probability, even in the case of low quality images. We also present some methods for detecting gradual transitions, like dissolves and wipes.
- Flicker correction. In order to solve this artifact, very common in black and white movies, two methods are proposed: the first one based on an affine intensity transformation which reduce the random and periodic variations of mean and variance, calculating the transformation parameters with a temporal variation model of both statistics. The second method tries to improve the visual results using a nonlinear transformation based on an histogram matching. The purpose is to change each image histogram into a target histogram calculated as average of the image histogram and its neighbours.
- Noise reduction inside each shot. We present a FIR filter, with a high computational efficiency, that reduces significantly the stationary noise. This filter is based on the spatio-temporal response of human visual system, achieving a good reduction factor, without introducing a visible degradation.
- Detection and interpolation of missing data (blotches). Three methods are presented based on morphological operators without motion estimation. The proposed methods are used as in the detection phase as in the interpolation one. In order to restore this kind of artifacts we have exploited the blotches characteristics: high temporal variation, high contrast and low spatial variance.

Resum

La present tesi proposa una sèrie de tècniques d'anàlisi de seqüències de vídeo a fi d'aplicar-les a la restauració de pel·lícules antigues. La dita restauració consisteix a eliminar, mantenint la màxima quantitat possible d'informació original en bon estat, tots aquells defectes que, deguts al pas del temps, la roïna conservació de les pel·lícules, i els errors mecànics, fan que la qualitat de la projecció quede degradada notablement.

De tots els defectes presents en una pel·lícula antiga, en aquesta tesi ens hem centrat en el tractament d'alguns d'ells, aportant solucions noves als problemes que ens han sorgit. Els aspectes tractats han sigut:

- Segmentació de les seqüències degradades en plans, com a pas previ a la restauració. S'han presentat una sèrie de mètodes, basats en diferències de lluminositat entre imatges de la seqüència comprimida, que aconsegueixen detectar les transicions abruptes o talls amb una elevada probabilitat de detecció i baixa probabilitat de falsa alarma, inclús en el cas d'imatges de baixa qualitat, que és el que ens ocupa. També es presenten mètodes per a la detecció de transicions graduals, com esvaïments i cortinetes.
- Correcció de les variacions d'intensitat entre fotogrames del mateix pla. Per a resoldre aquest defecte, típic de les pel·lícules en blanc i negre, es proposen dos mètodes: un basat en l'estimació de paràmetres per a modelar la mitjana i la variància del pla, i un altre basat en l'equalització d'histograma de les diferents imatges, per a aconseguir que el dit histograma no varie ràpidament al llarg del pla.
- Reducció del soroll de gra dins de cada pla. Es proposa un filtrat FIR molt eficient computacionalment que redueix notablement el soroll estacionari, basant-se en la resposta espai-temporal del sistema visual humà, i aconseguint-se un bon factor de reducció, sense introduir una degradació visible.
- Detecció i interpolació de defectes incorrelats en el temps (taques). Es presenten tres mètodes que es basen en operadors morfològics i no realitzen estimació de moviment. Els mètodes proposats s'utilitzen tant en la fase de detecció com en la d'interpolació. Per a restaurar aquest tipus de defectes s'han explotat les característiques que presenten els mateixos: alta variació temporal, alt contrast i baixa variància espacial.

Índice general

1. Introducción	1
1.1. La restauración de películas antiguas	1
1.1.1. Introducción	1
1.1.2. La degradación de una película	2
1.1.3. Estado del arte	3
1.2. Guía de la tesis	5
1.2.1. Objetivos	5
1.2.2. Metodología de trabajo	6
1.2.3. Estructura de la tesis	6
2. Segmentación de secuencias	9
2.1. Introducción	9
2.1.1. Necesidad de la segmentación temporal	9
2.1.2. Definición de plano	10
2.2. Estado del arte	11
2.2.1. Detección de cortes	12
2.2.2. Detección de transiciones graduales	18
2.2.3. Medida de la efectividad de los algoritmos	20
2.2.4. Mejoras de la detección	22
2.3. Esquema general del proceso de segmentación	22
2.3.1. Vídeo comprimido	23
2.4. Detección de cortes	26
2.4.1. Cálculo de métricas	26
2.4.2. Procesado temporal	37
2.4.3. Elección de los umbrales: Método experimental	47
2.5. Localización de fundidos	51
2.5.1. Definición	51
2.5.2. Método de detección de fundidos	52
2.6. Localización de cortinillas	58
2.6.1. Definición	58
2.6.2. Método de detección de cortinillas	61
2.7. Resultados obtenidos	72
2.7.1. Parámetros de medida	72

2.7.2.	Resultados de la detección de cortes	74
2.7.3.	Resultados de la detección de transiciones graduales	76
2.7.4.	Comparación con los resultados publicados de otros métodos	77
3.	Corrección de las variaciones de intensidad	79
3.1.	Introducción	79
3.1.1.	Planteamiento del problema	79
3.1.2.	Estado del arte	79
3.1.3.	Objetivos	81
3.2.	Método de modelado de la variación temporal	81
3.2.1.	Idea básica	81
3.2.2.	Estimación de parámetros	82
3.2.3.	Resultados obtenidos	87
3.3.	Método de cambio de histograma	90
3.3.1.	Idea básica	90
3.3.2.	Resultados obtenidos	93
3.4.	Conclusiones	93
4.	Reducción de ruido	95
4.1.	Introducción	95
4.2.	Estado del arte	97
4.2.1.	Filtros promediadores	97
4.2.2.	Filtros de estadísticos ordenados	101
4.2.3.	Filtros bayesianos	102
4.2.4.	Filtros multiresolución	103
4.3.	Filtrado espacio-temporal lineal	104
4.3.1.	Idea básica	104
4.3.2.	Implementación eficiente	111
4.3.3.	Cálculo del factor de reducción de ruido	117
4.3.4.	Resultados obtenidos	121
4.4.	Conclusiones	122
5.	Defectos incorrelados en el tiempo	125
5.1.	Introducción	125
5.2.	Estado del arte	126
5.2.1.	Técnicas de detección de manchas	127
5.2.2.	Técnicas de interpolación de manchas	131
5.3.	Detección de manchas	133
5.3.1.	Método general	133
5.3.2.	Apertura (cierre) temporal	135
5.3.3.	Apertura (cierre) por reconstrucción	141
5.3.4.	Apertura (cierre) por λ -reconstrucción	149
5.3.5.	Comparación entre los métodos de detección	153

5.3.6. Detección de zonas de alto contraste: <i>h</i> -reconstrucción	154
5.3.7. Obtención de la máscara de manchas final	157
5.4. Interpolación de manchas	159
5.4.1. Métodos temporales de interpolación	159
5.4.2. Métodos espaciales de interpolación	160
5.4.3. Comparación entre métodos espaciales y temporales	162
5.5. Resultados obtenidos	165
5.6. Conclusiones	165
6. Sistema Global	169
6.1. Introducción	169
6.2. Diagrama de bloques del sistema	169
6.3. Resultados	172
7. Conclusiones y líneas futuras	177
7.1. Introducción	177
7.1.1. Conclusiones a la segmentación de secuencias	177
7.1.2. Conclusiones de la corrección de defectos	178
7.2. Líneas futuras	180
A. Descripción de las secuencias utilizadas	183
A.1. Introducción	183
A.2. Secuencias y su descripción	183

Capítulo 1

Introducción

1.1. La restauración de películas antiguas

1.1.1. Introducción

Uno de los grandes problemas al que se enfrenta la sociedad actual es la preservación de su patrimonio cultural, el cual se presenta bajo diferentes formas: pintura, música, cine, literatura, fotografía, etc. Algunas de estas formas presentan problemas para su conservación debido al envejecimiento del soporte en que originalmente fueron almacenadas. De entre los soportes más sujetos a degradación se encuentran las películas de cine y las grabaciones de sonido.

El 90% de las películas hechas antes de 1930 y el 50% de las anteriores a 1950 se han perdido totalmente [19], 2,2 billones de metros de película según la UNESCO (1992) [127]. Gran parte de esta destrucción ha sido debida al mal almacenamiento del nitrato de celulosa, sustancia de la que estaban hechas las películas, ya que este medio perdía su elasticidad con el tiempo y acababa desmoronándose, proceso que era favorecido por las malas condiciones de conservación. De todas formas, aunque el almacenamiento hubiese sido correcto, el proceso de destrucción del nitrato de celulosa tarda unos 50 años en condiciones óptimas. Además, diversos archivos de películas de todo el mundo se han perdido debido al fuego, por ejemplo el archivo mejicano se perdió en 1982 y el japonés en 1984 [38]. Las películas deterioradas, pero salvadas, son lavadas cuidadosamente y transferidas a modernas cintas de acetato de celulosa o poliéster.

Actualmente, la tecnología ofrece la posibilidad de almacenar de forma digital gran cantidad de información a un coste reducido y con un volumen de almacenamiento pequeño, por lo que parece interesante realizar el almacenamiento de la información en estos nuevos formatos. En los casos en que la información de partida, almacenada en el soporte original, se encuentre ya degradada, resulta interesante realizar una *restauración* de la calidad previa al almacenamiento de la información en el soporte estable, tanto para reducir los errores que se produzcan al digitalizar y aumentar la relación señal a ruido, como para que estas películas puedan ser contenidos transferibles en el mercado de la televisión digital. Además, el interés del público por las películas antiguas se ha hecho patente en las reediciones recientes, como por ejemplo *La Guerra de las Galaxias*. El interés por la restauración está justificado en este

sentido, ya que el público responde de manera entusiasta a la oportunidad de ver nuevamente sus películas favoritas restauradas en el cine o en casa.

1.1.2. La degradación de una película

La degradación de una película cinematográfica es debida principalmente al deterioro de los materiales que la componen, ya sea por el paso del tiempo, las malas condiciones de conservación o la acción de agentes externos.

Estructura de la película

[38] Una película cinematográfica está compuesta por tres capas, que son de menos a más externas:

- Soporte transparente. El primer soporte utilizado fue el nitrato de celulosa, que se dejó de usar en los años 50 debido a su inestabilidad. Además, el nitrato de celulosa es muy inflamable y tiene una duración de unos 50 años a temperatura ambiente. Actualmente se utilizan dos tipos de soportes: el triacetato de celulosa y el poliéster. El triacetato es mucho más estable que el nitrato y no es inflamable, y puede conservarse unos cientos de años bajo condiciones óptimas de conservación. Sin embargo, sufre una descomposición en celulosa y ácido acético (*síndrome del vinagre*) que se acelera si las condiciones no son óptimas y que da lugar a la pérdida del material fotográfico. El otro soporte utilizado, el poliéster, es el más estable de todos, ya que se trata de un polímero sintético de gran estabilidad y resistencia. Se estima que una película en este soporte puede durar más de 500 años en condiciones óptimas de conservación.
- Capa de emulsión fotosensible. En el caso de película en blanco y negro se trata de una capa formada por gelatina y sales de plata, y para películas en color, gelatina y colorantes orgánicos.
- Capa antiabrasión. Se trata de una capa protectora antiabrasión para proteger las emulsiones fotosensibles.

Los factores responsables del deterioro de la película a lo largo del tiempo son, entre otros, la humedad, la temperatura, la acidez, la mala manipulación, etc. Así pues, para conservar bien una película hay que cuidar estos factores [38].

Catálogo de defectos

Los factores expuestos anteriormente y otros muchos producen numerosos defectos en las películas. Los algoritmos de restauración dependen de la clase de defecto a restaurar, y es por ello, que en primer lugar se deben catalogar los defectos existentes [127, 52, 140] para, posteriormente, pensar en soluciones que los minimicen o, en el mejor caso, los eliminen.

- **Parpadeo.** Se trata de variaciones globales bruscas de la intensidad de la imagen de un fotograma a otro (figura 1.1). Este tipo de defecto es típico de películas antiguas

en blanco y negro, y es debido principalmente al tiempo de obturación irregular de las primeras cámaras.

- **Vibración.** Son movimientos globales rápidos entre fotogramas consecutivos que se manifiestan como vibraciones indeseadas. Se originan debido a la precisión mecánica limitada de los sistemas de transporte de películas en las cámaras de cine o en el equipo de duplicado tanto de cine a cine como de cine a vídeo (telecine). Si las secuencias digitales han sido obtenidas directamente con un escáner, es posible que las imágenes no estén bien alineadas. Las vibraciones se superponen a los movimientos globales propios de la escena como barridos de cámara, *zooms*, rotaciones, etc.
- **Manchas.** Son el tipo de defecto más frecuente en las películas antiguas. Se manifiestan como manchas oscuras o claras aleatorias en el tiempo (figura 1.3). Son debidas al polvo, la suciedad, y el paso del tiempo junto con las malas condiciones de conservación que hacen que se produzcan pérdidas de gelatina en el material fotosensible.
- **Rayas o arañazos.** Aparecen en la dirección de arrastre de la película durante más de un fotograma (figura 1.2). Son debidas al frotamiento de partículas con la película o el negativo durante la fase de copia o proyección. Estas rayas, producidas por el transporte, son exactamente paralelas a las perforaciones para el arrastre.
- **Ruido de grano.** Se manifiesta como un ruido estacionario a lo largo de la secuencia, y es debido a las dimensiones de las sales de plata que forman el material fotosensible.
- **Fotogramas perdidos o fuertemente dañados.** Tiene su origen en el almacenamiento durante un largo periodo de tiempo de la película en malas condiciones de conservación.



Figura 1.1: Variación de brillo en dos fotogramas consecutivos de una secuencia.

1.1.3. Estado del arte

Existen numerosos sistemas manuales y semiautomáticos para la restauración digital de películas [127], realizados por diferentes empresas: sistema Cineon de Kodak, sistema Domino de Quantel, sistema Flame de Discreet Logic y el sistema Matador de Avid. El problema de estos sistemas es que necesitan una gran interacción humana durante el proceso de restauración por lo que el coste por segundo de película restaurada es muy elevado.



Figura 1.2: Defecto correlado en el tiempo: raya.



Figura 1.3: Defecto incorrelado en el tiempo.

En 1993 *Blancanieves y los siete enanitos* fue la primera película restaurada digitalmente, usando el sistema Cineon de Kodak. Sin embargo, para realizar esta proeza se necesitaron 60 personas trabajando durante más de un año, y costó más de 3 millones de dolares. Dicha restauración se llevó a cabo fotograma por fotograma arreglando arañazos y otros daños de la copia maestra. Este mismo año comienza el proyecto de conservación de archivos de noticias de la Fox (*Movietone News*), que contenían 13 millones de metros de película de 35 mm. Tres años después todo el archivo estaba digitalizado con una resolución de 1008×1018 .

La Comunidad Europea ha financiado diversos proyectos de investigación que trabajan en el campo de la restauración automática de películas:

- EUREKA LIMELIGHT(1994-1997): Se encargó de desarrollar útiles para digitalizar las películas y, además, el Instituto Joanneum Research (Graz) desarrolló el software LIMELIGHT, el cual fue aplicado con éxito para restaurar partes de la película austriaca *Opernball*. Finalmente se trabaja en la mejora de dicho software para conseguir su uso industrial, en el proyecto FRAME [4].
- AURORA(1995-1998)(AUTomated Restoration of ORiginal video and film Archives) [2, 119, 34]: Proyecto enmarcado dentro del programa ACTS (Advanced Communications Technologies and Services). En él participan distintas universidades (Cambridge y Delft) y diferentes instituciones (Institute National de l'Audiovisuel, etc.), así como empresas como la BBC y Snell & Wilcox, con el objetivo de crear nuevos algoritmos y hardware para restaurar rápida y eficazmente grandes cantidades de archivos audiovisuales. Una mejora de AURORA es el proyecto BRAVA [3], que tiene como objetivos mejorar los

resultados de la restauración conseguida por AURORA y además la eficiencia para poder restaurar grandes cantidades de archivos audiovisuales destinados a la radiodifusión.

- ESPIRIT LTS NOBLESSE (1996-1998): Destinado a desarrollar modelos no lineales orientados a objetos para la descripción de secuencias de imágenes. Además, la restauración de películas antiguas es una parte del proyecto NOBLESSE.

En el marco de estos proyectos se han desarrollado diversas tesis como [52, 89, 19, 122] y numerosas publicaciones, como [141, 48, 123].

Como símbolo del creciente interés en la restauración y el rápido avance producido en este tema han aparecido numerosas empresas que ofrecen material y servicios de restauración, como el software REVIVAL [36] realizado en la Singapore's Nanyang Technological University o ARCHANGEL realizado por Snell & Wilcox que restaura archivos en tiempo real [1].

Otras entidades también apoyan la iniciativa, como la *Commission Supérieure Technique de l'Image et du Son*, que ha publicado libros sobre el tema, entre ellos uno que sirva como guía al proceso de restauración de secuencias ([39]), sin especificar ningún tipo de algoritmo.

1.2. Guía de la tesis

1.2.1. Objetivos

La presente tesis se va a centrar en las técnicas de restauración de la imagen de películas antiguas de cine ya que además del interés de la preservación del patrimonio, existe en la actualidad una gran demanda en cuanto a contenidos multimedia que precisan una cierta calidad antes de ser difundidos. Un aspecto importante que debe tenerse presente para que un sistema de restauración de películas sea viable es que la restauración debe ser lo más automática posible, de modo que reduzca al máximo la intervención humana, resultando un proceso económico.

Este trabajo está financiado por un proyecto CICYT (TIC 98-0335 Técnicas de análisis de secuencias de imágenes. Aplicación a la restauración de películas antiguas), en el cual colabora Radio Televisión Valenciana Canal 9 que dispone de un amplio archivo en vídeo, el cual nos sirve como base para este trabajo.

El objetivo de esta tesis es realizar una restauración automática de una base de películas antiguas, en blanco y negro, suficientemente general. Dicha restauración consiste en eliminar todos aquellos defectos que, debidos al paso del tiempo y la mala conservación [38] de las películas, y a errores mecánicos, hacen que la calidad de la proyección quede degradada notablemente, manteniendo la máxima cantidad posible de información original en buen estado. Así pues, como hemos visto, existe un catálogo de defectos determinados cuya eliminación o minimización son los objetivos perseguidos a la hora de la restauración. Esta tesis se va a centrar en la consecución de los siguientes objetivos, con el fin de aportar soluciones novedosas a los problemas que se nos planteen:

- Segmentación de las secuencias degradadas en planos.
- Eliminación de las variaciones bruscas de brillo entre fotogramas del mismo plano.

- Eliminación del ruido de grano dentro de cada plano.
- Detección e interpolación de defectos incorrelados en el tiempo (manchas).

Debido al volumen de trabajo desarrollado para cumplir los objetivos anteriores, no nos vamos a ocupar, en esta tesis, de otro tipo de defectos como la vibración o los defectos correlados en el tiempo (rayas).

1.2.2. Metodología de trabajo

La metodología de trabajo a seguir en el desarrollo de esta tesis se ha dividido en tres fases diferenciadas:

1. **Digitalización de películas y creación de base de datos a restaurar.** Esta primera fase ha consistido en pasar las películas de formato analógico (SVHS) a formato digital, tanto las cedidas por Canal 9 como otras de origen comercial. De esta forma hemos podido pasar a trabajar con ellas realizando, en primer lugar, una base de datos de las secuencias en la que se ha guardado información como duración, número de planos diferentes, catálogo de daños que contienen, etc.
2. **Segmentación de la secuencia en planos.** Una vez que la secuencia puede ser tratada por el ordenador, el primer paso es segmentarla en diferentes planos, entendiéndose por plano como el conjunto de fotogramas que suceden entre que la cámara empieza y termina de grabar. La segmentación consiste en detectar las diferentes transiciones que permiten el paso de un plano a otro, como: cortes (transiciones bruscas), fundidos entre imágenes, fundidos a un color (negro o blanco) y cortinillas.

Es conveniente esta segmentación previa ya que todos los fotogramas de un plano tendrán una gran correlación temporal, hecho que nos va a permitir llevar a cabo el proceso de restauración.

3. **Proceso de restauración.** Esta fase trata de encontrar una solución para cada uno de los objetivos que se especificaron anteriormente. Para conseguirlo tendremos que centrar nuestro estudio en diferentes disciplinas como la morfología matemática, algoritmos de interpolación, métodos de eliminación de ruido, etc. A la hora de restaurar la película un objetivo primordial es mantener inalteradas las zonas no dañadas de la imagen e interpolar los trozos dañados. Por ello, resulta de interés obtener una máscara binaria de las zonas a actuar en cada fotograma.

1.2.3. Estructura de la tesis

Capítulo 2: Segmentación de secuencias

En este capítulo se estudian los diferentes métodos de detección de cambio de plano: detección de cortes, de fundidos y cortinillas. Se proponen métodos para detectar cortes teniendo en cuenta la calidad de las imágenes que contienen las secuencias que pretendemos segmentar. Se proponen diferentes métricas así como un postprocesado basado en algoritmos

morfológicos unidimensionales que nos permiten maximizar la probabilidad de detección minimizando la de falsa alarma. Los métodos presentados pretenden detectar tanto transiciones abruptas como graduales, sin embargo, son las primeras las que nos preocupa detectar con gran precisión, ya que son las únicas que supondrían una ruptura en la continuidad temporal dentro del plano si no se detectasen.

Capítulo 3: Corrección de las variaciones de intensidad

Se presentan dos métodos para la corrección del parpadeo: uno basado en la estimación de parámetros para modelar la media y la varianza de la secuencia, y otro basado en la equalización de histograma de las diferentes imágenes, para conseguir que el histograma no varíe rápidamente a lo largo de la secuencia.

Capítulo 4: Reducción de ruido

Una vez estudiados los diferentes métodos de reducción de ruido en secuencias propuestos en la literatura, se propone un filtrado FIR muy eficiente computacionalmente que reduce notablemente el ruido estacionario basándose en la respuesta espacio-temporal del sistema visual humano. Se muestran resultados y se analiza el factor de reducción de ruido conseguido.

Capítulo 5: Defectos incorrelados

Este capítulo se dedica a la eliminación de defectos incorrelados en el tiempo (o manchas), caracterizados por un alto contraste con el fondo en el que se encuentran y muy baja correlación temporal. Se presentan tres métodos que utilizan operadores morfológicos para la detección, en una primera fase, e interpolación de dichos defectos. Examinando diversos ejemplos se analizan las ventajas e inconvenientes de cada método y se exponen condiciones para su utilización.

Capítulo 6: Sistema global

Una vez presentados en los capítulos anteriores los algoritmos para la restauración de los diversos defectos que han sido tratados en la tesis, se presenta, de forma justificada, un esquema global de restauración, en el que se concatenan los diferentes algoritmos en un orden determinado justificado con diversos razonamientos. También se muestran resultados de secuencias restauradas utilizando este esquema.

Capítulo 7: Conclusiones y líneas futuras

Por último, se presentan una serie de conclusiones, las aportaciones realizadas con esta tesis al perseguir la consecución de los objetivos planteados, así como las líneas de continuación para un trabajo futuro.

Capítulo 2

Segmentación de secuencias

2.1. Introducción

En este capítulo, dedicado a la segmentación, vamos a dedicar esta introducción a analizar la necesidad de una segmentación temporal, y a definir qué es un plano, ya que detectar las fronteras entre los mismos será el objetivo de la segmentación. En la sección 2.2 vamos a hacer un estudio del estado del arte de la segmentación de secuencias. La siguiente sección, 2.3, presenta el esquema global del proceso de segmentación, deteniéndose en los conceptos del vídeo comprimido. Los métodos propuestos en esta tesis para la detección de transiciones son presentados en las siguientes secciones: en la 2.4, la detección de cortes, en la 2.5 la de fundidos y en la 2.6 la de cortinillas. Para finalizar se exponen, en la sección 2.7, los resultados obtenidos en las pruebas realizadas, así como una serie de conclusiones. Lo que se va a presentar en este capítulo es fruto del trabajo conjunto con Jesús Angulo en su proyecto fin de carrera [14]. Otras referencias relacionadas son [13, 10].

2.1.1. Necesidad de la segmentación temporal

El primer paso antes de restaurar una película consiste en dividir la misma en fragmentos que tengan contenido común, es decir, aquel conjunto de fotogramas entre los que no se producen cambios de escena, y que llamaremos en esta tesis “plano”. Este proceso es muy importante ya que, como veremos en los diferentes capítulos de esta tesis, todos los algoritmos de restauración se basan en los componentes comunes que tienen un fotograma y sus vecinos si pertenecen al mismo segmento, los cuales se diferencian de los de otro fragmento de la película. La característica principal que va a servirnos a la hora de restaurar es que en el segmento seleccionado exista continuidad temporal entre sus fotogramas, es decir, que la escena no varíe rápidamente. Por tanto, será muy importante detectar los cambios de escena que supongan una ruptura brusca de dicha continuidad. De los dos tipos de cambios de escena diferentes que existen, transiciones abruptas y transiciones graduales, será crítico, para el buen resultado de la restauración, determinar de manera muy precisa las transiciones abruptas que son las que suponen un cambio brusco de continuidad temporal. Para detectar las transiciones graduales también se presentarán algoritmos, pero la exigencia de buenos resultados en la segmentación es menos estricta, ya que no detectar una de estas transiciones

no supone un deterioro considerable en los resultados de la restauración posterior.

Esta segmentación será útil también en el almacenamiento de la película, ya restaurada, en formato digital. Actualmente existen grandes cantidades de información de vídeo en distintos formatos (cintas analógicas de diferentes sistemas, CD-ROM's con imágenes comprimidas, bases de datos de programas de T.V., filmotecas, etc). Esta diversidad de formatos y fuentes tiende a desaparecer con el establecimiento de los nuevos estándares de codificación de imagen y vídeo. El objetivo de estos estándares digitales es doble: uniformizar los sistemas y optimizar la cantidad de memoria necesaria para su almacenamiento.

Para soportar estas nuevas aplicaciones será indispensable que los interfaces con el usuario, además de intuitivos y fáciles de manejar, sean potentes. Pero lo más importante para poder dar respuesta a las peticiones de información, será la preparación previa de ésta. Las bases de datos deberán disponer de mecanismos eficientes de indexado de la información, que no serán más que sistemas para la descripción automática (o semiautomática, minimizando el trabajo manual de visionado) del contenido de la información audiovisual. Pensando en secuencias de vídeo, está claro que no es factible disponer de información sobre cada uno de sus fotogramas. Parece más útil poder dividir las secuencias en fragmentos, dentro de cada uno de los cuales exista conexión semántica.

2.1.2. Definición de plano

Las secuencias de vídeo se componen de fragmentos temporales cuyo contenido está relacionado. Habitualmente estos fragmentos representarán un evento o una secuencia de acciones continuas, y esto es conocido como plano temporal o *shot*. Un plano es la sucesión de imágenes capturadas por una cámara entre las operaciones de *grabar* y *parar*.

Las fronteras entre planos de vídeo son conocidos comúnmente como cambios de escena y el acto de segmentar una secuencia de vídeo en sus planos se llama detección de cambios de escena.

Hasta 1901, las películas de cine eran monoplanos. De esta forma su género era exclusivamente documental, no se podía contar una historia que tuviera más de un escenario o se desarrollara a lo largo del tiempo. En este año James Williamson en su película "*Fire!*", y con el fin de mostrar una casa en llamas cuando se inicia el fuego en el interior y cómo se consume vista desde fuera, puso su cámara en el interior y después la sacó al exterior. El resultado fueron dos planos que se yuxtaponían. El *cut* o corte estaba inventado. Más adelante se atribuye la invención del montaje (y por tanto de la estructuración de la narración audiovisual) a Levin Fitzhamon, que realizó en 1904 "*Rescued by Rover*", y que presentaba una historia estructurada en cinco planos. El corte funcionaba aquí como una elipsis espacio-temporal que permitía al realizador estructurar el argumento de la historia. Rápidamente, esta sintaxis se enriqueció con nuevos elementos de estructuración. Se atribuye por ejemplo, al gran cineasta David Wark Griffith, la invención del efecto de transición progresiva, algunos años más tarde. Esta riqueza de potencialidades ofrecidas por los efectos de transición es sin duda el origen de la siguiente cita de Orson Welles: "*La sola puesta en escena cobra una gran importancia a lo largo del montaje... Las imágenes en sí mismas no son suficientes: son muy importantes, pero no sólo las imágenes. Lo esencial es la duración de cada imagen, lo que es cada imagen; en esto consiste la elocuencia del cine y que se fabrica en la sala de*

montaje” [66].

Hoy en día, todavía se mantienen estos dos caminos para que un plano pueda realizar una transición al siguiente: transición abrupta o corte (figura 2.1) y transición gradual (figura 2.2). En las transiciones graduales, y con la aparición de los equipos digitales de edición, se han producido grandes avances y mejoras. Pero en cualquier caso, todavía se pueden clasificar en dos grandes grupos: fundidos o *dissolves* y efectos de incrustaciones, por ejemplo las cortinillas o *wipes*.



Figura 2.1: Ejemplo de transición abrupta: (a)Último fotograma de un plano. (b)Primer fotograma del plano siguiente.

En algunos casos también puede ser interesante encontrar las variaciones dentro del plano. Esto es conocido como detección de cambios de escena intraplano o *intra-shot*. De esta manera se podrían encontrar efectos tales como: *zooms*, panorámicas (*panning*), etc. Las técnicas utilizadas para conseguir esta segmentación más fina, se fundamentan en los vectores de movimiento, y no los hemos estudiado en esta tesis, ya que sólo nos interesan aquellos cambios bruscos que dificulten la restauración.

2.2. Estado del arte

La segmentación temporal o detección de cambios de escena en secuencias de vídeo es un tema sobre el que se ha trabajado mucho, desde que surge la idea de construir bases de datos de la información audiovisual. Al principio de los años 70, ya se introdujo hardware de detección de cambio de escena en los telecines. Las primeras publicaciones sobre técnicas de segmentación de vídeo aparecen con los 90, inicialmente los investigadores intentan detectar cambios bruscos y posteriormente se adentran en la detección de transiciones graduales. Desde entonces no ha cesado la introducción de novedades y mejoras.

En la bibliografía aparecen una serie de artículos que presentan las diferentes técnicas y hacen un estudio comparativo de las mismas utilizando modelos experimentales, como son las contribuciones de Boreczky *et al.* [20], Apostolos Dailianas [37], Irena Koprinska [90]



Figura 2.2: Ejemplo de transición gradual (fundido con dos fotogramas intermedios): (a)Último fotograma de un plano. (b)Fotograma intermedio del fundido. (c)Fotograma intermedio del fundido. (d)Primer fotograma del plano siguiente.

y Philippe Joly [66], entre otros. Otros artículos, como los de Ahanger y Little [6], Philippe Aigrain [8] y Brunelli *et al.* [28] están dedicados a la recopilación de técnicas de indexación de vídeo, y dedican un apartado a hacer un repaso de las técnicas de segmentación.

Las primeras técnicas de segmentación estaban enfocadas a la detección de cortes, pero más recientemente el trabajo se ha orientado a la detección de transiciones graduales, como las cortinillas y fundidos.

2.2.1. Detección de cortes

Las investigaciones sobre detección de cortes pueden agruparse en dos categorías: métodos para vídeo no comprimido y métodos que detectan los cortes directamente sobre la secuencia comprimida.

Los métodos de segmentación se basan en la idea de que las imágenes que pertenecen al mismo plano tienen un grado de similitud mayor que las pertenecientes a planos diferentes [60]. Basándose en esto, el funcionamiento de cualquier método de segmentación consistirá en realizar una medida de las diferencias entre fotogramas sucesivos, y fijar umbrales que determinen cuándo una diferencia constituye una transición. Lo que diferirá de unos casos a otros es la métrica elegida para computar las diferencias así como el método de establecer el umbral. A continuación vamos a hacer un repaso de las técnicas fundamentales.

Algoritmos para vídeo no comprimido

Estos métodos calculan las diferencias bien entre los niveles de luminosidad de los píxeles, las diferencias entre los histogramas o las diferencias estadísticas entre imágenes (enteras o subimágenes) consecutivas de una secuencia.

■ Diferencias entre los niveles de luminosidad

Los métodos más simples, y los que se han utilizado en esta tesis, son los que calculan las diferencias entre los niveles de luminosidad de los píxeles de fotogramas consecutivos. Hay algoritmos que utilizan sólo dos fotogramas en la comparación, y otros que utilizan tres. Pero todos se basan en calcular un valor a partir de la diferencia entre los niveles de luminosidad de los píxeles, el cual debe superar un umbral para que se considere que existe un corte.

Sean $F_i(x, y)$ y $F_j(x, y)$ dos fotogramas de una secuencia, uno situado en la posición temporal i y otro en la j , en los cuales $F(x, y)$ nos da el nivel de luminosidad del píxel (x, y) para ese fotograma. Denotaremos la diferencia entre ambos como $D_{i,j}$. A continuación se recogen algunas de las alternativas propuestas en la bibliografía para $D_{i,j}$.

- Comparación píxel a píxel

$$D_{i,i+1} = \sum_{x,y} |F_i(x, y) - F_{i+1}(x, y)|^\alpha. \quad (2.1)$$

Típicamente, $\alpha = 1$ o 2 . La medida D también puede definirse para contar el número de píxeles con valores diferentes,

$$D = \sum_{x,y} g(d),$$

siendo, por ejemplo

$$g(d) = \begin{cases} 1 & \text{Si } d \geq l \\ 0 & \text{Otro caso} \end{cases},$$

donde $d = |F_i(x, y) - F_{i+1}(x, y)|$.

Esta métrica es muy sensible al movimiento de la cámara y al movimiento de los objetos. Zhang, Kankanhalli y Smoliar [158] implementaron este método e intentaron

solucionar el problema de la sensibilidad al movimiento filtrando previamente con un filtro de media de 3×3 .

Este mismo método puede ser aplicado a imágenes en color, por ejemplo tomando la media de las tres diferencias (R,G y B), la cual es utilizada por Demarty [42] tanto para la detección de cortes como de transiciones graduales. Con esta métrica obtiene unas máscaras de transición (en niveles de gris), de las cuales, estudiando su volumen determinará si existe corte o fundido, y estudiando además su geometría se detectarán las cortinillas.

Otra contribución que utiliza esta métrica es la de Aigrain [7], en la que presentan un algoritmo basado en un modelo estadístico para calcular el número de píxeles que cambia de un fotograma a otro.

- Algoritmo con tres fotogramas [132]

La medida que se utiliza en estos métodos es la *Coherencia de movimiento observado*: *OMC*, definida del siguiente modo:

$$OMC = \left| \frac{D_{i-1,i} - D_{i,i+1}}{D_{i-1,i} + D_{i,i+1}} \right|, \quad (2.2)$$

la cual tendrá un valor cercano a 1 si existen muchos cambios entre los tres fotogramas bajo test y 0 en otro caso. Los autores proponen diferentes medidas para $D_{i,j}$ tales como:

$$D_{i,j} = \frac{\sum_{x,y} [F_i(x,y) - F_j(x,y)]^2}{\sqrt{\sum_{x,y} [F_i^2(x,y)] \sum_{x,y} [F_j^2(x,y)]}} \quad (2.3)$$

$$D_{i,j} = \frac{\sum_{x,y} |F_i(x,y) - F_j(x,y)|}{\sum_{x,y} |F_i(x,y) + F_j(x,y)|}. \quad (2.4)$$

■ Diferencias estadísticas

Los métodos estadísticos se basan en dividir la imagen en bloques y comparar medidas estadísticas de los píxeles de dichos bloques. Una de las medidas estadísticas utilizadas es la verosimilitud.

- Medida de verosimilitud [158]: Para cada bloque de la imagen t y de su correspondiente bloque en $t + 1$ se calcula la relación de verosimilitud L_i , definida del siguiente modo:

$$L_i = \frac{\left[\frac{\sigma_t^2 + \sigma_{t+1}^2}{2} + \left(\frac{\mu_t - \mu_{t+1}}{2} \right)^2 \right]^2}{\sigma_t^2 \sigma_{t+1}^2}, \quad (2.5)$$

donde μ es la media aritmética y σ la desviación típica de los histogramas de intensidad. Un corte se produce cuando L_i , para un número suficiente de bloques, supera un determinado umbral.

- Test de verosimilitud de Yakimovsky [132]. La relación que se calcula para realizar este test viene dada por:

$$y = \left(\frac{\sigma_0^2}{\sigma_{t-1}^2} \right) \left(\frac{\sigma_0^2}{\sigma_t^2} \right), \quad (2.6)$$

donde σ_{t-1}^2 y σ_t^2 representan las varianzas de los valores de la luminancia de los píxeles pertenecientes a los fotogramas anterior y actual respectivamente, mientras que σ_0^2 es la varianza de los datos conjuntos de ambos histogramas. Un cambio de plano se produce si y excede un determinado umbral.

- Test de Kolmogorov-Smirnov [132]. Este método se basa en calcular las distribuciones acumuladas $C_1(x)$ y $C_2(x)$ de la luminancia de los píxeles en dos fotogramas consecutivos y medir la máxima diferencia absoluta entre ellas. Las distribuciones pueden ser aproximadas por los histogramas acumulados.

Otros métodos basados también en la estadística están siendo propuestos más recientemente, como Vanconcelos y Lippman [149] que establecen modelos estadísticos bayesianos de repetición de transiciones, o Kim y Park [73] que proponen una nueva medida de diferencias entre fotogramas, la entropía cruzada.

■ Comparación entre histogramas

El método simple de comparación de histogramas calcula el histograma de grises o el de color de los dos fotogramas a comparar (normalmente consecutivos) determinando si la diferencia entre ambos (2.7) supera un determinado umbral, en cuyo caso se dice que se ha detectado un corte.

$$D = \sum_j |H_{t+1}(j) - H_t(j)|. \quad (2.7)$$

Para imágenes en color esta diferencia se determina mediante el histograma ponderado:

$$D = \frac{r}{s} D_{rojo} + \frac{g}{s} D_{verde} + \frac{b}{s} D_{azul}, \quad (2.8)$$

donde r , g y b es la luminancia media para las componentes roja, verde y azul, respectivamente, y $s = (r + g + b)/3$.

Una variante para determinar la diferencia entre dos fotogramas basada en los histogramas es el *test* χ^2 , el cual compara las dos distribuciones discretas de la siguiente forma:

$$\chi^2 = \sum_j \frac{(H_{t+1}(j) - H_t(j))^2}{H_{t+1}(j) + H_t(j)}. \quad (2.9)$$

Otra fórmula comúnmente usada es la *intersección de histogramas* dada por la ecuación 2.10.

$$\sum_j \frac{\min(H_{t+1}(j), H_t(j))}{\max(H_{t+1}(j), H_t(j))}. \quad (2.10)$$

Zhang, Kankanhalli y Smoliar [158], además de comparar diferencias entre píxeles y diferencias estadísticas, han utilizado diferentes métodos basados en el histograma, determinando que éstos últimos llegan a un buen compromiso entre la precisión en el resultado y la velocidad de cálculo. También Nagasaka y Tanaka [103] experimentan con varias técnicas de comparación, incluyendo la diferencia de la suma de niveles de gris, suma de las diferencias de los niveles de gris, diferencia de los histogramas de nivel de gris, comparación de paletas de color, diferencia de los histogramas de color y comparación χ^2 de histogramas de color. Cada una de las comparaciones son aplicadas sólo a una porción de la imagen. Los autores concluyen que el método más robusto es el de las comparaciones χ^2 de histogramas de color. Para protegerse frente a “ruidos” momentáneos, como *flashes* de cámara, dividen la imagen completa de cada fotograma en 4×4 regiones rectangulares de igual tamaño y comparan cada región par a par. Las grandes diferencias son desechadas, y la detección se basa en aquellas que permanecen. Su método es robusto frente a efectos como *zoom* o *panning*, pero falla al intentar detectar efectos especiales, del tipo fundido.

■ Diferencias de movimiento

Shahraray [135] detecta cambios abruptos y graduales basándose en un filtrado temporal del movimiento entre fotogramas consecutivos. Un proceso bloque a bloque se desarrolla para cada bloque de la primera imagen, con el fin de encontrar la región que mejor se ajusta de la segunda imagen. Un filtrado estadístico no lineal es utilizado para generar los valores de ajuste.

Oh et al. [64] proponen una idea basada en el seguimiento del movimiento de la cámara usando el fondo de las imágenes.

Algoritmos para vídeo comprimido

Con la aparición de MPEG, muchos investigadores se han centrado en la detección de los cambios de escena sobre vídeo comprimido. Un ejemplo de los primeros pasos en la segmentación utilizando vídeo comprimido es el de Arman, Hsu y Chiu [17]. En este método se utiliza un conjunto de coeficientes de la DCT de un conjunto de bloques de la imagen JPEG para detectar los cambios de escena. Estos coeficientes forman un vector de características del fotograma que será usado en la detección. El problema es que el éxito del algoritmo depende de la elección de los bloques. Zhang et al. [159] extendieron este método a imágenes MPEG, apareciendo a partir de aquí muchos métodos que trabajan sobre vídeo comprimido en MPEG.

Meng et al. [99] usan la varianza de los coeficientes *dc* (ver sección 2.3.1) en fotogramas *I* y *P* y la información de los vectores de movimiento para caracterizar los cambios de escena. Sethi y Patel [132, 110] han usado sólo los coeficientes *dc* de los fotogramas *I* para desarrollar un test de hipótesis usando el histograma de luminancia. Se asume que la separación entre

dos imágenes I es fija y pequeña. En su trabajo, experimentan con tres tests estadísticos: el de verosimilitud de Yamikovsky, el test χ^2 y el de Kolmogorov-Smirnov. El problema es la alta tasa de falsos positivos. Para solucionarlo incorporan dos tests adicionales sobre los histogramas de filas y columnas de las imágenes I [111]. De todas formas, al utilizar sólo imágenes I , la localización exacta de los cambios abruptos no puede ser detectada con este método. Este algoritmo es usado por Ceccarelli *et al.* [31] sobre los histogramas combinados de las dos señales de crominancia U y V .

Yeo y Liu [154] proponen algoritmos rápidos para operar sobre *Motion JPEG* o *MPEG* comprimidos. Sus algoritmos operan directamente sobre una pequeña fracción de los datos comprimidos (se realiza una descompresión parcial). Su método es efectivo para detectar cortes, transiciones graduales y escenas con *flashlight*, es decir, escenas donde se producen cambios bruscos de luminosidad debidos a la toma de fotos.

Shen [136] utiliza la distancia de Hausdorff obtenida comparando puntos pertenecientes a bordes en fotogramas sucesivos. La información sobre los bordes está tomada directamente de imágenes I de vídeo *MPEG*.

Kobla *et al.* [81] usan la información de los macrobloques (por ejemplo el tipo de predicción utilizado en la compresión) para detectar cortes. Para casos en que esta información no es suficiente se recurre al análisis de los coeficientes de la *DCT*. En su método, utilizan los vectores de movimiento para determinar también el movimiento de cámara con el fin de subdividir planos en subplanos.

Albiol *et al.* [10] utilizan la llamada imagen *dc* para reducir el tamaño de los datos analizados calculando diferencias de luminosidad mejoradas con técnicas de morfología matemática. Estos algoritmos están especialmente diseñados para secuencias dañadas (películas antiguas) en las que se producen grandes cambios de luminosidad aleatorios. También utilizan morfología matemática (*watershed*) para la detección de cortes Llach y Salembier [93].

Otro grupo de métodos propuestos trabajan sobre técnicas estadísticas en vídeo comprimido, como el de Bouthemy *et al.* [23]. Estos autores utilizan estimación de movimiento dominante, mediante un modelo afín 2D, para localizar los cortes y caracterizar los movimientos de cámara. Posteriormente, Bouthemy *et al.* [22] proponen como parte de su sistema de indexado de vídeo unos algoritmos para detectar cortes y otros para detectar transiciones graduales ambos sobre vídeo comprimido. El de detección de cortes se basa en métricas de comparación de histogramas, y en ir aproximándose en pasos sucesivos al punto donde se produce el corte, dentro de la secuencia *dc*. En cuanto al algoritmo para detectar fundidos, se fundamenta en la detección de movimiento global.

Ngo, Pong y Chin [105] proponen un método basado en el análisis de dos imágenes espacio-temporales conseguidas como la proyección horizontal y vertical de una secuencia de imágenes *MPEG*. Estas imágenes tienen regiones coherentes que indican la localización temporal de una cortinilla o un corte (el método no resuelve fundidos). La segmentación de estas imágenes, para lo cual utilizan campos aleatorios de Markov, equivale a la segmentación temporal. Para eliminar falsos positivos se recurre a un análisis de los vectores de movimiento *MPEG* y de los coeficientes *DCT*. Drew *et al.* [94] extienden el método de Ngo usando un algoritmo basado en la diferencia de los histogramas de cromaticidad para formar la imagen espacio temporal, con el fin de reducir ruido.

Cheng *et al.* [33] extraen una serie de características de la secuencia *dc*, como la inter-

sección de histogramas del fotograma i e $i - 1$, el número de macrobloques codificados como intra en el fotograma i , etc. Todas las características se posicionan en un vector que se denomina *traza generalizada (GT)* el cual se utiliza para determinar la probabilidad de que en un fotograma se haya producido un corte.

2.2.2. Detección de transiciones graduales

Las transiciones graduales (ediciones) son una clase importante de efectos, tales como fundidos entre fotogramas, fundidos (de entrada o salida) a o de un color (negro, blanco, etc.) y cortinillas. Alrededor de los años 50 las transiciones graduales fueron introducidas en el lenguaje cinematográfico; un fundido, por ejemplo, era muy utilizado para realizar un cambio de lugar o de época (*flashback*) dentro de la misma acción. Así pues estas transiciones pueden ser también utilizadas como indicios para detectar cambios de escena, además del simple corte del que hemos hablado anteriormente.

El desarrollo de la tecnología de la edición ha facilitado el uso de los efectos de edición en la producción de vídeo. Actualmente existe un gran número de efectos de transición (por ejemplo el Adobe Premiere, un programa para la edición de vídeo, proporciona más de 75 efectos de transición diferentes), lo cual dificulta la detección de este tipo de efectos. Sin embargo, las transiciones graduales ocurren con menos frecuencia que los cortes, reduciéndose la importancia que la detección de estos cambios de escena tienen en la mejora de los sistemas de análisis de vídeo. Concretamente, en el caso que ocupa a esta tesis, la restauración de películas antiguas, estas transiciones son bastante escasas y además no resultan críticas para la restauración, ya que no suponen un cambio brusco de escena, es decir, una ruptura de la continuidad temporal. Sin embargo, nos ocuparemos a continuación de analizar el estado del arte en la detección de transiciones graduales, y más adelante dedicaremos una sección a la explicación del método usado en esta tesis, así como a la presentación de los resultados obtenidos.

Detección de mesetas

Las comparaciones basadas solamente en fotogramas sucesivos no son adecuadas para la detección de transiciones graduales ya que los cambios entre un fotograma y el siguiente son pequeños en este tipo de efectos. Una alternativa es comparar el fotograma f_t con el f_{t-k} ; sin embargo, este tipo de comparación puede dar lugar a muchos falsos positivos, ya que dentro del mismo plano, dos fotogramas separados k pueden tener grandes diferencias estadísticas debidas, por ejemplo al movimiento de la cámara o de objetos dentro de la escena.

Yeo y Liu [154] presentan un detector de transiciones graduales que se basa en la comparación de los fotogramas i e $i + k$, obteniéndose una secuencia de distancias calculada como $\{D_i^k = d(f_i, f_{i+k})\}$. Si k se elige mayor que la duración de la transición, la secuencia D_i^k presentará una meseta de máxima anchura. Para reducir la detección de falsas transiciones, se exige que el valor de la meseta sea más alto que la media de las diferencias entre fotogramas para los cortes. Un método similar se presenta en [137], aquí se trabaja con imágenes *dc* y la distancia se calcula como la diferencia entre los histogramas de las componentes YUV de dos fotogramas consecutivos.

Detección por modelado

Otra importante aproximación para la detección de transiciones graduales ha sido propuesta en [60, 7] y está basada en el modelado de los efectos. Se formulan modelos matemáticos, basados en la producción de efectos de vídeo, que representarán cada transición. Por ejemplo, una operación de fundido desde la escena X a la Y es una secuencia de fotogramas en la que cada imagen se calcula escalando la intensidad de los píxeles de los fotogramas extremos de los planos X e Y (el último fotograma del plano X y el primero del plano Y) por medio de la siguiente expresión:

$$E(t) = \left[\frac{(t - T_2)}{L_2} Y \right]_{t \in [T_2, T_2 + L_2]} + \left[\left(1 - \frac{(t - T_1)}{L_1} \right) X \right]_{t \in [T_1, T_1 + L_1]}, \quad (2.11)$$

donde T_1 y T_2 representan los instantes en los cuales el escalado de las intensidades de X e Y comenzaron y L_1 y L_2 representan las duraciones de los escalados. Un fundido a negro es un caso especial de fundido, en el cual $X = 0$ si es fundido desde negro (*fade-in*) o $Y = 0$ si es hacia negro (*fade-out*).

Concretamente Hampapur, Jain y Weymouth [60], calculan lo que ellos llaman *imágenes cromáticas*, dividiendo la diferencia del nivel de gris de cada píxel entre dos imágenes, entre el nivel de gris de dicho píxel en la segunda imagen. Durante los fundidos, esta imagen cromática mantiene un valor razonablemente constante. También calculan una imagen similar para detectar cortinillas. Desafortunadamente esta técnica es muy sensible al movimiento de los objetos y de la cámara.

Fernando *et al.* [49] presentan un algoritmo para detectar fundidos basado en características estadísticas derivadas del modelo matemático elegido para la transición gradual, concretamente se basan en el hecho de que si la secuencia de vídeo es un proceso ergódico, durante un fundido la media tiene un comportamiento lineal y la varianza cuadrático. Implementan el método para vídeo no comprimido y para vídeo comprimido MPEG. En el primer caso obtienen mejores resultados ya que existen limitaciones en vídeo comprimido para calcular la varianza. Estos mismos autores proponen también [50] un método de detección de cortinillas que consiste en dividir cada fotograma en bloques de 16×16 píxeles y obtener una *imagen estadística* que contenga el valor de la media y la varianza de cada bloque. Las cortinillas se detectan encontrando los bloques en los que el error cuadrático medio supera cierto umbral. La dirección y el patrón de la cortinilla se determinan aplicando la transformada de Hough sobre la imagen estadística una vez que se ha puesto un umbral. Un método similar al de Fernando [49] para detección de fundidos es el de Lu, Zhang y Yao [62], sin embargo este algoritmo (basado en el carácter de la media y la varianza en el fundido) sí trabaja con vídeo comprimido. Para reducir las falsas alarmas provocadas por el movimiento de objetos y de cámara calculan la doble diferencia cromática. También presentan una mejora Truong, Dorai y Venkatesh [145], los cuales ponen ciertas restricciones a las transiciones detectadas con los métodos tradicionales de detección de fundidos basados en el modelo (variación de varianza cuadrática y de media lineal) para evitar falsos positivos.

Detección basada en características

El método propuesto por Zabih en [156, 157] es útil para detectar transiciones graduales así como cortes. Se basa en el hecho de que durante un corte o un fundido aparecen nuevos bordes en los fotogramas y desaparecen otros. Los píxeles que pertenecen a estos bordes se denominan píxeles de bordes de entrada (si aparecen) o de salida (si desaparecen). Los métodos que nos ocupan se basan en contabilizar los píxeles de bordes de entrada (ρ_{in}) y salida (ρ_{out}) para detectar cortes y fundidos o determinar la distribución espacial de los mismos para la detección de cortinillas. Los cambios de escena se detectan calculando $\rho = \max(\rho_{in}, \rho_{out})$. Un corte se detectará si existe un pico estrecho de ρ , mientras que para una transición gradual el pico de ρ será más ancho. Durante un fundido desde negro $\rho_{in} \gg \rho_{out}$, y al revés si el fundido es hacia negro. Un fundido se caracteriza por una predominancia de ρ_{in} durante una primera fase y de ρ_{out} durante la segunda.

Comparaciones dobles

El método de dobles comparaciones [158] se basa en la idea de que existen marcadas diferencias entre los histogramas de los fotogramas de antes y después de una transición gradual. Las diferencias entre los histogramas de dos fotogramas consecutivos pertenecientes a una transición son más pequeñas que los pertenecientes a un corte pero mayores que las de los histogramas de dos imágenes consecutivas que pertenecen a un mismo plano. Así pues, se utiliza un primer umbral más alto para detectar cortes, y uno menor para detectar transiciones graduales. Toda pareja de fotogramas que supere el segundo umbral será marcada como candidata a pertenecer a un efecto de edición. Para cada candidata se calcula una diferencia acumulativa con más parejas adyacentes. Si dicha diferencia supera el primer umbral (el utilizado para detectar cortes), se asume que se ha encontrado una transición gradual.

2.2.3. Medida de la efectividad de los algoritmos

La metodología para medir la efectividad de los algoritmos de detección de transiciones entre planos se basa en la comparación de los resultados obtenidos mediante los algoritmos con el conjunto de resultados obtenidos mediante la observación humana.

Si un algoritmo requiere la introducción de algún parámetro (por ejemplo un umbral óptimo) el conjunto de secuencias de las que disponemos ha de dividirse en tres subconjuntos disjuntos: uno de entrenamiento, uno para validación y otro de test. Los parámetros se seleccionan para optimizar la medida realizada con los conjuntos de entrenamiento y validación, y finalmente la medida de la efectividad del algoritmo se realiza sobre el conjunto de test. Una función usada comúnmente para la estimación de los parámetros óptimos es:

$$C(\alpha, v) = \alpha S_n(v) + (1 - \alpha)S_f(v), \quad (2.12)$$

donde $S_n(v)$ representa las transiciones no detectadas (falsos negativos) y $S_f(v)$ los falsos positivos. v son los parámetros del algoritmo y $\alpha \in [0, 1]$ regula el grado de compromiso entre precisión (no detección de falsos positivos) y el acierto o detección de todas las transiciones existentes. Si $\alpha = 1$ se da máxima importancia a la detección (a que no exista ninguna

transición sin detectar) y la secuencia será sobresegmentada, mientras que un valor pequeño de α favorecerá la infrasegmentación. Los parámetros óptimos de los algoritmos se pueden obtener minimizando $C(\alpha, v)$.

En [37, 20] se presentan los resultados, en porcentaje de verdaderos (correcto) y falsos positivos (falso), de diferentes algoritmos de detección de transiciones. Dichos resultados aparecen resumidos en la tabla 2.1. Los resultados de los algoritmos para vídeo comprimido (tabla 2.2) han sido presentados por Gargi [54]. En ambas tablas, los porcentajes de detección y falsos positivos están dados con respecto a las transiciones reales encontradas mediante observación. En estos tests se han utilizado algoritmos de detección de cortes para detectar no sólo éstos sino también las transiciones graduales, comparando fotogramas no sucesivos (más separados en el tiempo, es decir, diezmado temporalmente la secuencia) y bajando el umbral de detección. Esto provoca, como puede observarse en las tablas, que la detección sea buena a costa de aumentar mucho los falsos positivos.

En la tabla 2.1 el conjunto de test contiene tanto cortes como transiciones graduales. En la 2.2, dedicada a la segmentación de secuencias de vídeo comprimido, sólo se han tenido en cuenta las transiciones abruptas o cortes.

Algoritmo	Referencia	Correcto (%)	Falso (%)
Diferencia absoluta	[37]	73	56
Diferencia histograma rojo	[37]	94	175
Diferencia histograma ponderado	[37]	94	135
χ^2 diferencia histograma rojo	[37]	95	137
Intersección histograma rojo	[37]	94	174
Momentos invariantes	[37]	54	105
Rango de cambio de píxeles	[37]	95	68
Basado en características	[37]	92	59
Histograma de luminancia	[20]	99	521
Dobles comparaciones	[20]	98	151
Histograma de 16 regiones	[20]	96	205
Dif. con compensación de movimiento	[20]	95	541

Cuadro 2.1: Resultados de algunos algoritmos comunes en segmentación de vídeo

Algoritmo	Correcto (%)	Falso (%)
Arman [16], Zhang [159]	95	2490
Meng et al. [99]	65	338
Liu and Zich [92]	30	42
Yeo and Liu [154]	69	4
Seti and Patel [132]	45	274
Shen and Delp [137]	68	69

Cuadro 2.2: Resultados de la detección de cortes para vídeo comprimido

2.2.4. Mejoras de la detección

Además de todos los diferentes métodos presentados, los cuáles intentan minimizar el número de cambios de escena no detectados (falsos negativos) y el de falsos positivos, existen autores que mejoran los algoritmos en cuanto al coste computacional (hecho que ya se mejoraba al trabajar con vídeo comprimido). Hua *et al.* [63, 65, 107] proponen métodos no lineales para reducir el coste computacional. Su método se basa en no comparar todos los fotogramas para detectar los cambios, sino en empezar comparando los que están a determinada distancia (fija o adaptativa) para posteriormente ir afinando en la detección exacta del momento en que se produce la transición.

Además del coste computacional hay autores que dedican esfuerzos a unificar los diferentes métodos de detección como Bescós [30] que desarrolla una metodología para comparar la eficacia de las diversas técnicas de evaluación de disparidad entre fotogramas, tanto aisladamente como combinadas y plantea un modelo unificado de espacio de detección válido para la detección de cualquier tipo de cambio de plano, tanto abrupto como gradual.

2.3. Esquema general del proceso de segmentación

El objetivo del proceso de segmentación de vídeo es determinar, a partir de una secuencia de entrada, el intervalo de fotogramas que suceden entre el final de un plano y el principio del siguiente. A dicho intervalo se le conoce con el nombre de transición. Como ya se ha dicho anteriormente, existen dos tipos de transiciones, las abruptas o cortes y las graduales. En el caso de un corte, se define la transición *i-ésima* como la que sucede entre los fotogramas i e $i+1$, donde el fotograma *i-ésimo* es el final de un plano y el $i+1$ el principio del plano siguiente. Para las transiciones graduales, la transición *i-ésima* es la que ocurre entre los fotogramas $i - N_0$ (último fotograma del plano) e $i + N_0$ (primer fotograma del plano siguiente).

El proceso de segmentación de secuencias de vídeo sigue un esquema general común a la detección de transiciones abruptas (cortes) y de transiciones graduales (fundidos y cortinillas). Los pasos de dicho esquema son:

1. Cálculo de métricas

El primer paso de la detección es definir el parámetro o conjunto de parámetros (métricas) que, debidamente procesados, nos van a determinar el punto de la secuencia donde existe una transición. Este paso es muy importante, ya que de la buena elección de estos parámetros va a depender en gran medida el éxito de la segmentación. Para definir las métricas que se van a utilizar en la detección de cada tipo de transición habrá que estudiar previamente el tipo de transición, para determinar en qué características de la secuencia se manifiesta el hecho de que está ocurriendo un cambio de plano. Por ejemplo, una métrica elegida para detectar cortes será la diferencia de luminosidad entre los píxeles de fotogramas consecutivos, ya que cuando se produce un cambio de plano, el contenido de la escena cambia bruscamente, y con ello las luminosidades de los píxeles.

2. Procesado temporal

Una vez elegida la métrica, el siguiente paso es procesar la evolución de la misma a lo largo del tiempo. En el caso de los cortes y eligiendo como métrica la diferencia de luminosidad entre fotogramas consecutivos, antes de pasar a la detección, o sea antes de poner un umbral a la diferencia que determine qué transiciones lo sobrepasan, habrá que procesarla para quedarnos sólo con las diferencias más grandes que el umbral pero también de anchura temporal 1, ya que un corte es una transición abrupta que dura 1.

3. Paso a binario: umbral

Finalmente, a la señal procesada se le pone un umbral, decidiendo que ocurre una transición en todos aquellos instantes en que la métrica utilizada y procesada supera un determinado umbral.

Los métodos de detección de transiciones que serán presentados en este capítulo siguen todos el mismo esquema de procesado, por lo que para cada uno de ellos se particularizarán los pasos expuestos anteriormente. Además, todos ellos utilizan como entrada la *secuencia dc* de la original, así pues, como primer paso del esquema de procesado habría que obtener dicha secuencia. A continuación se repasan algunos conceptos sobre vídeo comprimido así como la formación de la *secuencia dc*.

2.3.1. Vídeo comprimido

La compresión de imágenes se plantea en el momento en que se pretende codificar una imagen como una señal digital. El problema reside en la cantidad de bits que se necesitan para realizar dicha codificación. La codificación MPEG [142, 139] consigue altas tasas de compresión de la señal de vídeo eliminando no sólo redundancias espaciales sino también temporales. Para conseguirlo se basa en dos filosofías fundamentales: la codificación para una imagen fija, es decir, eliminar redundancias dentro de una misma imagen y la determinación de vectores de movimiento que es la base para eliminar redundancias entre imágenes sucesivas.

Por el tipo de codificación que usa MPEG, las imágenes se tratan y codifican de distinta forma. Los distintos tipos de imágenes usados son:

- Imágenes *intraframe*, **I**: Se codifican como si fuesen imágenes fijas utilizando una norma similar a JPEG, por tanto, para decodificar una imagen de este tipo no hacen falta otras imágenes de la secuencia, sino sólo ella misma.
- Imágenes *interframe*, **P**: Es la predicción de una imagen con compensación de movimiento a partir de la I anterior o P anterior. Para decodificar una imagen de este tipo se necesita, además de ella misma, la I anterior o la P anterior decodificada.
- Imágenes *interframe* bidireccionales, **B**: Son imágenes que se codifican utilizando la I o P anterior y la I o P siguiente.
- Imágenes *intraframe* de baja resolución, **D**: Son imágenes de las mismas características que las I pero con menos resolución. Se usan en aplicaciones que no necesitan gran calidad, como en el avance rápido.

Lo interesante para un preprocesado de la información, como es la segmentación temporal, sería no tener que decodificar totalmente la secuencia comprimida en MPEG. A continuación describiremos el método que utiliza las llamadas imágenes *dc* y secuencias *dc*, cómo se pueden extraer de forma eficiente de vídeos comprimidos, y por qué es útil para rapidez y eficiencia en operaciones de análisis de escenas.

Imagen *dc* y secuencia *dc*

¹ Las imágenes *dc* son versiones reducidas espacialmente de las imágenes originales. Cada una de estas imágenes reducidas, una vez extraídas, pueden también ser usadas para otras aplicaciones, como por ejemplo la detección de cambios de escena. Algunos autores las han utilizado para una detección eficiente de los planos [154, 10], para generación automática de una representación compacta de la estructura y contenido de documentos de vídeo [71].

Una imagen se divide en bloques de $N \times N$ píxeles. El valor del píxel (i, j) de la *imagen dc* es la media de los valores de los píxeles del bloque (i, j) de la imagen original. La secuencia formada por las imágenes *dc* de fotogramas sucesivos de la secuencia original se llama *secuencia dc*.

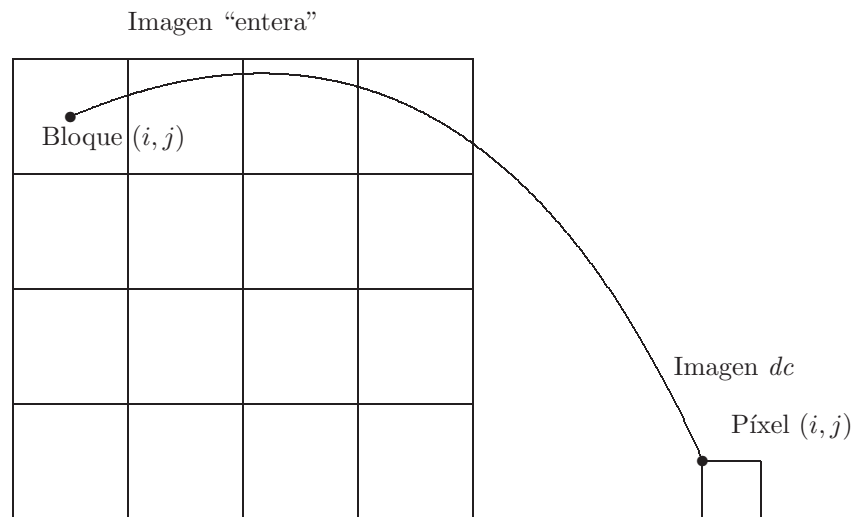


Figura 2.3: Imagen "entera" dividida en bloques frente a la imagen *dc*.

En la figura 2.4 se puede observar una imagen original de tamaño 512×512 y la correspondiente *imagen dc* de tamaño 64×64 , usando $N = 8$.

Las ventajas del uso de las imágenes *dc* y de las secuencias *dc* para el análisis de escenas es triple:

- Las operaciones se pueden desarrollar directamente sobre los datos comprimidos (más concretamente, realizando sólo una decodificación parcial), eliminando la necesidad de una descompresión del fotograma completo.

¹O también *DC-image* y *dc-sequence* respectivamente.



Figura 2.4: Imagen de *Lena*: (a) Imagen “entera” de 512×512 píxeles. (b) Imagen *dc* de tamaño 64×64 píxeles.

- Se trabaja con una pequeña fracción de los datos originales. Esto reducirá la carga computacional de los algoritmos y por consiguiente el tiempo de procesado.
- El proceso de promediado o filtrado paso bajo que sufre la imagen a la hora de obtener la imagen *dc* introduce otra mejora. Por una parte, eliminará gran número de falsas alarmas en la detección de cortes, al eliminar parte del movimiento de la cámara a lo largo de la secuencia. Y además, también minimizará el efecto del movimiento de los objetos en la imagen (en el caso de objetos menores de 8×8 píxeles, los eliminará totalmente).

Relación con la Transformada Discreta del Coseno

Para JPEG e imágenes I de MPEG, la imagen original se agrupa en bloques de $N \times N$ píxeles (normalmente $N = 8$), y el término *dc* o término de la “componente continua”, $c(0, 0)$ de la Transformada Discreta del Coseno 2-D, está relacionado con los valores de los píxeles, $f(i, j)$ de cada bloque de la imagen original a través de 2.13:

$$c(0, 0) = \frac{1}{N} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y), \quad (2.13)$$

lo cual es aproximadamente igual (salvo por un factor de escalado de N) a diezmar la imagen por un factor N habiéndola filtrado previamente con un filtro de media de tamaño $N \times N$.

Aunque la imagen *dc* es mucho más pequeña que la imagen original, contiene suficiente información. Esto sugiere que las variaciones de escena de naturaleza global (es decir, descartando el movimiento de objetos) que se producen en la imagen original, se siguen produciendo también en la imagen *dc*.

Para formar la *imagen dc* de una imagen original no comprimida se requieren $O(N^2)$ operaciones por bloque², mientras que en el caso de imágenes con compresión basada en la DCT, el término *dc* es una versión escalada del valor de la media de un bloque. Cuando $N = 8$, de 2.13, el valor de la media es $\frac{1}{8}$ del término *dc* de la DCT de un bloque.

Además de la extracción de la *imagen dc* para imágenes I será necesario obtenerla para el caso de imágenes P y B . Existen varios métodos para obtener la secuencia *dc* a partir de la secuencia comprimida. En esta tesis no vamos a ocuparnos de este tema, pero para obtener información sobre dichos métodos se puede consultar [155, 137].

2.4. Detección de cortes

A la hora de restaurar películas antiguas, ya se ha dicho anteriormente que se deben usar tantas propiedades de la señal de vídeo como sea posible. Una propiedad clave a utilizar es la fuerte correlación que existe entre los diferentes fotogramas a lo largo del tiempo. Sin embargo, esta propiedad desaparece cuando ocurre un cambio brusco de escena, es por ello que la detección de cortes es el primer paso y uno de los más importantes cuando se procede a restaurar. Es por esto, que la detección de cortes sea la más precisa que debemos realizar frente a la de otro tipo de transiciones, ya que es en los cortes donde se produce mayor ruptura de la continuidad temporal. Si en la segmentación nos equivocamos detectando un falso positivo es menos perjudicial que si no detectamos un corte existente, ya que en este caso todos los fotogramas del plano, y en general los más cercanos al corte no detectado, serán mal restaurados.

Siguiendo los pasos del esquema general para la segmentación presentado en la sección 2.3, primero se presentarán las métricas utilizadas para la detección de los cortes, concretamente dos tipos diferentes de ellas, las que son sensibles a cambios de luminosidad bruscos y aquellas que son robustas frente a los mismos, que son las que se han utilizado en la detección de cortes para películas antiguas, en las que estos cambios bruscos de luminosidad se producen. Posteriormente, se explicará el procesado temporal utilizado, el cual se basa en la morfología matemática. Finalmente, se expone el método experimental utilizado para la elección de un umbral óptimo.

2.4.1. Cálculo de métricas

Métricas sensibles a cambios de luminosidad: $d_c(n)$

Como hemos visto en la sección 2.2, han sido propuestas multitud de técnicas para detectar cortes en secuencias. La mayoría de ellas se basa en las diferencias existentes entre un fotograma y el siguiente, diferencias entre la luminosidad de los píxeles, diferencias en sus histogramas, etc. La primera métrica utilizada en nuestro algoritmo es una de las más utilizadas, la diferencia media absoluta entre los niveles de luminosidad de los píxeles de

²Las correspondientes al filtrado paso bajo y el diezmado.

fotogramas sucesivos:

$$d_c(n) = \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J |f_n(i, j) - f_{n-1}(i, j)|, \quad (2.14)$$

donde I y J son la dimensión horizontal y vertical en píxeles de la imagen de cada uno de los fotogramas (f_n), n es el índice temporal del fotograma e (i, j) son las coordenadas espaciales. Nuestro algoritmo calcula d_c sobre la secuencia dc .

En la figura 2.5 puede verse la imagen diferencia de dos fotogramas consecutivos cuando hay un corte y cuando no lo hay.



Figura 2.5: Diferencia de dos fotogramas consecutivos en la secuencia *Drama*: (a)Imagen diferencia de los fotogramas 363 y 364, pertenecientes al mismo plano. (b)Imagen diferencia de los fotogramas 364 y 365, que constituyen una transición abrupta.

Para imágenes en color el método equivalente consistirá en descomponer la imagen en sus componentes R, G y B, y calcular la diferencia ponderada en función del nivel de luminosidad que aporta cada componente³, de tal forma que la expresión de $d_c(n)$ quedará:

$$d_{c_{RGB}}(n) = \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J 0,3 |f_{n_R}(i, j) - f_{n-1_R}(i, j)| + 0,59 |f_{n_G}(i, j) - f_{n-1_G}(i, j)| + 0,11 |f_{n_B}(i, j) - f_{n-1_B}(i, j)|. \quad (2.15)$$

Esta expresión también nos permite (al trabajar con la diferencia de la luminancia) poder comparar resultados obtenidos para imágenes de grises y para imágenes RGB.

Así pues, d_c mide el grado de diferencia en la transición entre dos fotogramas. Las causas de estas diferencias pueden ser debidas a diferentes factores:

- Cambios de plano.

³Lo que se hace en la práctica es obtener la secuencia de luminancia y sobre ella actuar como si se tratase de una secuencia de grises.

- Movimiento de objetos.
- Movimiento de cámara.
- Variaciones de luminosidad (*flicker*).

En secuencias estándar (con buenas condiciones, no degradadas) las diferencias debidas a variaciones de luminosidad no son apreciables, excepto en algunos casos especiales como escenas en que se provocan cambios bruscos de intensidad, como la activación de un *flash* para tomar fotos (*flashlight*).

Los movimientos de objetos y cámara normalmente tienen lugar durante más de una transición, produciendo un pulso ancho en la señal d_c . Sin embargo, cuando se produce un corte aparece un pico de anchura 1 en d_c . Esta diferencia de anchura en los máximos locales de d_c será explotada, en la etapa de procesamiento temporal, para diferenciar movimientos en la escena (de cámara y de objetos) y cambios abruptos de plano.

En la figura 2.6, podemos observar el resultado obtenido al calcular d_c sobre la secuencia *Drama*. Cada uno de los máximos locales que aparecen indican un fotograma donde se pro-

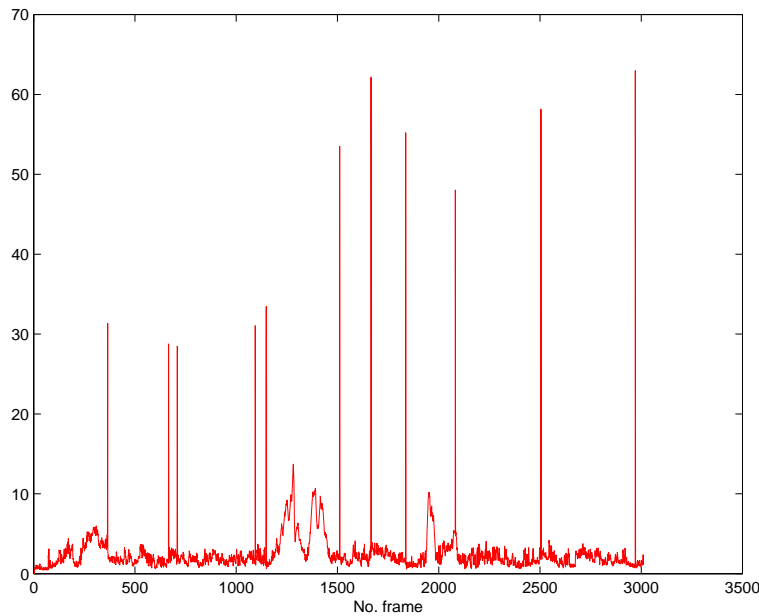


Figura 2.6: d_c calculado para la secuencia *Drama*.

duce una transición abrupta o corte. Llegados a este punto, las técnicas posibles a aplicar, para determinar que un cierto pico corresponde a un corte, son diferentes. Se podría poner directamente un umbral global, de tal forma que todas las diferencias que superen el umbral se consideren cortes. Esta técnica aplicada tal cual sobre d_c puede producir errores: en las zonas donde hay mucho movimiento aparecen, como se ha apuntado anteriormente, “mesetas” con algunos picos que, al superar el umbral, podrían considerarse transiciones cuando en realidad no lo son. Además, la elección de un valor del umbral u otro afectará de manera

importante, ya que los cambios de escena son una actividad “local” en el dominio temporal, y por tanto será más apropiado aplicar un umbral relativo, en función de dicha actividad. Para eliminar estos dos efectos negativos en la detección de cortes, como paso previo a la colocación de un umbral global, se aplica el procesado temporal, que consistirá en una serie de operadores morfológicos sobre la señal $d_c(n)$, los cuales pretenderán básicamente eliminar de la señal todos aquellos picos más anchos que 1, los cuales no corresponden a un corte. Estos operadores serán explicados más exhaustivamente en la sección 2.4.2.

Métricas robustas frente a cambios de luminosidad

Como hemos dicho anteriormente, un problema que tienen las secuencias que debemos restaurar son los cambios abruptos y no deseados de luminosidad, defecto que será tratado en el capítulo 3 de esta tesis, en el cual se presentarán métodos para minimizarlo. Sin embargo, a la hora de segmentar, debido al tipo de métricas utilizadas para la detección, este defecto afecta empeorando los resultados. Por ello, se presentan a continuación dos métricas que son robustas frente a estos cambios de brillo: $d(n)$, que tiene en cuenta las diferencias de luminosidad y además las diferencias entre las medias de dos fotogramas consecutivos; y $d_\rho(n)$, que calcula una especie de correlación determinando si la variación de un píxel con respecto a la media en un fotograma se mantiene en el siguiente.

Definición de $d(n)$

El algoritmo propuesto en el apartado anterior es suficiente para localizar más del 90% de los cortes, sin embargo adolece de ser muy sensible a las variaciones del nivel de brillo en los fotogramas, siendo éste un hecho muy presente en las secuencias a restaurar procedentes de películas antiguas. En la figura 2.7, aparecen tres fotogramas consecutivos de la secuencia *Chaplin*. Como podemos observar, entre los fotogramas 1498 y 1499, pertenecientes al mismo plano, se ha producido una variación brusca en el brillo. La consecuencia de esta variación no deseada, es la aparición de un pico en $d_c(n)$, como puede apreciarse en la figura 2.8. No existe ninguna forma de saber si los picos del módulo de la diferencia se han producido por la existencia de un corte o por un fenómeno relacionado con el brillo del fotograma, y por tanto, en el caso que nos ocupa detectaríamos una falsa transición.

Con el fin de solucionar este problema se piensa en nuevas métricas que sean independientes del brillo o nivel medio del fotograma. La idea básica de estas métricas es separar la diferencia entre dos fotogramas consecutivos en dos términos:

- Diferencias debidas al cambio en el valor del brillo medio de la escena.
- Diferencias debidas al cambio en el contenido de la escena.

Denotemos μ_n a la media del fotograma n de la secuencia, dada por

$$\mu_n = \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J f_n(i, j).$$



(#1498)

(#1499)



(#1500)

Figura 2.7: Fotogramas de la secuencia *Chaplin*.

Definimos también la diferencia entre las medias de dos fotogramas sucesivos, d_μ

$$d_\mu(n) = |\mu_n - \mu_{n-1}|.$$

Si recordamos la ecuación de d_c dada en 2.14 se observa que $d_\mu(n) \leq d_c(n)$, y por tanto podemos escribir

$$d(n) = d_c(n) - d_\mu(n) \quad d(n) \geq 0. \quad (2.16)$$

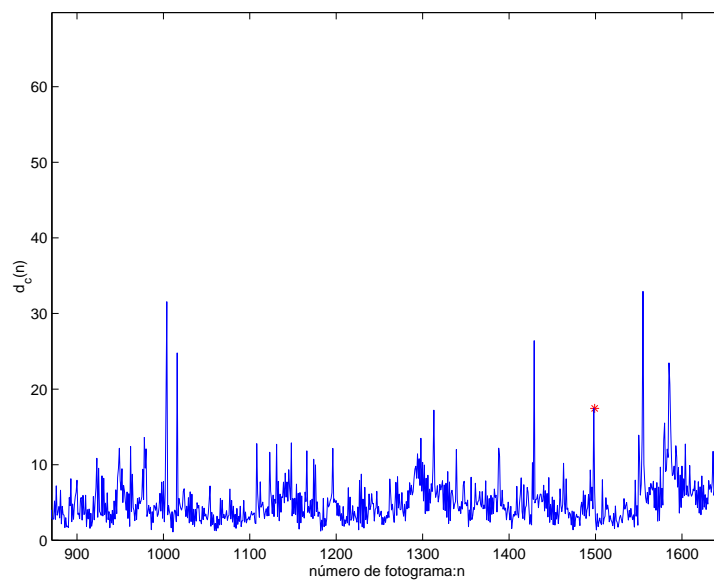


Figura 2.8: Módulo de la diferencia para una subsecuencia de *Chaplin*. La falsa transición aparece marcada con un “*”.

Las diferencias debidas a cambios en la media vienen dadas por d_μ , mientras que el término $d(n)$ representa el resto de las diferencias. En el caso de secuencias con variaciones bruscas en la media (con presencia de *flicker*) la mayoría de las diferencias son debidas a la contribución de $d_\mu(n)$ mientras que $d(n) \approx 0$.

La figura 2.9 muestra $d_c(n)$ y $d(n)$, así como indicaciones de dónde ocurren los cortes realmente. Se puede observar que la utilización de d_c produciría una gran cantidad de falsos positivos (picos estrechos en $d_c(n)$). En la figura 2.10 se puede observar un ejemplo de transición que podría ser considerada como corte si se utilizase $d_c(n)$ como métrica en la detección, sin embargo ese mismo punto en $d(n)$ ya no es un pico tan alto como para superar un umbral. La figura 2.11 muestra un ejemplo de transición que se detecta correctamente usando cualquiera de las dos métricas ($d_c(n)$ o $d(n)$). Por otro lado, el número de picos en $d(n)$ es considerablemente más reducido, reduciéndose las falsas detecciones, y además manteniéndose los máximos que correspondían a cortes reales, no bajando la probabilidad de acierto.

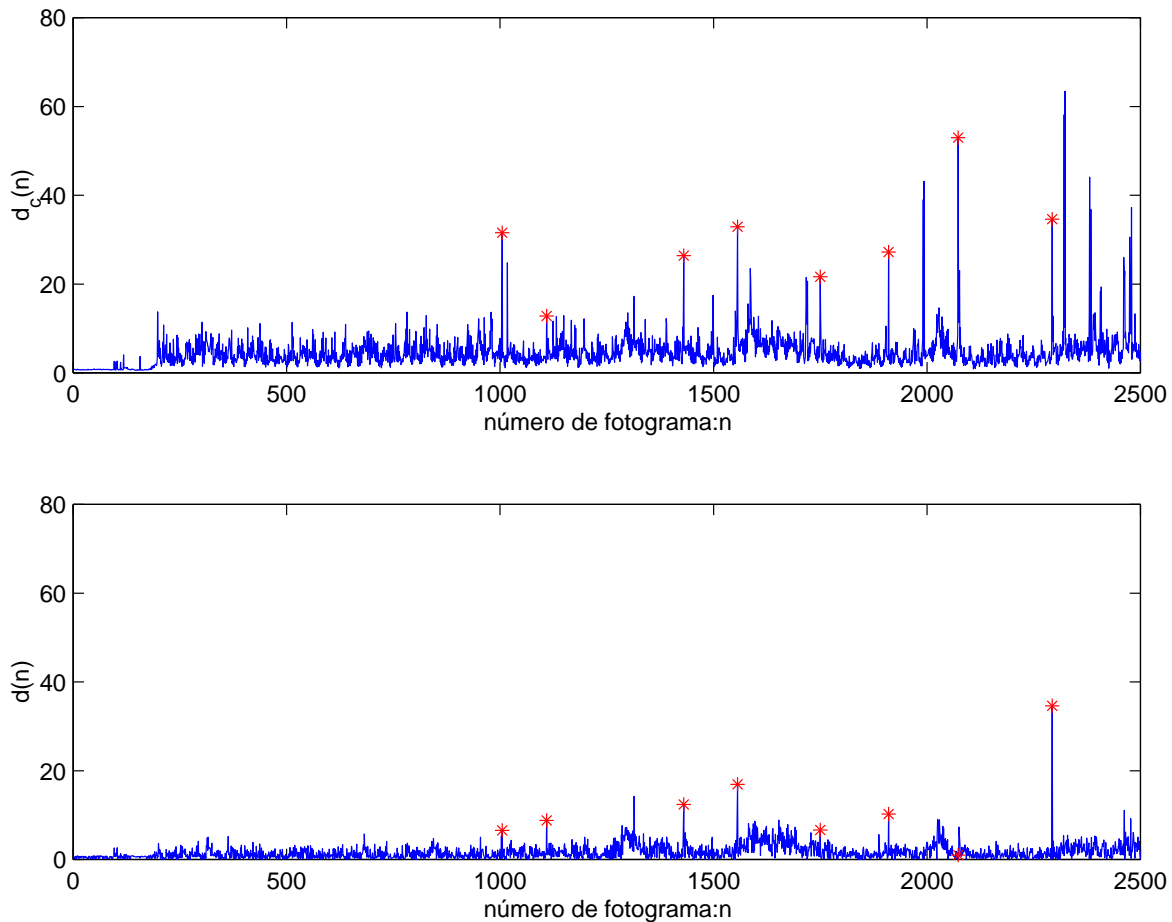


Figura 2.9: $d_c(n)$ y $d(n)$ de 2500 fotogramas de la secuencia *Chaplin*. El * muestra dónde existe un verdadero corte.



Figura 2.10: Fotogramas 2320 y 2321 de la secuencia *Chaplin*. Véase el pico correspondiente en $d_c(n)$ en la figura 2.9, y cómo no aparece dicho pico en la misma posición en $d(n)$.



Figura 2.11: Fotogramas 1555 y 1556 de la secuencia *Chaplin*. Véase los picos correspondientes en $d_c(n)$ y $d(n)$ en la figura 2.9.

Sin embargo, aún aparecen falsos positivos en $d[n]$ como, por ejemplo, el pico en el fotograma 1313. Si analizamos la comparación entre este fotograma y el siguiente puede observarse que se ha producido un movimiento de cámara muy brusco entre ambos fotogramas, hecho que provoca el que las diferencias entre ambos sean elevadas (ver figura 2.12), siendo por tanto la detección de esta falsa transición un problema del contenido de la película y no de la métrica utilizada.

El verdadero problema del uso de $d[n]$ es que algunos cortes no son detectados. Esto puede observarse, en la secuencia ejemplo que estamos analizando, entre los fotogramas 2072 y 2073 (figura 2.13). Si analizamos el caso se observa que la transición no se detecta porque el cambio de plano lleva implícito un gran cambio del brillo medio del fotograma. Para intentar resolver estas situaciones podemos utilizar diferentes métricas. La más inmediata sería añadir a los cortes detectados utilizando $d[n]$, aquellos valores de d_μ (diferencia entre las medias) que superan un cierto umbral.



Figura 2.12: Fotogramas 1313 y 1314 de la secuencia *Chaplin*.



Figura 2.13: Fotogramas 2072 y 2073 de la secuencia *Chaplin*.

Definición de $d_\rho(n)$

Una alternativa al uso de $d(n)$ es definir una nueva métrica que denotaremos por d_ρ . En primer lugar definiremos el signo modificado de un fotograma como:

$$\rho_n(i, j) = \begin{cases} 1 & f_n(i, j) > \mu_n + t \\ -1 & f_n(i, j) < \mu_n - t \\ 0 & \text{otro caso,} \end{cases} \quad (2.17)$$

donde t es un pequeño umbral. El aspecto de ρ_n para los pares de fotogramas de las figuras 2.10 y 2.13 se muestra en las figuras 2.14 y 2.15, y para cada píxel determina si el nivel de dicho píxel está por encima de la media de la imagen, por debajo o es casi igual.⁴

La idea de la métrica que definiremos a continuación es medir la diferencia (o el grado de parecido) en el signo modificado de dos fotogramas consecutivos, idea similar a la de la correlación, por lo que la llamaremos *correlación del signo modificado* $d_\rho(n)$, la cual viene dada por:

⁴El umbral t , típicamente de valor 2, se utiliza para desechar variaciones debidas al ruido



Figura 2.14: $\rho_n(i, j)$ para los fotogramas de la figura 2.10. Negro representa -1, blanco 1 y gris medio 0.



Figura 2.15: ρ para los fotogramas de la figura 2.13.

$$d_\rho(n) = \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \rho_n(i, j) \rho_{n-1}(i, j). \quad (2.18)$$

La correlación del signo modificado $d_\rho(n)$ es una señal unidimensional que está en el rango $-1 \leq d_\rho(n) \leq 1$. Un valor de 1 significa un parecido total entre los signos modificados ($\rho(i, j)$) de los fotogramas comparados, y un valor de -1, una diferencia total. En otras palabras, $d_\rho(n)$ mide la cantidad de píxeles que estando por encima (debajo) de la media en el fotograma $n-1$ lo siguen estando en el fotograma n . Así pues, para detectar un corte habrá que determinar mínimos locales estrechos en esta señal.

Al igual que ocurre para $d_c(n)$, será necesario redefinir la expresión 2.18, de tal forma que sea aplicable a imágenes RGB. La idea es la misma que utilizamos en $d_{c_{RGB}}(n)$: combinar linealmente la información que aportan cada uno de los espacios de color, obteniendo así resultados comparables en imágenes de grises y en imágenes de color, lo cual se puede calcular obteniendo la luminancia de la secuencia y calculando sobre ella $d_\rho(n)$.

La gráfica 2.16 muestra $d_\rho(n)$ para la secuencia *Chaplin*. La detección de valles estrechos se puede realizar fácilmente calculando el residuo del cierre morfológico, como se verá en la

sección siguiente, obteniéndose una señal como la de la figura 2.17, donde puede observarse que al poner un umbral adecuado, se detectan correctamente los cortes deseados, así como un par de falsas detecciones en los pares de fotogramas 1313-1314 y 2461-2462, los cuales aparecen en las figuras 2.12 y 2.18. Si pasamos a analizar el porqué de la aparición de mínimos en $d_\rho(n)$ en estos fotogramas, podemos observar en la figura 2.12 un gran movimiento hacia la izquierda, el cual se puede apreciar fácilmente comparando el lado derecho de ambas imágenes. En el caso de la figura 2.18 lo que ha ocurrido es un brusco movimiento en vertical (obsérvese la parte superior de la puerta), así como un cambio en la posición de los actores.

En toda la secuencia analizada (y todas las analizadas en esta tesis), el hecho que estamos comentando (falsas detecciones por la existencia de mínimos locales en d_ρ) ocurre siempre en casos similares: bruscos movimientos de un fotograma al siguiente. Una explicación posible a la aparición de estos movimientos tan bruscos podría ser la pérdida de fotogramas entre los dos “*consecutivos*” analizados, conclusión a la que se llega por simple observación de la secuencia. Otro problema de las películas antiguas para la falsa segmentación es la vibración indeseada de la secuencia debido a la vibración de la cámara al filmar, los empalmes manuales de cinta, o a la desalineación en el paso de la película de cine a vídeo en el telecine.

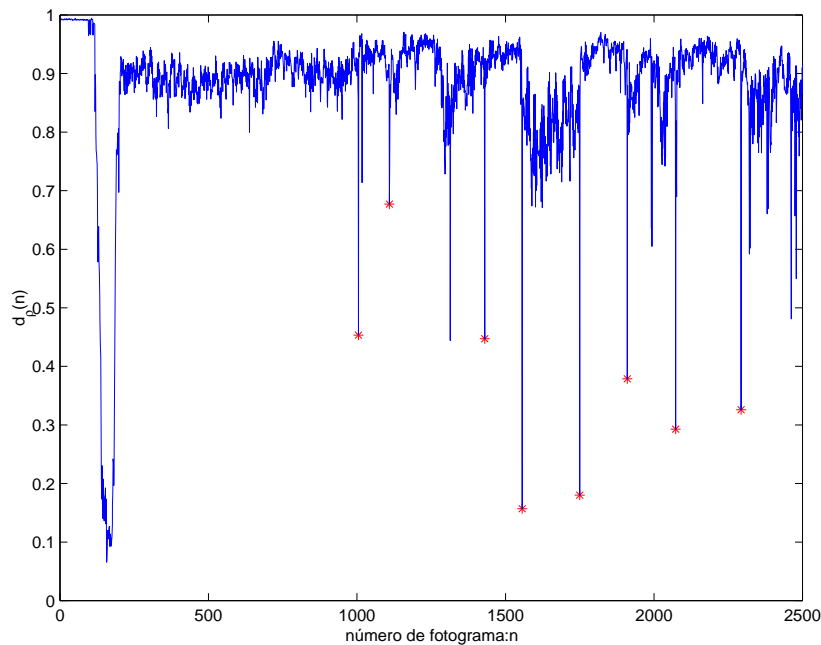


Figura 2.16: $d_\rho(n)$ para la secuencia *Chaplin*. El * aparece en los fotogramas donde ocurren cortes reales.

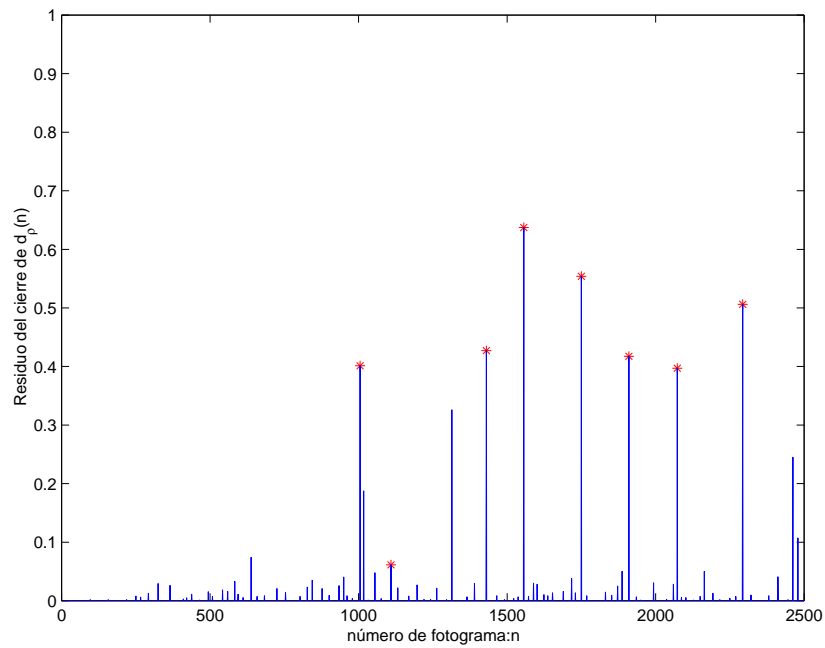


Figura 2.17: Residuo del cierre de $d_p(n)$ para la secuencia *Chaplin*. El * aparece en los fotogramas donde ocurren cortes reales.

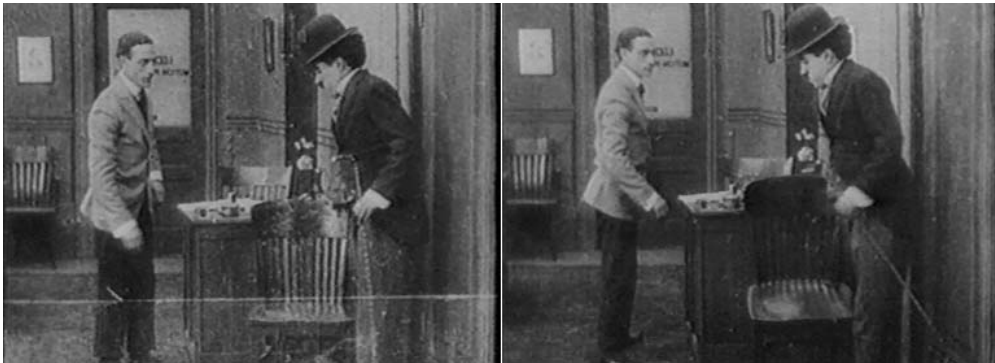


Figura 2.18: Fotogramas 2461 y 2462 de la secuencia *Chaplin*.

2.4.2. Procesado temporal

Se ha apuntado anteriormente que para detectar cortes correctamente, sin la aparición de un número elevado de falsos positivos, los umbrales para las métricas expuestas deben ser relativos, en función de la actividad local en la zona de la secuencia donde se encuentra el corte. Esto es complicado, ya que antes del proceso de umbralización habría que hacer un estudio de dicha actividad. Para resolver este problema se recurre a la morfología matemática posibilitando, como veremos a continuación, el uso de umbrales globales.

Operadores morfológicos: Apertura y Cierre

La morfología matemática (M.M.) [129, 130] es un procesado no lineal basado en máximos y mínimos. Se trata de una técnica aplicable a muchos problemas de tratamiento de señales y entre éstas las imágenes. En este apartado vamos a presentar algunos operadores morfológicos unidimensionales, ya que las señales que pretendemos filtrar ($d_c(n)$, $d(n)$ y $d_\rho(n)$) sólo tienen una dimensión, la temporal. Más adelante, ya en la restauración, se volverá a utilizar la morfología matemática sobre imágenes de grises.

En filtros lineales, la respuesta impulsional $h[n]$ caracteriza totalmente al sistema. En los operadores morfológicos, el *elemento estructurante* (EE) determina la ventana de análisis del operador, actuando como una sonda que analiza el entorno del punto de trabajo.

Los dos operadores básicos de la M.M. son la dilatación y la erosión. Se trata de dos operadores duales cuya combinación da lugar a gran cantidad de operadores morfológicos. La dilatación obtiene el máximo de los valores abarcados por el elemento estructurante, mientras que la erosión calcula el mínimo. Los operadores que nos interesan en el filtrado de nuestras señales, la apertura y el cierre, son combinaciones de la erosión y la dilatación.

- **Apertura:** Consiste en realizar primero una erosión, y sobre el resultado una dilatación, utilizando el mismo EE. Al aplicarse sobre funciones unidimensionales elimina los picos en que no cabe el elemento estructurante; el resto se mantiene inalterado.
- **Cierre:** En este caso se trata de una dilatación seguida por una erosión con el mismo EE. Eliminará los valles más estrechos que el elemento estructurante, y el resto se mantendrá inalterado.

La apertura y el cierre son, por tanto, dos operadores duales, que para el caso que nos ocupa, nos permitirán eliminar los picos o los valles del ancho que le indiquemos con el EE. Tal y como se puede observar en la figura 2.19, la elección de un EE u otro produce resultados totalmente diferentes.

Sin embargo, normalmente lo que desearemos, y de hecho en nuestro caso será así, es quedarnos con picos (valles) de determinada anchura; por ejemplo, para detectar los cortes desearemos quedarnos con los picos de $d_c(n)$ de anchura 1. En estos casos habrá que calcular el residuo (resta) entre la señal original y el resultado de la apertura para el caso de d_μ o d (o el cierre para el caso de d_ρ) con un elemento estructurante de tamaño 2.

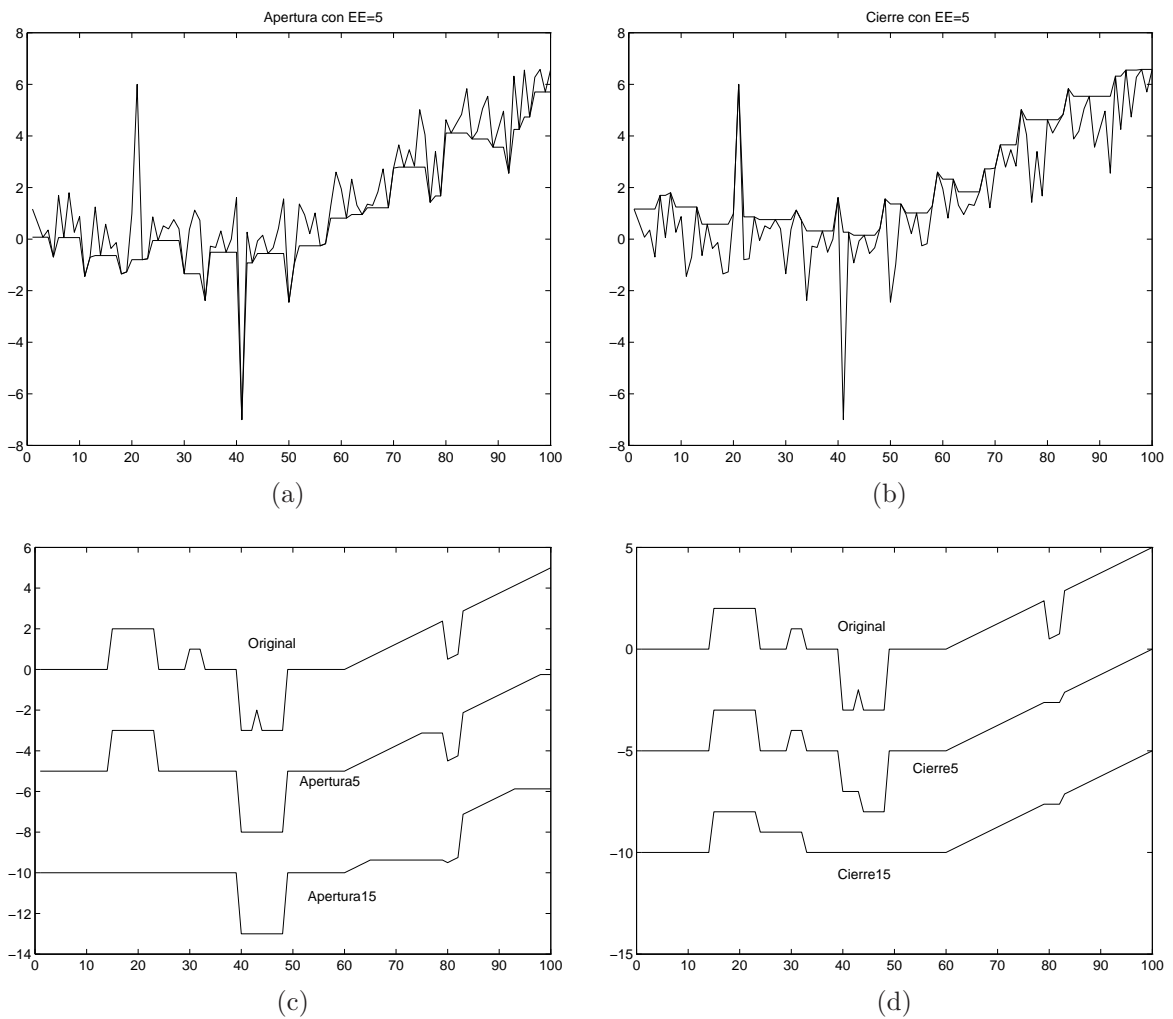


Figura 2.19: Resultados de la apertura y el cierre sobre funciones unidimensionales: (a)Apertura con EE de tamaño 5. (b)Cierre con EE de tamaño 5. (c)Comparación del efecto de la apertura con dos EE diferentes. (d)Comparación del efecto del cierre con dos EE diferentes.

Aplicación de la Apertura y el Cierre a $d_c(n)$.

$d_c(n)$ es una señal unidimensional de la que nos interesan sus picos. La primera idea para conseguirlo sería poner un umbral directamente a la señal, sin embargo ya comentamos el problema que se tiene al aplicar un umbral directamente: algún pico que constituye un máximo relativo de una meseta, pero de poca amplitud respecto a sus puntos adyacentes, puede ser considerado una transición cuando en realidad no lo es. Las mesetas se producen por dos razones: mucho movimiento de la imagen (ya sea por los objetos que hay en ella o por el movimiento de la cámara) o por la existencia de algún efecto de transición gradual. Este movimiento de la imagen dura varios fotogramas, por eso se producen las zonas con formas de meseta (conjunto de valores sucesivos relativamente altos). Para la detección de cortes, nuestro objetivo es eliminar estas mesetas que nos producen las falsas alarmas.

Así pues, deseamos quedarnos, de todos los picos de d_c , sólo con aquellos que tengan anchura 1, que son los que nos darán los cortes. El operador más indicado para esto es la apertura, con la que eliminamos los picos de anchura deseada, o si nos queremos quedar con ellos, realizamos la diferencia entre la señal original y la apertura ⁵. Como las transiciones abruptas o cortes serán picos de anchura uno, el elemento estructurante a aplicar será de tamaño 2, con lo que los picos más estrechos (los de anchura 1) no cabrán dentro del elemento estructurante y serán eliminados al realizar la apertura, apareciendo únicamente éstos en el residuo. En figura 2.20 se muestran los resultados para la secuencia *Newsb*. Se trata de la secuencia de un programa informativo (“Informe Semanal”) de TVE. Tiene cuatro cortes en los fotogramas 625, 744, 1247 y 1286 que son localizados sin problemas.

Los resultados obtenidos hablan por sí solos. Aunque esta secuencia no tenga mucho movimiento (salvo el plano entre el corte 625 y 744), con el residuo de la apertura nos quedamos exclusivamente con los picos de anchura uno. Ahora sí se puede aplicar un umbral fijo, u_{dc} , cuya elección no será muy crítica, por ejemplo $u_{dc} = 5$.

Sin embargo, la M.M. todavía nos puede ayudar más. En edición de vídeo, los cortes no son siempre abruptos. Se puede utilizar algún fotograma intermedio, de manera que el efecto podrá ser considerado como un fundido, pero muy corto. En la figura 2.21, se muestra un corte utilizando un solo fotograma de transición. Un recurso muy utilizado también, sobre todo en las películas antiguas, consiste en realizar el cambio de un plano a otro mediante un oscurecimiento (con incluso algún fotograma a negro) o un aclarado (con incluso algún fotograma a blanco). Si dicho efecto ocupa muy pocos fotogramas intermedios, podemos utilizar también $d_c(n)$ para su detección. Lo más habitual es tener uno o dos fotogramas intermedios, que equivalen a transiciones con picos de anchura dos o tres respectivamente, los cuales podrían detectarse aislando mediante residuos de aperturas, los picos de anchura dos y tres. La figura 2.22 muestra un ejemplo de este tipo de transición.

⁵En M.M., la diferencia entre la original y la apertura se conoce como residuo de la apertura o *Top-Hat*.

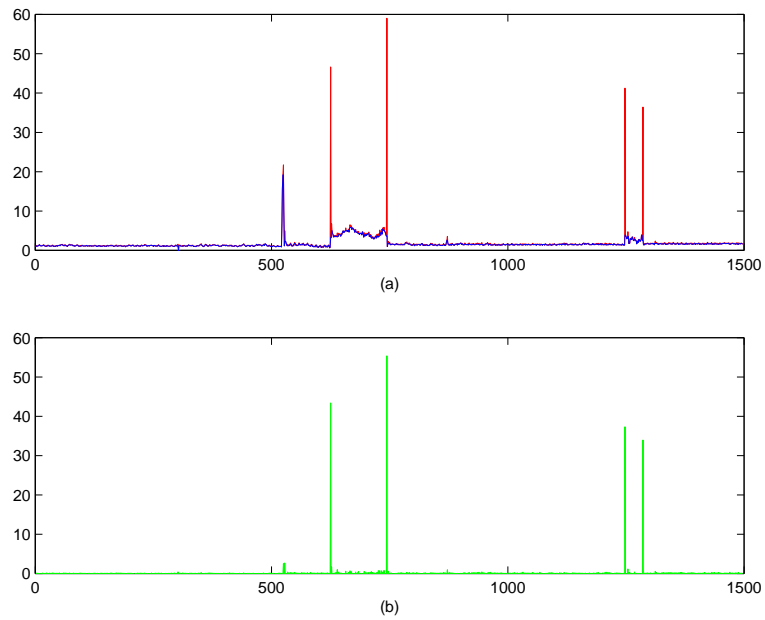


Figura 2.20: Aplicación de la apertura para la localización de los cortes en *Newsb*: (a) En rojo, aparece $d_c(n)$ y en azul el resultado de su apertura con $EE = 2$. (b) Diferencia entre $d_c(n)$ y la apertura.

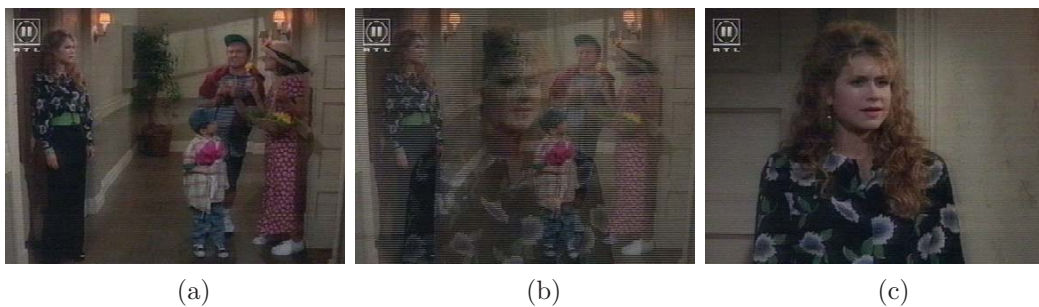


Figura 2.21: Ejemplo de transición abrupta utilizando un fotograma intermedio.

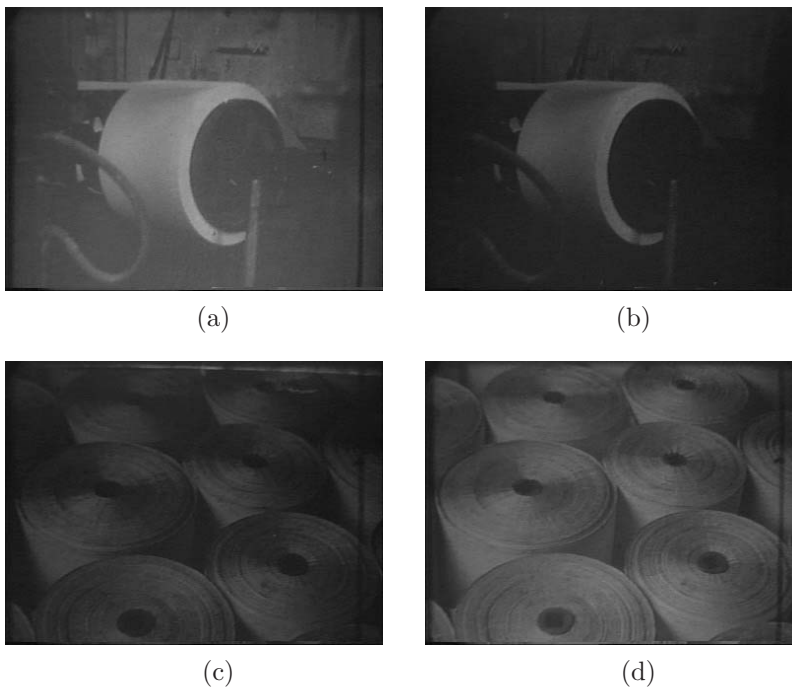


Figura 2.22: Ejemplo de transición utilizando dos fotogramas, (b) y (c), intermedios: (a)Fotograma 355. (b)Fotograma 356. (c)Fotograma 357. (d)Fotograma 358 de la secuencia *Alcoi*

Así pues, el método para detectar cortes o fundidos cortos se podría resumir de la siguiente manera:

- **Residuo de $d_c(n)$ y de la apertura con $EE = 2$:** Se quedará con los picos de anchura uno, que corresponden a las transiciones abruptas.
- **Residuo de la apertura con $EE = 2$ y de la apertura con $EE = 3$:** Se quedará con los picos de anchura dos, que corresponden a las transiciones de un plano a otro mediante un fotograma intermedio.
- **Residuo de la apertura con $EE = 3$ y de la apertura con $EE = 4$:** Se quedará con los picos de anchura tres, que corresponden a las transiciones que utilizan dos fotogramas intermedios.

Se podría pensar en continuar aplicando el mismo método para obtener los fundidos o transiciones graduales de distintos tamaños. El problema surge porque en transiciones graduales de más de tres fotogramas intermedios, la diferencia entre un fotograma y otro se reduce, obteniéndose picos de amplitud cada vez menor, con lo que los umbrales a fijar serán cada vez más bajos, pudiéndose introducir falsas alarmas. Así pues, con la información de $d_c(n)$ buscaremos transiciones de dos fotogramas intermedios o menos. En la figura 2.23 se muestran los resultados del proceso aplicado a la secuencia *Alcoi*. Ahora se aplicará el umbral, $u_{dc} = 3$, y obtendremos siete cortes (hay un corte entre los fotogramas 17 y 18 que no se

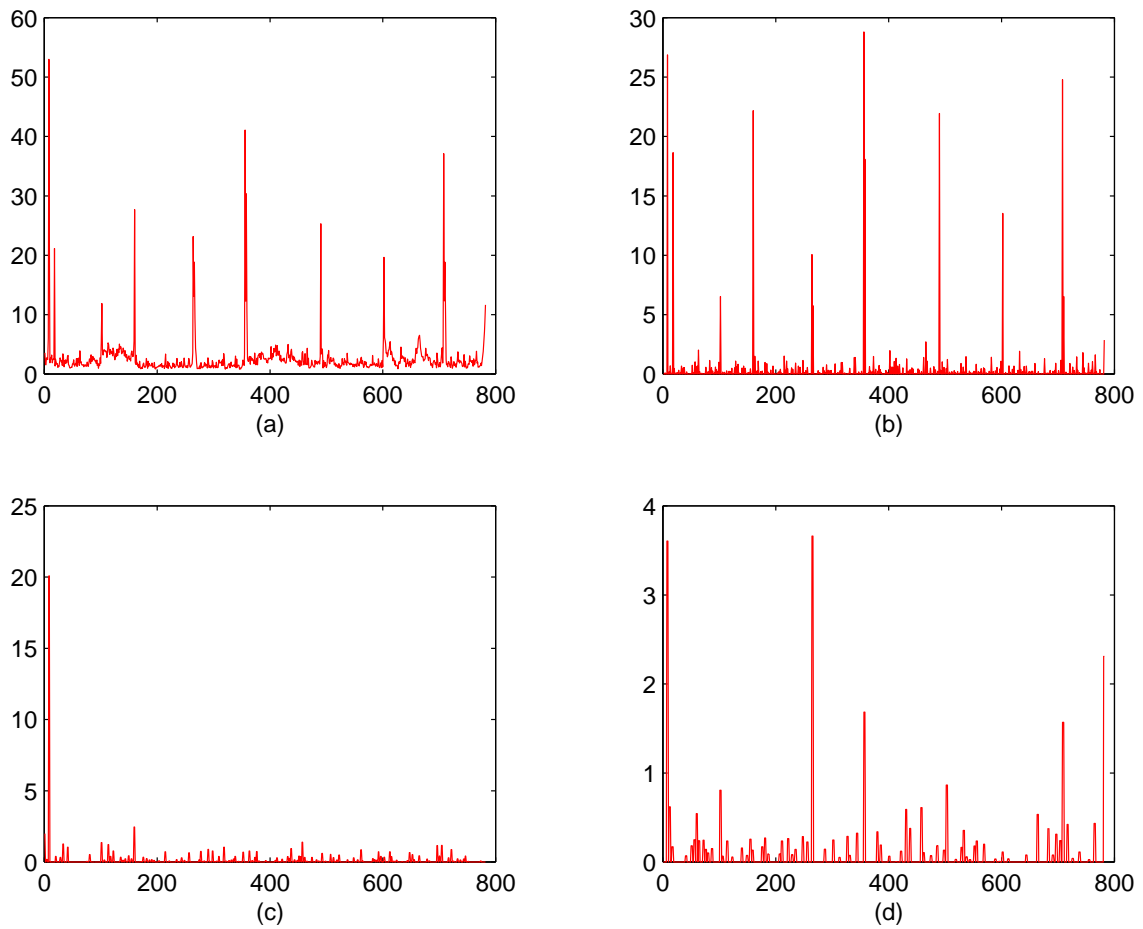


Figura 2.23: Aplicación de la apertura para la localización de los cortes en *Alcoi*: (a) Señal $d_c(X, Y)$. (b) Residuo entre $d_c(n)$ y la apertura con $EE = 2$. (c) Residuo entre la apertura con $EE = 2$ y la apertura con $EE = 3$. (d) Residuo entre la apertura con $EE = 3$ y la apertura con $EE = 4$.

localiza), una transición con fotograma intermedio y tres transiciones con dos intermedios. Los resultados son muy satisfactorios, al perder únicamente un corte.

Con las aperturas, según hemos visto, podemos obtener los picos de la anchura deseada. Vamos a ver que el cierre nos puede mejorar el proceso. Recordando, un cierre nos elimina los valles menores que el EE . Si aplicamos, por ejemplo, un cierre de $EE = 20$ a la señal $d_c(n)$ obtendremos una señal donde sólo quedará el pico máximo de cada 20 valores de $d_c(n)$. La utilidad surge inmediatamente: podemos fijar la duración mínima del plano. Es decir, si como en nuestro caso las secuencias han sido digitalizadas a 25 fotogramas/segundo, tomando un cierre inicial de $EE = 12$, limitamos a dos posibles cortes producidos en un segundo, o lo que es lo mismo, imponemos que el tamaño mínimo de un plano sea de 12 fotogramas o de 0,5 segundos. Mediante este cierre podemos descartar también algunas falsas alarmas. En la figura 2.24 se muestra un ejemplo de su aplicación.

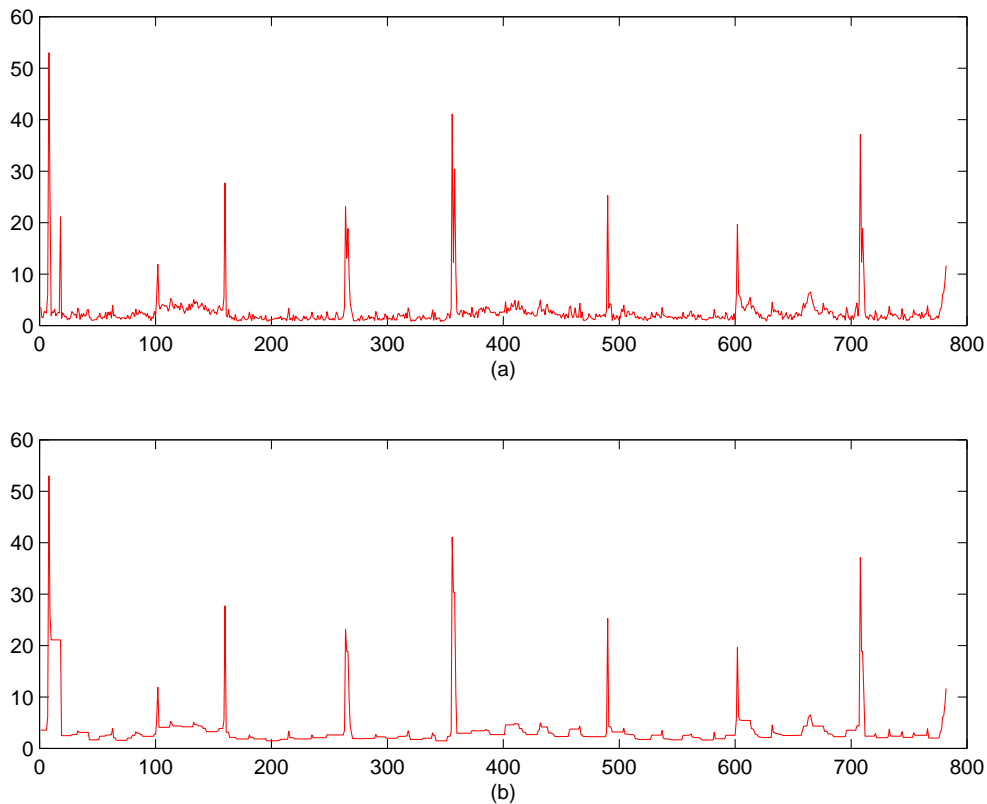


Figura 2.24: Resultado del cierre sobre la secuencia *Alcoi*: (a) $d_c(n)$. (b) Cierre de $d_c(n)$ con un $EE = 9$.

A modo de resumen, el proceso para la detección de cortes, partiendo de $d_c(n)$, consistirá en comenzar con un cierre, que nos limitará las posibles transiciones en un segundo, y después aplicar sucesivas aperturas para encontrar las transiciones que superen u_{dc} .

Aplicación de la Apertura y el Cierre a $d_\rho(n)$.

Las señales $d_c(n)$ y $d_\rho(n)$ son duales, es decir, en $d_\rho(n)$, las transiciones se corresponden con valles de anchura 1 (o de anchura 2 y 3 para el caso de fotogramas intermedios) mientras que en d_c se correspondían con picos. De esta forma, las operaciones a realizar serán complementarias: usaremos la apertura para limitar la dimensión temporal del plano y cierres sucesivos para determinar los distintos tipos de transiciones.

En la figura 2.25 aparece $d_\rho(n)$ y el resultado de aplicarle una apertura de tamaño 9, para la secuencia *Alcoi*. Para este ejemplo, se puede apreciar claramente el efecto que produce la

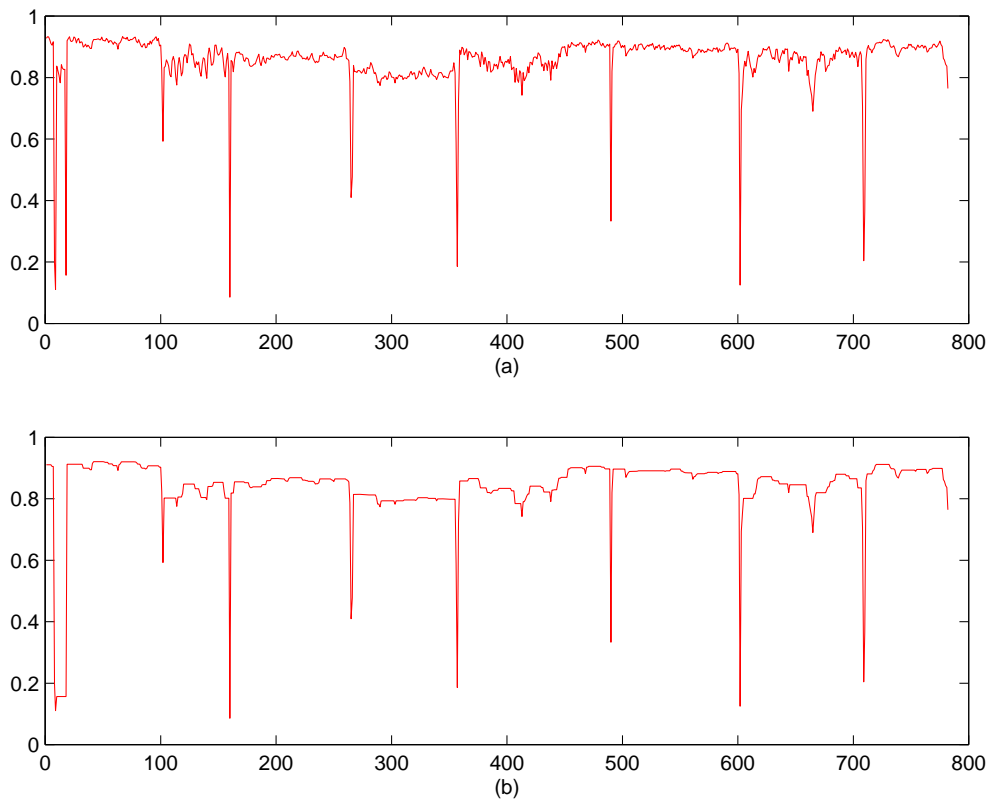


Figura 2.25: Resultado de la apertura sobre la secuencia *Alcoi*: (a) $d_\rho(n)$. (b) Apertura de $d_\rho(n)$ con un $EE = 9$.

apertura cuando tenemos dos valles separados una distancia menor del EE: se reduce a un solo valle.

A continuación, deberemos aplicarle sucesivos cierres con $EE = 2$, $EE = 3$ y $EE = 4$ y quedarnos con los residuos correspondientes. En este caso, los residuos serán entre el cierre aplicado y el anterior, por ejemplo, para obtener los picos de anchura 2, se realiza la diferencia entre el cierre con $EE = 3$ y el cierre con $EE = 2$. De esta forma los valles se convierten en picos, como se ve en la figura 2.26; después de esto, será posible aplicarles un umbral que llamaremos u_ρ .

De forma práctica, hemos comprobado que un buen umbral es $u_\rho = 0,15$, y usándolo

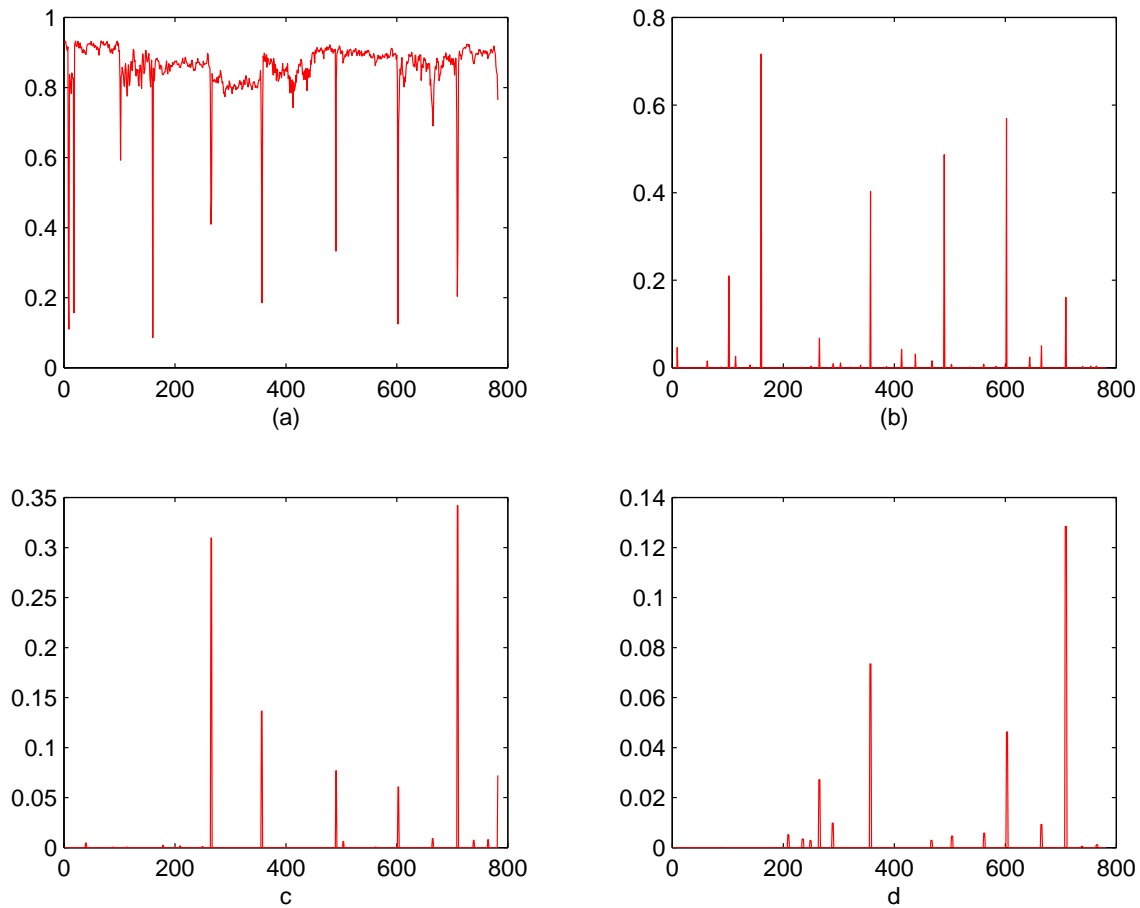


Figura 2.26: Aplicación del cierre para la localización de los cortes en *Alcoi* (los cierres se aplican sobre la apertura de $d_\rho(n)$ con $EE = 9$): (a) Señal $d_\rho(n)$. (b) Residuo entre el cierre con $EE = 2$ y la apertura con $EE = 9$. (c) Residuo entre el cierre con $EE = 3$ y el cierre con $EE = 2$. (d) Residuo entre el cierre con $EE = 4$ y el cierre con $EE = 3$.

para *Alcoi*, obtenemos seis cortes de los ocho que hay y captura los tres fundidos (con dos fotogramas intermedios). Los resultados con esta métrica también resultan satisfactorios.

Conclusiones de la aplicación de la Apertura y el Cierre

En este apartado se han desarrollado dos métodos para la detección de los cortes, utilizando como información básica $d_c(n)$ y $d_\rho(n)$, y aplicando operadores de M.M. para funciones. Se trata de un método automático, donde se deben configurar dos parámetros para cada caso:

- Cierre (ó Apertura) inicial: Tamaño mínimo que se exigirá al plano.
- Umbral, u_{dc} (ó u_ρ): Si un pico supera el umbral, se considerará que es una transición abrupta.

2.4.3. Elección de los umbrales: Método experimental

De forma práctica, utilizando las secuencias de prueba, hemos determinado que unos valores óptimos para los umbrales de detección son:

- Umbral para el módulo de la diferencia, $u_{dc} = 5$
- Umbral para la correlación, $u_\rho = 0,15$.

y un valor para el cierre o tamaño mínimo de plano que puede resultar adecuado es ⁶:

- Cierre inicial (o Apertura)= 9.

A continuación, vamos a desarrollar un estudio estadístico de todas las secuencias de prueba, para comprobar de forma cuantitativa si los valores sugeridos para los umbrales se corresponden con los óptimos. En primer lugar, ha sido necesario localizar de forma manual los cortes (y demás efectos de transición) de cada secuencia ⁷.

El trabajo se centrará en la elección del umbral para el módulo de la diferencia, u_{dc} ; se procedería de igual forma si se realizara el equivalente para u_ρ .

El objetivo buscado es obtener las funciones de densidad de probabilidad (fdp) estimadas, para las diferencias que constituyen un corte (las llamaremos transiciones) y para las que no (que llamaremos no transiciones), utilizando la información de las secuencias de prueba. De esta forma el problema de elección del umbral se parece bastante al esquema de elección del umbral óptimo en los sistemas radar [55]. Se dispone como información (al igual que en nuestro caso) de dos probabilidades a posteriori:

- $P_D \equiv$ Probabilidad de detección de blanco (entendiendo por blanco el corte buscado, en nuestro caso).
- $P_{FA} \equiv$ Probabilidad de falsa alarma.

⁶En las conclusiones comentaremos qué criterio se puede tomar para su elección.

⁷En el apéndice A se incluye la descripción de cada secuencia.

En la figura 2.27 aparecen las funciones de densidad de probabilidad condicional para estas dos decisiones y para la selección de un umbral de decisión. El criterio utilizado para la decisión es el de *Neyman-Pearson*: a diferencia de la regla de decisión de Bayes no se requiere información de la probabilidad de fuente a priori. Se trata de un criterio particularmente apropiado para la detección de pulsos en ruido gaussiano, como ocurre en aplicaciones radar, donde las estadísticas de fuente (por ejemplo probabilidad de presencia y ausencia, de un blanco) son desconocidas. También es apropiado cuando el coste de información de decidir erróneamente que hay un blanco cuando no lo hay es mucho menor que el coste de decidir erróneamente que no hay un blanco cuando sí está (aplicado al caso de los cortes, es preferible introducir alguna falsa alarma en la detección, a perder algún corte que haya).

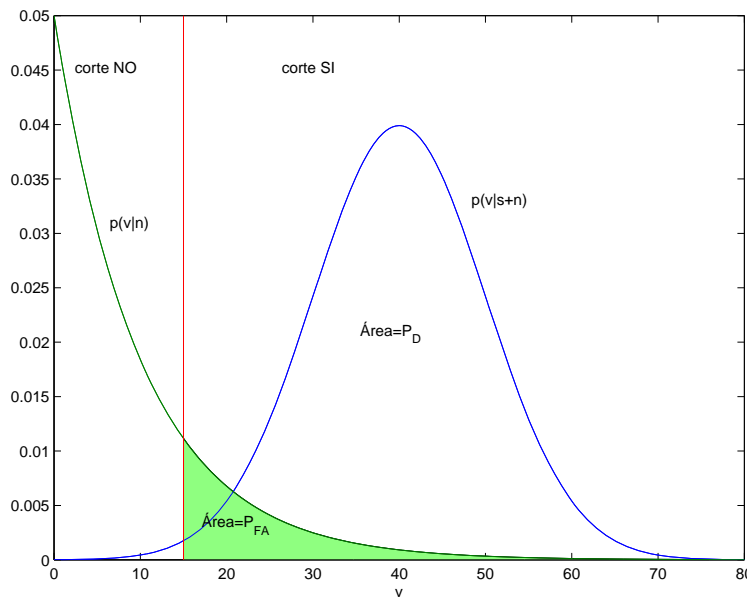


Figura 2.27: Modelo de elección del umbral según Neyman-Pearson.

Las probabilidades son:

$$P_D = \int_{v_{th}}^{\infty} p(v | s + n) dv$$

$$P_{FA} = \int_{v_{th}}^{\infty} p(v | n) dv$$

donde s denota señal y n denota ruido, y v_{th} es el umbral óptimo. En el detector de “Neyman-Pearson” el umbral de detección, v_{th} , es elegido para obtener una aceptable P_{FA} .

Siguiendo este modelo, deberemos empezar por fijar la fdp para las transiciones y para las no transiciones. Una forma cualitativa de elegir la función que mejor se aproxime es observar los histogramas para cada caso, es decir, dividir la señal $d_c(n)$ en dos: por un lado las diferencias que corresponden a transiciones $d_{c_{trans}}(n)$ y por otro las que corresponden a parejas de fotogramas que no son cortes $d_{c_{notrans}}$. Para estas dos nuevas señales se calculan los histogramas y se obtienen los de la figura 2.28. La normalización de estos histogramas,

el primero por el número de transiciones y el segundo por el número de no transiciones, da lugar a las funciones densidad de probabilidad de las transiciones y de las no transiciones, respectivamente. Se puede decir, que de forma aproximada, la fdp de las transiciones corre-

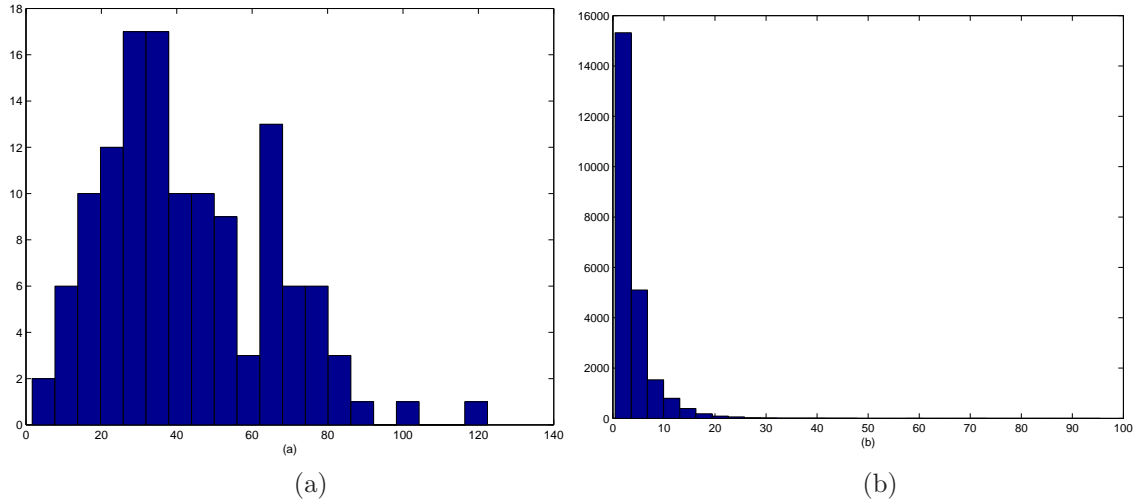


Figura 2.28: (a)Transiciones: fdp *normal*. (b)No transiciones: fdp *exponencial*.

sponde a una fdp *normal* y la fdp de las no transiciones a una *exponencial*. Sus expresiones matemáticas serían:

$$p_t(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

$$p_{nt}(x) = \lambda e(-\lambda x).$$

Una vez disponemos de $p_t(x)$ y $p_{nt}(x)$, para un determinado umbral, u_{dc} , se pueden calcular las siguientes probabilidades:

$$P_T = \int_{u_{dc}}^{\infty} p_t(x)dx = \int_{u_{dc}}^{\infty} \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx,$$

$$P_{NT} = \int_{u_{dc}}^{\infty} p_{nt}(x)dx = \int_{u_{dc}}^{\infty} \lambda e(-\lambda x)dx = e(-\lambda u_{dc}).$$

Con estas expresiones, haciendo un barrido para distintos valores de u_{dc} se obtienen los correspondientes valores de P_T y P_{NT} . En la figura 2.29 se muestran gráficamente los resultados obtenidos.

El estudio se ha realizado para distintos cierres: 3, 6, 9 y 12, de tal forma que podamos comprobar si la elección de 9 es buena. En (a), se muestran de forma conjunta la fdp para las transiciones y las no transiciones con un cierre de $EE = 3$. Se puede observar como $u_{dc} = 5$ puede ser considerado casi el umbral óptimo (según el criterio elegido, el óptimo se corresponde con la intersección de ambas fdp's). En las gráficas (c) y (d) se muestran las probabilidades de transición y de no transición frente a u_{dc} , y en (b) se muestran de forma conjunta P_T y P_{NT} . Es posible comprobar un resultado esperable a priori: cuando aumentamos el umbral, disminuye

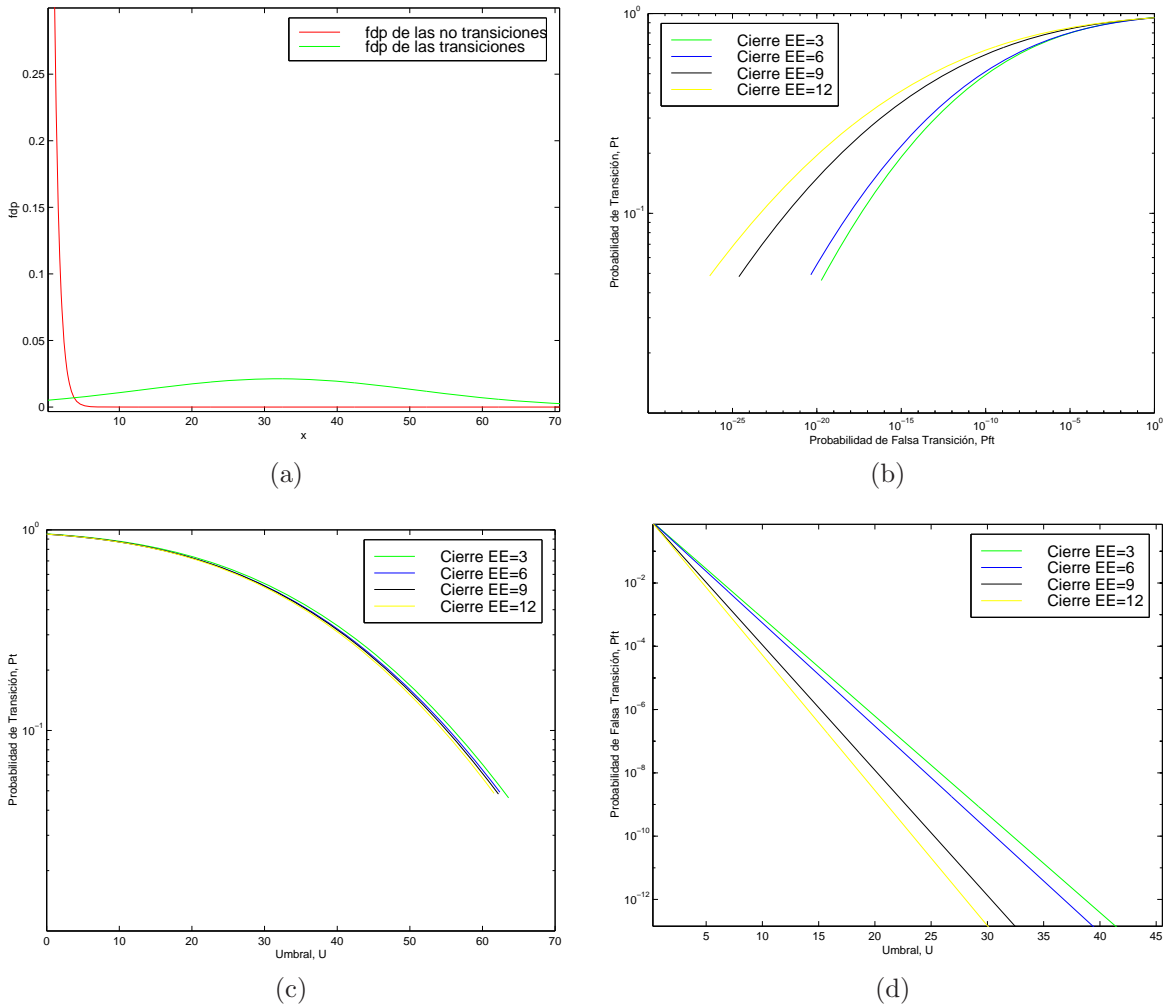


Figura 2.29: Estudio estadístico para la elección del u_{dc} óptimo: (a) fdp para las transiciones y las no transiciones. (b) P_T frente a P_{NT} . (c) P_T frente a u_{dc} . (d) P_{NT} frente a u_{dc} .

la probabilidad de falsa transición pero también la de detectar las transiciones correctas y de igual forma, al disminuirlo, aumenta la probabilidad de detección (como se desea), pero también la de no transición. Se debe tomar, por consiguiente, un valor de compromiso, y según las gráficas, el valor que proponemos: $u_{dc} = 5$ cuando $EE = 9$, es una buena elección.

2.5. Localización de fundidos

2.5.1. Definición

Se trata del efecto de transición gradual por excelencia. Consiste en la mezcla, siguiendo una ley lineal, de dos imágenes de partida, a través de varias imágenes intermedias. Si A y B son las imágenes inicial y final, y el fundido se realiza con N imágenes intermedias, C_1, \dots, C_N , cada una de éstas se obtendrá:

$$C_i = \frac{N + 1 - i}{N + 1} \times A + \frac{i}{N + 1} \times B$$

En la figura 2.31 aparece un fundido como el descrito donde $N = 4$.

El proceso de creación de un fundido, como se realiza de forma analógica en una mesa de mezclas, puede verse en la figura 2.30. Las imágenes de la fuente 1 se mezclan con las de la fuente 2 en un fundido que dura 5 imágenes intermedias. Las imágenes de ambas fuentes son filtradas con filtros que cada vez van aumentando la atenuación (cajas más oscuras en el gráfico) para la fuente 1 y disminuyéndola (cajas más claras) para la fuente 2. La opacidad de los filtros sucesivos sigue un proceso lineal, y al sumar el resultado de filtrar 1 y 2 se obtiene su combinación lineal. Las mesas de edición y montaje de vídeo permiten configurar la duración del fundido, generando de forma electrónica al valor de ponderación para cada imagen (mediante una rampa de tensión) y después haciendo la suma con amplificadores operacionales.

De igual forma, los programas de tratamiento digital de secuencias de vídeo también permiten mezclar secuencias diferentes, añadiendo todo tipo de efectos de transición entre ellas, como los fundidos. El método utilizado por estos programas se parece más al utilizado por nosotros para construir el ejemplo de la figura 2.31.

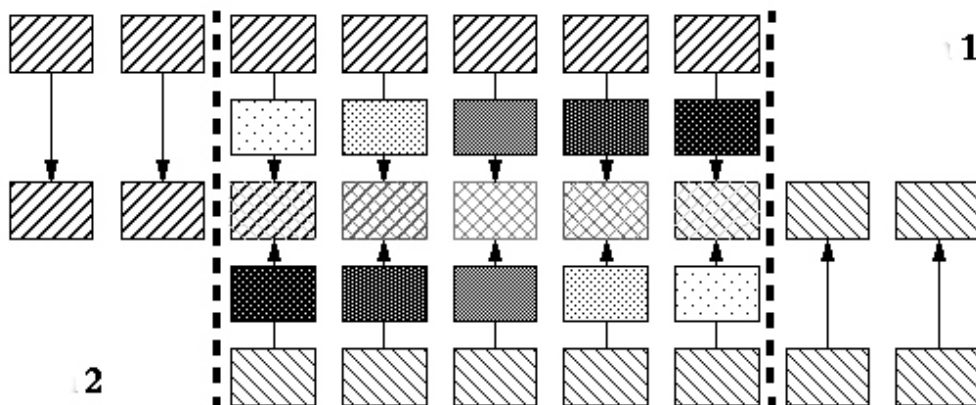


Figura 2.30: Proceso de obtención del fundido de dos fuentes vídeo 1 y 2.

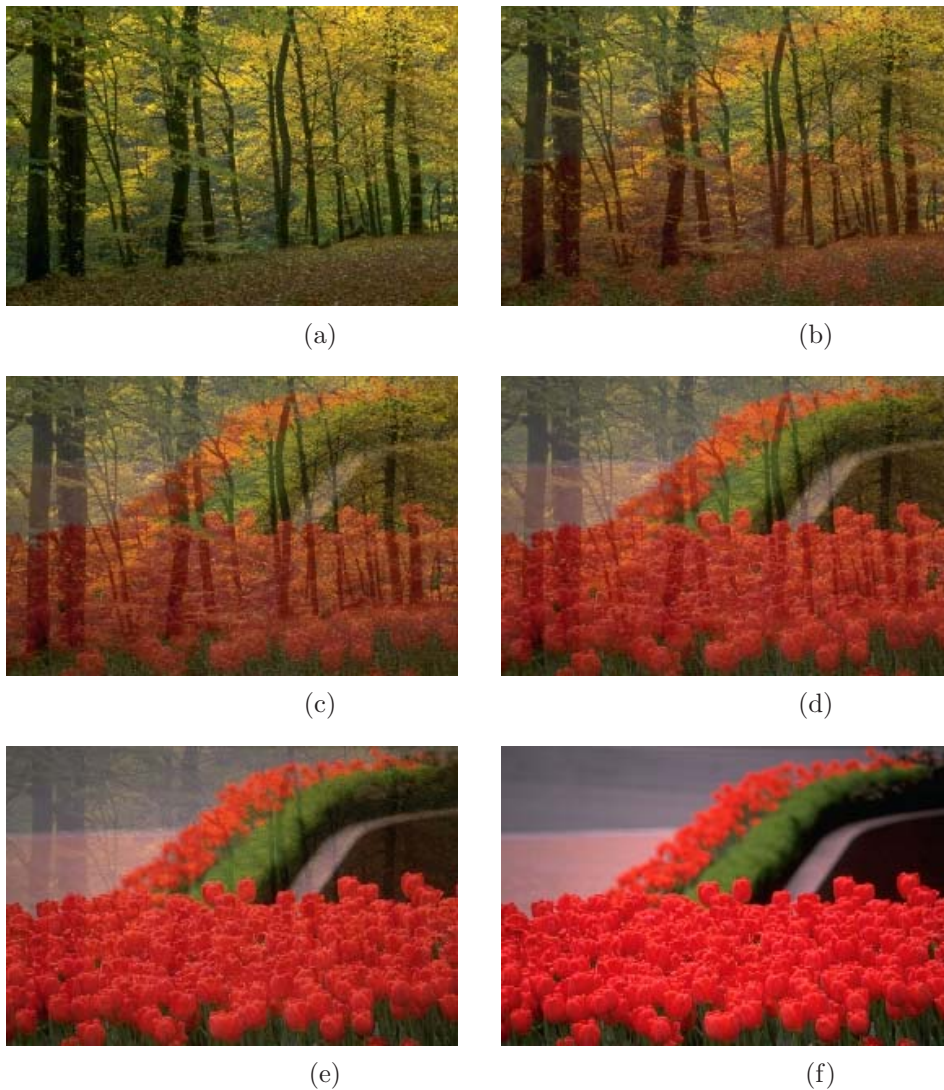


Figura 2.31: Ejemplo de fundido con cuatro fotogramas intermedios: (a)Fotograma inicial, A . (b)Fotograma intermedio, $C_1 = \frac{4}{5} \times A + \frac{1}{5} \times B$. (c)Fotograma intermedio, $C_2 = \frac{3}{5} \times A + \frac{2}{5} \times B$. (d)Fotograma intermedio, $C_3 = \frac{2}{5} \times A + \frac{3}{5} \times B$. (e)Fotograma intermedio, $C_4 = \frac{1}{5} \times A + \frac{4}{5} \times B$. (f)Fotograma final.

2.5.2. Método de detección de fundidos

A la hora de proponer un método para detectar fundidos el primer paso consistiría en determinar qué características tienen los píxeles de los fotogramas que componen un fundido, y que nos permita localizarlos. La característica en que nos basaremos en nuestra detección es en la varianza, método que fue propuesto por Alattar y que aparece descrito en [99].

Sean $f_1(t)$ y $f_2(t)$ dos secuencias ergódicas en las que la intensidad tiene una varianza σ_1^2 y σ_2^2 , respectivamente. En un fundido, la intensidad de $f_1(t)$ decrecerá linealmente a 0

(*fade-out*), y la de $f_2(t)$ se incrementará linealmente de 0 a un valor “normal” (*fade-in*). La región del fundido es la suma ponderada de $f_1(t)$ y $f_2(t)$ en el intervalo que dura la transición. Así pues, la intensidad de los fotogramas que componen el fundido viene dada por la siguiente expresión:

$$f(t) = f_1(t)[1 - \alpha(t)] + f_2(t)\alpha(t),$$

donde $t_1 \leq t \leq t_2$, siendo t_1 el fotograma de comienzo y t_2 el de fin, y donde α es el factor de ponderación $\alpha(t) = \frac{t-t_1}{t_2-t_1}$ y se incrementa linealmente de 0 a 100 %.

La varianza de $f(t)$ en la región del fundido, para el caso ideal en que $f_1(t)$ y $f_2(t)$ sean ergódicas, es:

$$\sigma^2(t) = (\sigma_1^2 + \sigma_2^2)\alpha^2(t) - 2\sigma_1^2\alpha(t) + \sigma_1^2.$$

En la figura 2.32 puede observarse que en un mismo plano la varianza se mantiene aproximadamente constante, mientras que en la región del fundido sigue una forma parabólica. En secuencias reales, con objetos en movimiento y movimientos de cámara, la varianza de $f_1(t)$ y $f_2(t)$ puede no permanecer constante en los periodos anteriores y posteriores a la transición (en el interior de un plano), pero se puede demostrar que en la región del fundido su forma se aproxima al caso ideal, es decir, se aproxima a una parábola. Así pues, buscando una forma parabólica en la evolución de la varianza a lo largo del tiempo podemos detectar un fundido.

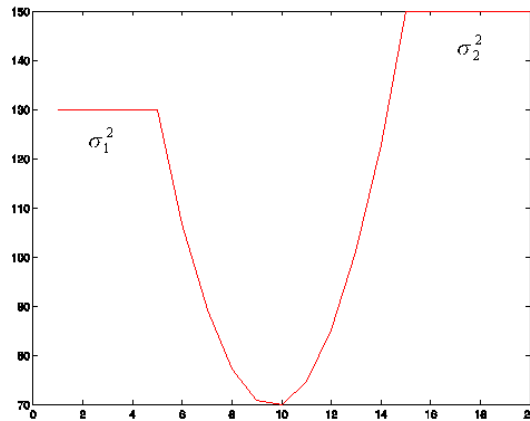


Figura 2.32: Curva de la varianza en un fundido.

Sin embargo, también pueden producirse variaciones de tipo parabólico en la varianza cuando no hay un fundido. En la figura 2.33, aparece la varianza para la secuencia *Newsb*. Esta secuencia incluye un fundido que se ha marcado y ampliado para poder observar su forma. También se puede observar que existen regiones, sin fundido, donde la variación se parece a una parábola.

Si añadimos a esto la dificultad de localizar zonas de tipo parábola en una señal unidimensional sin tener otra referencia, se hace necesario encontrar una métrica que al menos nos localice las regiones candidatas a fundido, para después comprobar en ellas si la varianza sigue la ley parabólica.

La métrica elegida es, $d_\rho^{fundidos}(n)$, y es similar a la utilizada en detección de cortes $d_\rho(n)$. Dicha métrica parte de la siguiente hipótesis:

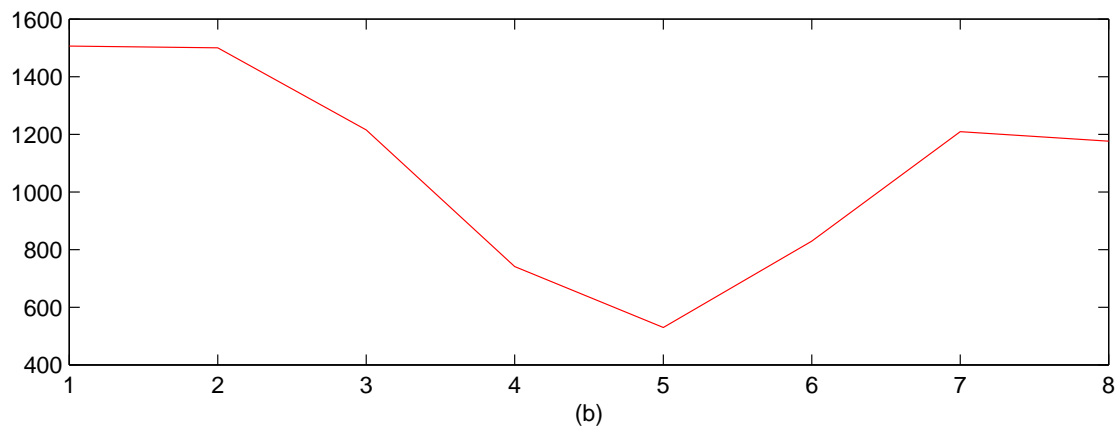
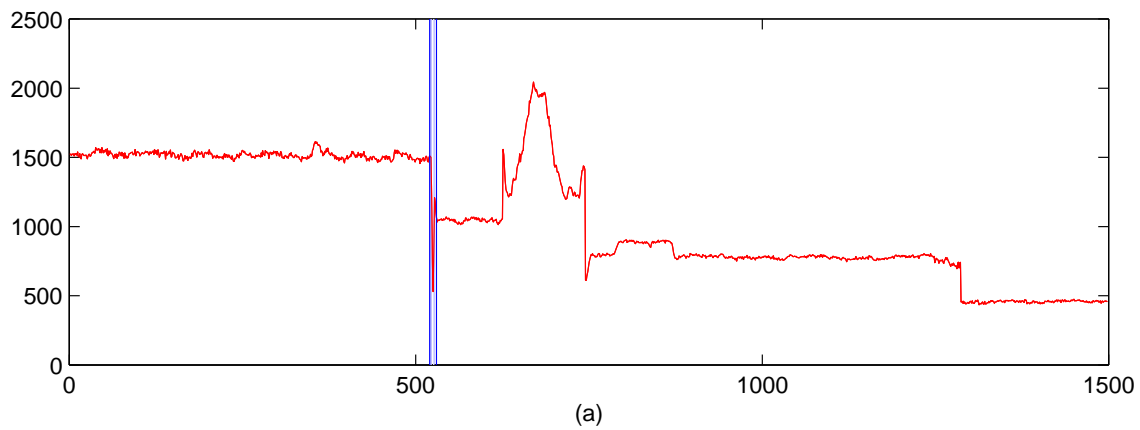


Figura 2.33: Varianza de la secuencia *Newsb*: (a) Varianza para todos los fotogramas. El fundido esta localizado entre las líneas verticales. (b) Varianza en la región del fundido.

”La luminosidad de los píxeles en los fotogramas de un fundido sigue una variación lineal y monótona”.

Para comprobar si se cumple esta hipótesis vamos a ver cómo varía la luminosidad a lo largo del tiempo para la secuencia *Newsb*, que tiene un fundido (de 5 fotogramas). Comenzaremos con tomar una subsecuencia de 50 fotogramas, que incluya el fundido en su zona central. En la figura 2.34 se muestran los cortes ortogonales de esta subsecuencia, tomados para un píxel del centro de la imagen. Estos cortes consisten en una imagen espacio-temporal en la que el eje y corresponde al tiempo, mientras que el x sigue siendo la dimensión horizontal de cada imagen, en el caso del corte horizontal. Para el vertical, el eje y es la dimensión vertical y el x el tiempo. Estas imágenes se construyen seleccionando la fila (para el corte horizontal) o la columna (para el vertical) del píxel elegido, para cada una de las imágenes de la secuencia; es como tener una ventana de anchura (altura) la de la imagen y de altura (anchura) un píxel, y rellenar una imagen copiando en ésta todo lo que nos deja ver la ventana a lo largo del tiempo. Una vez construidas las imágenes anteriores es fácil ver la evolución de

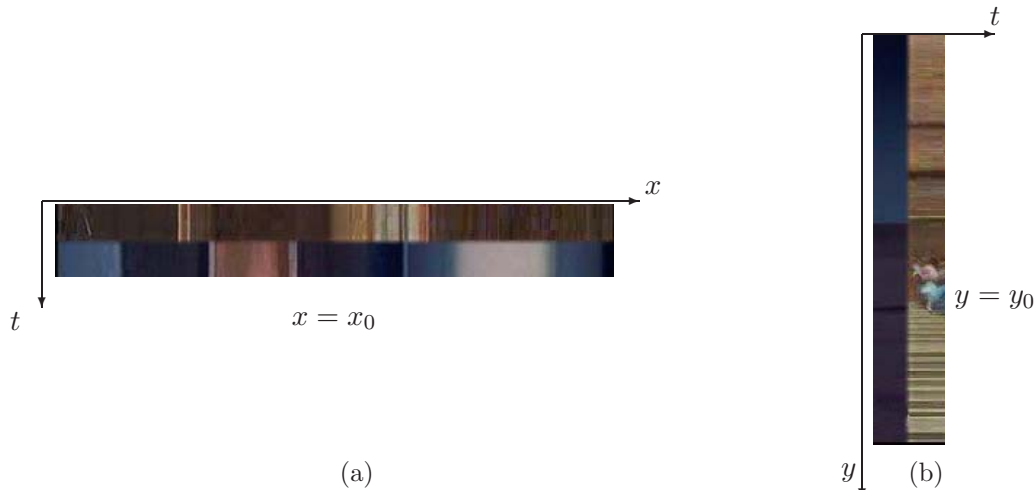


Figura 2.34: Cortes ortogonales de una secuencia, para el píxel (i, j) : (a) Corte Horizontal: fila x_0 - tiempo. (b) Corte Vertical: tiempo - columna y_0 .

la intensidad de un píxel a lo largo del tiempo, simplemente realizando un corte vertical de la primera imagen en la columna a la que pertenece el píxel del que queremos ver la evolución, es decir, representamos una gráfica donde el eje x es el tiempo y el y el nivel de luminosidad, los niveles de luminosidad a lo largo del tiempo que tiene un determinado píxel. El resultado obtenido sería el de la figura 2.35, en la que se muestra dicho corte para la componente R, ya que los de las otras dos son muy semejantes. En esta gráfica se observa la variación lineal de la luminosidad del píxel en los fotogramas del fundido, tal y como suponíamos en la hipótesis de partida. Se puede comprobar, realizando el mismo experimento para todos los píxeles (tanto del corte horizontal como del vertical), que este efecto se produce para todos los píxeles de la imagen.

Una vez verificada esta tendencia lineal, podemos definir una métrica para la localización de fundidos que se base en dicho conocimiento. Si x_{ij}^{n-1} , x_{ij}^n y x_{ij}^{n+1} son respectivamente el

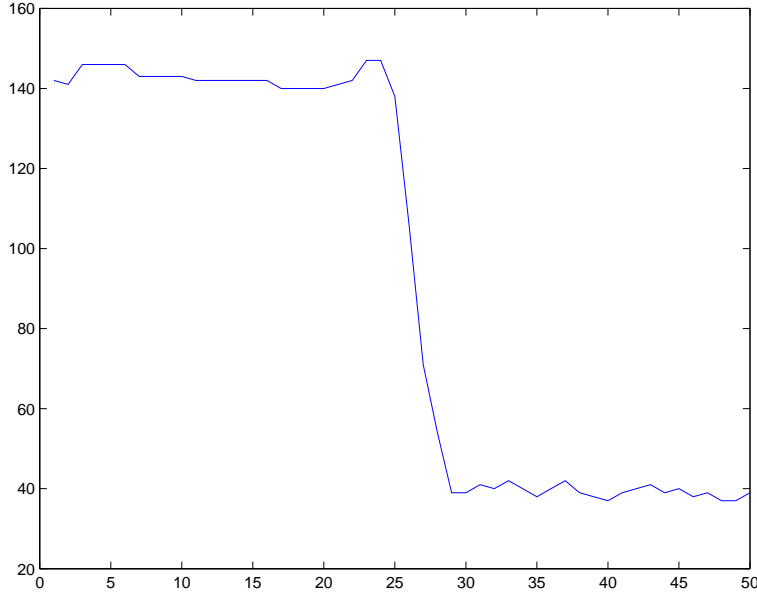


Figura 2.35: Corte vertical (para la componente R) de la imagen corte horizontal de una secuencia.

valor del píxel (i, j) de la imágenes $X_{n-1} = \{x_{ij}^{n-1}\}$, $X_n = \{x_{ij}^n\}$, y $X_{n+1} = \{x_{ij}^{n+1}\}$ (que a su vez se corresponden al fotograma $n - 1$, n y $n + 1$), y se definen la diferencias

$$d_{1ij}^n = (x_{ij}^n - x_{ij}^{n-1}),$$

$$d_{2ij}^n = (x_{ij}^{n+1} - x_{ij}^n),$$

se considerará que

$$\rho_{i,j}^{fundidos} = \begin{cases} 1 & \text{Si } |d_{1ij}^n| > t \text{ y } |d_{2ij}^n| > t, \text{ y además } \text{sign}(d_{1ij}^n) = \text{sign}(d_{2ij}^n) \\ -1 & \text{Si } |d_{1ij}^n| > t \text{ y } |d_{2ij}^n| > t, \text{ y además } \text{sign}(d_{1ij}^n) \neq \text{sign}(d_{2ij}^n) \\ 0 & \text{otro caso} \end{cases},$$

y a partir de esta medida por píxel, se define la siguiente métrica para cada fotograma:

$$d_{\rho}^{fundidos}(n) = \frac{\sum_{i,j} \rho_{i,j}^{fundidos}}{M N}. \quad (2.19)$$

Si tenemos una secuencia de L fotogramas ($n = 0, 1, \dots, L - 1$), el proceso se realizará a partir del segundo y hasta el penúltimo, definiéndose que $d_{\rho}^{fundidos}(0) = d_{\rho}^{fundidos}(N - 1) = 0$.

Lo que hace esta métrica para cada fotograma, n , es compararlo con el anterior, $n - 1$, y el siguiente, $n + 1$. Si la variación entre n y $n - 1$, y $n + 1$ y n tiene el mismo signo para un píxel, nos indicará una variación lineal en la luminancia en ese píxel. Cuando este fenómeno se produce para la mayoría de los píxeles de una imagen, obtendremos un valor elevado de

$d_{\rho}^{fundidos}(n)$. Durante todos los fotogramas de un fundido se deberán obtener valores altos de $d_{\rho}^{fundidos}(n)$.

Al igual que cuando definimos $d_{\rho}(n)$, aquí también se fija un umbral de $t = 2$ para no tener en cuenta las variaciones aleatorias (en forma de ruido).

En la figura 2.36, se muestra el resultado de calcular $d_{\rho}^{fundidos}(n)$ para la secuencia *Newsb*. Se observa claramente el pico producido por el fundido (fotogramas 522 – 525). Fijando

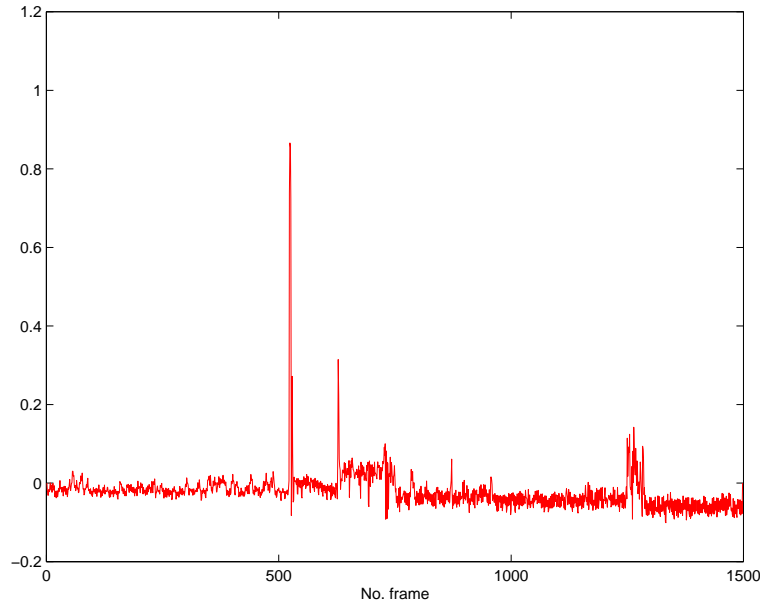


Figura 2.36: Resultado de calcular $d_{\rho}^{fundidos}(n)$ para la secuencia *Newsb*.

un umbral, $u_{\rho_{fundidos}} = 0,3$, capturaríamos este fundido, pero también una falsa alarma (fotograma 627).

Las falsas alarmas son debidas al movimiento de objetos grandes (ocupan muchos píxeles de la imagen) en imágenes sucesivas. Con el fin de minimizar estas falsas alarmas, nuestro método combinará la información obtenida con $d_{\rho}^{fundidos}$ y la información de la forma parabólica de la evolución de la varianza en un fundido. Los pasos seguidos por el algoritmo son:

1. Calcular $d_{\rho}^{fundidos}$ y σ^2 para cada fotograma de la secuencia.
2. Fijado un umbral, $u_{\rho_{fundidos}}$, todas las transiciones que lo superen serán candidatas a ser un fundido.
3. Se tomará la varianza de los fotogramas que forman parte de las transiciones candidatas a fundido y se compararán con las que debería obtenerse de forma ideal (según la expresión descrita anteriormente). Si el error cuadrático medio entre la varianza real y la estimada es menor que un cierto valor, que denominaremos umbral de varianza, $u_{\sigma^2_{fundidos}}$, se concluirá que la candidata a fundido lo es realmente.

2.6. Localización de cortinillas

2.6.1. Definición

Una cortinilla ⁸ es un efecto de transición entre dos planos que consiste en el desplazamiento espacial de una imagen sobre otra de forma gradual, a lo largo de varios fotogramas. De forma habitual, al igual que para los fundidos, el desplazamiento espacial se produce de una forma lineal. Las figuras 2.38 y 2.39 muestran una cortinilla vertical y otra horizontal.

Al igual que los fundidos, las cortinillas o efectos de incrustaciones, se utilizan para realizar transiciones graduales entre dos planos, utilizando fotogramas intermedios. Los programas de tratamiento digital de secuencias de vídeo también permiten mezclar secuencias diferentes, añadiendo todo tipo de efectos de transición entre ellas, como las cortinillas. Los métodos utilizados por éstos consisten en ir seleccionando la parte de la imagen de cada plano que formará cada imagen de la transición. Dicha selección se lleva a cabo utilizando máscaras binarias. Cada máscara seleccionará una porción cada vez mayor de la primera imagen, o la imagen en curso, del segundo plano a unir y una cada vez menor de la última imagen, o de la imagen en curso, del primero, como puede verse en la figura 2.37 que representa el proceso de formación de una cortinilla vertical para unir las fuentes de vídeo 1 y 2.

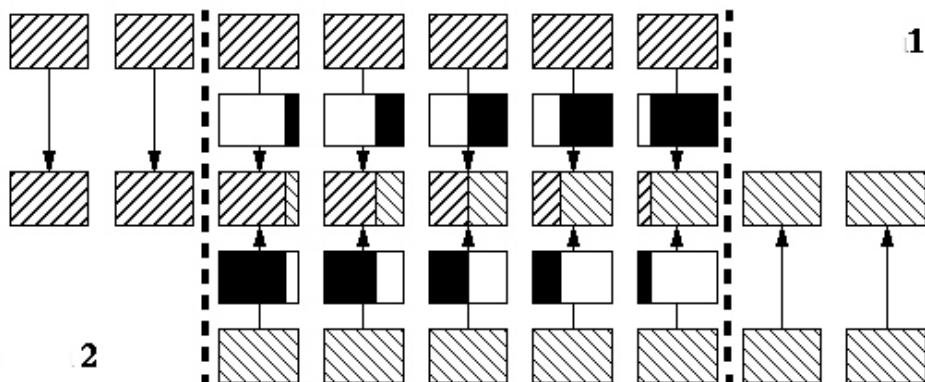


Figura 2.37: Proceso de realización de una cortinilla con dos fuentes vídeo 1 y 2.

⁸En francés, *volet*, y en inglés, *wipe*.



(a)



(b)



(c)



(d)



(e)



(f)

Figura 2.38: Ejemplo de cortinilla vertical (desplazándose hacia la derecha), con cuatro fotogramas intermedios.



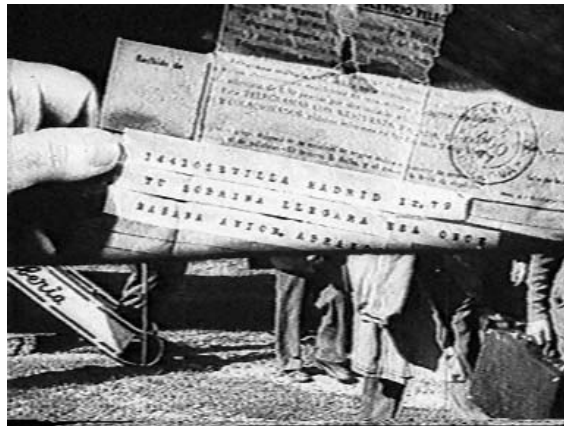
(a)



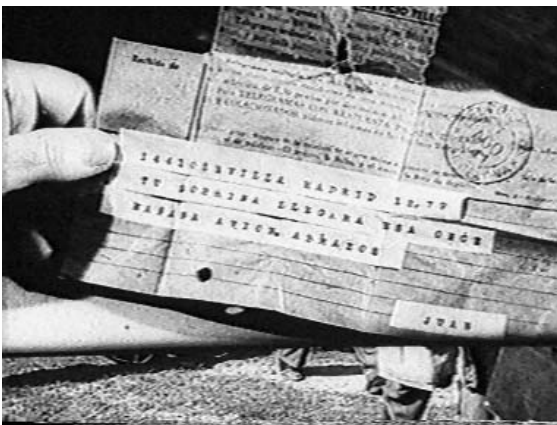
(b)



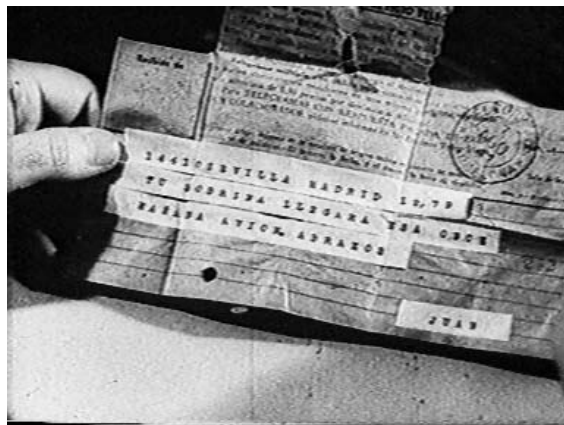
(c)



(d)



(e)



(f)

Figura 2.39: Ejemplo de cortinilla horizontal (desplazándose hacia abajo), con cuatro fotografías intermedias.

2.6.2. Método de detección de cortinillas

El método de detección de cortinillas se va a basar, al igual que los fundidos, en alguna característica que tengan los píxeles de los fotogramas que componen el efecto y que los diferencia del resto.

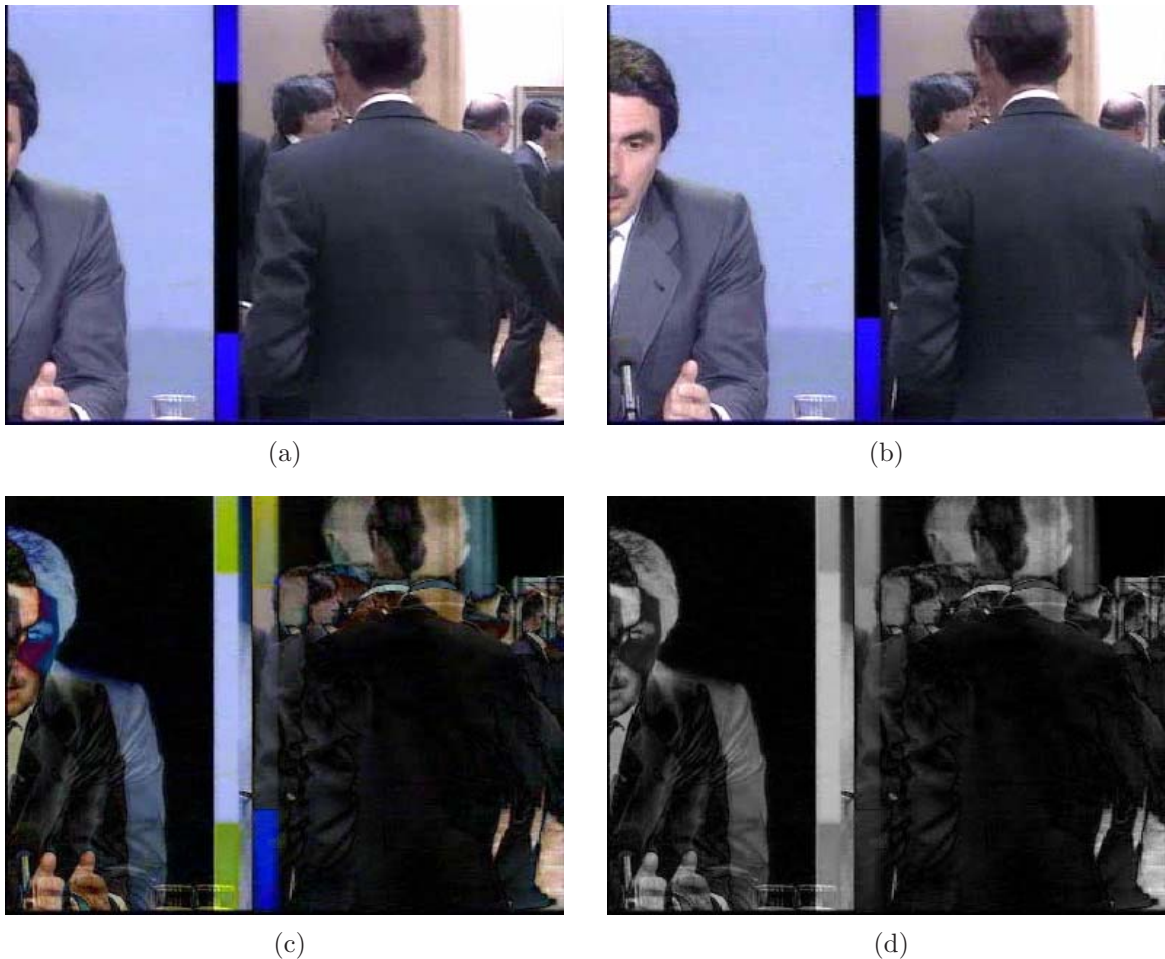


Figura 2.40: Dos fotogramas de la cortinilla vertical de *Newsa*: (a)Fotograma 1645. (b)Fotograma 1646. (c)Imagen diferencia de los fotogramas 1645 y 1646. (d)Imagen anterior convertida a grises.

Si nos fijamos en la figura 2.40 se pueden observar dos fotogramas consecutivos de una cortinilla vertical, y la imagen diferencia entre ellos. Al realizarse el desplazamiento espacial (anchura x píxeles), de una imagen sobre la otra, la imagen diferencia tendrá una región de anchura x con valores altos (zonas claras) a lo largo de toda la altura. Esta cortinilla también tiene una característica especial, se produce un desplazamiento del contenido de la imagen según avanzan los fotogramas. Por esto último, aunque el resto de la imagen diferencia pueda tener otras regiones claras, no ocuparán toda la imagen en la dimensión vertical, hecho que será el principio utilizado para la localización de las cortinillas.

Llamaremos $\phi[m, n]$ a la imagen diferencia, tal que $1 \leq m \leq M$ (filas) y $1 \leq n \leq N$ (columnas), siendo M y N el número de píxeles de alto y ancho, respectivamente. Tomemos ahora esta imagen y realicemos su proyección por columnas normalizada. Para la columna j tendremos que:

$$proyec_{col}[j] = \frac{1}{M} \sum_{i=0}^M \phi[j, i] \quad (2.20)$$

EL valor de esta proyección para cada una de las columnas de la imagen se muestra en la figura 2.41. Es fácil identificar el pico máximo como perteneciente a la región de anchura

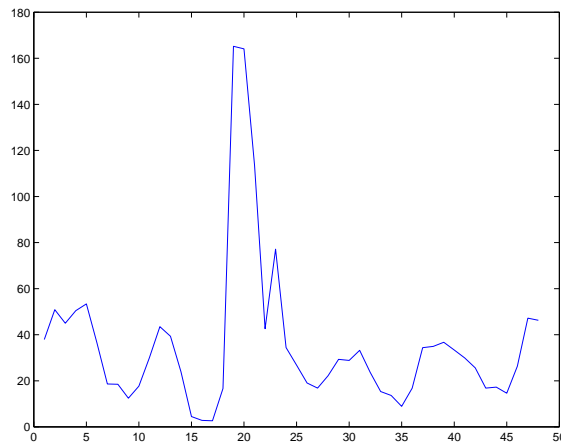


Figura 2.41: Proyección por columnas de la imagen diferencia entre dos fotogramas de una cortinilla vertical.

x que se ha desplazado la cortinilla al pasar de un fotograma a otro. Sin embargo, será más interesante quedarnos sólo con este máximo absoluto, eliminando los máximos regionales. Para ello utilizaremos la *reconstrucción geodésica* [80]⁹. Se trata de un operador morfológico que utiliza una señal *marcador* y otra de *referencia*, de tal forma que al aplicar el operador se obtendrán aquellos máximos de la señal *referencia* marcados con la señal *marcador*. En el caso que nos ocupa elegiremos como marcador una delta unidad en la posición del máximo de la señal de partida, *referencia*, nos permitirá eliminar todos los máximos relativos, figura 2.42.

Tenemos, por tanto, un máximo que nos indica la posición espacial de la cortinilla en la imagen diferencia. Al ir desplazándose la cortinilla del fotograma k al fotograma $k + 1$, el máximo de la imagen diferencia entre $k + 1$ y k también se irá desplazando espacialmente. En la figura 2.43 se muestra este desplazamiento para tres imágenes diferencia consecutivas.

Al realizar el proceso de proyección por columnas sobre la imagen diferencia de dos fotogramas que no pertenecen a una cortinilla vertical, y realizar después la reconstrucción, los máximos que se obtienen serán mucho más pequeños que cuando la cortinilla esté presente (la cortinilla ocupa toda la columna).

⁹La reconstrucción aplicada en nuestros métodos es un algoritmo eficiente que ha sido programado siguiendo las indicaciones de Vincent [151]

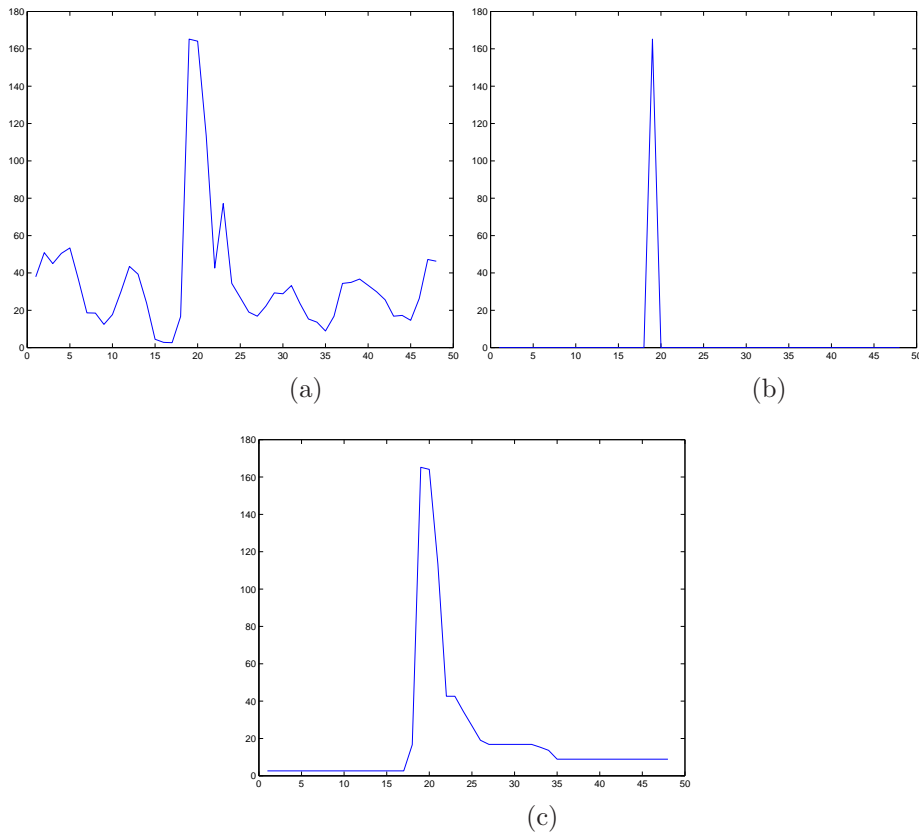


Figura 2.42: Reconstrucción geodésica: (a)Proyección por columnas de la imagen diferencia entre dos fotogramas de una cortinilla vertical. Es la señal de referencia. (b)Señal marcador. (c)Resultado de la reconstrucción.

Todo el proceso descrito hasta ahora para cortinillas verticales, es directamente trasladable a la cortinillas horizontales. En este caso, será necesario realizar una proyección por filas. Dicha proyección para la fila i vendrá dada por:

$$proyec_{fil}[i] = \frac{1}{N} \sum_{j=0}^N \phi[i, j] \quad (2.21)$$

Si calculamos el valor de esta proyección para cada una de las filas de la imagen diferencia entre dos fotogramas de una cortinilla horizontal y después realizamos la reconstrucción, el pico máximo se encontrará en la región de anchura y que se ha desplazado la cortinilla al pasar de un fotograma a otro. Al realizar el proceso de proyección por filas, sobre la imagen diferencia de dos fotogramas que no pertenecen a una cortinilla horizontal, y realizar después la reconstrucción, los máximos que se obtienen serán mucho más pequeños que cuando sí existe cortinilla (la cortinilla ocupa toda la fila).

A continuación mostraremos la forma de utilizar estos máximos deslizantes, para poder localizar cortinillas.

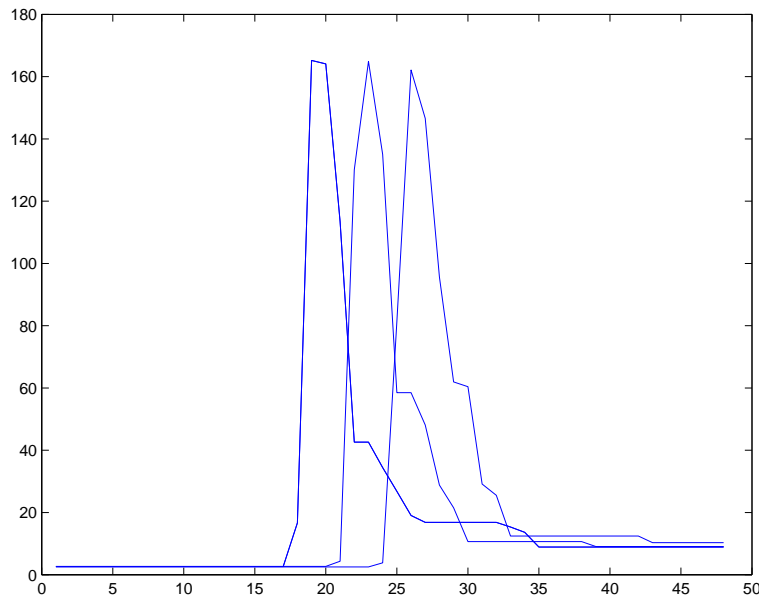


Figura 2.43: Proyecciones por columnas de tres imágenes diferencia entre fotogramas de una cortinilla vertical.

Imágenes “tira” para la localización de cortinillas

Así pues, una vez analizado el aspecto de una cortinilla, el método se basaría en:

- Para la localización de cortinillas horizontales: realizar la proyección por filas de la diferencia entre fotogramas consecutivos.
- Para la localización de cortinillas verticales: realizar la proyección por columnas de la diferencia entre fotogramas consecutivos.

Después de tener la proyección de cada imagen diferencia, se debe realizar la reconstrucción geodésica utilizando una delta en la posición del máximo como marcador.

Una forma de visualizar las sucesivas proyecciones a lo largo del tiempo es mediante una imagen de grises. Los valores de los píxeles de cada fila de la imagen se corresponderán con los valores de la reconstrucción de la proyección para cada imagen diferencia (normalizados a 256 niveles de gris). La fila superior de la imagen se corresponderá con la primera proyección.

Si tenemos una secuencia de L fotogramas, y dimensión en píxeles $M \times N$, la imagen de grises tendrá un tamaño de $L - 1$ píxeles de altura y N píxeles de anchura para la localización de cortinillas verticales, y $L - 1$ píxeles de altura y M píxeles de anchura para la localización de cortinillas horizontales.

Las secuencias con las que trabajamos estarán previamente diezmadadas (imagen dc) y además tendrán cientos, o incluso miles de fotogramas, por eso el aspecto de estas imágenes será parecido a unas tiras oscuras tan largas como sea la secuencia. De ahí el nombre de imagen “tira”.

Para poder observar un ejemplo, tomaremos una subsecuencia de *Newsa* (del fotograma 1350 al 1700) que incluya la cortinilla vertical. Se han calculado las proyecciones por columnas, después la reconstrucción y finalmente se ha construido la imagen descrita, figura 2.44.

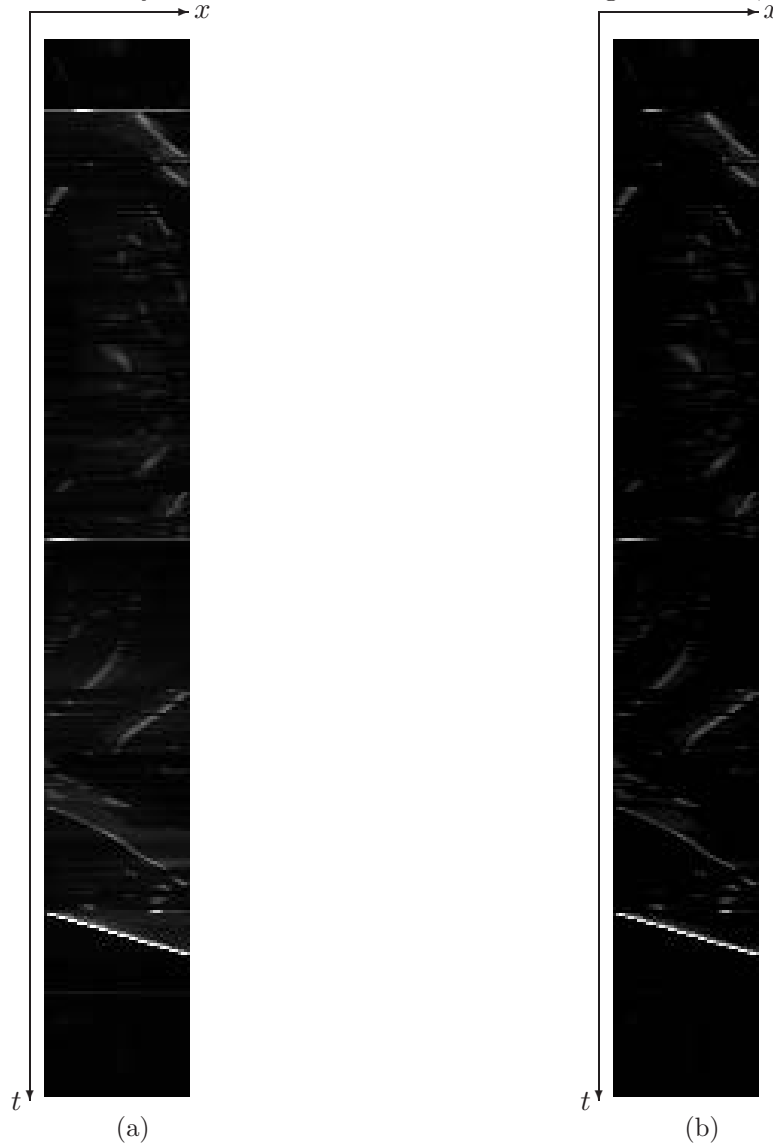


Figura 2.44: Resultado de búsqueda de cortinillas verticales para una subsecuencia de *Newsa*: (a)Imagen “tira” para proyecciones por columnas. (b)Residuo de la apertura horizontal de tamaño 20.

En (a) aparece la imagen de grises tal cual queda después de normalizar las proyecciones. En ella se pueden observar algunos fenómenos característicos:

- La cortinilla se corresponderá con una línea oblicua, cuya pendiente indicará la longitud temporal del efecto. Esta línea recta se origina al desplazarse espacialmente el máximo

de la proyección, y estará formada por pequeñas líneas horizontales de tamaño igual al número de píxeles que avanza la cortinilla en fotogramas sucesivos.

- Las transiciones abruptas o cortes, se corresponden con líneas horizontales. La explicación también es sencilla: la imagen diferencia de un corte tendrá valores muy altos en todos sus píxeles.
- Las zonas con mucho movimiento originarán zonas grises de forma irregular. Cuanto más importante sea el movimiento más clara será la zona correspondiente.

El objetivo del método será eliminar todo aquello que no sean las líneas correspondientes a las cortinillas. Para ello vamos a utilizar operadores morfológicos. Si el lector quiere profundizar en morfología puede consultar las obras de Serra [129, 130] y Matheron [97, 98]. Un resumen de operadores morfológicos puede encontrarlo en las tesis de Marqués [95] y Mossi [100].

El primer operador morfológico utilizado es la apertura. En la figura 2.44 (b) aparece el resultado del residuo de la apertura horizontal con un elemento estructurante de tamaño 20. Una apertura elimina las zonas claras de una imagen donde no cabe el elemento estructurante, en este caso todas aquellas líneas horizontales menores que 20. El residuo de la apertura (diferencia entre la imagen original y el resultado de la apertura) obtendrá dichas zonas claras y de esta forma conseguiremos eliminar las líneas rectas originadas por los cortes que no aparecerán en el residuo, ya que en la apertura sobreviven (son líneas horizontales mayores que 20). Para la elección del tamaño del elemento estructurante se deberá tener en cuenta el número de píxeles que la cortinilla descubre en cada fotograma (desplazamiento de la cortinilla) ya que si no podrán eliminarse los trozos de la línea oblicua que indica la cortinilla, por lo que el tamaño será mayor que el desplazamiento de la cortinilla y menor que la línea provocada por el corte.

Además de eliminar los cortes, será necesario eliminar el efecto producido por el movimiento. Una característica que tienen las líneas correspondientes a las cortinillas es que van de un extremo a otro. Explotando esta propiedad, obtendremos todas aquellas zonas claras y conexas de la imagen que empiecen a la izquierda (derecha), realizando una reconstrucción utilizando como marcador una imagen negra del mismo tamaño que la imagen “tira” con una zona blanca de varios píxeles de ancha, situada en el extremo izquierdo (derecho), y todo lo alto de la imagen. Para eliminar del resultado todo lo que empiece a la izquierda (derecha) pero no acabe en el extremo derecho(izquierdo) se realizará sobre el resultado anterior otra reconstrucción, pero ahora el marcador tendrá la zona blanca a la derecha (izquierda) de la imagen. De esta forma, finalmente sólo nos quedamos con las líneas que van de un lado a otro de la imagen. En la figura 2.45, se muestra el resultado obtenido para la subsecuencia de *Newsa*.

El siguiente paso será segmentar la imagen, quedándonos con cada una de las líneas oblicuas, que se corresponderán a cada una de las cortinillas que tiene la secuencia. Una posible alternativa de segmentación será aplicar un umbral y posteriormente realizar una segmentación binaria, es decir un proceso que clasifique todos los píxeles blancos conexas como pertenecientes al mismo objeto, y los marque con una etiqueta común (*binary labeling*). En la figura 2.45 (b) se muestra el resultado tras un umbral y tras la segmentación, resultando

el único objeto detectado marcado con un único color, la única cortinilla existente en esta secuencia.

Con todo lo expuesto anteriormente, se puede realizar el siguiente resumen del algoritmo:

1. Obtener la secuencia de imágenes diferencia, calculando la diferencia entre cada fotograma y el anterior.
2. Obtener las imágenes “tiras”, realizando para cada imagen diferencia, las siguientes proyecciones:
 - Proyección horizontal (por filas): para detectar cortinillas horizontales.
 - Proyección vertical (por columnas): para detectar cortinillas verticales.
 - Otras proyecciones (p.ejemplo $\pm 45^\circ$): para detectar cortinillas en la dirección de la proyección.
3. Para cada “tira” obtenida en el paso anterior, realizar el residuo de la apertura horizontal de tamaño 20. Con esto eliminamos el efecto de los cortes.
4. Al resultado anterior se le somete a la doble reconstrucción, primero con el marcador de selección de zonas que tocan el borde izquierdo, y posteriormente el derecho.
5. Conversion a binario del resultado de la reconstrucción y segmentación posterior. De todo este proceso podremos extraer la siguiente información del efecto:
 - Cuándo ocurre: de cada objeto resultado de la segmentación podremos obtener propiedades como la posición vertical inicial y final (y_{min} e y_{max}), puntos que determinan el inicio y fin de la cortinilla, y por tanto su duración.
 - Según la imagen “tira” (proyección) en la que aparezca algún objeto, la cortinilla tendrá una dirección u otra, determinando así el patrón del efecto. El problema es que la detección queda limitada a aquellos patrones de variación para los que se haya calculado proyecciones.

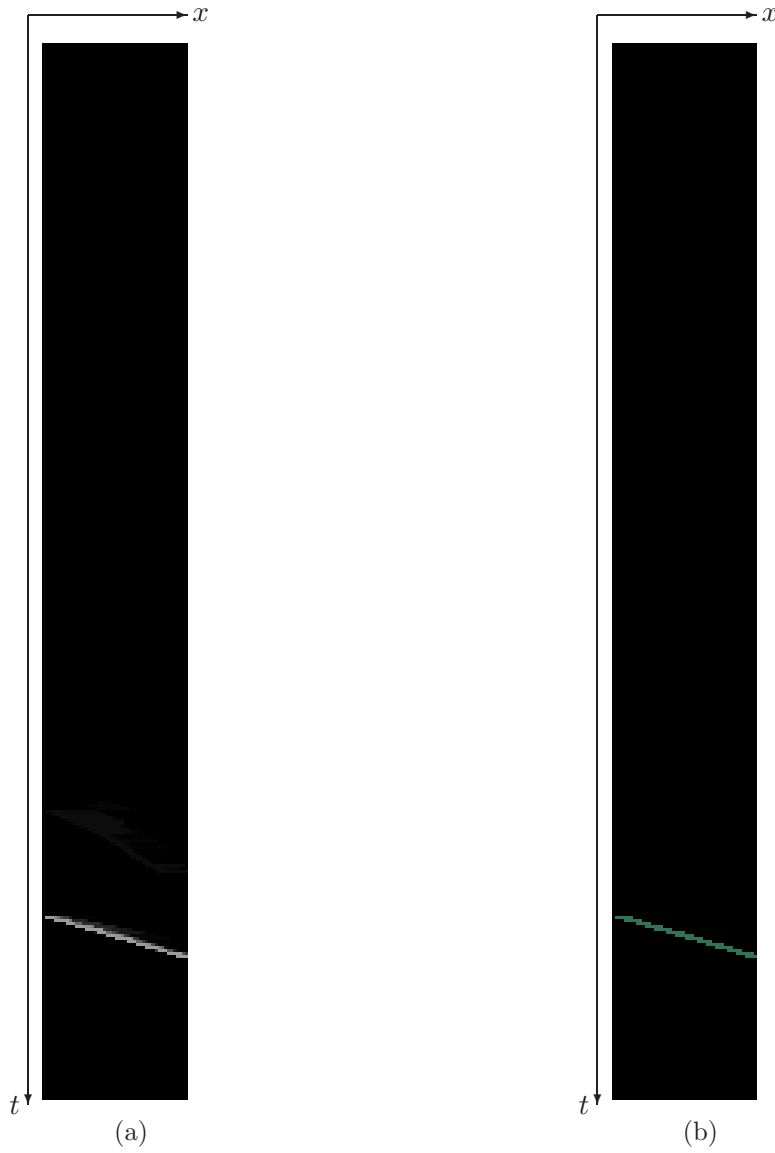


Figura 2.45: Localización de cortinillas verticales en la subsecuencia de *Newsa*: (a)Imagen “tira” después de procesado. (b)Segmentación mediante umbral de la imagen de (a).

Ejemplo de falso positivo en la localización de cortinillas

Con el ejemplo de la sección anterior, ha quedado demostrada la utilidad de proyecciones por filas o columnas para localizar las cortinillas horizontales o verticales. Sin embargo, como en la mayoría de las técnicas utilizadas para segmentar secuencias de vídeo, se pueden producir errores de detección si el contenido de las imágenes cumple algunas características muy específicas del efecto a detectar, que han sido utilizadas como determinantes para crear el método de detección.

Centrándonos en las cortinillas, si en las imágenes de los fotogramas sucesivos de una secuencia, aparece un objeto limitado por líneas horizontales (verticales) que ocupe gran parte de la imagen en vertical (horizontal) y la cámara realiza un movimiento panorámico en la dirección perpendicular a las líneas del objeto, se confundirá con una cortinilla. Vamos a verlo con un ejemplo.

En la figura 2.46 se muestran seis fotogramas consecutivos de una secuencia.

Si observamos la línea que produce en la imagen la unión de la puerta (muy oscura) con la pared (muy clara) se desplaza verticalmente a lo largo de la imagen. El efecto que produce es el mismo que si se tratase de una cortinilla vertical.

Por eso, al realizar todo el proceso de localización de cortinillas verticales, se obtiene un falso positivo en los fotogramas correspondientes a la región donde se mueve la puerta, figura 2.47.

Aunque no se producirán con mucha frecuencia, ya que es difícil que el patrón de variación coincida justo con el de la cortinilla buscada, si ocurre, no tiene solución ya que no hay forma de poder diferenciar estos Falsos Positivos de las auténticas cortinillas.



(a)



(b)



(c)



(d)



(e)



(f)

Figura 2.46: Ejemplo de secuencia que produce un falso positivo al detectar cortinillas verticales.

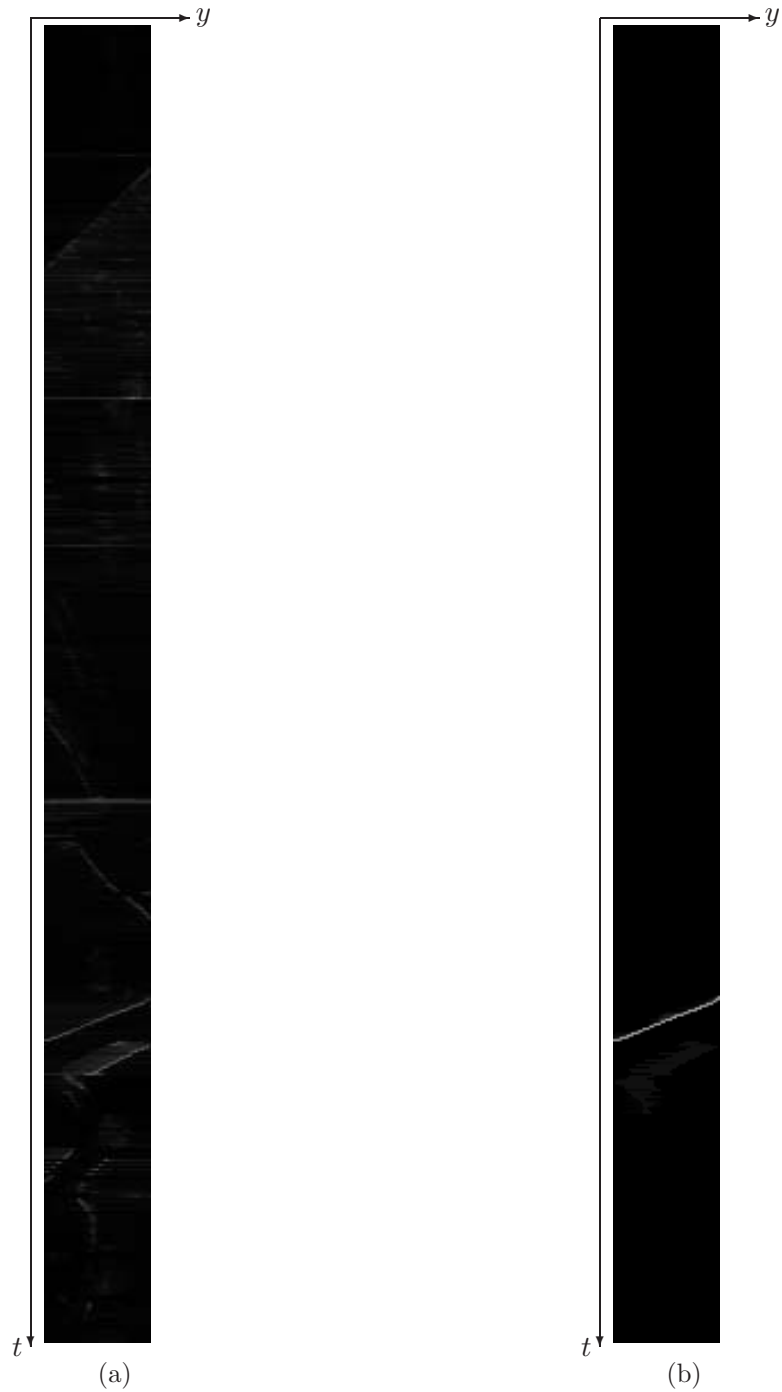


Figura 2.47: Localización de cortinillas verticales de la figura 2.46: (a)Imagen “tira” inicial. (b)Imagen “tira ” después de todo el procesado.

2.7. Resultados obtenidos

En los apartados anteriores de este capítulo, han sido presentados los diferentes métodos que se han utilizado en esta tesis a la hora de segmentar las secuencias en planos, para detectar cortes, fundidos y cortinillas. Aunque se ha expuesto un método para cada tipo de transición, el trabajo más exhaustivo y minucioso se ha llevado a cabo en la detección de cortes. Los motivos que han llevado a esto son diversos:

- Las transiciones más frecuentes entre planos son los cortes, sobre todo en el caso que nos ocupa, las películas antiguas.
- Para restaurar las películas (objetivo de esta tesis) la condición necesaria para que los algoritmos sean efectivos es continuidad de la información en el tiempo (que el nivel de gris de un píxel varíe lentamente), hecho que sólo no sucede cuando existe un corte (o en una zona dañada, la cual se restaurará). En el resto de transiciones, la continuidad temporal se cumple, ya que se trata de cambios graduales.

Es por todo ello que de los resultados obtenidos, los cuales se presentarán a continuación, es para el cumplimiento de nuestro objetivo mucho más exitosa la detección de cortes que la del resto de transiciones, y es por tanto muy importante que en esta detección se obtengan muy buenos resultados.

2.7.1. Parámetros de medida

Para presentar los resultados obtenidos, es interesante elegir algún parámetro que cuantifique la fiabilidad de la detección de los efectos de transición. En la literatura científica se proponen distintas herramientas. Antes de pasar a presentarlas, vamos a definir los siguientes parámetros:

- $N_{VP} \equiv$ Número de efectos correctos o verdaderos positivos. Es decir, una transición detectada por el método y que realmente es una transición.
- $N_{FP} \equiv$ Número de Falsos Positivos. Siendo un Falso Positivo (FP) un efecto detectado por el método, pero que realmente no existe.
- $N_{FN} \equiv$ Número de Falsos Negativos. Un Falso Negativo (FN) se corresponde con un efecto no detectado por el método, pero que realmente existe.
- $N_R \equiv$ Número de efectos reales $N_R = N_{VP} + N_{FN}$.

Algunas de las métricas que se utilizan para la medida de la bondad de un método de detección de transiciones son:

- Acierto y precisión.

La relación de acierto (*recall*) y precisión (*precision*) son dos parámetros comúnmente usados en la literatura [20] y vienen dados por las expresiones siguientes:

$$A = \frac{N_{VP}}{N_{VP} + N_{FN}}, \quad (2.22)$$

$$P = \frac{N_{VP}}{N_{VP} + N_{FP}}. \quad (2.23)$$

El acierto, también denominado en otros campos de la detección como especificidad, es la relación entre los aciertos (verdaderos positivos) y el número total de transiciones reales existentes. Por tanto, tiene relación con la cantidad de efectos que existen y que el algoritmo se ha dejado por detectar, o visto desde un punto de vista optimista, con la capacidad que tiene el método de detectar. La precisión, por su parte, es la relación entre los aciertos y el número total de transiciones detectadas por el método (correctas y falsas), y por tanto determina si el algoritmo se ha excedido a la hora de decidir transiciones. Cuanto más próximos a 1 son ambos parámetros, mejor segmenta el método. Si P disminuye la secuencia se sobresegmentará, mientras que si A disminuye, será infrasegmentada.

- Tasa de detección.

Otro parámetro usado en la literatura [66] es la tasa de detección, dada por:

$$\tau = \frac{N_R - (N_{FP} + N_{FN})}{N_R}. \quad (2.24)$$

Este parámetro trata de cuantificar el error global, tanto el producido por las falsas alarmas como por la no detección. Para el mejor de los casos (cuando no haya ningún error en la detección), τ deberá valer 1. A medida que aumentan los fallos, el valor de τ disminuye. Para casos muy críticos (los errores son mayores que los efectos reales), se pueden incluso obtener valores negativos.

- Porcentajes de detecciones correctas y falsas.

En [28] se presentan tablas en las que se comparan diferentes métodos utilizando los porcentajes de detecciones correctas y falsas calculados con respecto al número de transiciones correctas.

$$C = \frac{N_{VP}}{N_R} 100, \quad (2.25)$$

$$F = \frac{N_{FP}}{N_R} 100. \quad (2.26)$$

Lo ideal sería que el porcentaje de detecciones correctas fuera 100% y el de falsos positivos 0%. El porcentaje de detecciones correctas es equivalente al parámetro de acierto (A) definido anteriormente.

En las tablas de resultados que se presentan a continuación se muestran el número de todos los efectos: los reales (N_R), los verdaderos positivos (N_{VP}), los falsos positivos (N_{FP}) y los falsos negativos (N_{FN}). Posteriormente se calculan todos los parámetros A , P , τ , C (%) y F (%) de los resultados globales para cada método.

2.7.2. Resultados de la detección de cortes

Para la detección de los cortes se han utilizado los dos métodos presentados en la sección 2.4.

1. Cálculo de máximos en $d_c(n)$ y $d_\mu(n)$. El umbral para el módulo de la diferencia escogido ha sido $u_{d_c} = 5$ y el umbral para la diferencia de medias, $u_{d_\mu} = 15$.
2. Cálculo de mínimos en la señal correlación del signo modificado $d_\rho(n)$. El umbral utilizado ha sido $u_{d_\rho} = 0,15$.

Para ambos métodos se realiza el cierre (o apertura) inicial, que limita el tamaño mínimo del plano, con un elemento estructurante de tamaño 9.

En la tabla 2.3 se presentan los resultados obtenidos para todas las secuencias de prueba utilizando el primer método, el nombre de la secuencia, el número de fotogramas, el número de cortes reales, el de verdaderos positivos, el de falsos positivos y el de falsos negativos. La tabla 2.4 muestra los resultados para el segundo método, el de d_ρ .

Secuencia	Frames	N_R	N_{VP}	N_{FP}	N_{FN}
“DRAMA”	3012	11	11	0	0
“MOVIE”	3010	14	14	0	0
“NEWSB”	1497	4	4	0	0
“NEWSA”	1907	16	16	0	0
“BASKET”	1502	9	9	0	0
“CYCLING”	1998	0	0	1	0
“ALCOI”	783	7	5	0	2
“MALVA”	1782	5	5	0	0
“TARON”	5450	51	45	1	6
“CHAPLIN”	2713	10	7	3	3
“SANGA”	78	1	1	0	0
“SANGB”	1877	10	9	0	1
“SANGC”	1231	6	2	1	4
“TORBE”	2214	2	2	5	0
Total	29054	146	130	11	16

Cuadro 2.3: Resultados de la detección de cortes con el método 1.

En la tabla 2.5 aparecen reflejados todos los parámetros definidos en el apartado anterior; el acierto, la precisión, τ y los porcentajes de detecciones correctas y falsas, calculados con los resultados totales obtenidos para ambos métodos.

Como en los métodos presentados por otros autores las secuencias utilizadas son actuales, con imágenes de buena calidad, se han separado los resultados en dos tablas (2.6 y 2.7). En la primera se han utilizado para hallar los parámetros los resultados obtenidos utilizando películas de buena calidad (las 6 primeras de las tablas 2.3 y 2.4), para así poder comparar nuestros resultados con los de otros autores. En la tabla 2.7 están los resultados del resto de películas, las antiguas, que son el objeto de estudio de esta tesis.

Secuencia	Frames	N_R	N_{VP}	N_{FP}	N_{FN}
“DRAMA”	3012	11	11	0	0
“MOVIE”	3010	14	14	0	0
“NEWSB”	1497	4	4	0	0
“NEWSA”	1907	16	16	0	0
“BASKET”	1502	9	9	0	0
“CYCLING”	1998	0	0	0	0
“ALCOI”	783	7	7	0	0
“MALVA”	1782	5	5	0	0
“TARON”	5450	51	50	1	1
“CHAPLIN”	2713	10	9	3	1
“SANGA”	78	1	0	0	1
“SANGB”	1877	10	10	0	0
“SANGC”	1231	6	5	1	1
“TORBE”	2214	2	2	1	0
Total	29054	146	142	6	4

Cuadro 2.4: Resultados de la detección de cortes con el método 2.

Método	A	P	τ	$C(\%)$	$F(\%)$
d_c y d_μ	0,89	0,92	0,82	89	7,53
d_ρ	0,97	0,96	0,93	97,26	4,1

Cuadro 2.5: Valores de acierto, precisión, τ , % de correctos y % de falsos para los métodos 1 y 2 de detección de cortes.

Para la primeras, los valores obtenidos son excelentes para los dos métodos, de todas las secuencias analizadas (varios miles de fotogramas) y más de 50 transiciones abruptas reales, sólo se ha detectado un falso positivo para el método de $u_{dc}(n)$ debido a un cambio de iluminación gradual. En el caso de d_ρ los resultados son sorprendentes.

Para las películas antiguas los resultados son peores, pero bastante buenos, como comprobaremos cuando comparemos con otros autores. El hecho de que los resultados sean peores se debe a la peor calidad de la secuencia. La mayoría de los errores de detección de falsos positivos son producidos por desplazamientos temporales de la imagen (vibraciones) o por pérdidas de fotogramas intermedios. Los falsos negativos aparecen en zonas de la secuencia con iluminación muy pobre (imágenes muy oscuras) como ocurre, por ejemplo, en el caso de la secuencia “SANGC”, la cual es una toma nocturna en la calle y la porción más grande de la imagen, el cielo, no cambia al cambiar de plano. Estos falsos negativos podrían ser detectados bajando los umbrales, pero ello provocaría la aparición de muchos falsos positivos.

Nuestro objetivo en la segmentación de películas antiguas era determinar todas aquellas transiciones que rompiesen la continuidad temporal en la secuencia, cosa que nos facilitará posteriormente la restauración. Tanto en el caso de los errores obtenidos por falsos positivos como por falsos negativos, esta premisa se cumple. En el caso de los falsos positivos,

Método	A	P	τ	C(%)	F(%)
d_c y d_μ	1	0,98	0,98	100	1,85
d_ρ	1	1	1	100	0

Cuadro 2.6: Valores de acierto, precisión, τ , % de correctos y % de falsos para los métodos 1 y 2 de detección de cortes para el conjunto de películas actuales.

Método	A	P	τ	C(%)	F(%)
d_c y d_μ	0,83	0,88	0,72	83	11
d_ρ	0,96	0,94	0,89	96	6,5

Cuadro 2.7: Valores de acierto, precisión, τ , % de correctos y % de falsos para los métodos 1 y 2 de detección de cortes para el conjunto de películas antiguas.

al faltar fotogramas intermedios o existir una vibración brusca, esta continuidad temporal se rompe, y lo más conveniente para la restauración sería segmentar, que es lo que hacen nuestros métodos. Los falsos negativos obtenidos han sido precisamente debidos a la continuidad temporal en una gran porción de la imagen, cosa que no nos perjudica tanto el no haber segmentado.

Haciendo un balance de todos los resultados obtenidos, ambos métodos de detección de cortes funcionan relativamente bien, incluso con películas de mala calidad, aunque es un poco mejor el método de d_ρ .

2.7.3. Resultados de la detección de transiciones graduales

Tras un estudio experimental, como el realizado en la elección del umbral para la detección de cortes, se han utilizado los siguientes valores para los parámetros configurables para la detección de fundidos:

- Umbral para la correlación fotogramas $n - 1$, n y $n + 1$, $u_{\rho_{fundidos}} = 0,3$.
- Umbral para el error en la estimación de la varianza, $u_{\sigma_{fundidos}^2} = 350$.

Secuencia	Frames	N_R	N_{VP}	N_{FP}	N_{FN}	τ
“NEWSB”	1497	1	1	0	0	1
“NEWSA”	1907	1	1	0	0	1
“CYCLING”	1998	3	2	0	1	0,66
“ALCOI”	783	4	4	11	0	-1,75
“MALVA”	1782	1	0	1	1	-1
“CHAPL”	2713	2	0	100	2	↓↓
“CHAPLC”	2713	2	0	20	2	↓↓

Cuadro 2.8: Resultados de la detección de fundidos.

Secuencia	Frames	N_R	N_{VP}	N_{FP}	N_{FN}	τ
“NEWSA”	1907	2	1	0	1	0,5
“MALVA”	1782	1	1	0	0	1

Cuadro 2.9: Resultados de la detección de cortinillas.

Al igual que para los cortes, la calidad de la imagen influye en los resultados obtenidos en la detección de los fundidos y las cortinillas. Centrándonos en los fundidos, para secuencias con muy mala imagen, los resultados no son lo que en principio podíamos esperar.

Otra posible causa de Falsos Negativos en los fundidos, es la duración de éstos. Para fundidos muy largos (muchos fotogramas intermedios), el método falla. El problema se encuentra en que cuando el fundido se prolonga temporalmente a través de muchos fotogramas, las variaciones entre un fotograma y el siguiente serán muy pequeñas (valores bajos de $d_{\rho}^{\text{fundidos}}$). Podría pensarse en reducir el umbral de detección de posibles fundidos, pero el problema que surge inmediatamente es la aparición de gran número de Falsos Positivos (debidos a movimientos de la imagen y de objetos grandes en ella). La elección adecuada del valor de $u_{\sigma_{\text{fundidos}}^2}$ también influye en gran medida. Por todo esto, la detección de fundidos largos tiene una difícil solución.

Las cortinillas son un efecto de transición poco utilizado; en todas las secuencias con las que hemos trabajado sólo hemos encontrado tres o cuatro casos. El método programado sólo funciona con cortinillas verticales y horizontales que empiecen en uno de los extremos, aunque esto no es un gran problema: la inmensa mayoría de las cortinillas son así (la cortinilla no detectada de “NEWSA” es una excepción, se trata de una cortinilla horizontal que comienza en el centro de la imagen). El método de detección de cortinillas, tal y como esta programado, aunque efectivo, es un método lento y laborioso; y su utilización debería limitarse a las secuencias donde a priori se sepa que contiene alguna. Debido al procesado que se realiza sobre las imágenes “tira” utilizadas en la detección de las cortinillas, las variaciones de brillo no influyen de manera negativa en los resultados, como ocurre con los cortes y los fundidos.

Dado el bajo número de efectos de transición gradual analizados, aunque han sido todos los encontrados en nuestra base de datos, no se puede considerar el resultado significativo estadísticamente.

2.7.4. Comparación con los resultados publicados de otros métodos

La información sobre los valores de τ obtenidos por otros métodos aparece recopilada en Joly [66]. En este artículo, se presentan cada uno de los algoritmos publicados y los valores medios de τ que se pueden obtener.

En base a estos datos, se puede decir que:

- Para los cortes, se obtienen resultados que van desde 0,75 para los más sencillos (algunos incluso no pasan del 0,6) hasta el 0,85 o el 0,99 (esto último, para un método muy sofisticado, basado en el error residual de energía después de una transformación afín). Los métodos más usados (se corresponden con los más eficientes computacionalmente, es

decir, los más rápidos) se mueven en torno al 0,8, lo que demuestra que nuestro método se encuentra a un buen nivel. Además, en ningún caso se especifica si se han probado con imágenes de mala calidad, donde probablemente todos los métodos fallarían, como también lo hace el nuestro.

Otra cuestión a favor de la utilización de $d_c(n)$ y $d_\rho(n)$, es su sencillez. Como ya describimos, se trata de métodos computacionalmente sencillos y eficientes (se realizan restas y sumas de píxeles), frente a otros métodos que obtienen resultados similares, pero que implican procesamiento complejo (obtención de histogramas, procesado estocástico, transformaciones afines, etc.).

- En el caso de las transiciones graduales, fundidos y cortinillas, la mayoría de las propuestas que hay en la bibliografía parten de una idea básica: son efectos de escasa aparición y además difíciles de localizar. Por tanto, no se debe desperdiciar mucha computación en su localización. Incluso se da el caso en que efectos de incrustación, como las cortinillas, raramente son tenidos en cuenta. Al igual que para los cortes, los valores de τ obtenidos, dependen de la complejidad del método. Los mejores valores se mueven en torno al 0,8 (para un método basado en el cálculo de la distancia de Hausdorff entre histogramas calculada para contornos de imagen dividida en bloques, se llega al 0,92). Podemos decir, por consiguiente, que nuestros métodos para fundidos y cortinillas también se encuentran dentro de lo aceptable. En el caso de las películas antiguas, aunque los resultados son peores, para nuestro objetivo no es tan problemática la detección de transiciones graduales porque no supone un cambio brusco en la continuidad temporal.

Para comparar otros parámetros, como C y F, podemos comparar nuestros resultados con los presentados en [28], los cuales aparecen reflejados en las tablas 2.1 y 2.2 de la sección 2.2. Al igual que al comparar el parámetro τ , nuestros métodos, sobre todo el de detección de cortes, que es el que más nos interesa, mejora los resultados con respecto a los que obtienen otros métodos, incluso teniendo en cuenta que nuestro conjunto de secuencias de test es mucho peor, ya que se trata de secuencias procedentes de películas antiguas con imágenes de mala calidad.

Capítulo 3

Corrección de las variaciones de intensidad

3.1. Introducción

3.1.1. Planteamiento del problema

Las variaciones de intensidad indeseadas, efecto llamado parpadeo o en inglés *flicker* es un fenómeno muy común que aparece en películas antiguas en blanco y negro. El observador lo percibe como fluctuaciones artificiales en la intensidad de la imagen, las cuales no pertenecen a la escena original. Las causas que pueden provocar el parpadeo son variadas: la antigüedad de la película, el polvo, el proceso químico, el copiado, errores al muestrear (*aliasing*), y en el caso de antiguas cámaras de cine, variaciones en el tiempo de obturación.

Así pues, el problema que se nos plantea es eliminar las variaciones globales de intensidad que aparecen de manera artificial en la secuencia y que producen un efecto muy molesto a la hora de visualizar la película, que se traduce en un parpadeo de luminosidad. Dichas variaciones de intensidad temporales no siempre duran lo mismo, siendo mucho más perjudiciales cuanto más rápidas son. Además, a la hora de eliminar (o por lo menos reducir) este efecto debemos mantener intactas las variaciones de intensidad debidas a la propia escena, las cuales suelen ser más lentas.

La figura 3.1 muestra dos fotogramas consecutivos de un plano de la película de Charles Chaplin *His new job* (1915), donde aparece una variación brusca en la intensidad global entre ambas imágenes.

3.1.2. Estado del arte

De todos los defectos que presentan las películas antiguas y que se abordan a la hora de la restauración, el parpadeo ha sido el menos tratado. Quizás, el dejar de lado este problema, ha sido debido a que a simple vista parece que tiene fácil solución. Sin embargo, no es tan simple, ya que un cambio en el brillo del fotograma, como podría parecer la solución, no es suficiente para resolver el problema.

Últimamente se han dedicado esfuerzos a resolver el defecto que nos ocupa, contribuciones



Figura 3.1: Dos fotogramas consecutivos un plano de la secuencia *Chaplin*.

que analizamos en esta sección.

En [140, 141] se apunta la dificultad de la reducción del parpadeo, pero no se presenta ninguna solución específica sino que se indica que el uso de técnicas de supresión de ruido puede reducir también el parpadeo. Como esta solución no es acertada, los restauradores buscan otras soluciones que se engloban en dos tipos de métodos: los que modelan la variación del parpadeo y los que utilizan métodos de ecualización de histogramas.

- **Modelo de variación temporal.**

Estos métodos se basan en transformar la intensidad de la secuencia de manera lineal. Para ello modelan el efecto del parpadeo de la siguiente forma:

$$h(x, y, t) = \alpha(x, y, t) f(x, y, t) + \beta(x, y, t)$$

donde $h(x, y, t)$ representa la secuencia observada y $f(x, y, t)$ la secuencia original libre de parpadeo. $\alpha(x, y, t)$ se le llama ganancia de parpadeo y $\beta(x, y, t)$ es el *offset*. La expresión del parpadeo expuesta arriba obedece a una ley de variación afín como la que propone Ferrandiere en [52, 51]. El objetivo es obtener $f(x, y, t)$, hecho que se consigue al estimar α y β utilizando estadísticos (media y varianza) de la secuencia observada.

Roosmalen [122, 121, 120] realiza la estimación de dichos parámetros en dos situaciones, primero para el caso de secuencias estacionarias, y posteriormente introduce el movimiento, obteniendo para ambos casos los parámetros óptimos minimizando la función lineal del error cuadrático medio (LMMSE). Debido al ruido y al movimiento, hay parámetros que no pueden ser calculados con los métodos propuestos, por lo que el autor presenta tres formas de interpolarlos a partir de los hallados (interpolación por dilatación, por filtrados paso bajo sucesivos y con polinomios de relleno 2D).

Ohuchi *et al.* [108, 124] proponen un método jerárquico para la corrección de estas variaciones de intensidad. Exponen cuatro modelos en los que la ganancia y el *offset* se consideran sólo dependientes del tiempo (para el modelo más simple) o además

dependientes ambas de la posición (para el modelo más sofisticado). El método consiste en estimar los parámetros considerando la hipótesis del modelo más simple e ir refinando dichos parámetros con los modelos más complejos.

- **Ecualización de histograma.**

Otro tipo de métodos son los que utilizan una transformación no lineal sobre la intensidad de la secuencia, normalmente se basan en la ecualización del histograma. La idea, como proponen Naranjo *et al.* [104], Schaulauer *et al.* [127, 101], y Richardson y Suter [115], es transformar de manera no lineal cada imagen para conseguir cambiar el histograma de cada fotograma por el de otro que no posea parpadeo, llamado de referencia. Este tipo de métodos será explicado con más profundidad cuando pasemos a presentar nuestro método de ecualización de histograma, en la sección 3.3.

3.1.3. Objetivos

En este capítulo nuestro objetivo es minimizar el efecto de las variaciones globales de intensidad indeseadas, respetando las variaciones debidas a la iluminación de la escena. Se proponen dos métodos. El primero utiliza una transformación afín cuyos parámetros se obtienen en base al modelado temporal de la media y la varianza. El segundo propone una transformación no lineal, basada en la ecualización de histograma, para conseguir que la intensidad de la imagen transformada sea similar estadísticamente a la de las imágenes vecinas.

En la sección 3.2 presentamos el método que corrige la media y la varianza de la secuencia basándose en modelos temporales. La sección 3.3 explica la técnica usada para obtener la transformación no lineal. Finalmente, en la sección 3.4 se analizan los resultados obtenidos y se presentan una serie de conclusiones.

3.2. Método de modelado de la variación temporal

3.2.1. Idea básica

El método que se propone en este apartado se basa en el análisis de las variaciones temporales de la media y la varianza. Si obtenemos la media y la varianza de cada fotograma y representamos su evolución a lo largo del tiempo, obtenemos gráficas como las de las figuras 3.4 y 3.5 que nos muestran la media y la varianza para los planos diferentes de la secuencia *Chaplin* de las figuras 3.2 y 3.3. En ambas gráficas, tanto para el caso de la media como el de la varianza, podemos observar que aparecen varias componentes:

- **Tendencia lineal:** debida a la evolución real de la intensidad en la escena. Esta componente nos interesará conservarla.
- **Variaciones periódicas:** es la componente sinusoidal que presentan la media y la varianza la cual puede ser debida al proceso de obturación poco preciso que se llevaba a cabo en la toma de imágenes con una cámara antigua. Este tipo de componentes no lo esperábamos en un principio, pero al analizar las señales vemos que está presente.

- Variaciones aleatorias: son las variaciones rápidas que presentan ambas señales y que pueden tener su origen, como dijimos al principio, en la antigüedad de la película, la suciedad debida al polvo, etc.



Figura 3.2: Doce primeros fotogramas del plano de la secuencia Chaplin del fot. 1314 al 1429.

El objetivo de este método es estimar las componentes de la media y la varianza y, una vez estimadas, obtener la corrección a realizar, de tal forma que la media (y la varianza) de la secuencia corregida sea una combinación lineal de todas las componentes estimadas de la secuencia original. Dicha combinación lineal dará un peso de 1 a la componente lineal, que es la deseada, y un factor pequeño¹ a las variaciones no deseadas, las aleatorias y las periódicas.

3.2.2. Estimación de parámetros

Sea $x[n]$ la secuencia correspondiente a los valores de la media (o la varianza) para cada fotograma del plano que vamos a restaurar, de la cual queremos estimar sus componentes. Definimos $\hat{x}[n]$ como la secuencia que queremos estimar, dada por la expresión:

$$\hat{x}[n] = An + B + C \cos(\omega_0 n) + D \sin(\omega_0 n)$$

donde $An + B$ representa la tendencia lineal de los datos y $C \cos(\omega_0 n) + D \sin(\omega_0 n)$ la componente sinusoidal, mientras que las variaciones aleatorias vendrán dadas por $x[n] - \hat{x}[n]$.

¹Será muy cercano a cero. La solución teórica, según lo planteado, es que los factores de las variaciones no deseadas sean cero, sin embargo permitir que contribuyan mínimamente a la media de la secuencia corregida hace que la película restaurada no resulte artificial, sino que conserve el *sabor* antiguo.



Figura 3.3: Doce primeros fotogramas del plano de la secuencia *Chaplin* del fot. 2073 al 2292.

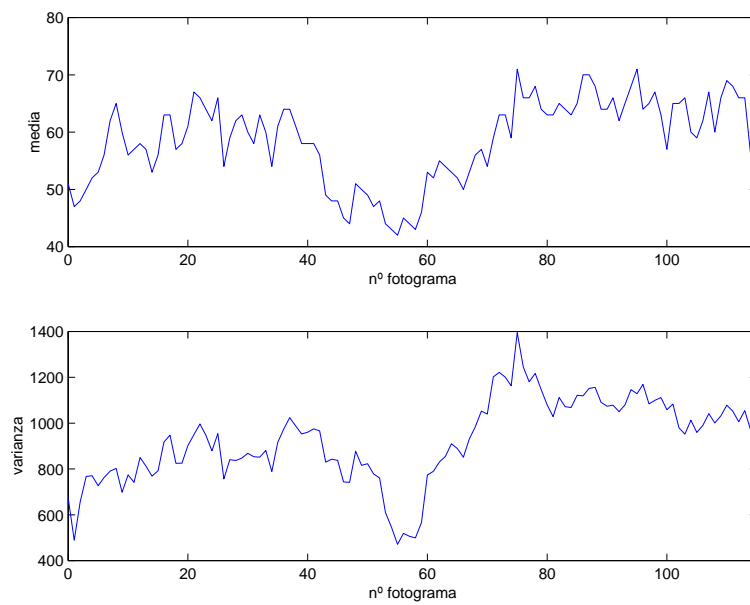


Figura 3.4: Media y varianza del plano de *Chaplin* de la figura 3.2.

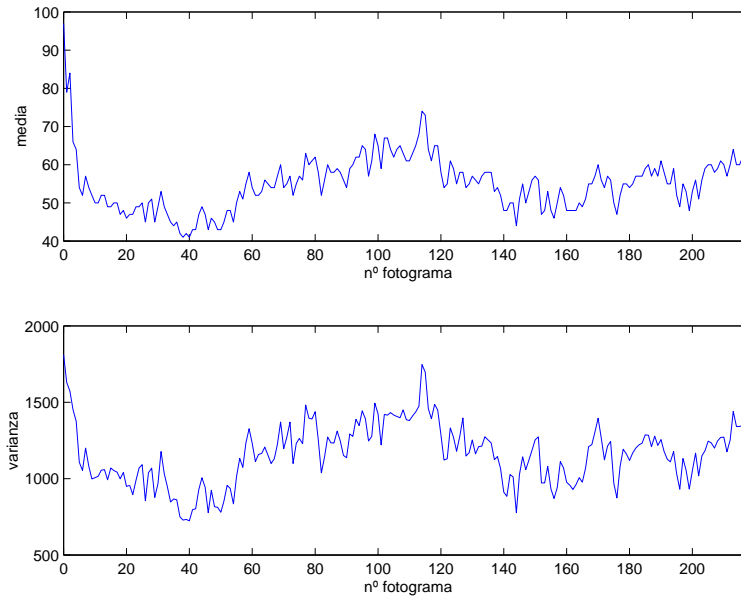


Figura 3.5: Media y varianza del plano de *Chaplin* de la figura 3.3.

Para estimar el vector de parámetros $v = [A, B, C, D]$ se va a utilizar el método de estimación del *error cuadrático medio mínimo*, para lo cual minimizaremos la expresión:

$$\xi = \sum_n (\hat{x} - x)^2.$$

Minimizando la expresión anterior en función de los parámetros

$$\nabla \xi = \left[\frac{\partial \xi}{\partial A}, \frac{\partial \xi}{\partial B}, \frac{\partial \xi}{\partial C}, \frac{\partial \xi}{\partial D} \right] = 0,$$

obtenemos una ecuación matricial que nos permitirá determinarlos:

$$\begin{pmatrix} \sum_n n^2 & \sum_n n & \sum_n n \cos \omega_0 n & \sum_n n \sen \omega_0 n \\ \sum_n n & \sum_n 1 & \sum_n \cos \omega_0 n & \sum_n \sen \omega_0 n \\ \sum_n n \cos \omega_0 n & \sum_n \cos \omega_0 n & \sum_n \cos^2 \omega_0 n & \sum_n \cos \omega_0 n \sen \omega_0 n \\ \sum_n n \sen \omega_0 n & \sum_n \sen \omega_0 n & \sum_n \cos \omega_0 n \sen \omega_0 n & \sum_n \sen^2 \omega_0 n \end{pmatrix} \begin{pmatrix} A \\ B \\ C \\ D \end{pmatrix} = \begin{pmatrix} \sum_n n x \\ \sum_n x \\ \sum_n x \cos \omega_0 n \\ \sum_n x \sen \omega_0 n \end{pmatrix}.$$

Resolviendo la ecuación anterior estimamos los parámetros que determinan la tendencia lineal (A y B), así como los que nos permiten obtener la fase y la amplitud de la componente sinusoidal (C y D). Para poder resolver el sistema anterior hay que determinar la pulsación de la sinusoide, ω_0 . El método óptimo para resolverlo sería estimar ω_0 a la vez que los demás parámetros del modelo, ya que se obtienen resultados óptimos cuando se estiman todos los

parámetros de un modelo conjuntamente. Sin embargo, este método nos proporcionaría ecuaciones no lineales más difíciles de resolver, por lo que utilizaremos un método más simple con el que obtendremos unos resultados parecidos. Para ello, previamente a la resolución del sistema anterior de ecuaciones lineales, obtendremos la pulsación (ω_0) de la senoide mediante la Transformada de Fourier. Como la frecuencia de la senoide es relativamente baja, previamente *limpiamos* la baja frecuencia de la señal sustrayendo la componente lineal de los datos, para que no nos enmascare la senoide. Una vez sustraída la media, calculamos el máximo del módulo de la Transformada Discreta de Fourier, ya que la frecuencia correspondiente a ese máximo es la de la senoide, como se muestra en la figura 3.6, en la que podemos ver el espectro de la media representada en la figura 3.4.

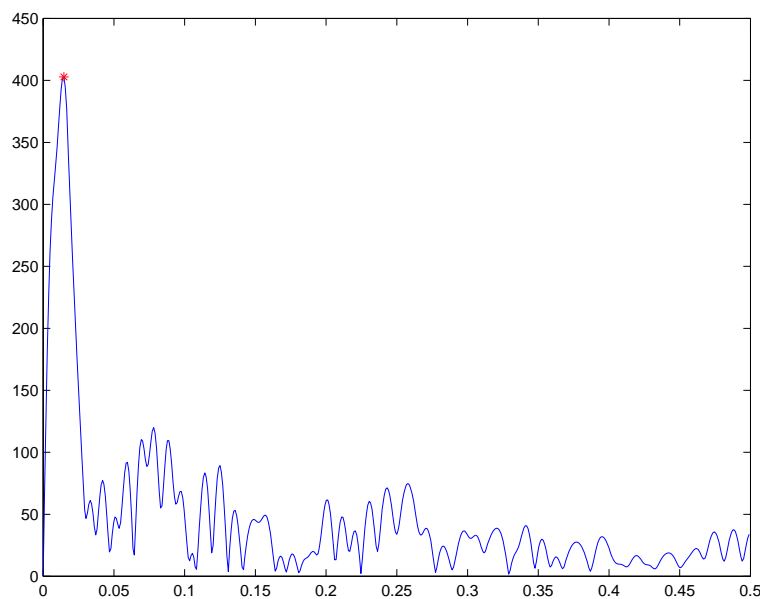


Figura 3.6: Espectro de la media de *Chaplin* de la figura 3.2. Se ha marcado con un * el máximo.

Se puede demostrar que el procedimiento anterior corresponde al estimador de máxima verosimilitud de la frecuencia de dicha senoide en el caso en que ésta se encuentre inmersa en un fondo de ruido blanco y gaussiano [128]. En nuestro caso, podemos asumir que una vez eliminada la tendencia lineal de los datos, lo que tenemos es una senoide sumada a unas variaciones aleatorias que pueden considerarse como ruido blanco gaussiano.

Las figuras 3.7 y 3.8 muestran los resultados obtenidos en la estimación de parámetros para las secuencias de test que estamos utilizando. Podemos observar la media y la varianza de ambas secuencias sobre las que se han superpuesto las componentes lineales y sinusoidales (en realidad, sinusoidal + lineal) obtenidas en la estimación para cada caso. En estos resultados podemos comprobar que la hipótesis hecha para el ruido es correcta ya que la frecuencia de la senoide se estima con bastante exactitud.

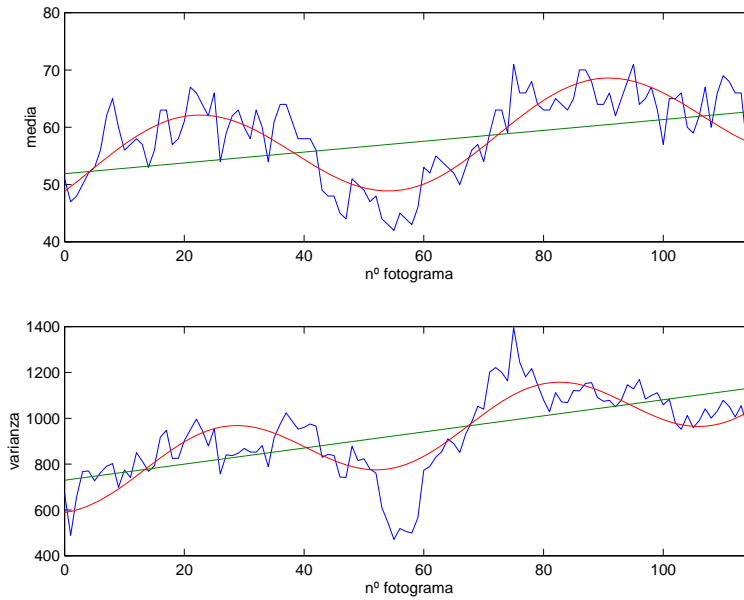


Figura 3.7: Componentes lineal y sinusoidal de la media y la varianza del plano de *Chaplin* de la figura 3.2.

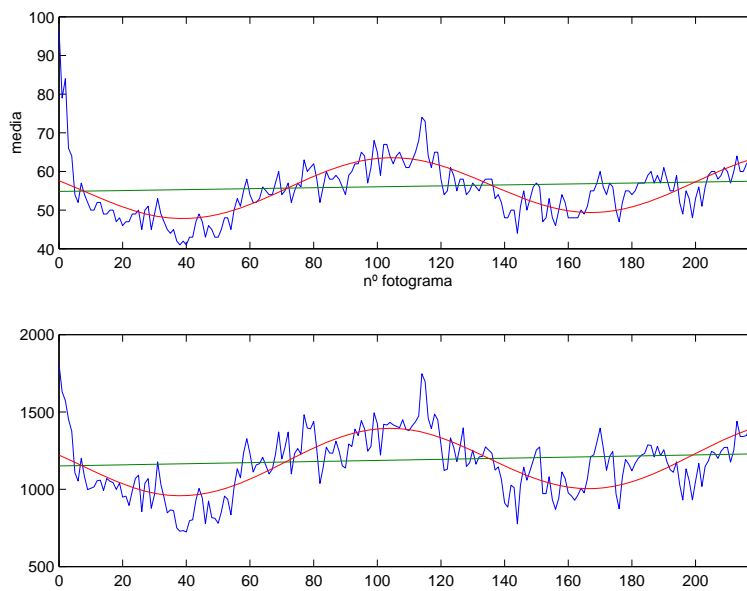


Figura 3.8: Componentes lineal y sinusoidal de la media y la varianza del plano de *Chaplin* de la figura 3.3.

3.2.3. Resultados obtenidos

Según lo visto, el método de modelado de la variación temporal consta de los siguientes pasos:

1. Extraer la media y la varianza de cada fotograma del plano sobre el que se desea eliminar el parpadeo.
2. Estimar las componentes de ambas señales: lineales ($\mu_l(n)$ y $\sigma_l^2(n)$), periódicas ($\mu_s(n)$ y $\sigma_s^2(n)$) y aleatorias ($\mu_e(n)$ y $\sigma_e^2(n)$).
3. Eliminar o reducir de la media y la varianza las componentes perjudiciales: componentes sinusoidal y aleatoria. Así, se obtiene para la media un nuevo valor dado por:

$$\hat{\mu}_n(n) = \mu_l(n) + \alpha \mu_s(n) + \beta \mu_e(n),$$

donde α y β son valores menores que la unidad e indican el porcentaje de las componentes indeseadas que queremos dejar en la media final. De forma análoga se procede para la varianza:

$$\hat{\sigma}_n^2(n) = \sigma_l^2(n) + \alpha \sigma_s^2(n) + \beta \sigma_e^2(n).$$

4. Cambiar la media y varianza de la secuencia con los nuevos valores mediante la expresión:

$$s(i, j, n) = \frac{(r(i, j, n) - \hat{\mu}_o(n)) \sqrt{\hat{\sigma}_n^2(n)}}{\sqrt{\hat{\sigma}_o^2(n)}} + \hat{\mu}_n(n),$$

donde $r(i, j, n)$ y $s(i, j, n)$ representan las intensidades del píxel, situado en la posición espacial (i, j) para el fotograma n , original y restaurada, respectivamente. $\hat{\mu}_o(n)$ y $\hat{\mu}_n(n)$ son la media original y restaurada estimadas, y $\hat{\sigma}_o^2(n)$ y $\hat{\sigma}_n^2(n)$ las varianzas.

Los resultados obtenidos para los dos planos de la secuencia Chaplin pueden verse en las figuras 3.9 y 3.10, las cuales nos muestran la media y la varianza originales y corregidas. Ambos parámetros son bien corregidos en los dos ejemplos, obteniendo lo que deseábamos, sólo la contribución de la variación lineal. Las figuras 3.11 y 3.12 muestran los fotogramas de las dos secuencias corregidas. Vemos que se ha producido un cambio de contraste y brillo, aunque visualmente, las alteraciones de la intensidad no han sido eliminadas. Esto es mucho más evidente cuando se pasa la película a la velocidad a la que el espectador va a verlas, produciéndose aún un ligero parpadeo que el ojo detecta. La conclusión que podemos extraer de estos resultados es que sólo un cambio de contraste y de brillo, para que la media y la varianza se mantengan constantes (o varíen linealmente) a lo largo del plano no es suficiente para eliminar el parpadeo, sino que será necesario además hacer que la distribución de los niveles de intensidad (el histograma) también varíe lentamente a lo largo del tiempo, dentro del mismo plano.

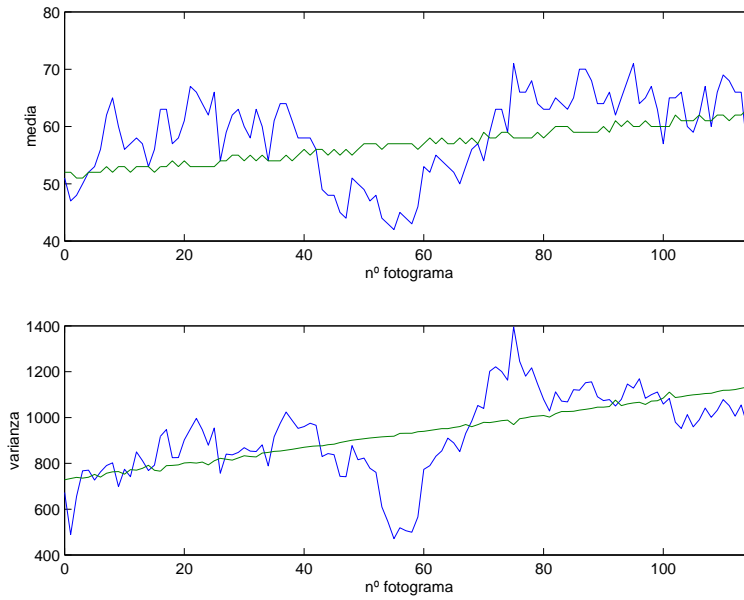


Figura 3.9: Media y varianza originales y corregidas del plano de la figura 3.2.

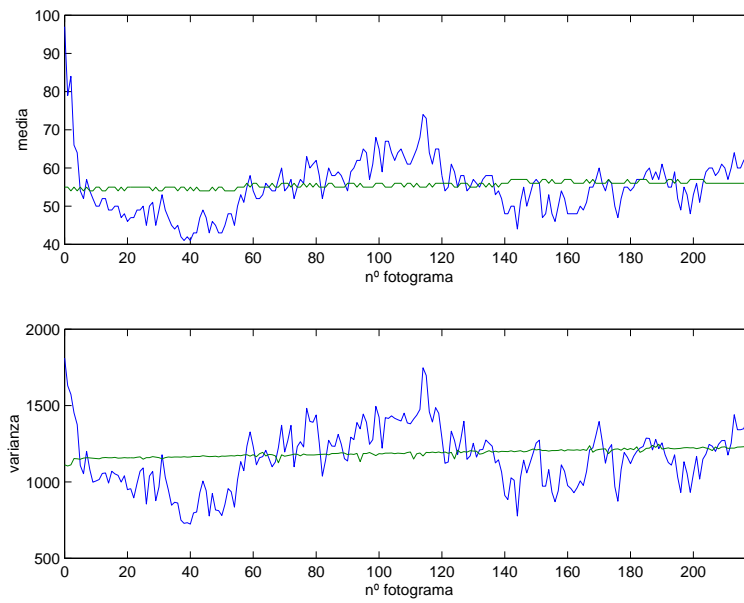


Figura 3.10: Media y varianza originales y corregidas del plano de la figura 3.3.



Figura 3.11: Fotogramas de la figura 3.2 corregidos con el método de variación temporal.



Figura 3.12: Fotogramas de la figura 3.3 corregidos con el método de variación temporal.

3.3. Método de cambio de histograma

3.3.1. Idea básica

Este método está basado en la técnica de ecualización de histograma [114, 58], que consiste en transformar el histograma de la imagen en un histograma uniforme, con lo que se consigue aumentar el contraste de la misma.

La idea de nuestro método es transformar el histograma de cada fotograma en otro que llamaremos histograma de referencia, y para conseguirlo se hará uso del algoritmo de ecualización de la siguiente manera. Sean dos imágenes I_1 e I_2 cuyos histogramas son $h_1[r]$ y $h_2[r]$, donde la variable r representa el nivel de intensidad. Definiremos $H[r]$ como una transformación T tal que el histograma de $T[I_1]$ sea igual al de I_2 . Los pasos a seguir para conseguirlo son:

1. Se ecualiza el histograma de la primera imagen utilizando la transformación $T_1[r]$, que actuando sobre la intensidad de los píxeles de I_1 cambiará su histograma, $h_1[r]$, en un histograma uniforme.
2. Se calcula la transformación $T_2[r]$ que habría que utilizar para ecualizar el histograma de I_2 .
3. Sobre la primera imagen ya ecualizada se aplica la transformación inversa de $T_2[r]$, $T_2^{-1}[r]$.

Así pues, si llamamos $T_1[r]$ a la transformación de intensidad que cambia el histograma de la primera imagen por uno uniforme y llamamos $T_2[r]$ a la transformación de intensidad utilizada para cambiar el histograma de la segunda imagen por uno uniforme, la transformación global a aplicar sobre la primera imagen para transformar su histograma en el de la segunda será:

$$H[r] = T_2^{-1}[T_1[r]]$$

Para realizar la transformación $H[r]$ de manera eficiente se construye una tabla que contenga, para cada nivel de la imagen origen, el valor en el que ha de transformarse en función de la transformación $H(r)$.

Vamos a ver un ejemplo de este método en el que vamos a transformar dos imágenes de tal forma que sus histogramas se transformen en un tercer histograma, el de referencia. El objetivo es que las dos imágenes transformadas tengan el mismo histograma. Las figuras 3.13 muestran los histogramas de las imágenes de la figura 3.1. Se puede ver fácilmente que una simple transformación afín no puede transformar un histograma en el otro. En la figura 3.14 se muestra el histograma de referencia (el mismo para ambos fotogramas), que en este caso es el histograma promedio de los de las dos imágenes. La transformación a aplicar a cada imagen se puede ver en la figura 3.15, en ella puede observarse que la primera curva está por debajo de la diagonal, lo cual significa que la primera imagen hay que oscurecerla, mientras que la siguiente curva está por encima ya que en este caso la imagen resultante de la transformación será más clara que la original. Finalmente, la figura 3.16 muestra los resultados de este método para los fotogramas que se han utilizado para mostrar cada paso del algoritmo (los

de la figura 3.1). En estos resultados se puede observar una apreciable mejora en la apreciación visual.

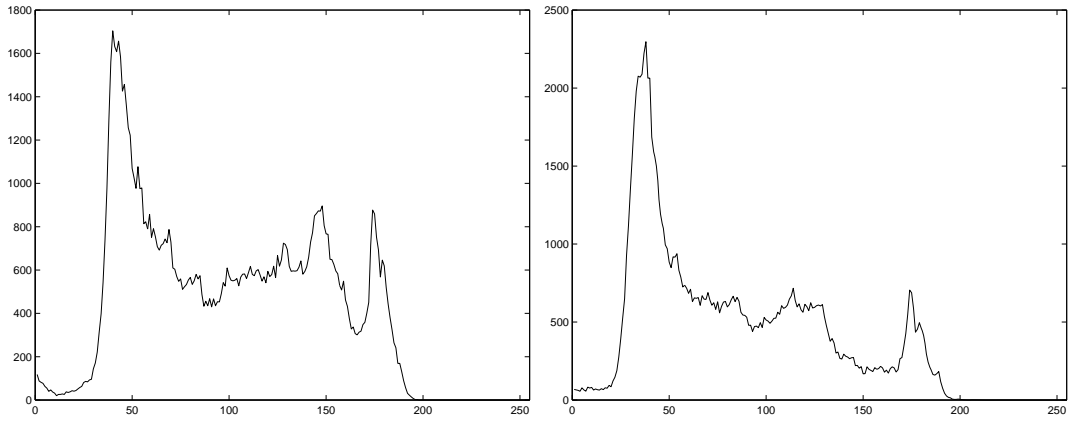


Figura 3.13: Histogramas de la imágenes de la figura 3.1.

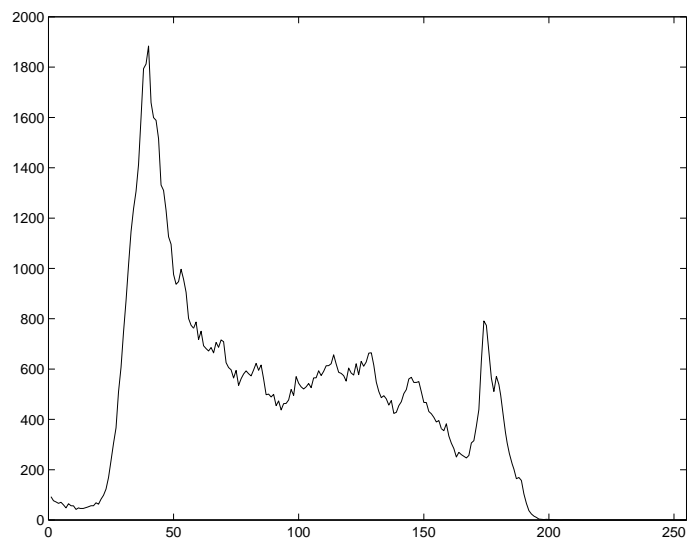


Figura 3.14: Histograma de referencia.

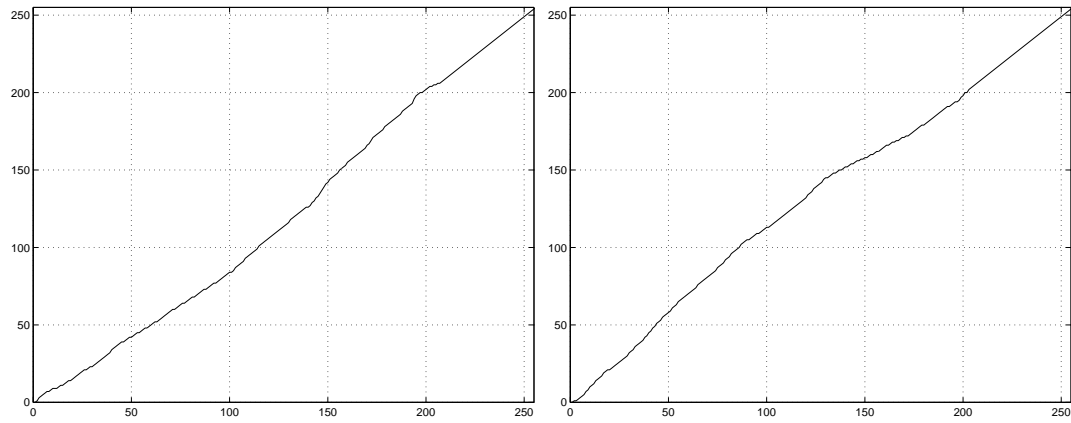


Figura 3.15: Relación E/S del método de cambio de histograma.



Figura 3.16: Resultado del método de cambio de histograma aplicado sobre los fotogramas de la figura 3.1.

3.3.2. Resultados obtenidos

Para restaurar el brillo en las películas, que es el objetivo que se pretende, para cada fotograma a restaurar se calcula el histograma de referencia y también la transformación (la tabla) para cambiar el histograma del fotograma por el de referencia. Dicho histograma de referencia podría calcularse, por ejemplo, como el histograma promedio de todos los del plano, sin embargo, esto eliminaría la evolución natural del histograma a lo largo de la escena. Con el fin de preservar esta evolución, para cada fotograma se calcula el de referencia como el promedio de los histogramas de las imágenes que entran dentro de una ventana centrada en el fotograma a restaurar.

Las figuras 3.17 y 3.18 muestran el resultado de aplicar el método de cambio de histograma sobre las secuencias de las figuras 3.2 y 3.3. La ventana utilizada para calcular los histogramas de referencia ha sido de tamaño 9 (un valor suficientemente grande para que las variaciones bruscas de intensidad se eliminen, pero lo suficientemente pequeño para que la evolución natural no resulte muy alterada).

En estos resultados se aprecia cómo mejora con respecto a los resultados del método de variación temporal (figuras 3.11 y 3.12), ya que la luminosidad se mantiene más uniforme a lo largo del plano, mejorando la apreciación visual.



Figura 3.17: Fotogramas de la figura 3.2 corregidos con el método de cambio de histograma.

3.4. Conclusiones

En este capítulo han sido presentados dos métodos para reducir el parpadeo en las películas antiguas. El primero está basado en una transformación afín de la intensidad para reducir las variaciones aleatorias y sinusoidales de la media y la varianza, calculando los parámetros



Figura 3.18: Fotogramas de la figura 3.3 corregidos con el método de cambio de histograma.

de la transformación con un modelo de variación temporal de los dos estadísticos. El método consigue reducir estas variaciones, aunque los resultados visuales obtenidos no son los esperados. El segundo método presentado trata de mejorar los resultados visuales mediante una transformación no lineal basada en cambios en el histograma. El objetivo es transformar el histograma de cada imagen en un histograma de referencia calculado como la media del histograma del fotograma en restauración y el de sus vecinos.

Al comparar los resultados obtenidos con ambos métodos podemos observar una mejora de la apreciación visual de la secuencia cuando se utiliza el segundo método. En la restauración de la base de datos de secuencias utilizada en esta tesis, así como en el software final, se utiliza como método de eliminación del parpadeo el de cambio de histograma, debido a sus mejoras en la apreciación visual.

Ejemplos de las secuencias restauradas pueden encontrarse en la página WEB <http://gpiserver.dcom.upv.es/restoration.html>.

Capítulo 4

Reducción de ruido

4.1. Introducción

El ruido es una de las distorsiones más comunes que presentan las películas antiguas. El grano, tamaño de las sales de plata, del material fotosensible influye de manera significativa en el proceso de formación de la película, haciendo que ésta presente ruido conocido como de grano. El nivel de este tipo de ruido es muy superior en las películas de archivo, ya que la degradación física del material hace que el nivel del ruido aumente considerablemente. En la figura 4.1 se muestran unos fotogramas de la secuencia *Taronja* en los que puede observarse el deterioro de los mismos debido en gran parte al ruido. Es por esto muy importante conseguir una buena técnica de reducción de ruido ya que buena parte del éxito de la restauración será debida a ella; sin embargo, como exigencia importante de la técnica que se proponga debe ser respetar el máximo posible de la secuencia original, es decir, limpiar la secuencia pero sin degradar lo que está en buen estado, y además no introducir en exceso efectos secundarios como el suavizado.

En este capítulo vamos a presentar la técnica de reducción de ruido que se propone en esta tesis, un filtrado espacio-temporal eficiente basado en las características del sistema visual humano. Antes de pasar a exponer nuestro método, en la sección 4.2 se va a hacer un repaso de las distintas técnicas de reducción de ruido existentes. En la sección 4.3 entramos en la presentación de nuestro filtrado espacio-temporal, así como en el análisis del factor de reducción que se consigue y además se muestran los resultados obtenidos. Finalmente, se dedica la sección 4.4 a analizar los resultados y presentar una serie de conclusiones.



Figura 4.1: Ejemplo de un plano de la secuencia *Taronja*.

4.2. Estado del arte

En esta sección vamos a hacer una revisión de las técnicas de reducción de ruido que aparecen en la literatura. Revisiones detalladas pueden encontrarse en [91, 25, 79]. En primer lugar vamos a definir el modelo analítico de la degradación que se pretende corregir, y a hacer una clasificación de las diferentes técnicas.

El modelo de la secuencia ruidosa que vamos a usar es el siguiente:

$$g(i, j, k) = f(i, j, k) + n(i, j, k)$$

donde $f(i, j, k)$ y $g(i, j, k)$ representan la intensidad de la imagen original y observada, respectivamente, para el fotograma k -ésimo en la posición espacial (i, j) , y $n(i, j, k)$ denota el término de ruido estacionario.

El objetivo de los filtrados que presentan los distintos autores es recuperar, a partir de la secuencia ruidosa $g(i, j, k)$, la secuencia original $f(i, j, k)$, o al menos una aproximación a la misma $\hat{f}(i, j, k)$ lo más parecida posible a la original. Existen contribuciones al filtrado de secuencias que utilizan filtros temporales o 1D, los cuales filtran sólo en la dimensión temporal pero permiten una carga computacional menor que los espacio-temporales que son los que filtran en las tres dimensiones. Las técnicas espacio-temporales, muy extendidas en la literatura, tienen la ventaja de que aprovechan las correlaciones que existen tanto en la dimensión temporal como en las espaciales, y por contra requieren mayor carga computacional. El problema que presentan estos filtros, tanto los temporales como los espacio-temporales, es que si la secuencia tiene movimiento emborronan bordes en la dirección temporal (además de la espacial si es un filtro espacio-temporal). Para solucionar este problema se presentan dos alternativas, hacer que los coeficientes del filtro sean adaptativos, o filtrar con compensación de movimiento.

Aunque el problema anterior se disminuye con la compensación de movimiento, aparecen otros inconvenientes, como la carga computacional, ya que hay que hacer estimación de movimiento, y la dependencia del resultado del filtrado con la bondad del estimador de movimiento. Dicha estimación es complicada, sobre todo en el caso de secuencias ruidosas. Es por todo esto por lo que en la literatura no sólo aparecen técnicas que compensan en movimiento, sino también otras que no lo hacen y consiguen resultados comparables utilizando otro tipo de estrategias.

A continuación se presenta un estudio de los diferentes tipos de filtros que aparecen en la literatura, los cuales hemos clasificado en cuatro grupos: filtros promediadores (FIR y recursivos), filtros basados en estadísticos ordenados, las técnicas bayesianas y los filtros multiresolución.

4.2.1. Filtros promediadores

Muchos de los filtros espacio-temporales propuestos en la literatura han sido desarrollados generalizando las técnicas de filtrado 2D. Por ejemplo, una técnica simple es obtener, a partir del filtro FIR 2D, el filtro 3D, añadiendo la dimensión temporal en el filtrado:

$$\hat{f}(i, j, k) = \sum_{p, q, l \in S} w(p, q, l) g(i - p, j - q, k - l). \quad (4.1)$$

Si el dominio del filtro dado por la ecuación 4.1 se restringe a la dimensión temporal obtenemos un filtro FIR temporal:

$$\hat{f}(i, j, k) = \sum_{l=-N}^{l=N} w(l) g(i, j, k - l). \quad (4.2)$$

De todos los filtros FIR dados por la ecuación 4.1, el que reduce más ruido es el promediador, es decir, aquel en que los coeficientes tienen el mismo peso $w(p, q, l) = 1/(MNL)$, donde M , N y L son el tamaño del filtro en cada dimensión. Este filtro promediador eliminará una cantidad de ruido dependiente de la longitud del filtro. Esta técnica tiene como ventaja la fácil implementación que, además, es poco costosa computacionalmente. El problema es que en áreas donde existe mucho movimiento y existen bordes espaciales se produce un gran emborronamiento. La solución podría ser hacer que los coeficientes fuesen adaptativos, es decir, en zonas con movimiento se da más peso a la imagen actual y esta importancia disminuye con la distancia en el tiempo. De la misma forma se da más importancia al píxel actual que a los vecinos cuando existe un borde espacial. Así pues, intentando solucionar el problema del que adocelen estos filtros aparecen multitud de técnicas en la literatura.

Como se ha dicho anteriormente, para obtener todo el partido posible de las correlaciones existentes en la dirección temporal es necesario realizar una estimación de movimiento y una compensación posterior. Realizando este paso antes de filtrar, se consigue eliminar el problema del suavizado de bordes debidos al movimiento, aunque no soluciona el problema del suavizado de bordes debido al filtrado espacial. Sin embargo, estos filtros requieren una mayor carga computacional, así como el resultado de los mismos depende en una gran medida del estimador de movimiento. Han aparecido muchas técnicas de filtrado espacio-temporal compensado en movimiento, desde modificaciones de los filtros anteriores (los no compensados), hasta filtros que llevan implícita dicha compensación. Por ejemplo, la modificación del filtro 3D de la ecuación 4.1 para obtener el filtro compensado en movimiento sería:

$$\hat{f}(i, j, k) = \sum_{p,q,l \in S} w(p, q, l) g(i - p - \hat{d}_x^{k,k-l}(i - p, j - q), j - q - \hat{d}_y^{k,k-l}(i - p, j - q), k - l).$$

donde $\hat{d}_x^{k,k-l}(i, j)$ y $\hat{d}_y^{k,k-l}(i, j)$ representan el desplazamiento horizontal y vertical estimados entre los fotogramas k y $k - l$.

La versión compensada en movimiento de la ecuación 4.2 viene dada por

$$\hat{f}(i, j, k) = \sum_{l=-N}^{l=N} w(l) g(i - \hat{d}_x^{k,k-l}(i, j), j - \hat{d}_y^{k,k-l}(i, j), k - l).$$

Los filtros promediadores se clasifican en dos tipos, atendiendo a la duración de la respuesta al impulso del mismo. Así pues tenemos los de respuesta finita, filtros FIR, y los de respuesta impulsional infinita, los IIR o recursivos.

Filtros FIR

Los coeficientes $w(p, q, l)$, estacionarios o adaptativos, se eligen de tal forma que se consiga cierto objetivo. Una opción común es hacer que se optimicen en el sentido de error cuadrático

medio

$$w(p, q, l) \leftarrow \min_{w(p,q,l)} E[(f(i, j, k) - \hat{f}(i, j, k))^2].$$

En este caso, 4.1 se convierte en el filtro de Wiener 3D. Özkan *et al.* [160] presentan un algoritmo eficiente que implementa el filtro de Wiener (filtro de Wiener multitrama con correlación cruzada, CCMF). Sin embargo este tipo de filtro tiene dos desventajas principales: la primera es que la función de autocorrelación 3D para la secuencia original debe ser conocida a priori, y la segunda es la suposición de estacionariedad 3D en sentido amplio, la cual nunca se cumple debido al movimiento de objetos y a los cambios de escena. Estas dos razones hacen que los resultados del filtro de Wiener no sean muy satisfactorios.

Chan *et al.* [32] presentan una solución diferente al filtro de Wiener con el fin de modelar las no estacionariedades de la secuencia, el filtro de Volterra, que consiste en un filtro no lineal de la siguiente forma:

$$\hat{f}(i, j, k) = \sum_{p,q,l \in S} w_1(p, q, l) g(i - p, j - q, k - l) + \sum_{p,q,l \in S} \sum_{p',q',l' \in S} w_2(p, q, l, p', q', l') g(i - p, j - q, k - l) g(i - p', j - q', k - l')$$

donde los coeficientes w_1 y w_2 se determinan también minimizando el error $E[\{\hat{f} - f\}^2]$. Este filtro tiene el problema de que para determinar los coeficientes hace falta conocer momentos de orden superior a priori.

Existen versiones compensadas en movimiento de muchos de los filtros anteriores, por ejemplo del filtro de Wiener [160] o del de Volterra [32]. Dekeyser *et al.* [41, 40] proponen un filtro de Wiener compensado en movimiento, estimando dicho desplazamiento mediante un modelo afín. Para el filtro de Wiener estudian diferentes aproximaciones para reducir la carga computacional, como por ejemplo la aproximación de Kokaram [84], que propone una simplificación usando una transformada discreta de Fourier 3D.

Boyce [24] propone un filtro promediador temporal que cambia adaptativamente de un promedio temporal compensado en movimiento a no compensado en movimiento. Usando un algoritmo de bloque Boyce propone el cambio entre un promediador y otro en base a la diferencia media absoluta (entre dos bloques, uno de cada fotograma) antes y después de la compensación. Así pues la compensación de movimiento se usará si

$$MAD_o < \beta MAD_{mc} \quad \text{o si } MAD_o < \tau ,$$

donde MAD_{mc} y MAD_o son las diferencias medias absolutas obtenidas con y sin compensación de movimiento, respectivamente. Los parámetros τ y β se calculan para minimizar la probabilidad de detectar movimiento en un bloque cuando no lo hay, es decir, que la falsa alarma se debe al ruido.

Borghys y Acheroy [21] proponen también un filtro FIR temporal compensado en movimiento, para lo cual utilizan un estimador de movimiento robusto frente al ruido utilizando restricciones de conservación de características y de vecindad. Con esta compensación consiguen que las no estacionariedades de la secuencia debidas al movimiento se eliminen, eliminándose el emborronamiento cuando se filtra temporalmente. Para evitar también el emborronamiento

no debido al movimiento adaptan los coeficientes del filtro en función de la correlación espacial. Con todo ello consiguen mejorar los resultados del filtro temporal no compensado en movimiento, y del compensado con coeficientes no adaptativos.

Otros autores se inclinan por utilizar una técnica clásica de filtrar señales no estacionarias que consiste en descomponer la señal en dos partes, la estacionaria y la no estacionaria. Si el ruido es estacionario, estará enteramente contenido en la parte estacionaria. De esta forma, cualquier filtro lineal puede ser usado para filtrar esta parte. Para descomponer la señal es necesaria la estimación de estadísticos. Kleihorst *et al.* [79, 77] utilizan un estimador adaptativo basado en estadísticos ordenados para estimar la media ($\mu_g(i, j, k)$) y la varianza ($\sigma_g(i, j, k)$) de la secuencia observada. Con la estima de estos estadísticos, la señal no estacionaria, la secuencia, puede descomponerse en una parte no estacionaria y otra estacionaria, que denotaremos como $y(i, j, k)$ y que viene dada por:

$$y(i, j, k) = \frac{g(i, j, k) - \hat{\mu}_g(i, j, k)}{\hat{\sigma}_g(i, j, k)},$$

donde $\hat{\mu}_g(i, j, k)$ y $\hat{\sigma}_g(i, j, k)$ son las estimaciones de la media y la varianza respectivamente de la secuencia observada. La secuencia filtrada mediante este método quedará como:

$$\hat{f}(i, j, k) = \hat{\mu}_g(i, j, k) + \hat{\sigma}_g(i, j, k)F\{y(i, j, k)\},$$

donde el operador $F\{\cdot\}$ representa el filtrado, que puede ser, como propone Kleihorst, un filtro FIR temporal causal donde los coeficientes se estiman por mínimos cuadrados.

Filtros recursivos IIR

Los filtros FIR, tanto temporales como espacio-temporales, requieren un compromiso entre la cantidad de ruido suprimido y la cantidad de memoria necesaria para su implementación, ya que ambas cosas son proporcionales a la longitud del filtro. Una alternativa de los filtros FIR para evitar este compromiso son los filtros recursivos IIR, filtrado que viene dado por la expresión:

$$\hat{f}(i, j, k) = [1 - \alpha(i, j, k)] \hat{f}_b(i, j, k) + \alpha(i, j, k) g(i, j, k), \quad (4.3)$$

donde $\hat{f}_b(i, j, k)$ denota una predicción de la secuencia original antes de ser actualizada por el filtro, y $\alpha(i, j, k)$ es la ganancia del filtro. De la ecuación 4.3 se han derivado muchos algoritmos dependiendo de la forma de predecir $\hat{f}_b(i, j, k)$ y de la actualización de $\alpha(i, j, k)$.

Cano y Benard [29] utilizaron un campo de Markov 3D para modelar la secuencia y generar la predicción $\hat{f}_b(i, j, k)$, y la ganancia fue calculada para conseguir un error cuadrático medio mínimo. El filtro resultante es una extensión 3D del filtro de Kalman 2D. Los resultados obtenidos por Cano y Benard indican que el filtro de Kalman reduce el ruido visiblemente, sin embargo los suavizados introducidos en el movimiento son inaceptables, por lo que concluyen que el filtro de Kalman requiere compensación de movimiento. La misma idea utilizan Kim y Woods [72], quienes evitan la carga computacional asociada con el filtro de Kalman utilizando una actualización reducida. Este filtro (RUKF) también utiliza un modelo de vídeo basado en campos aleatorios de Markov, y también se concluye que se necesita una compensación de movimiento, por lo que proponen el filtro anterior con compensación de movimiento

(MRUKF). El filtro de Kalman compensado en movimiento propuesto por Patti *et al.* [112] es de orden reducido (ROMKF) y utiliza un modelo AR. También Signes [138] propone un filtro recursivo temporal compensado en movimiento utilizando un modelo AR-1 para estimar $\hat{f}_b(i, j, k)$. El factor $\alpha(i, j, k)$, en una primera versión, es adaptativo en función de la cantidad de movimiento mientras que en una versión posterior se calcula en función de los estadísticos de la señal para conseguir unos coeficientes óptimos que minimicen el error cuadrático medio.

Otro tipo de filtro consiste en reemplazar la predicción de la secuencia original antes de la actualización, $f_b(i, j, k)$, por una estimación de la media de la secuencia observada [75, 134]:

$$f_b(i, j, k) = \hat{\mu}_g(i, j, k).$$

Con esta elección, la ganancia o parámetro de control del filtro de Kalman $\alpha(i, j, k)$ óptima para conseguir un error cuadrático medio mínimo (MMSE) es:

$$\alpha(i, j, k) = 1 - \hat{\sigma}_n^2(i, j, k) / \hat{\sigma}_g^2(i, j, k),$$

donde $\hat{\sigma}_n^2(i, j, k)$ y $\hat{\sigma}_g^2(i, j, k)$ son las estimaciones de la varianza del ruido y de la secuencia observada, respectivamente. Este filtro se conoce con el nombre de filtro lineal de error cuadrático medio mínimo local (LLMMSE) y es el que utiliza Sezan [134] *et al.* compensado en movimiento, para lo cual utilizan el algoritmo de estimación de movimiento desarrollado por Fogel [53]. Usan una ventana espacio-temporal para estimar $\hat{\sigma}_g^2(i, j, k)$ y $\hat{\mu}_g(i, j, k)$, lo cual no es muy preciso cuando la ventana contiene los bordes de un objeto. El efecto que produce esta imprecisión en la estimación es que la varianza tenga un valor muy alto, lo que hace que el filtro no actúe y que el ruido no sea eliminado en los bordes espacio-temporales.

Kleihorst *et al.* [75, 74] usan un estimador basado en estadísticos ordenados para evitar la imprecisión en la estimación de $\hat{\sigma}_g^2(i, j, k)$ y $\hat{\mu}_g(i, j, k)$ en los bordes. Consideran dos estimadores, el espacio-temporal no compensado en movimiento [75] y el recursivo espacio-temporal compensado en movimiento [74], donde utilizan un método de estimación de movimiento de bloques en el que el criterio de búsqueda está basado en estadísticos de orden 3, para hacerlo más robusto frente al ruido.

Dugad y Ahuja [47, 45] presentan un método que combina los resultados obtenidos al filtrar la secuencia con un filtro de Kalman temporal y un filtro de Wiener de coeficientes adaptativos para preservar los ejes espaciales. Para el filtro temporal se estima previamente el movimiento usando un método de bloques [47, 45] o de flujo óptico [45], utilizando el fotograma actual, sobre el que previamente se ha realizado el filtrado espacial, y el anterior ya filtrado. Además presentan varias formas de combinar los resultados de ambos filtrados, concluyendo que la mejor forma de hacerlo es promediando los resultados del filtro de Wiener y los del filtro de Kalman. Posteriormente presentan una mejora del método [46] introduciendo fotogramas futuros en el filtro temporal y utilizando como filtro espacial un filtro de Wiener sobre la descomposición en *wavelets* de la imagen.

4.2.2. Filtros de estadísticos ordenados

Estos filtros se diferencian de los anteriores en que ordenan los valores de la intensidad de los píxeles dentro de la ventana de análisis antes de operar con ellos. Los filtros de estadísticos

ordenados más usados son los de mediana. El problema que presentan es que objetos estrechos que se mueven deprisa son eliminados. Para solucionar este problema Arce [15] y Alp y Neuvo [11] proponen métodos basados en filtros de mediana espacio-temporales multietapa y multinivel, respectivamente. Un filtro de mediana multietapa (MMF) es un método que combina la salida de varias etapas para dar lugar al resultado final. Cada etapa, a su vez, puede estar formada por una estructura de filtros en cascada, donde cada filtro se diseña para preservar una característica de la secuencia (líneas, esquinas, ejes,..) en una dirección. Si el conjunto de máscaras que determinan uno de estos filtros trabajan sólo en una dimensión (espacial o temporal) hablamos de estructura unidireccional, mientras que si toman en cuenta píxeles en ambas direcciones recibe el nombre de bidireccional. Arce [15] considera dos conjuntos de filtros, el primero unidireccional (primera etapa) y un segundo grupo de filtros (segunda etapa) bidireccionales. El resultado del filtrado final se calcula como la mediana entre el máximo encontrado en la primera etapa, el mínimo obtenido en la segunda, y el valor del píxel central.

$$\hat{f}(i, j, k) = \text{mediana}\{\text{máx}(\text{primera etapa}), g(i, j, k), \text{mín}(\text{segunda etapa})\};$$

además concluye que si en la primera etapa también se utilizan filtros bidireccionales se consigue mayor supresión de ruido.

Alp y Neuvo [11] proponen dos filtros de mediana multinivel. El primero (P3D) consiste en calcular la mediana de los resultados de tres filtros de mediana trabajando cada uno en dos planos, es decir, en el primer filtro la ventana de análisis (de 5 puntos) se extiende en el plano $x - y$, en el segundo en el $x - t$ y en el tercero en el $y - t$. El segundo filtro (ML3D), calcula la mediana de los resultados de dos filtros donde cada uno de ellos obtiene la mediana de los píxeles que están dentro de una ventana 3D ($x - y - t$) de 7 puntos, centrada en el píxel a filtrar; donde para cada uno de los dos filtros la ventana es diferente. Este método mejora los resultados obtenidos por Arce.

El método propuesto por Norell *et al.* [106] es un filtrado, en tiempo real, temporal de mediana con una máscara de 7 elementos temporales centrada en el fotograma a filtrar y un tamaño espacial ajustable. Previamente al filtrado se detecta si existe movimiento y si así no se realiza el filtrado temporal. También presentan la implementación ASIC del filtrado.

Kleihorst *et al.* [76, 78] presentan un método para filtrar secuencias de imágenes que han sido tomadas con una cámara, por lo que han sufrido la corrección gamma. Debido a esto el ruido se ha hecho dependiente de la señal. Para filtrar este tipo de ruido presentan un filtro FIR espacio-temporal basado en estadísticos ordenados cuyos coeficientes se adaptan de manera óptima a la función densidad de probabilidad del ruido.

4.2.3. Filtros bayesianos

Otro tipo de métodos que han sido propuestos más recientemente para filtrar secuencias de imágenes se basan en maximizar una función de verosimilitud, de la siguiente forma:

$$\hat{f}(i, j, k) \leftarrow \max_{f'(i, j, k)} p\{g(i, j, k) | f'(i, j, k)\}.$$

El estimador obtenido de esta forma encuentra la secuencia original más probable que ha dado lugar a la secuencia observada que se tiene. Muchos de los métodos bayesianos modelan

la secuencia con un campo aleatorio de Markov [95, 100], con lo que la expresión que intentan resolver es la siguiente:

$$\hat{f}(i, j, k) \leftarrow \max_{f'(i, j, k)} p\{g(i, j, k)|f'(i, j, k)\}p\{f'(i, j, k)|\text{píxeles vecinos}\}.$$

Usando las propiedades de los campos de Markov, la segunda parte de la ecuación anterior puede escribirse como una distribución de Gibbs:

$$p\{f'(i, j, k)|\text{píxeles vecinos}\} = \frac{1}{k} e^{-U\{f'(i, j, k), \hat{f}(i, j, k-1)\}},$$

donde k es una constante de normalización y $U\{.\}$ una función de energía.

Como hemos dicho anteriormente los filtros compensados en movimiento tienen el problema de que el resultado depende mucho del estimador de movimiento. El nivel de ruido en una secuencia provoca que el estimador de movimiento no sea preciso, por lo que habría que eliminar el ruido de la secuencia para conseguir un buen estimador. Sin embargo, estamos calculando el movimiento para poder filtrar bien y así eliminar el máximo ruido posible. Es por ello que existen métodos en la literatura que realizan simultáneamente ambas cosas, filtrar y estimar el desplazamiento. Brailean y Katsaggelos [26, 27] proponen un estimador de movimiento pel-recursivo [143]. Utilizan un campo aleatorio de Markov para modelar el desplazamiento y la intensidad original, y proponen maximizar una función densidad de probabilidad conjunta a posteriori con respecto al campo de movimiento y a la intensidad original, lo cual da lugar a un conjunto de ecuaciones consistentes en dos filtros de Kalman acoplados. Otros autores, como Dubois y Konrad [44] también proponen utilizar campos aleatorios de Markov para modelar el campo de movimiento, concretamente ellos proponen un método de estimación de movimiento basado en la minimización de funciones objetivo modeladas como campos aleatorios de Markov, y presentan algoritmos que simplifican dicha maximización.

4.2.4. Filtros multiresolución

Las técnicas multiresolución aplicadas a imágenes 2D se han convertido en un método muy popular de análisis, y son muy utilizadas en aplicaciones como la compresión. Esta representación de la señal es útil también para restauración. La idea fundamental es que si se realiza una descomposición en bandas de diferentes resoluciones, la energía de la señal quedará concentrada en determinadas bandas mientras que el ruido se extenderá por todas ellas.

Amer y Schröder [12] y Joustschulte *et al.* [69] proponen un método orientado a los requerimientos del sistema visual humano y a su percepción del ruido. Así pues el algoritmo pretende limpiar la secuencia sin emborronar bordes espaciales ni temporales. Para conseguir lo primero, se descompone cada imagen en altas y bajas frecuencias; como el ojo es menos sensible al ruido en altas frecuencias, éstas se promedian menos, consiguiendo de esta forma también no emborronar los bordes espaciales, que son componentes de alta frecuencia. Para evitar el suavizado de bordes temporales se hace una detección de movimiento, aumentando el suavizado a medida que la secuencia se hace más estacionaria. Los autores presentan el algoritmo para vídeo progresivo y entrelazado. Además proponen un diferente tipo de filtro

espacial para eliminar ruido de baja frecuencia: un filtro recursivo [69] o uno de mediana con máscaras de diferentes orientaciones para preservar bordes [67, 68].

Una técnica multiresolución usada en eliminación de ruido en imágenes es la llamada *Coring*. Dicha técnica consiste en descomponer la señal en subbandas, utilizando una transformada como la *wavelet* (DWT), la transformada del coseno (DCT), etc., que tiene como misión separar lo mejor posible la señal del ruido, por lo que es ventajoso usar transformadas que compacten lo máximo posible la energía de la señal. A los coeficientes transformados se les aplica como umbral una determinada función llamada función *core*. Roosmalen *et. al* [122] extienden esta técnica a 3D, realizando los siguientes pasos:

1. Se compensan en movimiento los fotogramas.
2. Se realiza la descomposición 2D para cada fotograma compensado. Proponen dos transformadas: la DWT sin diezmar, método que se conoce como algoritmo *à trous* y la pirámide de Simonelli [148] que es una descomposición en subbandas basada en filtros direccionales.
3. Se aplica la DWT en la dirección temporal, sobre cada conjunto de coeficientes en las correspondientes posiciones de las diferentes escalas.

Este tipo de técnica, la implementada con el algoritmo *à trous*, es comparada por Roosmalen *et al.* [117] con un filtro de estadísticos ordenados y uno de Wiener para filtrar secuencias que van a ser codificadas en MPEG-2 intentando mejorar la eficiencia de codificación reduciendo el ruido, y concluyen que se consigue mayor eficiencia de codificación, sobre todo con el algoritmo *à trous*.

4.3. Filtrado espacio-temporal lineal

4.3.1. Idea básica

El filtro propuesto en esta tesis para eliminar ruido, y que describiremos a continuación, se basa en la respuesta espacio-temporal del sistema visual humano. Muchos investigadores [5, 147, 153, 146, 152] han estudiado esta respuesta y han intentado establecer un modelo para aprovechar todas las propiedades del mismo a la hora de, por ejemplo, comprimir secuencias de vídeo, entre otras aplicaciones. Dicha respuesta espacio-temporal, de manera muy simplificada [43] y representando sólo la frecuencia temporal y una de las frecuencias espaciales, es la que vemos en la figura 4.2. En ella podemos observar la característica fundamental que va a utilizar nuestro filtro, y es que el ojo no es capaz de captar a la vez una alta frecuencia espacial y una alta frecuencia temporal, es decir, los objetos de la imagen que se mueven rápido podemos verlos con poco detalle, y por el contrario, a medida que el objeto se mueve más lentamente vamos siendo capaces de apreciar mejor el detalle espacial. Por tanto, el objetivo del filtro espacio-temporal que veremos en esta sección será eliminar el máximo ruido posible siempre y cuando las degradaciones introducidas al hacerlo (emborronamiento) no sean apreciables por el espectador.

Implementar un filtro que presente como respuesta espacio-temporal la misma que el sistema visual humano (figura 4.2) es costoso computacionalmente, por lo que se simplifica

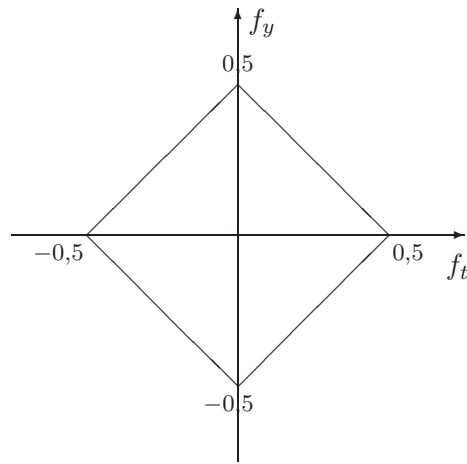


Figura 4.2: Modelo simplificado de la respuesta espacio-temporal del sistema visual humano.

de tal forma que el efecto del filtrado sea parecido, siendo la implementación más sencilla. En la figura 4.3 podemos ver la respuesta del sistema visual y la respuesta del filtro que se va a realizar. Si trasladamos la respuesta del filtro a tres dimensiones, las dos espaciales f_x y f_y y la temporal f_t , obtenemos la respuesta completa del filtro espacio-temporal 3D propuesto, la cual se muestra en la figura 4.4, donde W_x y W_y son los anchos de banda en las dos dimensiones espaciales, y W_t el ancho de banda en la dimensión temporal.

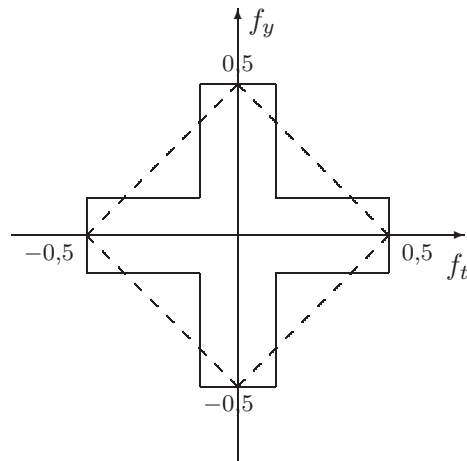


Figura 4.3: Respuesta espacio-temporal del sistema visual humano (- - -) y respuesta del filtro a implementar (—).

A continuación, vamos a explicar como funciona el filtro, y para ello vamos a fijarnos sólo en uno de los planos espacio-temporales del mismo, por ejemplo el $f_y - f_t$, el cual se representa en la figura 4.5.

Comprobaremos que el filtro que vamos a implementar cumple con el objetivo propuesto, funcionar de forma análoga al sistema visual humano. Para ello vamos a establecer para cada

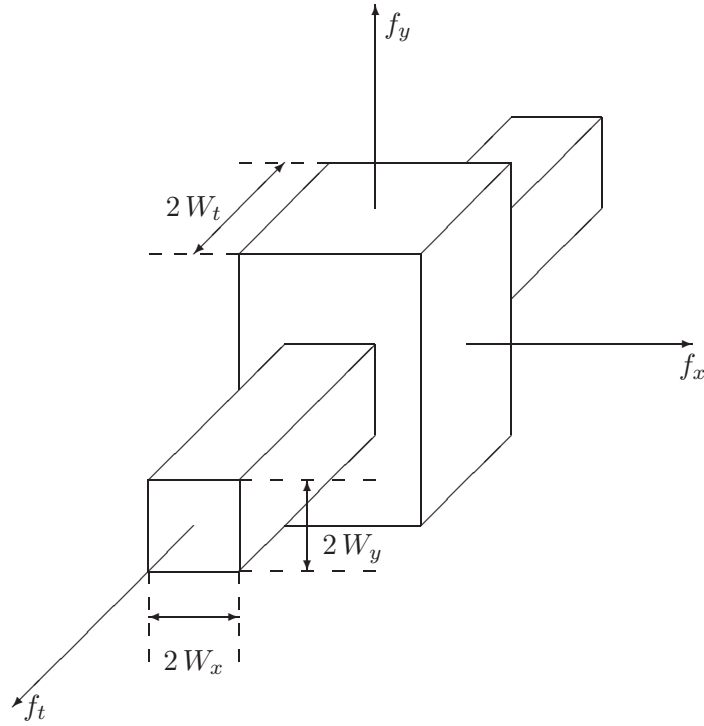


Figura 4.4: Respuesta en frecuencia del filtro espacio-temporal.

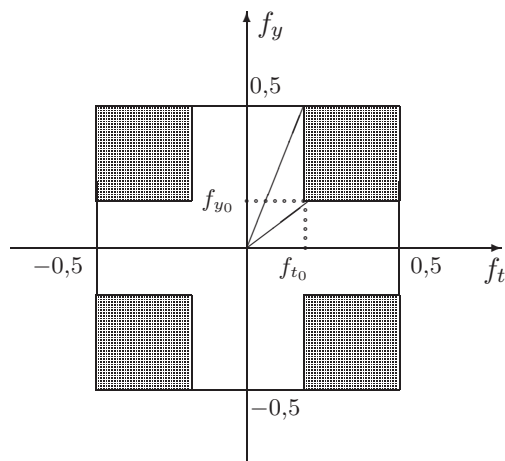


Figura 4.5: Corte por el plano $f_y - f_t$ de la respuesta en frecuencia del filtro espacio-temporal ideal. La zona sombreada corresponde a la banda eliminada del filtro.

tamaño de un objeto (resolución espacial) cuál es la velocidad máxima para que no sea emborronado, es decir, para que no sean eliminadas sus altas frecuencias. Para ello comenzaremos comprobando cómo a partir de la respuesta en frecuencia de una secuencia espacio-temporal se puede determinar la velocidad de la misma.

Relación entre la frecuencia espacio-temporal y la velocidad de una señal

Vamos a ver cómo a partir del espectro espacio-temporal de una señal podemos obtener la velocidad para cada frecuencia espacial de la misma. Para simplificar la explicación empezaremos demostrando cómo se puede obtener la velocidad para una señal de una sola frecuencia. Sea $s(y, t)$ una señal sinusoidal espacio-temporal que varía únicamente en el espacio, según la dimensión y ; es decir, una señal estática la cual viene dada por la expresión:

$$s(y, t) = \cos(2\pi f_{y_0} y).$$

Esta secuencia correspondería a una secuencia de imágenes todas iguales (estática), formada por franjas horizontales que varían su intensidad sinusoidalmente con y . El espectro espacio-temporal ($f_y - f_t$) de dicha secuencia serán dos deltas situadas en $f_t = 0$ y $f_y = \pm f_{y_0}$, como muestra la figura 4.6.

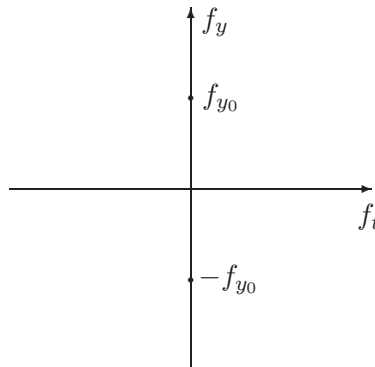


Figura 4.6: Espectro espacio-temporal de la secuencia $s(y, t) = \cos(2\pi f_{y_0} y)$.

Supongamos ahora que la senoide se mueve a lo largo del tiempo con una velocidad v , con lo que la expresión anterior se transforma en:

$$s(y, t) = \cos(2\pi f_{y_0}(y - vt)) = \cos(2\pi f_{y_0}y - 2\pi f_{t_0}t),$$

donde

$$f_{t_0} = v f_{y_0}. \quad (4.4)$$

La nueva secuencia serán imágenes formadas por las franjas horizontales cuya intensidad varía con y , pero ahora estas franjas se mueven de una imagen a otra, con una velocidad v . El espectro de la nueva señal se muestra en la figura 4.7, y se trata de dos deltas con la misma frecuencia espacial (f_{y_0}) que en el caso estático, pero de frecuencia temporal f_{t_0} .

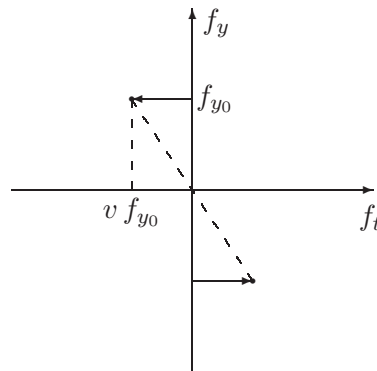


Figura 4.7: Espectro espacio-temporal de la secuencia $s(y, t) = \cos(2\pi f_{y0}(y - vt))$.

Según lo visto, la velocidad de toda componente frecuencial de una señal, dado su espectro espacio-temporal, se puede calcular como la pendiente de la recta que une la frecuencia $(0, 0)$ con el punto donde se sitúa dicha componente. O lo que es lo mismo, el cociente entre la frecuencia temporal y la espacial, relación que viene dada en la ecuación 4.4.

Funcionamiento del filtro

Una vez visto cómo se puede obtener la velocidad a la que se mueve una determinada frecuencia, ese mismo resultado se puede extrapolar al caso de una secuencia formada por objetos con diferentes frecuencias espaciales. En un objeto de una secuencia que se mueve a una velocidad v , todas las frecuencias espaciales que lo componen se moverán a la misma velocidad, y todas cumplirán la relación $\frac{f_{t0}}{f_{y0}} = v$.

Observando la respuesta espacio-temporal del filtro que se muestra en la figura 4.5 se puede establecer cómo afectará el filtro a cada objeto de la secuencia según su velocidad:

- Objetos estáticos: $f_t = 0$. Sean cuales sean las frecuencias espaciales del objeto, estarán en la banda de paso del filtro, por lo que el objeto no sufrirá ninguna variación al filtrar. Para estos objetos estáticos el ojo ve mejor el detalle, por lo que interesa no degradarlos.
- Objetos que se mueven a una velocidad $v < \frac{f_{t0}}{0,5}$ (píxeles/fotograma) no se emborronan, sea cual sea su tamaño, ya que están dentro de la banda de paso del filtro, es el mismo caso que los objetos estáticos. En la práctica, el emborronamiento que sufren estos objetos no es nulo, ya que la respuesta del filtro no es plana en la banda de paso, sino que tiene cierta atenuación.
- Objetos con su velocidad en el rango $\frac{f_{t0}}{0,5} < v < \frac{f_{t0}}{f_{y0}}$ (píxeles/fotograma) empiezan a emborronarse, ya que empieza a disminuir el ancho de banda espacial, a medida que aumenta la velocidad.
- A partir de $v > \frac{f_{t0}}{f_{y0}}$ (píxeles/fotograma) ya no se emborrona más y comienza a aumentar el ancho de banda temporal. Objetos con velocidades mayores de este límite, se mueven tan rápidamente que el ojo podría apreciar el suavizado temporal.

La figura 4.8 muestra el ancho de banda espacial en función de la velocidad con el fin de ilustrar el comportamiento descrito anteriormente. Puede observarse como el ancho de banda espacial es constante y máximo hasta la velocidad $f_{t_0}/0,5$ (píxeles/fotograma). En este tramo estarían los objetos estáticos o con velocidad menor que $f_{t_0}/0,5$ (píxeles/fotograma) los cuales no se emborronan. A partir de este punto va decreciendo el ancho de banda hasta un límite para la velocidad f_{t_0}/f_{x_0} (píxeles/fotograma) en el que ya permanece constante, punto donde se produce el emborronamiento máximo permitido, donde el ancho de banda espacial se reduce hasta f_{y_0} .

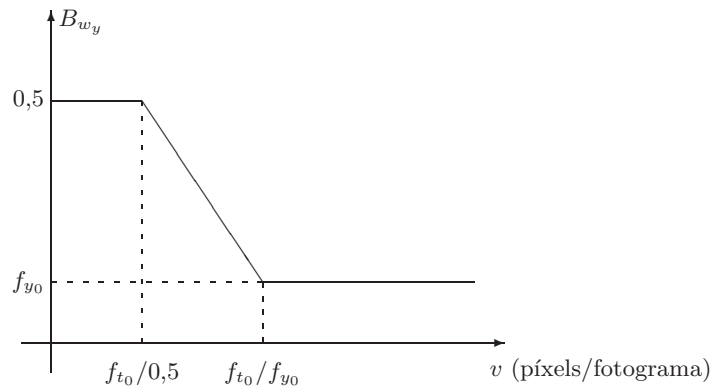


Figura 4.8: Ancho de banda espacial en función de la velocidad para el filtro espacio-temporal.

Si comparamos este filtrado con uno temporal, cuyo espectro espacio-temporal se muestra en la figura 4.9, la reducción de ruido es menor ya que la banda de paso es menor. Este hecho lo comprobaremos más adelante cuando se presenten los resultados. Sin embargo, el filtro espacio-temporal es más respetuoso con la imagen, emborrona menos los bordes espaciales en el caso de que exista movimiento. Si nos fijamos en la relación ancho de banda espacial-velocidad para el filtro temporal, representada en la figura 4.10, observamos que inicialmente tiene el mismo comportamiento el ancho de banda que en el caso del filtro espacio-temporal, pero una vez sobrepasado el límite de velocidad $f_{t_0}/0,5$ decrece tendiendo asintóticamente a 0, en lugar de permanecer constante en ese límite. Es en este punto donde se producen las diferencias entre ambos filtros, es decir, para aquellos objetos con velocidades mayores que este límite, el filtro temporal emborronará cada vez más a medida que aumente la velocidad, reduciendo también cada vez más el ruido, mientras que el espacio-temporal, para objetos con velocidades mayores que este límite, el emborronamiento máximo que sufrirán será el debido a la reducción del ancho de banda espacial producido a esta velocidad, pero no más.

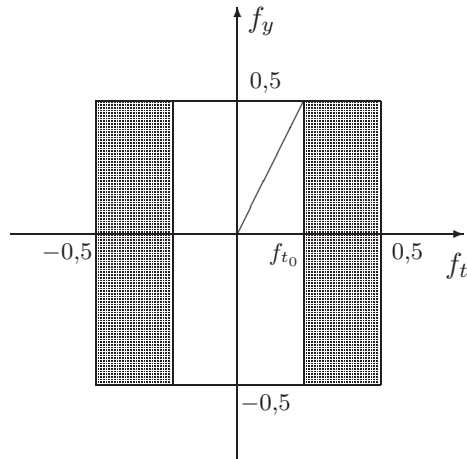


Figura 4.9: Corte por el plano $f_y - f_t$ de la respuesta en frecuencia de un filtro temporal.

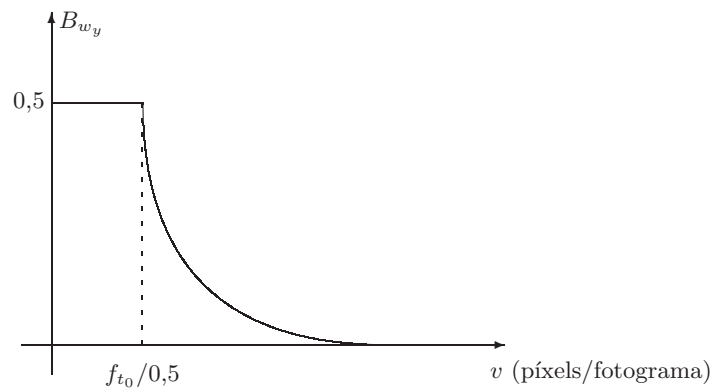


Figura 4.10: Ancho de banda espacial en función de la velocidad para el filtro temporal.

4.3.2. Implementación eficiente

El filtro espacio-temporal que vamos a implementar es el que, como hemos visto, tiene como respuesta en frecuencia la de la figura 4.4. Como no vamos a realizar el filtrado en frecuencia, sino que lo vamos a realizar mediante convolución, debemos obtener la respuesta impulsiva $h[x, y, t]$. Si observamos la respuesta tridimensional de la figura 4.4 ($H(e^{j\omega_x}, e^{j\omega_y}, e^{j\omega_t})$), vemos que es separable y que se puede obtener como combinación lineal de tres respuestas, de la siguiente forma:

$$H(e^{j\omega_x}, e^{j\omega_y}, e^{j\omega_t}) = H_1(e^{j\omega_x}, e^{j\omega_y}, e^{j\omega_t}) + H_2(e^{j\omega_x}, e^{j\omega_y}, e^{j\omega_t}) - H_3(e^{j\omega_x}, e^{j\omega_y}, e^{j\omega_t})$$

donde $H_1(e^{j\omega_x}, e^{j\omega_y}, e^{j\omega_t})$ es la respuesta en frecuencia de un filtrado sólo espacial, es decir, que no limita el ancho de banda temporal (constante en todas las frecuencias temporales), respuesta que se muestra en la figura 4.11; $H_2(e^{j\omega_x}, e^{j\omega_y}, e^{j\omega_t})$ es un filtrado sólo temporal con respuesta la de la figura 4.12 y $H_3(e^{j\omega_x}, e^{j\omega_y}, e^{j\omega_t})$ es un filtrado espacio-temporal cuya respuesta se muestra en la figura 4.13.

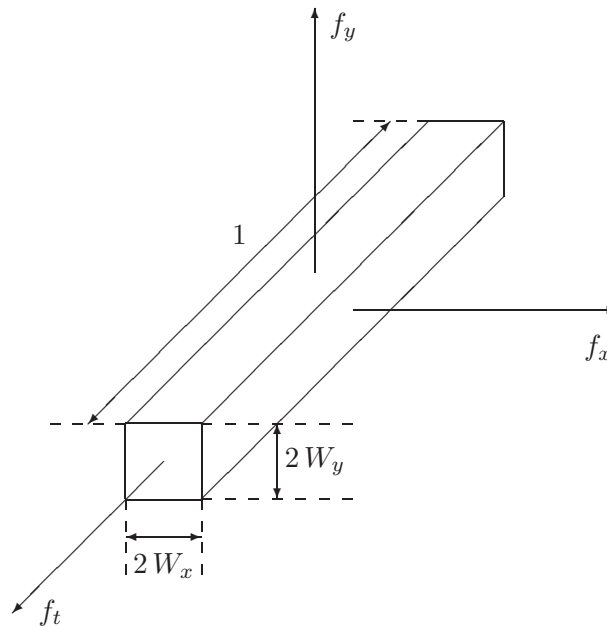


Figura 4.11: Respuesta en frecuencia del filtro espacial, $H_1(e^{j\omega_x}, e^{j\omega_y}, e^{j\omega_t})$.

Para obtener la respuesta impulsiva $h[x, y, t]$ habría que obtener $h_1[x, y, t]$, $h_2[x, y, t]$ y $h_3[x, y, t]$, que serían las transformadas de Fourier inversas de $H_1(e^{j\omega_x}, e^{j\omega_y}, e^{j\omega_t})$, $H_2(e^{j\omega_x}, e^{j\omega_y}, e^{j\omega_t})$ y $H_3(e^{j\omega_x}, e^{j\omega_y}, e^{j\omega_t})$, respectivamente. Estas respuestas son imposibles de implementar en la práctica, debido a sus bandas de transición abruptas, por lo que se implementa una aproximación eligiendo como $h_1[x, y, t]$, $h_2[x, y, t]$ y $h_3[x, y, t]$ las respuestas impulsivas dadas por las expresiones siguientes.

$$h_1[x, y, t] = h_e[x, y] \delta[t], \quad (4.5)$$

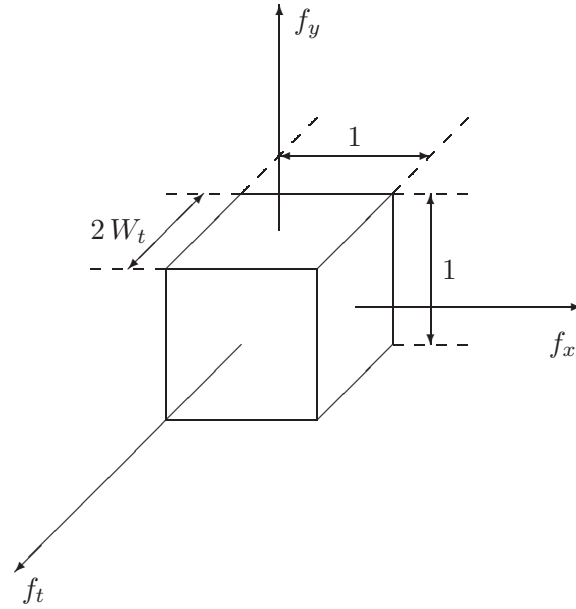


Figura 4.12: Respuesta en frecuencia del filtro temporal, $H_2(e^{j\omega_x}, e^{j\omega_y}, e^{j\omega_t})$.

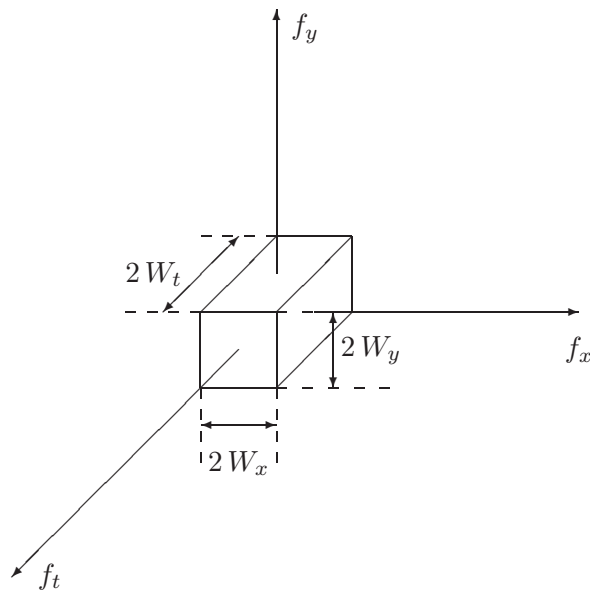


Figura 4.13: Respuesta en frecuencia del filtro espacio-temporal, $H_3(e^{j\omega_x}, e^{j\omega_y}, e^{j\omega_t})$.

siendo $h_e[x, y]$:

$$h_e[x, y] = \begin{cases} 1/(MN) & \text{si } |x| < \frac{M-1}{2} \text{ y } |y| < \frac{N-1}{2} \\ 0 & \text{otro caso} \end{cases},$$

lo que corresponde a un promediado espacial con una máscara de tamaño $M \times N$.

$$h_2[x, y, t] = h_t[t] \delta[x, y] \quad (4.6)$$

siendo $h[t]$:

$$h_t[t] = \begin{cases} 1/L & \text{si } |t| < \frac{L-1}{2} \\ 0 & \text{otro caso} \end{cases},$$

que es un promediado en la dimensión temporal con una máscara de tamaño L . Por último, la respuesta $h_3[x, y, t]$ viene dada por:

$$h_3[x, y, t] = h_1[x, y, t] * h_2[x, y, t], \quad (4.7)$$

que por la propiedad de separabilidad es:

$$h_3[x, y, t] = h_e[x, y] h_t[t]. \quad (4.8)$$

Es conveniente que los tamaños M , N y L sean impares, para que las máscaras tengan simetría par, y así poder obtener filtros de fase cero, los cuales no desplazan la salida con respecto a la entrada. Así pues, la respuesta impulsiva de nuestro filtro se obtendrá mediante:

$$h[x, y, t] = h_1[x, y, t] + h_2[x, y, t] - h_1[x, y, t] * h_2[x, y, t]. \quad (4.9)$$

Realmente, la respuesta que se obtiene es una aproximación de la respuesta en frecuencia mostrada en la figura 4.4, ya que las respuestas en frecuencia de cada filtro implementado no son paralelepípedos sino *sincs*. Sin embargo, esta aproximación es válida, ya que la mayor parte del ancho de banda del filtro obtenido, el que queda determinado por los primeros nulos de las *sincs*, es aproximadamente la respuesta buscada. En la figura 4.15 se muestra el plano $f_y - f_t$ de la respuesta real, y en la figura 4.14 el mismo plano para la respuesta ideal.

La implementación de la respuesta buscada se muestra en la figura 4.16 y consiste en la suma de dos promediados, uno temporal y otro espacial, y restando el resultado del promediado temporal en cascada con el espacial. Sin embargo, este filtro es todavía optimizable, teniendo en cuenta que $h_1[x, y, t]$ podemos separarla como $h_{1_x} * h_{1_y}$, donde $h_{1_x} = h_x[x] \delta[y]$ siendo $h_x[x]$:

$$h_x[x] = \begin{cases} 1/M & \text{si } |x| < \frac{M-1}{2} \\ 0 & \text{otro caso} \end{cases},$$

es decir, un promediado con una ventana de tamaño M en la dirección x ; y $h_{1_y} = h_y[y] \delta[x]$, donde $h_y[y]$:

$$h_y[y] = \begin{cases} 1/N & \text{si } |y| < \frac{N-1}{2} \\ 0 & \text{otro caso} \end{cases},$$

un promediado con una ventana de tamaño N en la dirección y .

Así pues, el filtro FIR lineal a implementar será muy eficiente, ya que todos los filtrados a realizar son unidimensionales. Los pasos a seguir en el proceso de filtrado para cada fotograma sería:

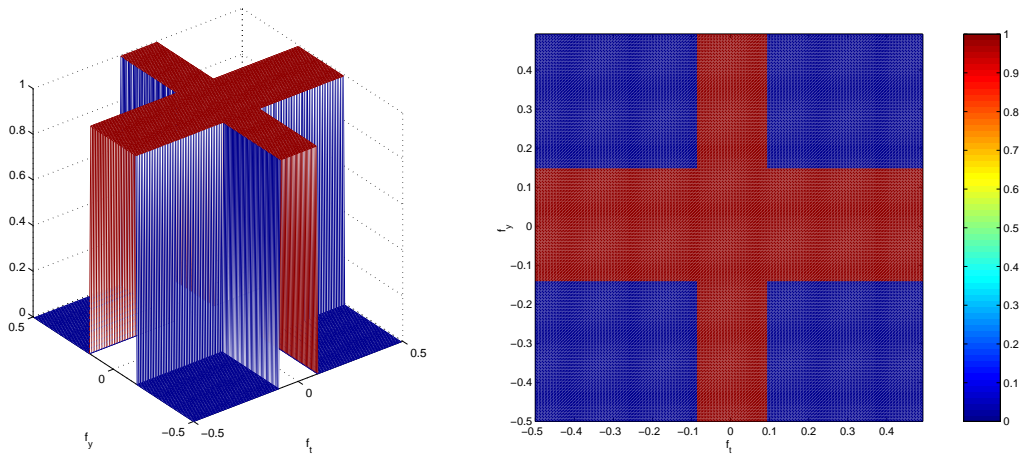


Figura 4.14: Plano $f_y - f_t$ de la respuesta ideal $H(e^{j\omega_x}, e^{j\omega_y}, e^{j\omega_t})$.

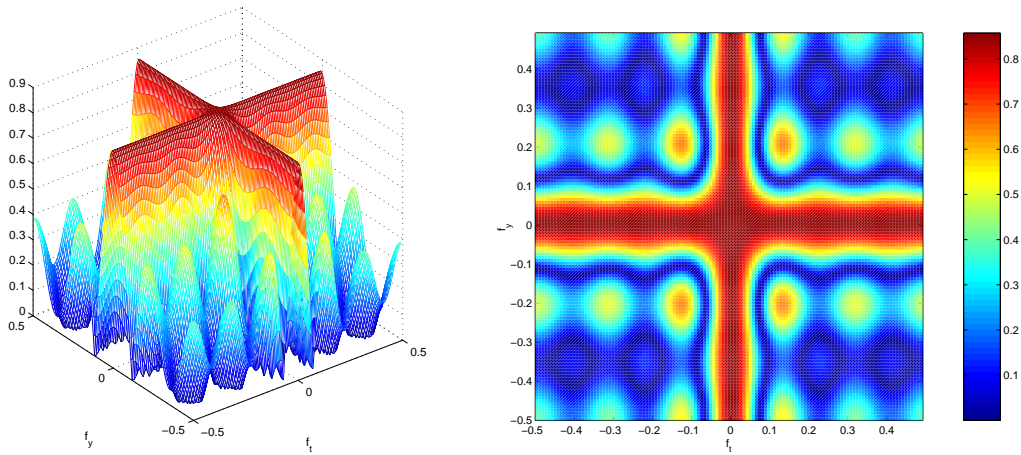


Figura 4.15: Plano $f_y - f_t$ de la respuesta real $TF\{h[x, y, t]\}$.

1. Filtrado espacial:
 - a) Promediado del fotograma en la dirección x .
 - b) Sobre el resultado anterior realizar el promediado en la dirección y .
2. Filtrado temporal: sobre el fotograma original se realiza un promediado temporal, es decir, para cada píxel se calcula el promedio de la intensidad con la de los píxeles que ocupan su misma posición espacial en los $(L - 1)/2$ fotogramas anteriores y los $(L - 1)/2$ posteriores (excepto en los bordes temporales, es decir al principio y al final de la secuencia, donde en vez de con $L - 1$ fotogramas vecinos se promedia con los que haya).
3. Sobre el resultado del filtrado espacial se realiza el filtrado temporal, de la forma indicada en el punto anterior, obteniéndose el filtrado espacio-temporal.

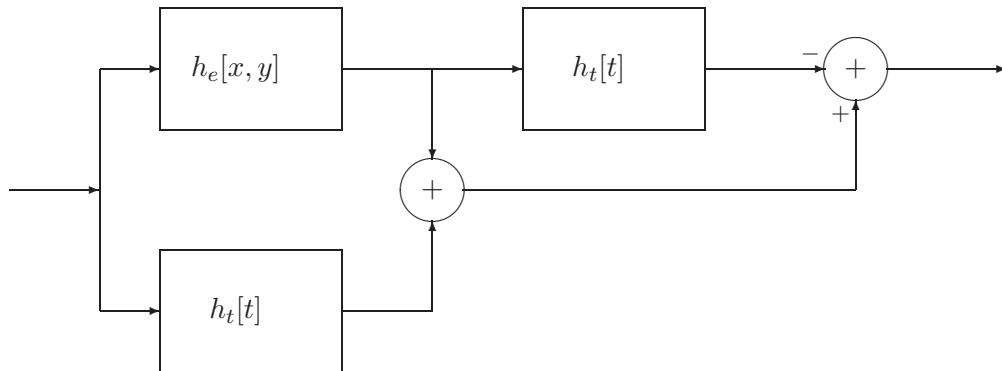


Figura 4.16: Proceso para filtrar la secuencia con $h[x, y, t]$.

4. Finalmente se realiza la suma de la intensidad de cada píxel de cada fotograma del resultado del filtrado espacial con la del píxel situado en la misma posición $[x, y, t]$ del resultado del filtrado temporal, y se resta con la del píxel $[x, y, t]$ del filtrado espacio-temporal.

Efectos de borde

Al implementar el filtrado espacio-temporal de la forma descrita anteriormente, en los bordes de la secuencia, tanto espaciales como temporales, se produce un transitorio en el filtrado. Las zonas donde se produce dicho transitorio son aquellas donde la máscara no cabe íntegramente al filtrar un determinado píxel.

- Bordes espaciales: la figura 4.17 muestra la zona que representa un borde debido al filtrado espacial con una máscara de tamaño $M \times N$; dicha zona la formarán las $\frac{M-1}{2}$ primeras y últimas filas de cada imagen y las $\frac{N-1}{2}$ primeras y últimas columnas, ya que cuando vamos a filtrar los píxeles de esas filas y columnas la máscara sale de la imagen. Así pues, las $\frac{M-1}{2}$ primeras y últimas filas y $\frac{N-1}{2}$ primeras y últimas columnas de cada imagen de la secuencia el filtrado no se implementará exactamente como en el resto de la imagen. En estas zonas el promediado espacial se realizará sólo de los píxeles de la ventana de promediado que están dentro de la imagen.
- Bordes temporales: la figura 4.18 muestra la zona que representa un borde debido al filtrado temporal con una máscara de tamaño L . El borde lo formarán las $\frac{L-1}{2}$ primeras y últimas imágenes de la secuencia. Para cada píxel de esas imágenes que forman los bordes temporales, en el promedio temporal sólo participarán los píxeles de las imágenes que están dentro de la máscara, por lo que tendrá más peso la imagen a filtrar en un borde temporal que en el resto de la secuencia. Por ejemplo, supongamos que vamos a filtrar una secuencia de N fotogramas utilizando una máscara temporal de tamaño 5 ($L = 5$), los bordes temporales serán las imágenes $I_0, I_1, I_{N-2}, I_{N-1}$, para cada una de ellas en el promediado temporal participarán las siguientes imágenes¹:

¹Para el filtrado temporal que sigue al espacial (figura 4.16) hay que tener en cuenta que las imágenes de

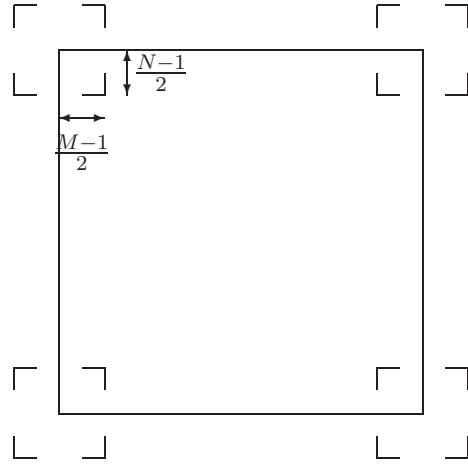


Figura 4.17: Bordes espaciales del filtrado espacio-temporal.

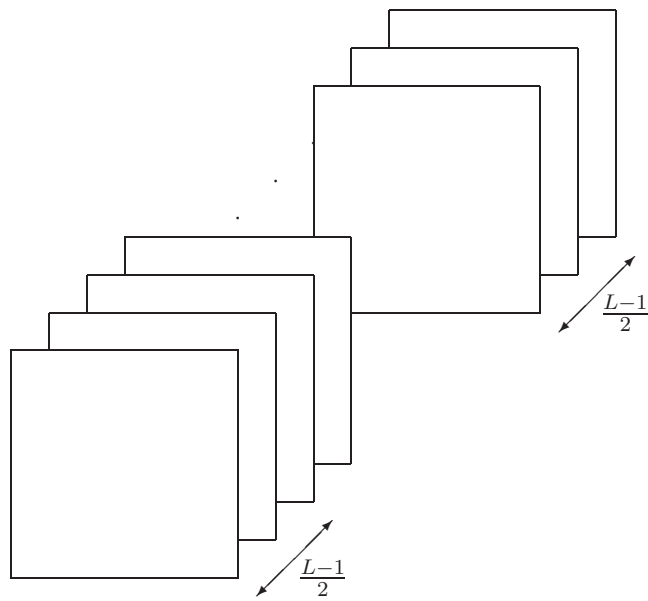


Figura 4.18: Bordes temporales del filtrado espacio-temporal.

- Resultado del filtrado temporal para I_0 : I_0 .
- Resultado del filtrado temporal para I_1 : promediado de I_0, I_1, I_2 .
- Resultado del filtrado temporal para I_{N-2} : promediado de $I_{N-3}, I_{N-2}, I_{N-1}$.
- Resultado del filtrado temporal para I_{N-1} : I_{N-1} .

La máscara, incluso en los bordes temporales, debe ser simétrica y es por ello por lo que en los bordes temporales el filtrado se implementa de la forma especificada. Esta simetría de la máscara debe mantenerse para que, en las operaciones que se llevan a cabo entre imágenes filtradas para implementar el filtrado espacio-temporal (figura 4.16), los retardos sean iguales.

Parámetros de diseño del filtro

A la hora de filtrar espacio-temporalmente una secuencia de imágenes, el usuario podrá controlar dos aspectos del filtrado los cuales vienen condicionados por los parámetros que el usuario puede seleccionar: el tamaño del filtrado temporal L y el tamaño del filtrado espacial $M \times N$. Estos aspectos aparecen representados en la figura 4.8, y son:

1. A partir de qué velocidad los objetos empezarán a emborronarse: $f_{t_0}/0,5$, valor que viene condicionado por el tamaño del filtrado temporal L . A medida que L crece f_{t_0} disminuye, y por tanto, la velocidad que debe tener un objeto para ser emborronado será menor.
2. El otro aspecto controlable es el emborronamiento máximo producido, que está relacionado con f_{x_0} , que es el ancho de banda espacial mínimo que filtrará frecuencias espaciales con velocidades mayores o iguales a f_{t_0}/f_{y_0} . Este valor variará al cambiar el tamaño del filtrado espacial $M \times N$. Si aumentamos el tamaño de la máscara espacial f_{x_0} disminuirá, aumentando el emborronamiento máximo permitido.

Así pues, variando el valor de L controlaremos la velocidad máxima a la que un objeto pasa intacto por el filtro; y con el valor de $M \times N$ estableceremos el emborronamiento máximo permitido en el filtrado espacio-temporal, sea cual sea su velocidad.

4.3.3. Cálculo del factor de reducción de ruido

Vamos a determinar el factor de reducción de la varianza del ruido que se consigue filtrando espacio-temporalmente con el filtro diseñado en el apartado anterior. La señal a filtrar se suele modelar como ruido blanco de media nula que se distribuye en las tres dimensiones, x , y y t , al cual denotaremos como $n[x, y, t]$. Siendo la potencia de ruido de entrada al filtro σ_i^2 , la potencia a la salida será:

$$\sigma_o^2 = \sigma_i^2 \sum_x \sum_y \sum_t |h[x, y, t]|^2, \quad (4.10)$$

donde $h[x, y, t]$ es la respuesta impulsiva del filtro 3D espacio-temporal. Como hemos visto en el apartado anterior, este filtrado se implementa de forma separable según la ecuación 4.9,

entrada son las filtradas espacialmente.

en la que $h_1[x, y, t]$ es una máscara cuadrada, un plano en las dos dimensiones espaciales, de tamaño $M \times N$ donde todos los elementos valen $1/(MN)$, $h_2[x, y, t]$ es una máscara lineal en la dirección temporal de tamaño L donde cada elemento vale $1/L$ y el filtrado espacial seguido del temporal supondría el filtrado con una máscara cúbica de tamaño $M \times N \times L$ donde cada elemento vale $1/(MNL)$.

El factor de reducción de ruido se define como el cociente entre la potencia de ruido de salida y la potencia de ruido de entrada, que valdrá, despejando de la ecuación 4.10:

$$f.r.r. = \sum_x \sum_y \sum_t |h[x, y, t]|^2. \quad (4.11)$$

Sustituyendo la ecuación 4.9 en 4.11 se obtiene la expresión final que nos va a permitir calcular el factor de reducción de ruido:

$$f.r.r. = \sum_x \sum_y \sum_t |h_1[x, y, t] + h_2[x, y, t] - h_1[x, y, t] * h_2[x, y, t]|^2. \quad (4.12)$$

Para obtener este factor vamos a suponer que la máscara espacial es cuadrada ($M = N$). Además vamos a descomponer el cálculo, el sumatorio de la ecuación 4.12, en varias sumas parciales, cada una de ellas contendrá las contribuciones de los elementos que pertenecen a cada filtro. Así tendremos elementos que sólo están en el cubo ($h_1 * h_2$), otros que están en el plano (h_1) y el cubo, los que están en la línea (h_2) y el cubo, y finalmente los que están en el cubo, en la línea y en el plano.

- Elementos que están en el plano y en el cubo, son los del plano menos el del origen:

$$(N^2 - 1) \left(\frac{1}{N^2} - \frac{1}{N^2 L} \right)^2.$$

- Elementos que están en la línea y en el cubo, son los de la línea menos el del origen:

$$(L - 1) \left(\frac{1}{L} - \frac{1}{N^2 L} \right)^2.$$

- Elementos que están en el plano, en la línea y en el cubo, es decir, el del origen:

$$\left(\frac{1}{N^2} + \frac{1}{L} - \frac{1}{N^2 L} \right)^2.$$

- Elementos que están sólo en el cubo, es decir, todos los elementos menos los de la línea, los del plano y el del origen:

$$(N^2 L - (N^2 - 1) - (L - 1) - 1) \left(-\frac{1}{N^2 L} \right)^2.$$

El factor de ruido será la suma de todas las contribuciones anteriores, expresión que simplificando quedará de la siguiente forma:

$$f.r.r. = \frac{L + N^2 - 1}{N^2 L} = \frac{1}{N^2} + \frac{1}{L} - \frac{1}{N^2 L}. \quad (4.13)$$

La figura 4.19 representa las curvas de nivel del factor de reducción en decibelios en función de las dimensiones de las máscaras espacial (N) y temporal (L). En ella puede observarse que si no se filtra espacial y temporalmente ($N = 1$ o $L = 1$) el filtro no tiene ningún efecto, y el factor de reducción vale 0 dBs, es decir, que el ruido es igual a la entrada que a la salida del filtro. A medida que aumentamos N y L el factor de reducción de ruido va disminuyendo.

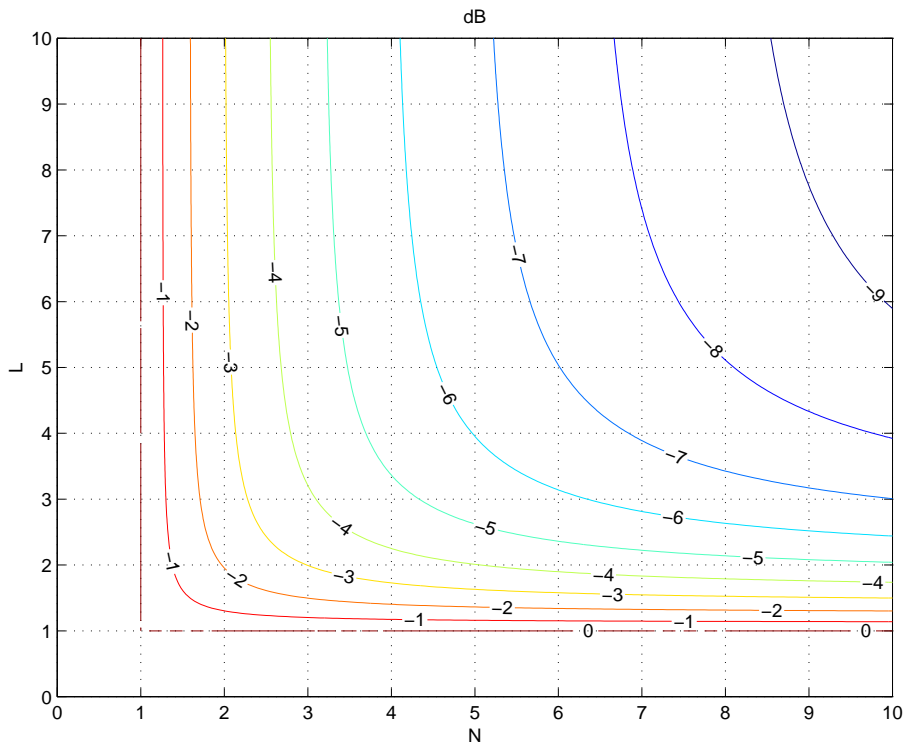


Figura 4.19: Factor de reducción de ruido en función del tamaño del filtrado.

En el caso de un filtrado temporal con una máscara de promediado de tamaño L , el factor de reducción según la ecuación 4.11, será:

$$f.r.r. = \sum_{-L/2}^{L/2-1} \frac{1}{L^2} = \frac{1}{L}, \quad (4.14)$$

y para un filtrado espacial de tamaño $N \times N$:

$$f.r.r. = \sum_{-M/2}^{M/2-1} \sum_{-N/2}^{N/2-1} \frac{1}{N^4} = \frac{1}{N^2}. \quad (4.15)$$

Comparando las ecuaciones 4.13 (factor de reducción para el filtro espacio-temporal) y 4.14 (factor de reducción para el filtro temporal), se puede comprobar que para el filtro temporal dicho factor es menor que para uno espacio-temporal con la misma L , por lo que reduce más ruido, resultado que ya se apuntaba en el apartado anterior. Sin embargo, hay que recordar que el filtro temporal emborrona más la imagen cuando existe movimiento. Lo mismo ocurre con el caso de un filtrado espacial (ecuación 4.15), en el que la reducción de ruido es mayor, pero el emborronamiento, esta vez en el caso de imagen estática, es mayor para el filtro espacial que para el espacio-temporal diseñado. En las tablas 4.1 y 4.2 se presentan los resultados del factor de reducción para valores típicos de máscaras temporales y espaciales, y se comparan con la reducción conseguida para filtrados espaciales o temporales del mismo tamaño. En la primera tabla se obtiene la reducción de ruido de nuestro filtro para un tamaño de máscara temporal fijo de valor 3 y diferentes tamaños de máscara espacial; y se compara con el que nos daría un filtro espacial de tamaño variable. Para ambos filtros el factor de reducción disminuye a medida que aumentamos la máscara espacial (o sea, se reduce más el ruido). Para el caso del filtrado espacial el factor es mucho más pequeño que para el espacio-temporal, a igualdad de tamaño de máscara, pero la degradación visual que presentará la imagen (figura 4.20) será mucho mayor. En la tabla 4.2 lo que se mantiene fijo es el filtrado espacial y se va aumentando el temporal, consiguiendo cada vez una reducción mayor y significativa. Comparada dicha reducción con el caso de filtrado temporal, vemos que la mejora que consigue éste con respecto a nuestro filtro no es tan grande, mientras que si observamos la degradación visual introducida, en el caso de movimiento, es muy notable (figura 4.20).

$M \times N$	$f.r.r_{e-t}$	$f.r.r._e$
3×3	-3,89	-9,54
5×5	-4,44	-13,98
7×7	-4,59	-16,9
9×9	-4,66	-19,08

Cuadro 4.1: Comparación del factor de reducción de ruido (en dB): filtro espacio-temporal $M \times N \times 3$ y filtro espacial $M \times N$.

L	$f.r.r_{e-t}$	$f.r.r._t$
3	-4,44	-4,77
5	-6,34	-6,98
7	-7,51	-8,45
9	-8,33	-9,54

Cuadro 4.2: Comparación del factor de reducción de ruido (en dB): filtro espacio-temporal $5 \times 5 \times L$ y filtro temporal L .

4.3.4. Resultados obtenidos

Se han realizado experimentos con secuencias ruidosas sintéticas, es decir, secuencias a las que se les ha introducido ruido de las mismas características que se suponía en el modelo; además se han realizado pruebas con secuencias reales ruidosas.

La figura 4.20 muestra el resultado de realizar tres filtrados sobre una secuencia sintética: el filtrado espacio-temporal, el espacial del mismo tamaño que la máscara espacial y el temporal del mismo tamaño que la máscara temporal del filtro espacio-temporal. El resultado del filtrado espacial es el que presenta mayor grado de emborronamiento ya que emborrona por igual la parte estática de la imagen (asiento del coche) como la parte en movimiento (personaje o fondo que pasa por la ventanilla). En los otros dos filtrados, la parte estática no se emborrona nada, mientras que la zona en movimiento (personaje) sufre un emborronamiento temporal que es mucho más acusado en el caso del filtro temporal, aunque por el contrario se elimina más ruido.



Figura 4.20: Comparación entre filtrados: (a)Secuencia original ruidosa: carphone+ruido. (b)Filtrado espacio-temporal ($M \times N = 3 \times 3$ y $L = 9$). (c)Filtrado temporal ($L = 9$). (d)Filtrado espacial ($M \times N = 3 \times 3$). Para comparar el resultado del filtrado espacio-temporal (b) y el temporal (c) podemos fijarnos en la cortinilla del fondo, que es estática, y en ambos resultados está nítida, mientras que la cara del personaje, que es una zona en movimiento es la que está emborronada, mucho más en el caso del filtrado temporal (c). Al comparar ahora con el filtrado espacial, vemos que éste emborrona por igual zonas estáticas (cortinilla) y dinámica (cara), y este emborronamiento es el máximo que introduce el filtrado espacio-temporal (caso cuando existe movimiento); esto puede comprobarse en la cara del personaje, que para ambos filtros (espacial y espacio-temporal) está igual de borrosa.

Las figuras 4.21 y 4.22 muestran el resultado del filtrado con una máscara $5 \times 5 \times 5$ del

plano mostrado en la figura 4.1, un plano de una película real degradada. En ella aparece, en la parte izquierda, la imagen original ruidosa y a la derecha, la filtrada. Puede observarse que el ruido se ha reducido notablemente, emborronando sólo aquello que se mueve. Sin embargo hay ciertas manchas, como en la primera imagen de la figura 4.22, que no han sido eliminadas por el filtrado. Estas manchas son ruido no estacionario en la dimensión temporal, pero con un tamaño espacial grande.

Nuestro filtro eliminará todo ruido o mancha que tenga menor tamaño espacial que la mitad del tamaño de la máscara espacial y que dure como mucho, en el tiempo, la mitad de la máscara temporal. Para eliminar este tipo de manchas incorreladas en la dimensión temporal, que el filtro no elimina por ser demasiado grandes espacialmente, hay que recurrir a métodos de restauración de muestras perdidas, que veremos en el capítulo 5.

4.4. Conclusiones

El filtro diseñado en esta tesis, y que ha sido presentado en este capítulo, consigue el objetivo propuesto: eliminar ruido intentando conseguir que la degradación introducida no sea muy visible por el espectador. Se trata de un filtro lineal y FIR implementado de manera muy eficiente y basado en las propiedades del sistema visual humano, que reduce ruido por un factor considerable, emborronando sólo aquellos bordes espaciales donde el ojo no lo aprecie, en las zonas con movimiento. El factor de reducción conseguido es menor que el que obtiene un filtro espacial o temporal del mismo tamaño, pero también degrada menos la secuencia. Dicho factor es comparable al obtenido por otros autores [25], pero la carga computacional es menor.

Sin embargo, este filtro no elimina las manchas, es decir ruido incorrelado temporalmente, con dimensiones mayores que las de la máscara espacial utilizada. Para solucionar este problema se implementa un método de restauración de este tipo de defectos, al cual se dedica el capítulo siguiente de esta tesis.



(a)

(b)

Figura 4.21: Secuencia *Taronja* (a)Original (b)Resultado filtro espacio-temporal ($M \times N = 5 \times 5$ y $L = 5$).



Figura 4.22: Secuencia *Taronja* (a)Original (b)Resultado filtro espacio-temporal ($M \times N = 5 \times 5$ y $L = 5$). En el primer fotograma se observa que la mancha original no ha desaparecido después del filtrado.

Capítulo 5

Defectos incorrelados en el tiempo

5.1. Introducción

Las manchas son defectos que aparecen en las películas antiguas frecuentemente. La pérdida de gelatina y las partículas de polvo que cubren la película hacen que la intensidad de ciertas zonas desaparezca, traducándose dicha pérdida en una mancha, que será clara si la película perdió gelatina en esa zona, y oscura si se ha depositado sobre la película una partícula de polvo u otra sustancia como hongos debidos a la humedad. Este defecto tiene la particularidad de que esta incorrelado en el tiempo, es decir, si aparece una mancha en un fotograma lo más probable es que el mismo defecto no aparezca en el fotograma anterior y siguiente. Gracias a esta propiedad se facilita la restauración de estos datos perdidos.

El proceso de restauración de manchas tiene dos fases: detección de manchas e interpolación de las mismas. El objetivo de la primera fase será obtener una máscara para cada fotograma que determine qué píxeles corresponden a la mancha o manchas que presenta dicha imagen. En la segunda fase, la intensidad de estos píxeles es restaurada a partir del vecindario dentro del mismo fotograma, pero principalmente con información de los fotogramas vecinos. Los métodos propuestos en este capítulo para la detección e interpolación de manchas se basan en las características del defecto que se pretende restaurar:

- Incorrelación a lo largo del tiempo.
- Alto contraste en intensidad con el vecindario.
- Poca varianza.

Es decir, se trata de manchas más claras que su entorno o manchas oscuras en un vecindario más claro, uniformes y que varían rápidamente con el tiempo. Basándonos en estas propiedades y en algoritmos de morfología matemática presentaremos una serie de métodos que permitirán minimizar el efecto de estas manchas en la visualización de la película sin necesidad de estimar el movimiento en la secuencia, proceso que se complica cuando en la secuencia faltan datos, como es el caso que nos ocupa. Tanto las bases como el desarrollo de nuestros métodos se presentarán en la sección 5.3, donde también se presenta el proceso de detección de manchas. En la sección 5.4 se muestra la utilización de los métodos propuestos

en la sección anterior para la interpolación de defectos aleatorios. Previamente se hará un repaso por las técnicas de restauración de defectos aleatorios existentes, en la sección 5.2. Finalmente se presentarán una serie de resultados 5.5 y conclusiones 5.6.

En la figura 5.1 aparecen tres fotogramas consecutivos de un mismo plano de una secuencia, en los que se puede observar cómo es el defecto que se pretende restaurar. Vemos que se trata de una mancha más clara que el fondo donde se encuentra (alto contraste), que no aparece en los fotogramas vecinos en el tiempo (incorrelación temporal), y cuya intensidad se mantiene constante en toda la extensión de la mancha (poca varianza espacial).



Figura 5.1: Fotogramas consecutivos de un plano de la secuencia *Taronja*. Mancha en el fotograma central.

5.2. Estado del arte

Los métodos para restauración de manchas que se van a presentar asumen el siguiente modelo de degradación:

$$z(x, y, t) = (1 - d(x, y, t)) y(x, y, t) + d(x, y, t) r(x, y, t),$$

donde $z(x, y, t)$ e $y(x, y, t)$ son las secuencias observada y original, respectivamente. $d(x, y, t)$ es la máscara binaria que indica, para cada fotograma, si un píxel está corrupto ($d(x, y, t) = 1$) o no ($d(x, y, t) = 0$). La intensidad de los píxeles erróneos viene dada por la señal $r(x, y, t)$. La señal restaurada vendrá dada por la siguiente expresión:

$$\hat{y}(x, y, t) = (1 - d(x, y, t)) z(x, y, t) + d(x, y, t) c(x, y, t),$$

siendo $z(x, y, t)$ la secuencia observada y $d(x, y, t)$ es la máscara de manchas. Para los píxeles erróneos se restaurará su intensidad con $c(x, y, t)$, siendo $c(x, y, t) \neq y(x, y, t)$, y los píxeles en buen estado mantendrán su información original, $y(x, y, t)$. Por tanto, el objetivo de un método de detección e interpolación de manchas será obtener $d(x, y, t)$ en la fase de detección y $c(x, y, t)$ en la de interpolación.

A continuación se presentará una revisión de las técnicas publicadas de detección e interpolación de manchas. La base de dicho resumen puede encontrarse en [122].

5.2.1. Técnicas de detección de manchas

El objetivo de la fase de detección, como hemos dicho anteriormente, consiste en determinar una máscara binaria para cada fotograma, $d(x, y, t)$, que indique con $d(x, y, t) = 1$ que dicho píxel pertenece a una mancha. El dato de partida para obtener $d(x, y, t)$ es la secuencia degradada $z(x, y, t)$.

▪ Detectores basados en diferencias entre píxeles

• Detector $SDIa$

Índice a de detección de picos (*Spike detection index-a*). Se trata de un método heurístico para detectar discontinuidades temporales en la intensidad de la imagen. Presentado por Kokaram en [83, 89, 84], compara la intensidad de cada píxel en el fotograma actual con la del píxel situado en la misma posición en el fotograma siguiente y anterior compensados en movimiento (z_{mc}), calculando las mínimas diferencias cuadráticas:

$$SDIa(x, y, t) = \min[(z(x, y, t) - z_{mc}(x, y, t + 1))^2, (z(x, y, t) - z_{mc}(x, y, t - 1))^2].$$

Valores altos del índice $SDIa(x, y, t)$ indican discontinuidades grandes en la intensidad de la imagen en ambas direcciones temporales, hacia delante y hacia detrás. Así pues, un píxel pertenecerá a una mancha si $SDIa(x, y, t)$ supera un determinado umbral T_1 :

$$d_{SDIa}(x, y, t) = \begin{cases} 1 & \text{Si } SDIa(x, y, t) > T_1 \\ 0 & \text{otro caso} \end{cases}$$

donde $T_1 \geq 0$ es seleccionado por el usuario, y establece la sensibilidad del detector. Si T_1 aumenta, aumenta la probabilidad de detección pero también la de falsa alarma. Para reducir las falsas alarmas debidas al ruido, existe una modificación de este detector, el $SDIp$, que además exige que los signos de las diferencias $z(x, y, t) - z_{mc}(x, y, t + 1)$ y $z(x, y, t) - z_{mc}(x, y, t - 1)$ sean iguales para decidir que se ha detectado una mancha.

• Detector de tres diferencias

Schallauer *et al.* [127] calculan tres diferencias entre el fotograma actual y el anterior y el siguiente compensados en movimiento y sólo deciden que un píxel pertenece a una mancha si las tres diferencias superan ciertos umbrales. Analíticamente:

$$d_{-1}(x, y, t) = |z(x, y, t) - z_{mc}(x, y, t - 1)|,$$

$$d_{+1}(x, y, t) = |z(x, y, t) - z_{mc}(x, y, t + 1)|$$

y

$$d_{\pm 1}(x, y, t) = |z_{mc}(x, y, t + 1) - z_{mc}(x, y, t - 1)|.$$

La máscara se construirá mediante la expresión siguiente:

$$d(x, y, t) = \begin{cases} 1 & \text{Si } d_{-1}(x, y, t) > T_1 \text{ y } d_{+1}(x, y, t) > T_1 \text{ y } d_{\pm 1}(x, y, t) > T_2 \\ 0 & \text{otro caso} \end{cases}$$

Rosenthaler [123] propone también un algoritmo basado en las diferencias d_{-1} y d_{+1} , pero previamente divide las imágenes en bloques, y además tiene en cuenta la varianza de la luminosidad para cada píxel, que en el caso de pertenecer a una mancha debe ser más pequeña. Una vez obtenida la máscara se segmenta, etiquetando cada píxel como perteneciente a una mancha (objeto) u otro, con el fin de eliminar píxeles aislados (falsas alarmas debidas al ruido).

- **Detector de continuidad**

Saito *et al.* [125, 133] proponen un detector de manchas que detectan la discontinuidad espacio-temporal de la intensidad que supone este tipo de defectos. Primero calculan un parámetro de continuidad robusto frente al movimiento:

$$C(x, y, t) = \text{mín} (|d_f(k', l')|, |d_b(k', l')|),$$

siendo

$$d_f(k', l') = z(x + k', y + l', t + 1) - z(x, y, t),$$

y

$$d_b(k', l') = z(x, y, t) - z(x - k', y - l', t - 1),$$

donde (k', l') representa el vecindario del píxel actual. Este vecindario tiene un tamaño espacial de 5×5 centrado en dicho píxel. Con esto se evitan estimar el movimiento pero discriminan un objeto en movimiento de una mancha, ya que el primero, en la imagen anterior o siguiente, debe encontrarse en dicho vecindario. A continuación calculan la probabilidad de incorrupción $\mu(x, y, t)$ en función de $C(x, y, t)$ que es la probabilidad de que un píxel pueda no ser corrompido por una mancha. Finalmente restauran la secuencia siguiendo la expresión:

$$\hat{y}(x, y, t) = M(x, y, t) + \mu(x, y, t) [z(x, y, t) - M(x, y, t)],$$

donde $M(x, y, t)$ es un filtro de mediana para el que proponen varias soluciones.

- **Detector basado en estadísticos ordenados: ROD**

Diferencias ordenadas por rango (*Rank-ordered differences*). Nadeau y Mitra [102] proponen este detector, también heurístico, basado en estadísticos ordenados. El método consiste en lo siguiente:

- Para cada píxel $z(x, y, t)$ se seleccionan seis vecinos $k = 1..6$: $p_k = \{z_{mc}(x, y, t - 1), z_{mc}(x, y - 1, t - 1), z_{mc}(x, y + 1, t - 1), z_{mc}(x, y, t + 1), z_{mc}(x, y - 1, t + 1), z_{mc}(x, y + 1, t + 1)\}$.
- Este conjunto de píxeles se ordenan de menor a mayor, obteniendo el conjunto ordenado r_k con $k = 1, 2, \dots, 6$.
- Se calculan tres diferencias de la siguiente manera:

$$ROD_l(x, y, t) = \begin{cases} r_l - z(x, y, t) & \text{si } z(x, y, t) \leq r_{mean} \\ z(x, y, t) - r_{7-l} & \text{si } z(x, y, t) > r_{mean} \end{cases} \quad \text{con } l = 1, 2, 3,$$

donde r_{mean} es la media de los valores ordenados:

$$r_{mean} = \frac{r_3 + r_4}{2}$$

Se detecta una mancha si al menos una de las tres diferencias supera su umbral correspondiente T_l :

$$d_{ROD}(x, y, t) = \begin{cases} 1 & \text{si } ROD_l(x, y, t) > T_l \\ 0 & \text{otro caso} \end{cases} \quad \text{con } 0 \leq T_1 \leq T_2 \leq T_3 \text{ y } l = 1, 2, 3.$$

Los umbrales T_l , como en los métodos anteriores, son elegidos por el usuario y determinan la sensibilidad del detector.

Roosmalen *et al.* [122, 116] simplifican este detector haciendo $T_2, T_3 \rightarrow \infty$, obteniendo de esta forma el detector S-ROD, cuya función de transferencia es:

$$d_{S-ROD}(x, y, t) = \begin{cases} \min(p_k) - z(x, y, t) & \text{si } \min(p_k) - z(x, y, t) > 0 \\ z(x, y, t) - \max(p_k) & \text{si } z(x, y, t) - \max(p_k) < 0 \\ 0 & \text{otro caso} \end{cases} .$$

Como puede observarse en la expresión anterior, este detector no necesita ordenar los valores de los píxeles vecinos del actual, reduciéndose de esta forma el coste computacional. Además, Roosmalen presenta un método de postprocesado compuesto por tres pasos para disminuir el número de falsas alarmas sin bajar la probabilidad de detección del detector S-ROD. La primera operación de postprocesado detecta y elimina las posibles falsas alarmas determinando la probabilidad de que un píxel haya sido detectado como mancha erróneamente debido al ruido. El segundo paso consiste en reconstruir las manchas que han sido detectadas sólo parcialmente. Por último, se rellenan pequeños huecos en las manchas detectadas. La clave del método propuesto por estos autores es usar un detector píxel a píxel (el S-ROD) seguido de un postprocesado que considera las manchas como objetos.

■ Detector basado en campos aleatorios de Markov: MRF

Este método de detección, propuesto por Kokaram [83, 89, 84], está basado en campos aleatorios de Markov y define una función de probabilidad a posteriori para un mapa de oclusiones binario. Dicho mapa indica dónde los objetos presentes en el fotograma actual son visibles en un fotograma de referencia. La máscara de oclusiones se determina maximizando la siguiente función de probabilidad:

$$P[d_k(x, y, t) | z(x, y, t), z_{mc}(x, y, t+k)] \propto P[z(x, y, t) | d_k(x, y, t), z_{mc}(x, y, t+k)] P[d_k(x, y, t)].$$

Una mancha es detectada cuando se detectan oclusiones en las dos direcciones temporales; hacia delante y hacia atrás, $k = 1$ y $k = -1$. Así pues, la máscara viene determinada por:

$$d_{MRF}(x, y, t) = \begin{cases} 1 & \text{si } d_1(x, y, t) = 1 \text{ y } d_{-1}(x, y, t) = 1 \\ 0 & \text{otro caso.} \end{cases}$$

Chong *et al.* [35] reducen el número de falsas alarmas introduciendo en la formulación de la probabilidad el resultado de un detector de ejes que permita diferenciar las discontinuidades temporales debidas al movimiento de las producidas por la existencia de una mancha.

▪ **Detector basado en un modelo AR**

La hipótesis en la que se basa este detector es que las imágenes originales, sin defecto, se pueden modelar como un proceso AR, y pueden estimarse bien a partir de los fotogramas compensados en movimiento precedentes o/y siguientes [83, 89, 84]. Así, la secuencia puede ser modelada como:

$$\begin{aligned} z(x, y, t) &= \sum_{l=1}^n a_l z_{mc}(x + q_{lx}, y + q_{ly}, t + k) + e(x, y, t + k) \\ &= \hat{z}_k(x, y, t) + e(x, y, t + k), \end{aligned}$$

donde a_l son los coeficientes del modelo AR estimados de la secuencia observada, (q_{k_x}, q_{k_y}) denota la posición relativa de los píxeles de referencia con respecto al píxel actual, y $e(x, y, t + k)$ es el error de predicción.

En ausencia de manchas y oclusión, el error de predicción es muy pequeño. Una mancha es detectada cuando el error excede un determinado umbral, T_1 , en las dos direcciones, hacia delante ($k = 1$) y hacia atrás ($k = -1$):

$$d_{AR}(x, y, t) = \begin{cases} 1 & \text{si } e^2(x, y, t + 1) > T_1 \text{ y } e^2(x, y, t - 1) > T_1 \\ 0 & \text{otro caso.} \end{cases} \quad \text{con } T_1 \geq 0.$$

▪ **Detectores morfológicos**

Tenze *et al.* [144] observando la propiedad de las manchas de que son mucho más claras que el fondo o mucho más oscuras, es decir, que son mínimos o máximos de la imagen, proponen utilizar como detector de manchas la combinación (OR) de dos operaciones morfológicas sobre el fotograma a restaurar, que detecte precisamente estas regiones máximas:

$$D_{dark} \ OR \ D_{bright},$$

donde D_{dark} es el residuo del cierre de la imagen original con un elemento estructurante y dicha imagen, lo que permite obtener los mínimos (zonas oscuras) de la imagen en los que no cabe el elemento estructurante. D_{bright} es el operador dual, es decir, el residuo de la apertura, que obtiene las zonas claras de la imagen.

Este detector tiene el problema de que todos aquellos máximos o mínimos de la imagen los etiquetará como manchas, aunque no lo sean, es decir, un objeto mucho más luminoso que el fondo será restaurado. Además, el éxito de la detección dependerá de la forma y tamaño de la mancha y del elemento estructurante. Este problema surge al no haber aprovechado la segunda propiedad, muy importante, que tienen las manchas, su incorrelación temporal. En nuestro método, que presentaremos en la siguiente sección, también se utilizan operadores morfológicos que tienen en cuenta este hecho.

E. Decènciere [52] propone operadores morfológicos que tienen en cuenta la propiedad de aleatoriedad temporal, y también ciertas propiedades espaciales de las manchas, como su área, pero con esto se presupone que la mancha tiene una determinada forma y tamaño, lo cual en muchos casos no se cumple. Además el método presenta numerosas falsas alarmas debido al movimiento de objetos o cámara.

Pardàs *et al.* [109] utilizan la reconstrucción geodésica, operador morfológico conexo, para eliminar ruido impulsivo en secuencias de televisión.

Finalmente, Harvey y Marshall [61] proponen utilizar operadores morfológicos suaves, que son operadores que tienen la misma filosofía que los morfológicos pero están basados no sólo en operaciones de máximos y mínimos sino en otras de estadísticos ordenados, como la mediana.

5.2.2. Técnicas de interpolación de manchas

Existen multitud de técnicas de restauración de zonas pérdidas en imágenes aisladas, de las cuales un resumen puede encontrarse en [113]. Estas técnicas se basan solamente en información espacial, ya que sólo se dispone de una imagen a restaurar; Veldhuis, por ejemplo, propone utilizar modelos AR [150], Albiol y Prades [9] presentan un algoritmo basado en procesamiento priorizado para interpolar manchas de cualquier geometría respetando los contornos. Otros muchos autores presentan sus técnicas para restaurar bloques perdidos en imágenes comprimidas [113]. Otra contribución más reciente en este tema es la de Masnou [96], que presenta un método para reconstruir los contornos de un objeto que ha sido deteriorado por una oclusión.

Las técnicas de interpolación de manchas en secuencias siguen otro camino diferente a las de interpolación de imágenes fijas, ya que en el caso del vídeo se dispone de información muy importante que debe ser aprovechada: la redundancia temporal. A continuación se hace un pequeño resumen de las diferentes técnicas de interpolación de vídeo.

- **Interpolador basado en estadísticos ordenados: MMF**

Se trata de un filtro de mediana multietapa, propuesto por Kokaram *et al.* [82], que consiste en la concatenación de varias operaciones de filtrado de mediana. La primera etapa está compuesta de cinco filtros de mediana, con máscaras diferentes, que tienen en cuenta el píxel de la imagen actual y vecinos de la imagen actual y de la precedente y siguiente compensadas en movimiento. El resultado final es la mediana de los resultados de la primera etapa.

- **Interpolador basado en campos aleatorios de Markov: MRF**

Este método, propuesto también por Kokaram *et al.* [82, 85] intenta encontrar una estimación de máxima verosimilitud para los datos perdidos, $\hat{y}(x, y, t)$, dadas la máscara de manchas (posiciones donde faltan los datos), y los fotogramas actuales y posterior y anterior compensados en movimiento. Esta estima se consigue maximizando:

$$P[\hat{y}(x, y, t) | d(x, y, t), z_{mc}(x, y, t-1), z(x, y, t), z_{mc}(x, y, t+1)] \propto$$

$$\exp \left(- \sum_{(x,y,t):d(x,y,t)=1} \left[\sum_{(x',y') \in S_S(x,y)} (\hat{y}(x,y,t) - \hat{y}(x',y',t))^2 + \sum_{(x',y') \in S_T(x,y)} \lambda [(\hat{y}(x,y,t) - z_{mc}(x',y',t-1))^2 + (\hat{y}(x,y,t) - z_{mc}(x',y',t+1))^2] \right] \right),$$

donde S_S y S_T son los vecindarios espacial y temporal, respectivamente, y λ es el peso que se da al vecindario temporal. Esta probabilidad sólo es optimizada en las zonas de la imagen que indica la máscara, es decir, la zona manchada.

Godsill y Kokaram proponen detectar e interpolar conjuntamente, mediante campos aleatorios de Markov. Kokaram *et al.* [87, 88] presentan un algoritmo, el JOMBADI, que mediante campos aleatorios permite detectar e interpolar conjuntamente y además estimar el movimiento.

■ Interpolador basado en un modelo AR

Otro algoritmo presentado por Kokaram *et al.* en [82, 86]. Está basado en un modelo AR 3D. Se estima un conjunto de parámetros AR para cada zona de la imagen manchada, para lo cual se dispone de las imágenes actual, previa y posterior compensadas en movimiento. Se asume que los datos en la zona manchada son estacionarios. Las intensidades que faltan son interpoladas minimizando el error de predicción cuadrático medio obtenido con los parámetros estimados.

Goh *et al.* [56] proponen una modificación de este modelo AR para no asumir estacionariedad en las oclusiones. Para ello en vez de tener en cuenta tres imágenes (actual, posterior y anterior) sólo tienen en cuenta dos, la actual y la posterior o la actual y la anterior. Se elige aquella pareja de imágenes que tenga menor diferencia cuadrática media.

Kalra *et al.* [70] presentan otra mejora que consiste en dividir cada región a interpolar en varias subregiones e interpolar los datos de cada subregión de manera independiente. Esto se hace porque un solo conjunto de parámetros AR no es capaz de modelar adecuadamente una gran región cuando la mancha es grande.

Armstrong [19, 18] indica que los modelos AR suavizan excesivamente los bordes a la hora de interpolar, por eso propone sustituirlo por un modelo no lineal basado en funciones min-max.

Roosmalen *et al.* [118, 122] simplemente copian en la zona perdida la información del fotograma anterior o siguiente compensados en movimiento. La elección de dónde se extrae la información a copiar está gobernada por una máscara de oclusión que indica si está a 1 que la información ha de extraerse del fotograma anterior; y si está a 0, del siguiente. Esta máscara de oclusión se obtiene modelando la función de distribución conjunta del fotograma correcto y dicha máscara con un campo aleatorio de Markov, utilizando para conseguirlo un modelo AR de orden 2 que modela la estimación de la imagen original. Así pues, este autor utiliza dos métodos de interpolación, los modelos AR y los campos aleatorios de Markov, aunque ninguno de ellos los utiliza explícitamente para la interpolación.

5.3. Detección de manchas

El objetivo de la detección de manchas será obtener una máscara binaria para cada fotograma de la secuencia; en esta máscara se indicará la presencia de mancha en blanco y la ausencia de la misma en negro.

En esta tesis se proponen varios métodos de detección de defectos incorrelados, los cuales se presentan a continuación. Todos ellos siguen un esquema general, y se basan en las características que presentan los defectos que se pretenden detectar, las cuales favorecen el uso de operadores morfológicos. En esta sección se presentan tres métodos para obtener la máscara, que explotan únicamente la propiedad de incorrelación temporal que tienen las manchas, así como las ventajas e inconvenientes de cada uno de ellos, ilustrados con ejemplos. Además, para mejorar la detección, se propone un método que complementa a los anteriores utilizando el resto de propiedades que tienen estos defectos: el alto contraste y la varianza pequeña dentro de la mancha.

5.3.1. Método general

El método general de detección consta de dos partes: la detección de manchas claras, que obtiene como resultado una máscara indicando para cada píxel de cada fotograma si pertenece o no a una mancha de este tipo, y el mismo procedimiento para la detección de las manchas oscuras. La máscara final se calcula, como puede verse en la figura 5.2, como la suma binaria de ambas máscaras. Si se sabe a priori que una película no tiene manchas de un tipo, el proceso de restauración del tipo de manchas que no posee no se lleva a cabo, evitando la introducción de artefactos que el proceso de restauración de dicho tipo de manchas puede conllevar.

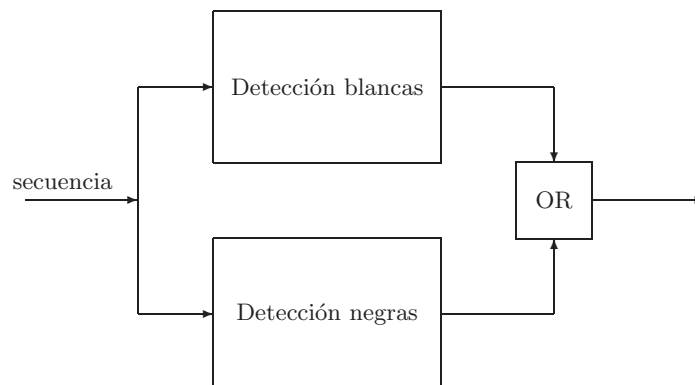


Figura 5.2: Esquema general del proceso de detección de manchas.

Características de un defecto incorrelado

Un defecto incorrelado o mancha lo definiremos como un conjunto de píxeles conexos que representa, casi con total seguridad, un máximo regional si la mancha es clara, ya que

proviene de una pérdida de gelatina en el soporte, siendo en esa zona la película totalmente transparente. Si la mancha es oscura, dependerá de la opacidad de la partícula depositada sobre el material el nivel de luminosidad de dicha mancha, pero probablemente se tratará de un mínimo regional. En cualquier caso, una mancha, para nosotros, será una región conexas muy contrastada, cuya intensidad varía poco espacialmente (poca varianza dentro de la mancha) y además varía rápidamente en el tiempo.

Estas características favorecen el uso de la morfología matemática a la hora de resolver el problema de la restauración. Existe un conjunto de operadores morfológicos que permiten extraer las regiones conexas máximas y mínimas de una imagen [126], y algunos de estos operadores son los que vamos a utilizar adaptándolos a nuestro problema.

Antes de pasar a la descripción de los métodos, conviene resaltar que un problema al que nos enfrentamos a la hora de la detección es el movimiento. Hemos dicho, en la definición del defecto que pretendemos encontrar, que una de sus características es la variación rápida en el tiempo. Cualquier objeto pequeño, más claro o más oscuro que el fondo en el que se encuentra y que se mueva rápidamente será detectado como mancha, produciéndose una falsa alarma. Para que esto se produzca, el objeto se tiene que mover muy rápidamente, y aunque no suele ser lo habitual, estas falsas alarmas se producen, ya que en las películas antiguas existen vibraciones indeseadas que dan lugar a movimientos bruscos entre fotogramas consecutivos. Por tanto, en los métodos que se van a proponer, en la medida de lo posible, hay que tener en cuenta que este problema se puede dar.

Métodos propuestos

Como hemos dicho anteriormente, los métodos que se van a proponer se basan en operadores morfológicos. Si el lector quiere profundizar en morfología puede consultar las referencias de Serra [129, 130, 126] y Matheron [97, 98]. Un resumen de operadores morfológicos puede encontrarlo en [100, 95]. Una breve descripción de los métodos, que serán explicados con más profundidad en los puntos siguientes, es:

▪ Métodos para detección de incorrelación temporal

- Apertura (cierre) temporal: para detectar las manchas claras (oscuras), se realiza una apertura (cierre) temporal con un elemento estructurante plano de tamaño 2. Con esto, se eliminarán todas aquellas zonas claras (oscuras) de la imagen que aparezcan en el fotograma actual pero no estén en el siguiente o en el anterior. Tomando el residuo entre el fotograma actual y la apertura (cierre), y aplicando un umbral a este resultado obtendremos la máscara binaria correspondiente a las manchas claras (oscuras). El problema que presenta es que si un objeto pequeño y claro (oscuro) se mueve rápidamente puede ser deformado o incluso eliminado.
- Apertura (cierre) con reconstrucción [80, 131, 126]: se trata de un operador que preserva intactas sólo aquellas zonas claras (oscuras) del fotograma actual que le marcamos mediante una imagen marcador. Las zonas claras (oscuras) que no aparecen en el marcador son eliminadas. Del mismo modo que en el caso anterior, la máscara se obtendrá realizando el residuo de la imagen reconstruida con la original y aplicando un umbral al resultado.

- Apertura (cierre) por λ -reconstrucción [57]: El método es el mismo que el anterior, pero el elemento estructurante que se utiliza para realizar la apertura (cierre) no es plano.

▪ **Método para detección de zonas de alto contraste**

Para mejorar la detección que llevan a cabo los tres métodos anteriores, se propone un método complementario que tiene en cuenta otra propiedad de las manchas: el alto contraste. El algoritmo está basado en la h -reconstrucción, un operador morfológico que elimina de una imagen las zonas que tienen un contraste menor que un nivel h . Toda esta información se combinará con la obtenida con el método elegido para detectar la incorrelación temporal.

5.3.2. Apertura (cierre) temporal

A continuación definiremos la apertura y el cierre aplicados a la dimensión temporal, pero antes definamos los dos operadores básicos: dilatación y erosión temporales.

Dada una secuencia de imágenes que denotamos como $f(x, y, t)$, la dilatación temporal de un píxel (x, y, t) de una imagen perteneciente a dicha secuencia con un elemento estructurante plano (B), es el valor máximo de los niveles de los píxeles que están dentro de la ventana de observación definida por el elemento estructurante B , centrado en dicho píxel.

$$\delta_B(f)(x, y, t) = \sup\{f(x, y, t + k), k \in B\}.$$

Igualmente se define la erosión temporal como:

$$\varepsilon_B(f)(x, y, t) = \inf\{f(x, y, t - k), k \in B\}.$$

Los operadores que nos interesan en los métodos que vamos a presentar son la apertura, para la detección de manchas claras, y el cierre para la detección de manchas oscuras; operadores que son combinaciones de la erosión y la dilatación.

- **Apertura temporal:** Consiste en realizar primero una erosión temporal, y sobre el resultado una dilatación, utilizando el mismo elemento estructurante.

$$\gamma_B(f(x, y, t)) = \delta_B(\varepsilon_B(f(x, y, t))).$$

El efecto de la apertura temporal sobre la secuencia es que elimina todas aquellas zonas claras que no tengan una duración temporal mayor que el tamaño del elemento estructurante.

- **Cierre temporal:** Se trata del operador dual a la apertura, es decir, una dilatación seguida por una erosión.

$$\varphi_B(f(x, y, t)) = \varepsilon_B(\delta_B(f(x, y, t))).$$

El efecto sobre la secuencia es el mismo que la apertura pero en este caso sobre zonas oscuras.

Descripción del método

El método propuesto para la detección de manchas claras se muestra en la figura 5.3. Consiste en realizar una apertura con un EE de tamaño 2 en la dimensión temporal, de la siguiente forma:

- Se realiza la erosión temporal, es decir, se toma el mínimo píxel a píxel de toda la secuencia tomando las imágenes en parejas de dos: el fotograma actual I_n y el anterior I_{n-1} , obteniéndose:

$$E_n = \text{mín}(I_n, I_{n-1})$$

- Sobre la secuencia resultante se realiza la dilatación temporal, o sea, se toma el máximo píxel a píxel de toda la secuencia, tomando ahora las parejas de imágenes en sentido contrario: el fotograma actual erosionado temporalmente (E_n) y el siguiente (E_{n+1}), obteniéndose:

$$A_n = \text{máx}(E_n, E_{n+1})$$

- Posteriormente se realiza el residuo entre la imagen original I_n y el resultado de la apertura temporal.
- Finalmente se aplica un umbral al resultado, obteniéndose de esta forma la máscara de las manchas claras correspondiente al fotograma I_n , M_{c_n} . Este paso selecciona sólo aquellas manchas que aparecen en el residuo con una intensidad mayor que dicho umbral, es decir, sólo aquellas que están contrastadas con respecto a su fondo. Las que no superan el umbral son defectos poco contrastados que no serán apreciables por el espectador.

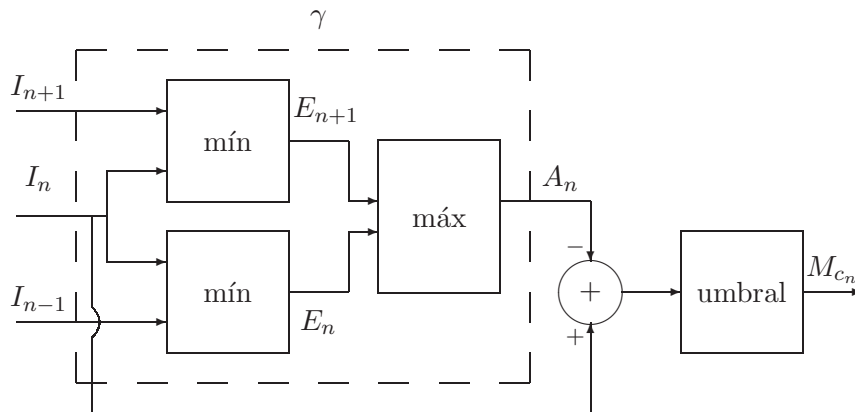


Figura 5.3: Obtención de máscara de manchas claras mediante el método de la apertura temporal.

El método para la obtención de la máscara de manchas oscuras, M_{o_n} es el dual del anterior, es decir, se realiza el cierre temporal en lugar de la apertura (máximo seguido de mínimo), como puede verse en la figura 5.4.

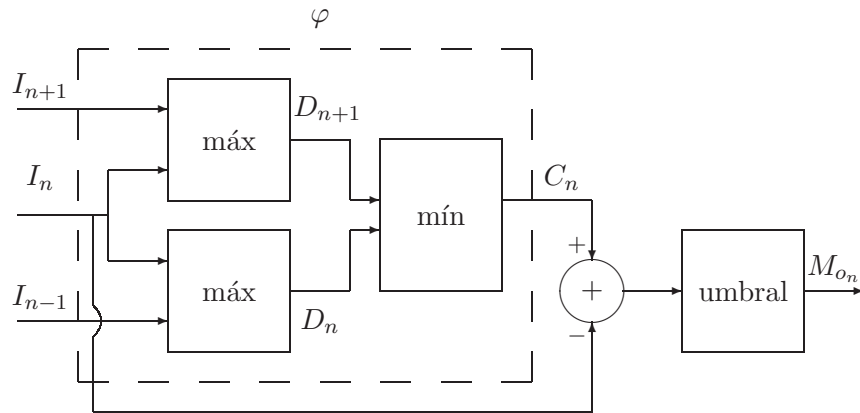


Figura 5.4: Obtención de máscara de manchas oscuras mediante el método del cierre temporal.

Ventajas e inconvenientes de la apertura (cierre) temporal

Al realizar la apertura (cierre) temporal, se eliminarán todas aquellas manchas claras (oscuras) que aparecen en el fotograma actual y que no están en el fotograma anterior ni en el siguiente. Como contrapartida, aquellos objetos claros (oscuros) que se mueven más que la mitad de su tamaño por fotograma, se rompen. En otras palabras, un objeto claro debe estar en la misma posición al menos durante dos fotogramas para que no sea eliminado ni deformado por el método. Este problema aparece ilustrado en las figuras 5.5 y 5.6; en la primera se muestra la evolución a lo largo del tiempo (en tres fotogramas consecutivos) de un objeto oscuro sobre fondo claro. Se observa que de un fotograma al siguiente se ha movido una cantidad mayor que la mitad de su tamaño (concretamente el desplazamiento es de D píxeles/fotograma). Las zonas sombreadas representan la intersección del objeto en dos fotogramas consecutivos, zonas que serán las únicas que sobrevivan al cierre temporal, ya que están presentes a la vez en dos fotogramas consecutivos. La segunda figura, la 5.6, muestra el resultado para el fotograma I_n del cierre temporal de tamaño 2, y como podemos ver, el objeto se ha partido en dos.

Este inconveniente puede ser producido no sólo por el movimiento de objetos sino también por un movimiento de cámara o una vibración indeseada y brusca, hecho que se produce frecuentemente en las películas antiguas.

Resultados de la apertura (cierre) temporal

En la figura 5.7 se muestra un fotograma original de la secuencia *Taronja*. Concretamente este plano es un tranvía que se mueve en horizontal, y como puede observarse, la imagen está deteriorada con unas manchas claras bastante extensas. En esta figura se muestran los resultados del proceso de detección de manchas claras: la apertura temporal de tamaño 2, el residuo de dicha apertura y a este residuo se le aplica un umbral de 30. Como vemos en el resultado de la apertura, la mayoría de manchas claras han sido eliminadas apareciendo, por tanto, en el residuo y la máscara, siendo detectadas correctamente.

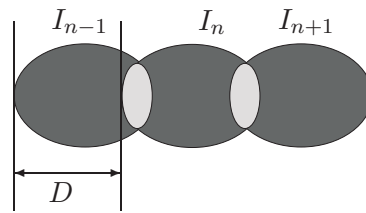


Figura 5.5: Objeto que se desplaza una cantidad igual a D en horizontal. Posición en los fotogramas I_{n-1} , I_n e I_{n+1} . El área sombreada es la intersección del objeto entre dos fotogramas adyacentes.



Figura 5.6: Resultado del cierre temporal de tamaño 2 sobre la secuencia de la figura 5.5 para el fotograma I_n . Vemos como el objeto se ha partido en dos ya que su desplazamiento por fotograma es mayor que la mitad de su tamaño.

Este ejemplo es ilustrativo del rompimiento de objetos: la barra vertical del tranvía se rompe al realizar la apertura, ya que el movimiento entre dos fotogramas consecutivos es mayor que la mitad de la anchura de la barra, produciéndose en esa zona una diferencia grande en el residuo y, por tanto, una falsa alarma en la detección.

La figura 5.8 muestra el proceso de detección de manchas oscuras para un fotograma de la secuencia *Sanga*. Vemos como las manchas oscuras desaparecen al hacer el cierre temporal, detectándose correctamente. Aparecen falsas alarmas debidas a la vibración de la película en los bordes de los objetos oscuros, como el traje de un personaje.

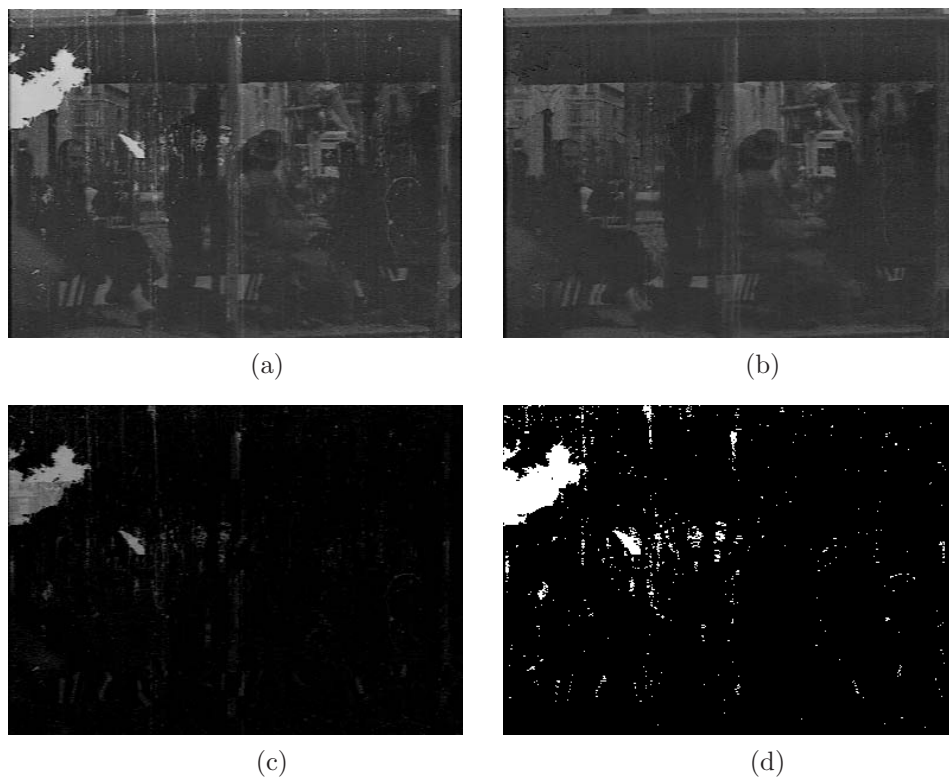


Figura 5.7: Detección de manchas claras: (a)Fotograma original de la secuencia *Taronja*. (b)Mismo fotograma una vez realizada la apertura temporal. Observar como desaparecen las manchas blancas, pero la barra vertical del tranvía se rompe por la mitad. (c)Residuo de la apertura con el fotograma original . (d)Máscara: residuo tras el umbral de 30.

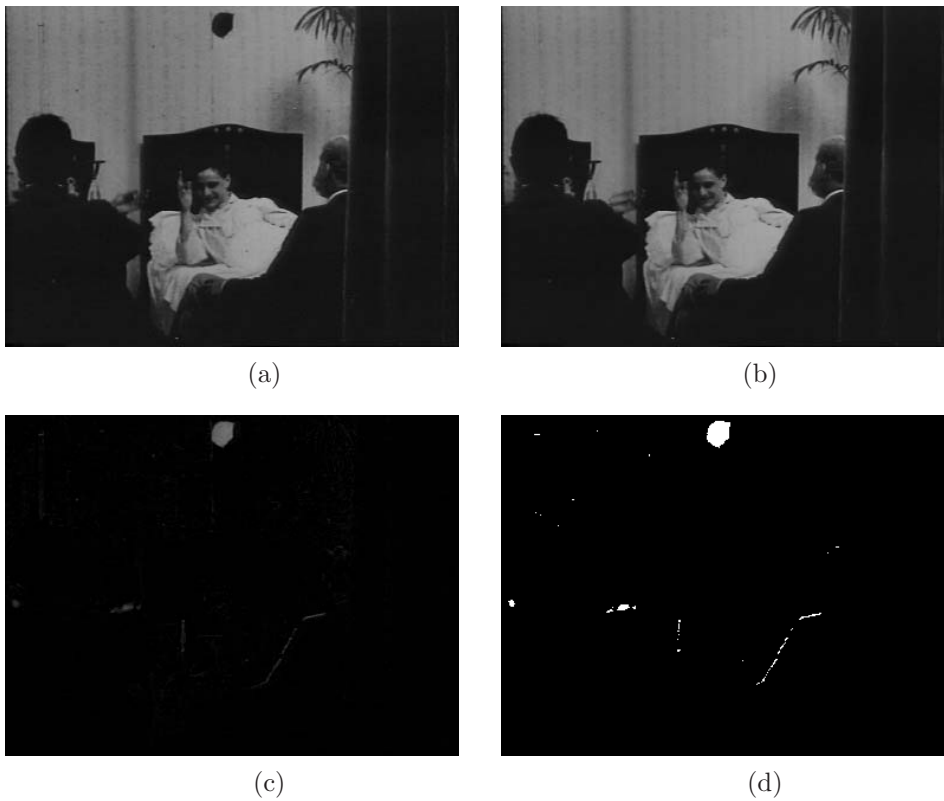


Figura 5.8: Detección de manchas oscuras: (a)Fotograma original de la secuencia *Sangre* (película *Sangre y arena*). (b)Mismo fotograma una vez realizado el cierre temporal. (c)Residuo del cierre con el fotograma original . (d)Máscara: residuo tras un umbral de 30.

5.3.3. Apertura (cierre) por reconstrucción

Como hemos visto en el apartado anterior, un problema que tiene la apertura (cierre) temporal es que elimina no sólo las manchas claras (oscuras) sino además aquellas zonas claras de un objeto que no aparecen en la misma posición en fotogramas consecutivos, debido al movimiento, produciéndose de esta forma falsas alarmas (se rompen objetos o desaparecen). Lo que desearíamos es que se eliminasen las manchas pero el resto permaneciese inalterado, y es por ello por lo que vamos a introducir la apertura (cierre) por reconstrucción.

La apertura y el cierre por reconstrucción derivan de las transformaciones llamadas geodésicas. Una transformación geodésica es la transformación morfológica correspondiente de una imagen llamada marcador (m) pero dentro de otra llamada de referencia (f), siendo $m \leq f$. Las operaciones básicas son la dilatación y la erosión geodésicas.

La dilatación geodésica de tamaño 1 del marcador m dentro de la imagen de referencia f , viene dada por:

$$\delta_f^1(m) = \inf(\delta_B(m), f),$$

es decir, es el ínfimo entre la dilatación morfológica del marcador con el elemento estructurante B y la imagen de referencia. La dilatación geodésica de tamaño n se define como:

$$\delta_f^n(m) = \underbrace{\delta_f^1(\delta_f^1(\dots \delta_f^1(m)))}_{n \text{ veces}}.$$

De igual forma se define la erosión geodésica de tamaño 1 como:

$$\varepsilon_f^1(m) = \sup(\varphi_B(m), f),$$

es decir, es el máximo entre la erosión morfológica del marcador con el elemento estructurante B y la imagen de referencia. La erosión geodésica de tamaño n será:

$$\varepsilon_f^n(m) = \underbrace{\varepsilon_f^1(\varepsilon_f^1(\dots \varepsilon_f^1(m)))}_{n \text{ veces}}.$$

Una vez definidas las transformaciones geodésicas básicas, podemos pasar a definir la apertura y el cierre por reconstrucción, que son los operadores que vamos a usar en nuestro método. La reconstrucción geodésica apareció por primera vez en la morfología, de manera experimental en la tesis de Klein [80]; las razones de su buen comportamiento y sus propiedades fueron presentadas por Serra y Salembier en [131].

- **Apertura por reconstrucción:** la apertura por reconstrucción de f tomando como marcador m es la dilatación geodésica de m dentro de f hasta que el resultado no cambie (idempotencia)¹.

$$\gamma^{rec}(f, m) = \delta_f^\infty(m),$$

Las zonas claras de la imagen referencia donde no cabe el elemento estructurante (B) desaparecen, se transforman en zonas planas que toman el valor máximo que tiene la

¹Para aplicarla a nuestros algoritmos la reconstrucción ha sido implementada de manera eficiente siguiendo las indicaciones de Vincent [151].

imagen marcador donde la de referencia era una región máxima. El resto de la imagen de referencia se reconstruye de manera eficiente y precisa. La figura 5.9 muestra la apertura por reconstrucción de una señal unidimensional de referencia (señal azul) con un marcador (señal rosa). El resultado de la apertura por reconstrucción (área sombreada verde) no presenta los máximos que no aparecen en la señal marcador, mientras que el resto permanece igual que en la señal referencia.

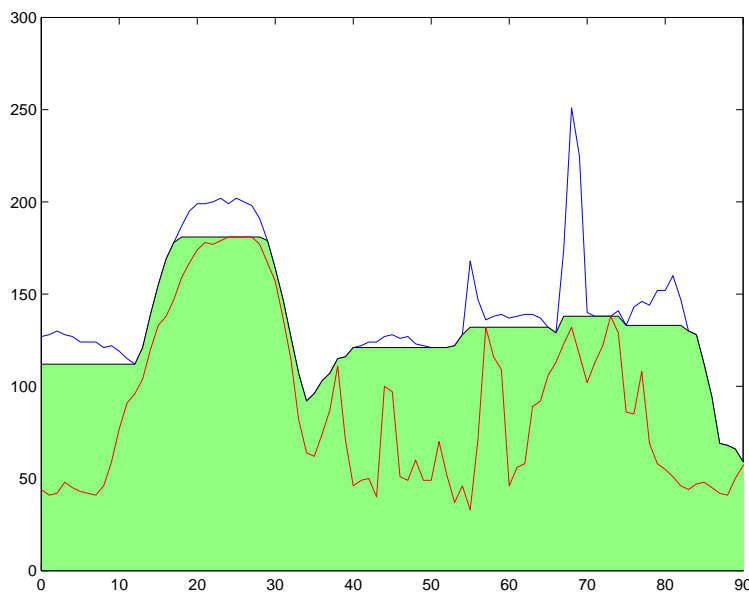


Figura 5.9: Apertura por reconstrucción.

- **Cierre por reconstrucción:** Es el operador dual a la apertura por reconstrucción, es decir, es la erosión geodésica de m dentro de f hasta la idempotencia.

$$\varphi^{rec}(f, m) = \varepsilon_f^\infty(m),$$

El efecto sobre la imagen es el mismo que la apertura por reconstrucción pero en este caso sobre zonas oscuras (mínimos regionales).

Descripción del método

Como se ha dicho anteriormente, la apertura (cierre) por reconstrucción elimina de la imagen referencia aquellas zonas claras (oscuras) que no aparecen en la imagen marcador. Siendo la imagen de referencia I_n , es decir el fotograma actual, el marcador debe ser una imagen donde no aparezcan las manchas que sí están en la de referencia, para que al reconstruir desaparezcan.

La figura 5.10 muestra el método propuesto para obtener la máscara de manchas blancas para el fotograma I_n :

- La imagen marcador se obtiene como el máximo píxel a píxel entre los fotogramas adyacentes al actual, I_{n-1} e I_{n+1} . Basándonos en la propiedad de incorrelación temporal que tienen las manchas, aquellas que aparezcan en I_n no aparecerán en los fotogramas adyacentes. Se toma el máximo entre el fotograma anterior y siguiente al actual, para que sí aparezca en el marcador un objeto claro que se mueva rápidamente. Con esto lo que se exige es que un objeto, para que no sea considerado defecto, deba estar en posiciones muy próximas al menos en alguno de los dos fotogramas adyacentes al actual, de tal forma que coincida, al menos en algún punto, en la imagen marcador y en la de referencia. Este desplazamiento máximo será el tamaño del objeto en la dirección del desplazamiento.

Antes de hacer la reconstrucción, se toma el mínimo píxel a píxel entre la imagen de referencia y el marcador, para asegurarnos de que el marcador sea siempre menor que la señal de referencia, siendo el resultado la nueva imagen marcador. Para el ejemplo de la figura 5.5, el marcador sería la imagen mostrada en la figura 5.6, es decir, las zonas comunes del objeto en los tres instantes de tiempo, por lo que el objeto se reconstruiría para la imagen I_n en el cierre por reconstrucción, y además preservando su forma.

- Se reconstruye el fotograma I_n con el marcador.
- Se aplica un umbral al residuo entre la imagen de referencia y el resultado de la apertura, obteniéndose la máscara de manchas claras para el fotograma I_n , M_{c_n} .

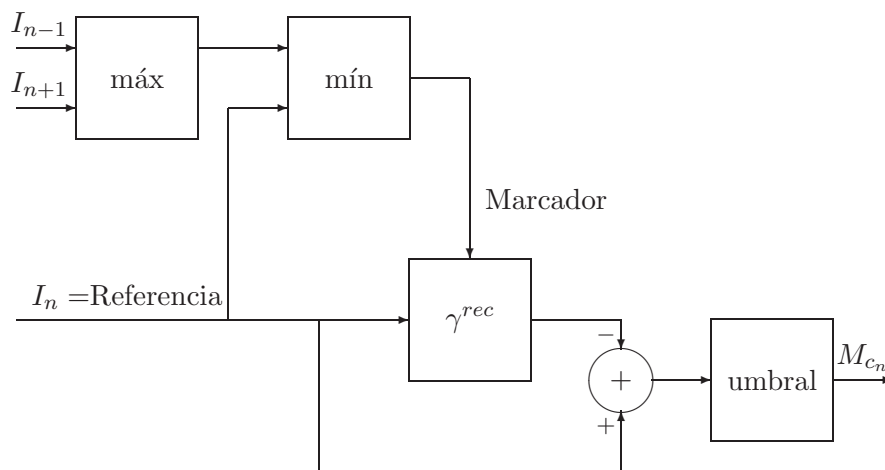


Figura 5.10: Obtención de máscara de manchas claras mediante el método de la apertura por reconstrucción.

El método para la obtención de la máscara de manchas oscuras, M_{o_n} es el dual del anterior, es decir, se toma el mínimo entre I_{n-1} e I_{n+1} como marcador y se realiza el cierre por reconstrucción en lugar de la apertura, como se observa en la figura 5.11.

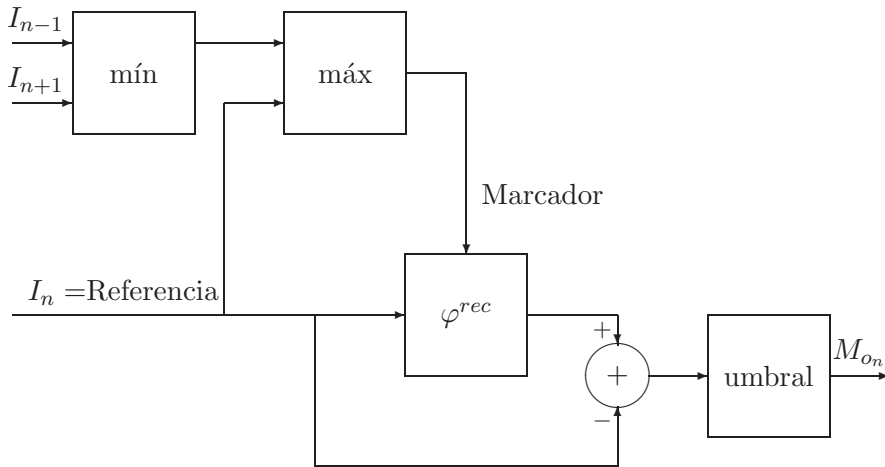


Figura 5.11: Obtención de máscara de manchas oscuras mediante el método del cierre por reconstrucción.

Ventajas e inconvenientes de la apertura (cierre) por reconstrucción

Este método soluciona el problema que tiene el método anterior, es decir, la eliminación o rotura de objetos que se mueven más píxeles por fotograma que la mitad de su tamaño, ahora los objetos pueden moverse una cantidad igual a su tamaño para que no sean deformados, las formas conexas se preservan si están en el marcador. En contraposición, existen manchas que no son detectadas ya que no son eliminadas por la apertura (o el cierre). Esto ocurre cuando la mancha está unida a una zona clara (oscura) de la imagen.

Así pues, el desplazamiento máximo de un objeto entre fotogramas, para que no sea considerado defecto, se duplica en este método con respecto al de la apertura (cierre) temporal. Sin embargo, este problema se produce frecuentemente, ya que si el objeto es pequeño el desplazamiento máximo permitido será también muy pequeño y si se mueve más rápido será eliminado. Para evitar este artefacto, antes de pasar a la reconstrucción, el marcador se dilata, en el caso de la apertura, o se erosiona, si se está realizando un cierre, para que si existe algún objeto claro (oscuro) que se haya movido rápidamente, la posición en que aparece en el marcador tenga más posibilidades de tocar en algún punto al objeto en la imagen de referencia. Con esto se amplía el desplazamiento máximo permitido al objeto, para no ser eliminado, a su tamaño más el tamaño del elemento estructurante.

El tamaño del elemento estructurante con que se realiza esta dilatación (erosión) depende de la cantidad de movimiento esperado, sin embargo, cuanto mayor sea, la probabilidad de detección de manchas disminuye, ya que la dilatación de una mancha en el marcador puede coincidir en algún punto con otra mancha diferente en la imagen referencia, o una zona clara de la misma, siendo de esta forma reconstruida la mancha de la imagen referencia. La opción utilizada en nuestros ejemplos es un EE cuadrado de tamaño 3×3 . Los diagramas completos de ambos métodos, apertura y cierre con reconstrucción, se muestran en las figuras 5.12 y 5.13.

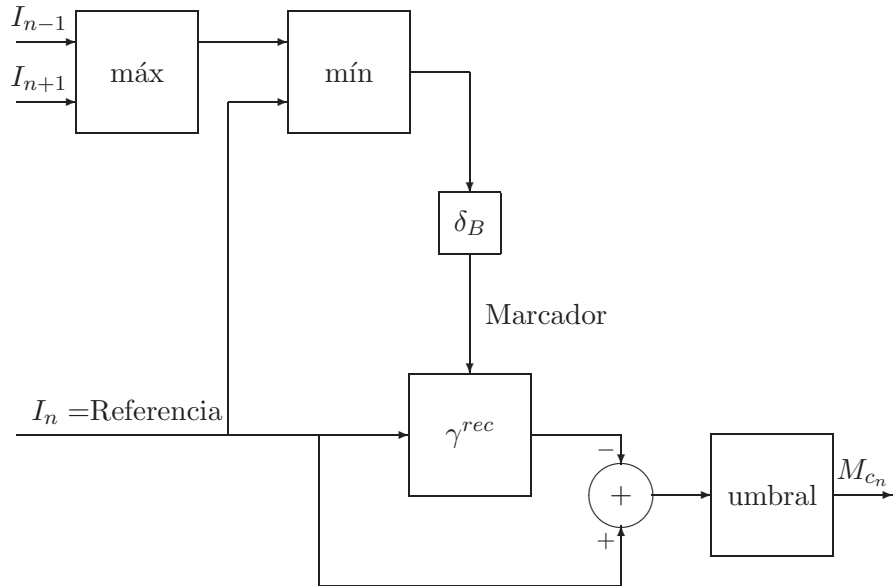


Figura 5.12: Obtención de máscara de manchas claras mediante el método de la apertura por reconstrucción con dilatación del marcador.

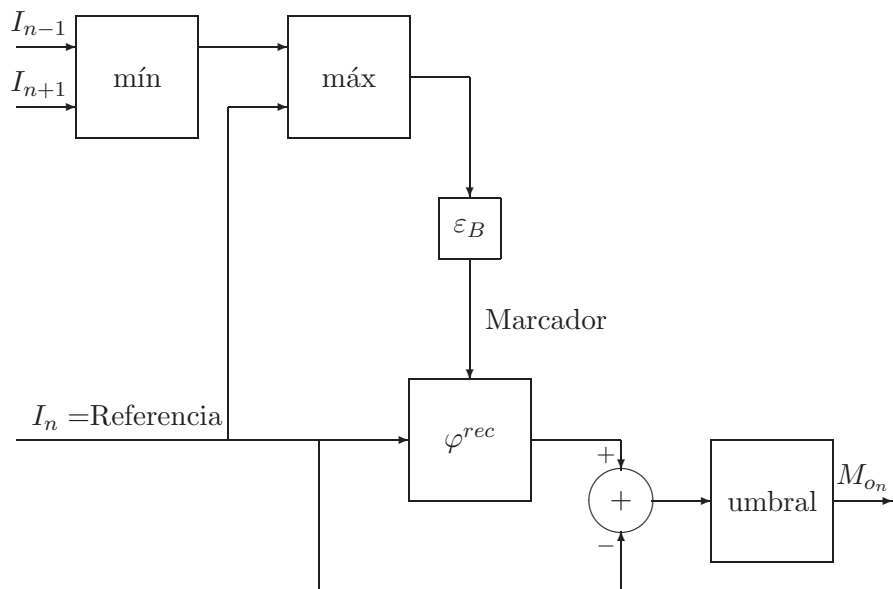


Figura 5.13: Obtención de máscara de manchas oscuras mediante el método del cierre por reconstrucción con erosión del marcador.

Resultados de la apertura (cierre) por reconstrucción

La figura 5.14 muestra el proceso de detección de manchas claras con este método para el fotograma I_n (imagen central de la primera fila, (b)). En esta figura aparecen los tres fotogramas involucrados en la detección: I_{n-1} , I_n e I_{n+1} ; el marcador obtenido según la figura 5.12, el resultado de la reconstrucción, el residuo entre la imagen original y la apertura por reconstrucción y finalmente, el resultado de aplicar un umbral a dicho residuo, es decir, la máscara. Como puede observarse, las manchas del fotograma I_n no aparecen en los adyacentes y, por tanto, tampoco en el marcador, por lo que dichas manchas no serán reconstruidas tal y como están en la imagen actual, sino que se reconstruirán con el valor máximo que tiene el marcador en dicha zona, apareciendo en el residuo.

Además de detectar las manchas, desaparece el problema que tenía la apertura temporal al detectar una falsa alarma en la barra vertical del tranvía. Este objeto en movimiento en el fotograma I_n tiene una intersección distinta de cero con dicho objeto en el marcador (se mueve menos por fotograma que su tamaño más el tamaño de la dilatación), por lo que es reconstruido preservando su forma.

La figura 5.15 muestra los resultados del proceso para manchas oscuras. Aquí la mayoría de manchas no son detectadas, ya que son reconstruidas con una intensidad similar a la de la mancha original, y en el residuo tienen un valor pequeño. Esto es debido a que hay muchas manchas conectadas a zonas oscuras, como el cabezal de la cama, que es una zona oscura de la imagen, por lo que a la hora de realizar el cierre son reconstruidas. Para reducir este efecto vamos a presentar el método siguiente, donde los operadores se realizan utilizando λ -reconstrucción.

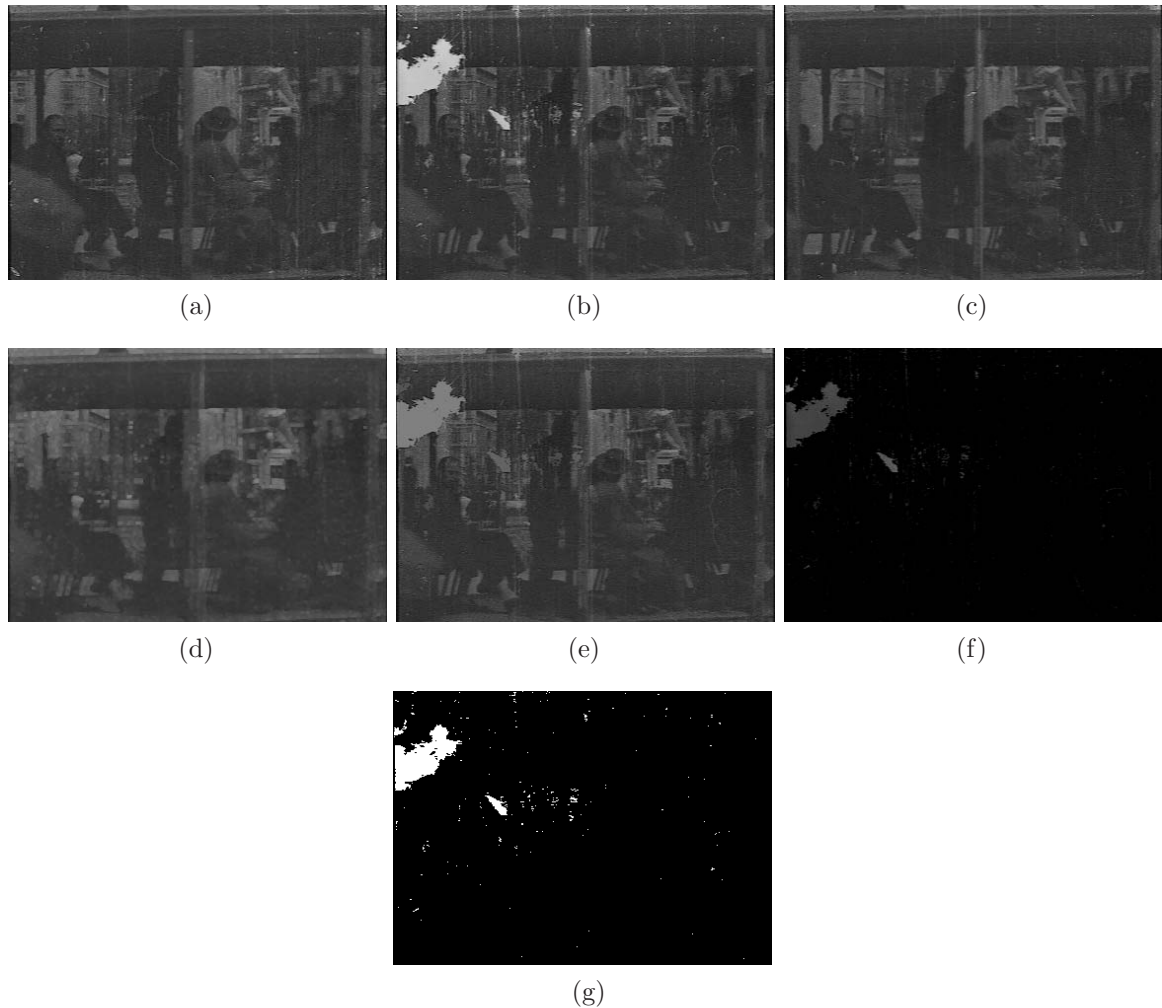


Figura 5.14: Detección de manchas claras: (a) Fotograma original de la secuencia *Taronja*, I_{n-1} . (b) Fotograma original de la secuencia *Taronja*, I_n . (c) Fotograma original de la secuencia *Taronja*, I_{n+1} . (d) Marcador obtenido según el proceso de la figura 5.12. Se puede observar que las manchas no aparecen en el marcador, por lo que no aparecerán en la reconstrucción con la misma intensidad que en la imagen de referencia, y por tanto, aparecerán en el residuo. Por otro lado, la barra del tranvía en la imagen I_n tiene algún punto en común con dicha barra en la imagen marcador, por lo que se reconstruirá. (e) Resultado de la apertura por reconstrucción de la imagen (b) con el marcador (d). (f) Residuo de la apertura con el fotograma original. (g) Máscara: residuo tras un umbral de 30. Observar cómo se detectan todas las manchas blancas, y no aparece el artefacto en la barra del tranvía.

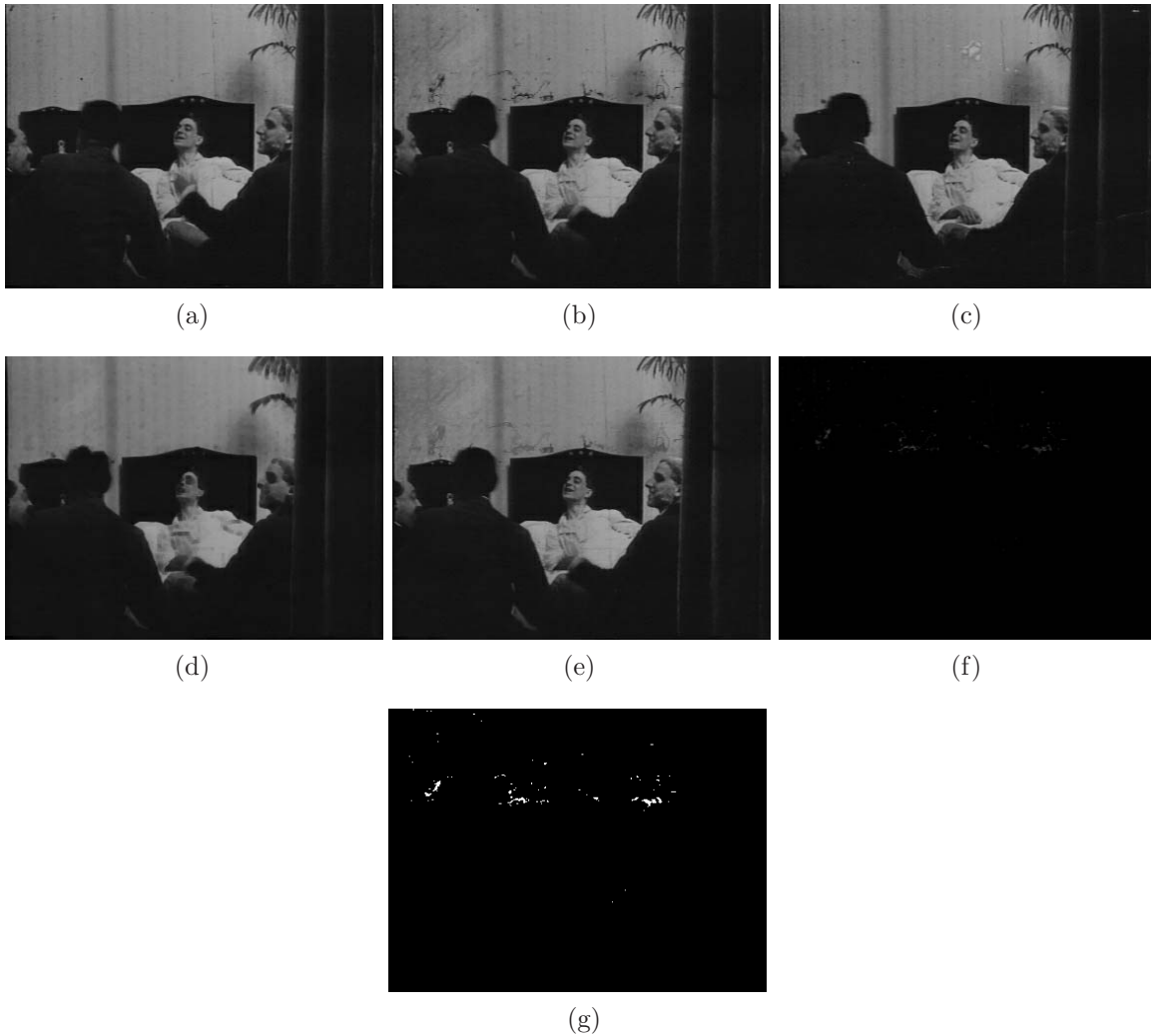


Figura 5.15: Detección de manchas oscuras: (a)Fotograma original de la secuencia *Sanga*, I_{n-1} . (b)Fotograma original de la secuencia *Sanga*, I_n . (c)Fotograma original de la secuencia *Sanga*, I_{n+1} . (d)Marcador obtenido según el proceso de la figura 5.13. Se puede observar que las manchas no aparecen en el marcador. (e)Resultado del cierre por reconstrucción de la imagen (b) con el marcador (d). Se puede observar que muchas manchas son reconstruidas, ya que están conectadas a zonas oscuras de la imagen. (f)Residuo de la apertura con el fotograma original. (g)Máscara: residuo tras un umbral de 15. Hay manchas que no son detectadas ya que han sido reconstruidas al nivel de intensidad de la zona oscura a la que están conectadas, y dicha intensidad es similar a la de la mancha.

5.3.4. Apertura (cierre) por λ -reconstrucción

El problema que presenta la apertura (cierre) por reconstrucción es que las manchas claras (oscuras) conectadas a zonas claras (oscuras) de la imagen son reconstruidas con el valor máximo (mínimo) que tiene el marcador en la zona. Si dicha zona tiene una intensidad parecida a la de la mancha, no aparece en el residuo y no es detectada, como en el ejemplo de la figura 5.15. Una mejora de este método es utilizar un elemento estructurante no plano a la hora de realizar la apertura (cierre) por reconstrucción, operación que se conoce como apertura (cierre) por λ -reconstrucción [57].

Definimos la apertura por λ -reconstrucción de una imagen de referencia f usando como marcador la imagen m de la siguiente forma:

$$\gamma_{\lambda}^{rec}(f, m) = \delta_{f, \lambda}^{\infty}(m),$$

donde $\delta_{f, \lambda}^{\infty}(m)$ es la λ -dilatación geodésica de tamaño 1 iterada hasta la idempotencia. La λ -dilatación geodésica es la dilatación geodésica con un elemento estructurante no plano, que se puede calcular mediante la expresión:

$$\delta_{f, \lambda}^1(m) = \inf(\delta_{B, \lambda}(m), f),$$

siendo $\delta_{B, \lambda}$, la λ -dilatación morfológica de la imagen m , dada por:

$$\delta_{B, \lambda}(m) = \sup(\delta_B(m) - \lambda, m)$$

De manera análoga, el cierre por λ -reconstrucción viene dado por:

$$\varphi_{\lambda}^{rec}(f, m) = \varepsilon_{f, \lambda}^{\infty}(m),$$

siendo $\varepsilon_{f, \lambda}^{\infty}(m)$ la λ -erosión geodésica de tamaño 1 iterada hasta la idempotencia. La erosión geodésica de tamaño 1 viene dada por la expresión:

$$\varepsilon_{f, \lambda}^1(m) = \sup(\varepsilon_{B, \lambda}(m), f),$$

siendo $\varepsilon_{B, \lambda}$, la λ -erosión morfológica de la imagen m , dada por:

$$\varepsilon_{B, \lambda}(m) = \inf(\varepsilon_B(m) + \lambda, m)$$

Lo que consigue este operador es que las zonas planas que aparecían con la reconstrucción decrezcan (crezcan) con una pendiente igual a λ . La figura 5.16 muestra los resultados de la apertura por reconstrucción, las aperturas por λ -reconstrucción para $\lambda = 1$ y para $\lambda = 5$ para una señal unidimensional.

Descripción del método

El método sigue el mismo esquema que la apertura (cierre) por reconstrucción (figuras 5.10 y 5.11) pero utilizando aperturas y cierres por λ -reconstrucción.

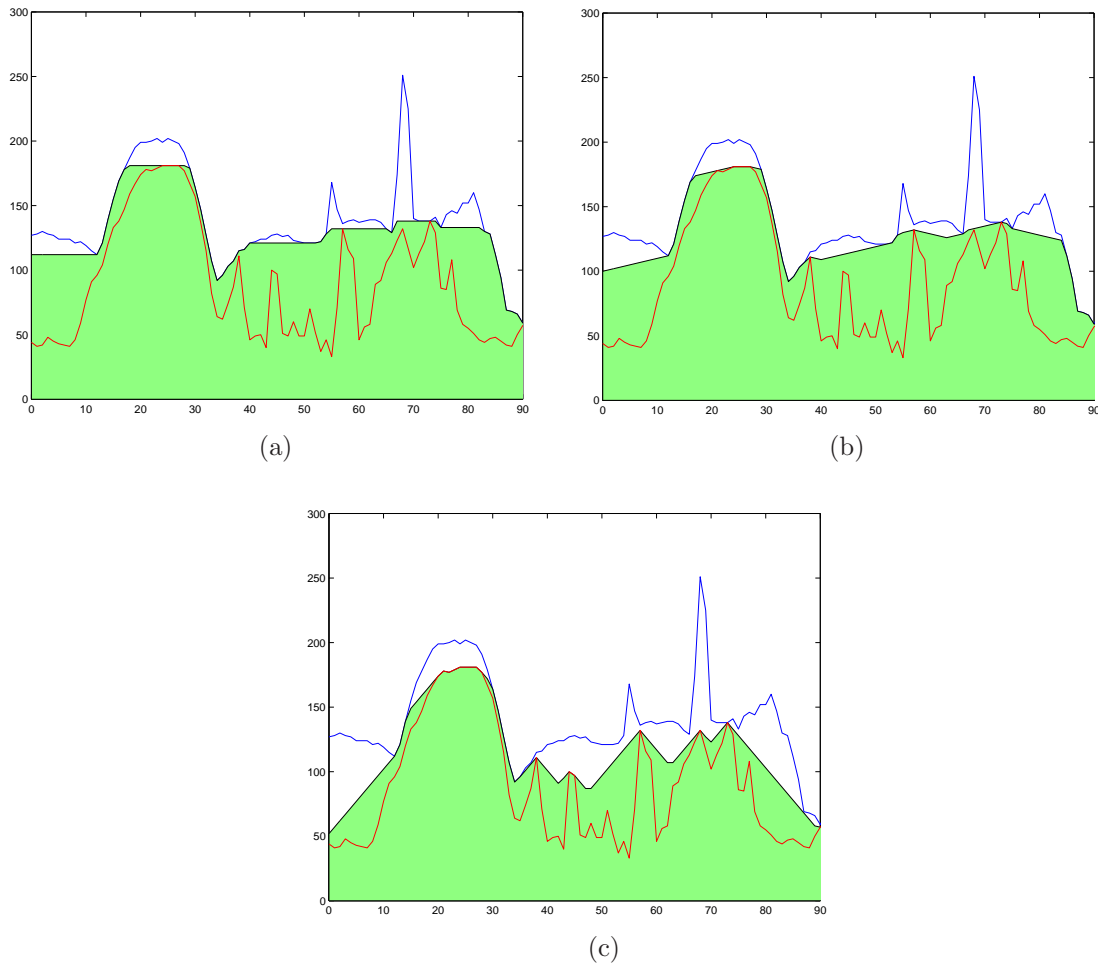


Figura 5.16: (a) Apertura por reconstrucción. (b) Apertura por λ -reconstrucción $\lambda = 1$. (c) Apertura por λ -reconstrucción $\lambda = 5$.

Ventajas e inconvenientes de la apertura (cierre) por λ -reconstrucción

Este método es una solución de compromiso entre el aumento de la probabilidad de detección de manchas y la disminución de la probabilidad de falsa alarma. Es un punto intermedio entre el método de la apertura (cierre) temporal, que consigue mayor probabilidad de detección, y el de la apertura (cierre) por reconstrucción, que consigue menor probabilidad de falsa alarma.

Si aumentamos λ la probabilidad de detectar manchas aumenta, pero por contra, la probabilidad de falsa alarma también, es decir, es más probable que un objeto en movimiento sea detectado como mancha; a medida que λ aumenta el método se asemeja más al de la apertura (cierre) temporal. Por el contrario, si λ disminuye la probabilidad de falsa alarma disminuye pero también la de detección. En el caso extremo de $\lambda = 0$, el método se convierte en el de la apertura (cierre) por reconstrucción.

Resultados de la apertura (cierre) por λ -reconstrucción

Las figuras 5.17 y 5.18 muestran los resultados para la detección de manchas oscuras en el caso de cierre por reconstrucción para $\lambda = 5$ y $\lambda = 20$, respectivamente. Comparando ambas figuras con el resultado del cierre por reconstrucción (figura 5.15), podemos ver que en el caso de la λ -reconstrucción las manchas oscuras, a medida que nos alejamos del punto de conexión con la zona oscura de la imagen con la que están conectadas, se van rellenando del color del fondo; además, comparando ambas figuras entre sí vemos que este efecto se potencia a medida que aumentamos λ , es decir, la intensidad del fondo rellena cada vez más cantidad de mancha, y por tanto aumentamos la probabilidad de detección. También podemos observar que se detecta una falsa alarma debida a uno de los personajes que mueve el brazo bruscamente, en el cierre por reconstrucción esta falsa alarma no se detectaba, mientras que en este caso sí, y a medida que aumenta λ esta zona aumenta en el residuo, siendo más fácilmente detectada al aplicar un umbral.

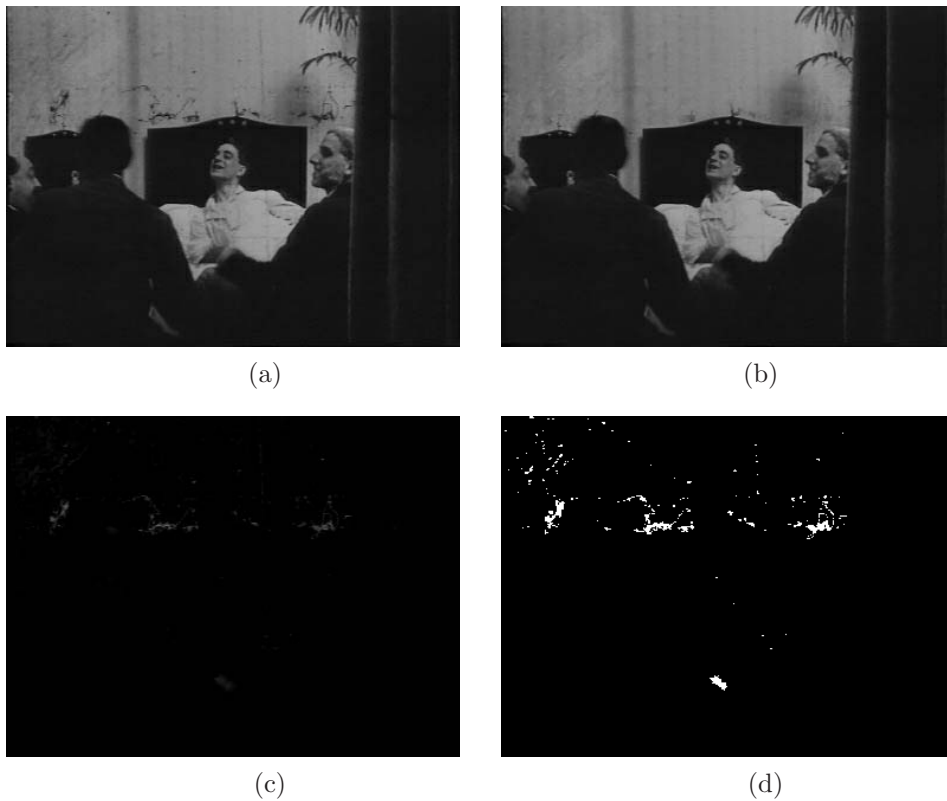


Figura 5.17: Detección de manchas oscuras: (a)Fotograma original de la secuencia *Sanga*. (b)Mismo fotograma una vez realizado el cierre por λ -reconstrucción con $\lambda = 5$. (c)Residuo del cierre con el fotograma original. (d)Máscara: residuo tras un umbral de 15. Observar como se detectan más manchas negras que en el caso del cierre por reconstrucción, aunque no todas. Nótese también como aparece la falsa alarma en el brazo del personaje.

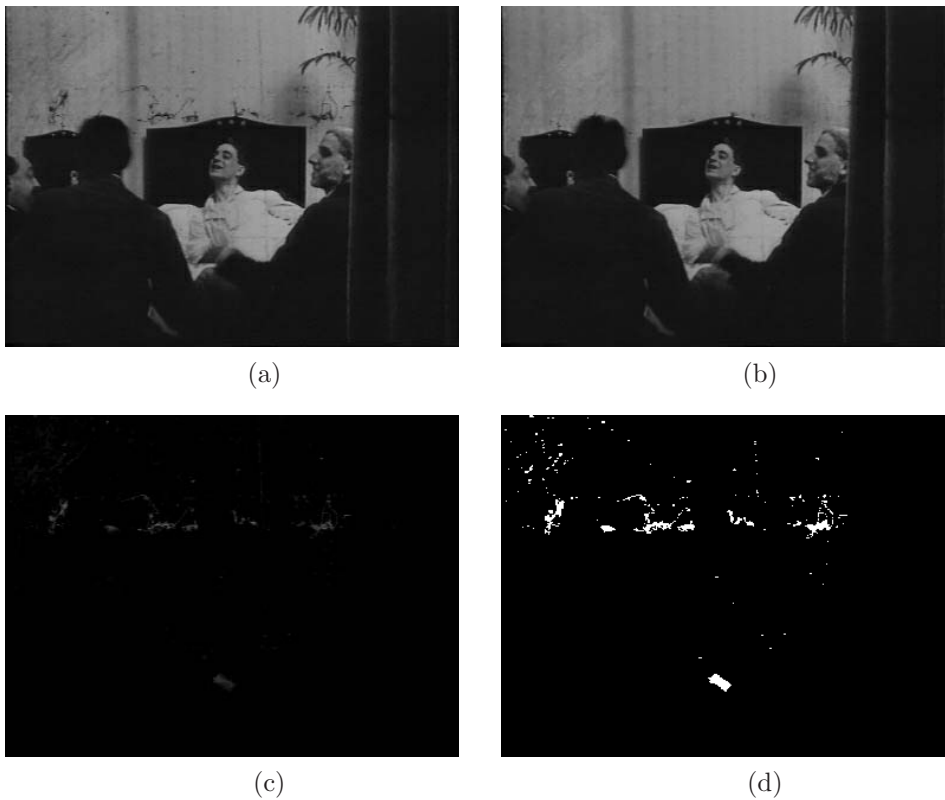


Figura 5.18: Detección de manchas oscuras: (a)Fotograma original de la secuencia *Sanga*. (b)Mismo fotograma una vez realizado el cierre por λ -reconstrucción con $\lambda = 20$. (c)Residuo del cierre con el fotograma original . (d)Máscara: residuo tras un umbral de 15. Observar como se detectan más manchas negras que en el caso de $\lambda = 5$, aunque también aparecen más puntos aislados, falsa alarma debida al ruido; además la falsa detección debida al brazo se incrementa.

5.3.5. Comparación entre los métodos de detección

La figura 5.19 muestra una imagen original de la secuencia *Sanga* y las máscaras obtenidas utilizando los diferentes cierres: temporal, por reconstrucción y por λ -reconstrucción ($\lambda = 5$). Lo dicho anteriormente se comprueba en este ejemplo, es decir, el cierre temporal es el que más manchas detecta, pero también es el que más falsas alarmas produce, hecho que puede comprobarse en este ejemplo porque el brazo del personaje al moverse, que es un objeto oscuro en movimiento, es detectado como mancha. El cierre por reconstrucción es el que menos manchas detecta, sobre todo las conectadas a zonas oscuras de la imagen, de nivel de intensidad similar al de la mancha (mancha conectada al cabezal de la cama), aunque por contra, es el que menos falsas alarmas produce, no detecta el brazo. Por último, la λ -reconstrucción es una solución intermedia controlable con el valor de λ . Estos métodos se utilizarán para la interpolación, pero no directamente los resultados de las aperturas (cierres) sobre toda la imagen, sino sólo en aquellas zonas dañadas, que serán las que indicará la máscara.

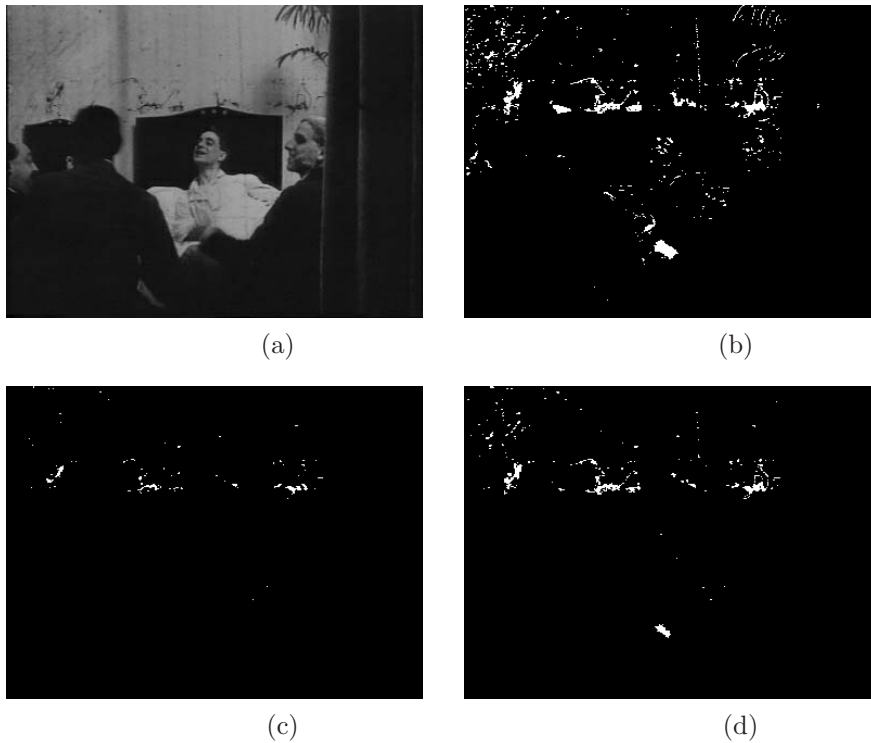


Figura 5.19: Comparación entre métodos de detección de manchas oscuras: (a) Fotograma original de la secuencia *Sanga*. (b) Máscara obtenida con el método del cierre temporal. (c) Máscara obtenida con el método del cierre por reconstrucción. (d) Máscara obtenida con el método del cierre por λ -reconstrucción con $\lambda = 5$.

Una vez visto el funcionamiento y los resultados de los diferentes métodos de detección podemos concluir que su funcionamiento dependerá de la cantidad de movimiento existente en la secuencia, de la siguiente forma:

- Si no existe movimiento o el movimiento es pequeño, el método de la apertura (cierre) temporal es el que mejor funciona, ya que tiene mayor probabilidad de detección, y sólo falla en el caso de objetos pequeños que se mueven rápido o cuando hay movimientos globales de escena.
- Si existe movimiento, la mejor opción son los operadores por λ -reconstrucción aumentando el valor de λ si el movimiento es menor (método que se aproxime al de los operadores temporales), y disminuyendo dicho valor si el movimiento aumenta (a medida que disminuye λ) se va acercando a la apertura (cierre) por reconstrucción ($\lambda = 0$). Estos métodos puede que no detecten todas las manchas, pero respetan la información original sana.

Si una secuencia tiene mucho o poco movimiento será una información que debe dar el usuario, ya que la determinación de la cantidad de movimiento de una secuencia se basa en diferencias entre fotogramas, y en el caso que nos ocupa las diferencias pueden deberse al deterioro de las imágenes, es decir, imágenes con muchas manchas, las cuales harían que se obtuviese una información de movimiento errónea.

El objetivo de determinar la máscara de daños es poder restaurar sólo aquello dañado, sin modificar la información correcta de la secuencia. De todo lo dañado, lo que nos interesará modificar será aquello que se aprecie a la hora de la visualización: manchas de un tamaño considerable, con un contraste notable con respecto a su fondo y una varianza espacial pequeña y que varían rápidamente en el tiempo. De todas estas características, con los métodos anteriores hemos impuesto sólo la propiedad de variación rápida en el tiempo. Para mejorar la detección, a los píxeles etiquetados como mancha en la máscara, vamos a exigirles que cumplan el resto de propiedades: alto contraste y varianza espacial pequeña.

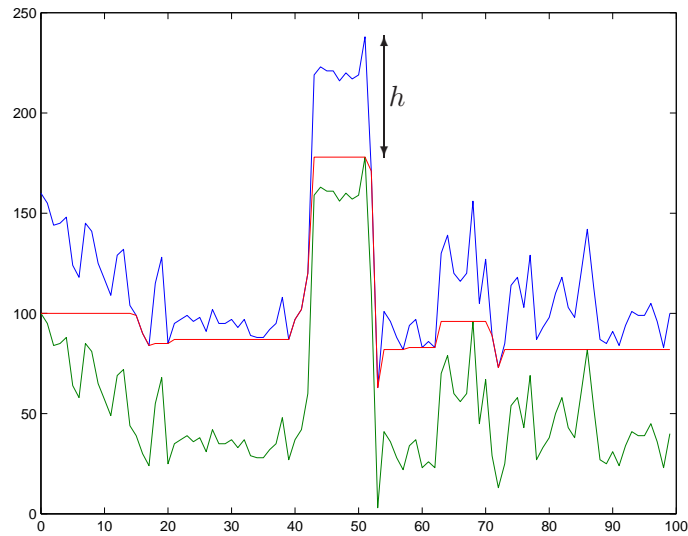
5.3.6. Detección de zonas de alto contraste: h -reconstrucción

De todas las zonas detectadas como manchas mediante los métodos anteriores sólo nos interesan aquellas que pertenezcan a una zona de la imagen con alto contraste. Para ello vamos a aplicar un método basado en la h -reconstrucción.

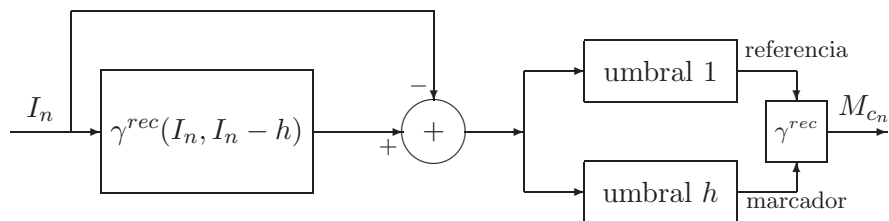
La h -reconstrucción [59] es un operador conexo que elimina de una imagen todas aquellas zonas claras con un contraste menor o igual que h , recortando por este valor los que sí lo superan. Este operador se implementa realizando la apertura por reconstrucción de una imagen tomando como marcador la misma imagen a la que se le ha reducido el brillo por un factor h . La figura 5.20 muestra las señales involucradas así como el resultado obtenido, para el caso de señales unidimensionales. También existe el operador dual, es decir, el que elimina aquellas zonas oscuras de la imagen que no superan en contraste el valor especificado.

Lo que buscamos son aquellas zonas claras que tienen un contraste como mínimo igual a h . Para obtenerlas se propone el siguiente método, el cual aparece esquematizado en la figura 5.21:

1. h -reconstrucción del fotograma actual (imagen A), obteniendo la imagen B.
2. Residuo entre las imágenes A y B (original y h -reconstruida); se obtiene C.

Figura 5.20: h -reconstrucción

3. Aplicación de un umbral al residuo con un valor de 1; se obtienen todas las zonas en las que la imagen original tienen alguna diferencia con la reconstruida (imagen D). De estas zonas sólo nos interesarán aquellas que tengan algún punto con diferencia mayor o igual que h , para ello debemos extraer estos puntos, y reconstruir de la imagen D sólo aquellas zonas que contengan alguno de estos máximos. Esto lo conseguiremos mediante los dos pasos siguientes.
4. Se aplica un umbral al residuo con un valor de h , obteniendo la imagen E, que contendrá los máximos que tengan un contraste en la imagen original mayor o igual que h .
5. Reconstrucción de la imagen D tomando como marcador E.

Figura 5.21: Obtención de máscara de zonas claras mediante la apertura con h -reconstrucción.

La figura 5.22 muestra una imagen original de la secuencia *Taronja*, la imagen resultado de la h -reconstrucción con un valor de $h = 70$, el residuo, el residuo tras un umbral de 1 y tras un umbral de 70, y el resultado final del proceso, una máscara de las zonas claras de la imagen original con un contraste mayor o igual que 70.

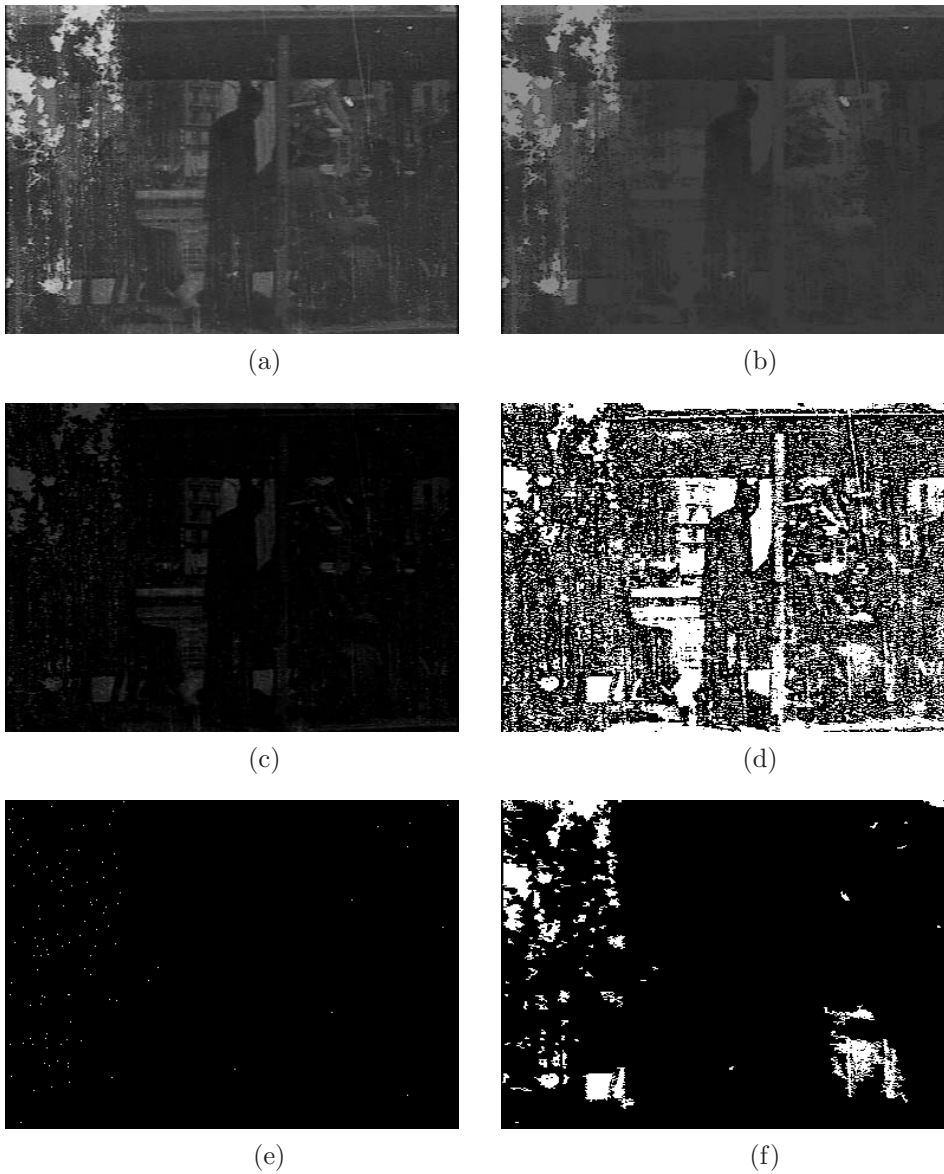


Figura 5.22: Detección de zonas claras mediante h -reconstrucción: (a)Fotograma original de la secuencia *Taronja*. (b)Resultado de la h -reconstrucción con $h=70$. En esta imagen se han eliminado todas las zonas claras con menor contraste que 70, y las que tienen mayor contraste se han transformado en una zona plana de ese valor. (c)Residuo de la h -reconstrucción. (d)Umbral del residuo de valor 1. Aparecen todas las zonas que en la imagen h -reconstruida tienen alguna diferencia con la original. (e) Umbral del residuo de valor 70. Máximos de la imagen original con un contraste mayor o igual que 70. (f) Resultado de la reconstrucción de la imagen obtenida en (d) utilizando como marcador la obtenida en (f). Zonas de la imagen original con un contraste superior a 70.

5.3.7. Obtención de la máscara de manchas final

Hasta ahora han sido obtenidas, para el fotograma actual (I_n), dos máscaras para cada tipo de mancha:

- Máscara que indica las zonas claras de la imagen con gran incorrelación temporal y la que indica las zonas oscuras, obtenidas utilizando la apertura (cierre) temporal, la apertura (cierre) por reconstrucción o la apertura (cierre) por λ -reconstrucción.
- Máscara que indica las zonas claras de la imagen con alto contraste y la que indica las zonas oscuras.

La máscara final será la combinación de ambas. Es decir, de la máscara que determina las zonas con gran variación temporal sólo nos interesarán aquellas que tengan un alto contraste espacial. Para ello, reconstruiremos la primera máscara de zonas claras utilizando como marcador la segunda de zonas claras, y el mismo procedimiento se llevará a cabo con las máscaras de manchas oscuras. Además, de todas las zonas detectadas mediante este proceso eliminaremos de la máscara aquellas cuya varianza espacial supere un determinado umbral. El proceso completo para la detección de manchas claras se muestra en la figura 5.23.

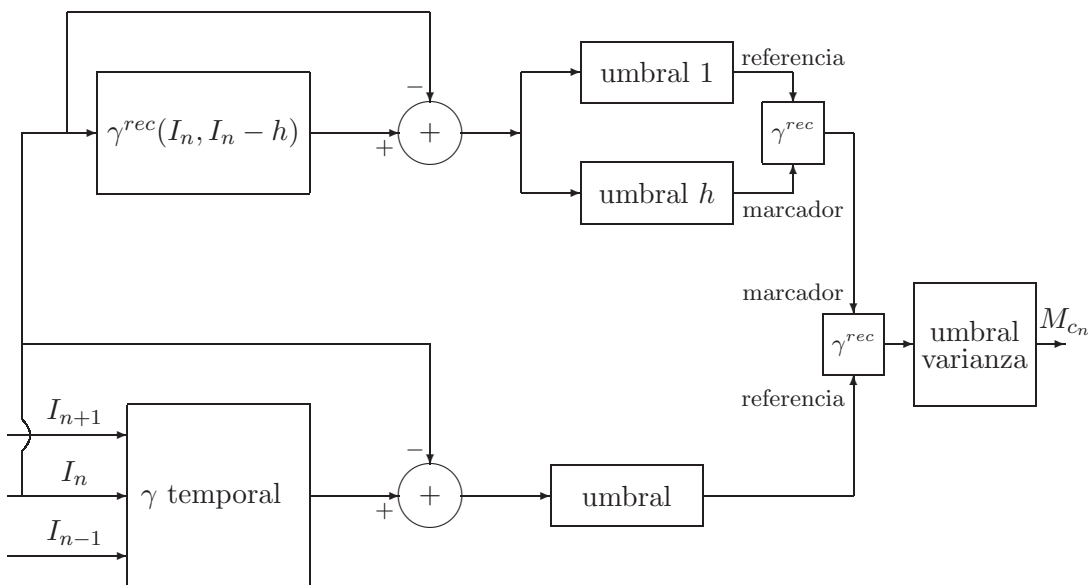


Figura 5.23: Procedimiento global de obtención de máscara de manchas claras.

En la figura 5.24 podemos ver una imagen original de la secuencia *Taronja* y las máscaras obtenidas para cada método, es decir, la que contiene zonas con gran variación temporal, y la que contiene zonas con alto contraste. La primera ha sido obtenida utilizando el método de la apertura temporal, y como puede observarse, aparecen falsas alarmas en la barra vertical del tranvía que se mueve rápidamente, artefacto que ya vimos al estudiar este método. La segunda máscara ha sido obtenida mediante la h -reconstrucción con un umbral de contraste

$h = 70$. Por último se presenta la máscara resultante de todo el esquema global de detección, y en ella puede observarse que muchas falsas alarmas de la máscara temporal han desaparecido.

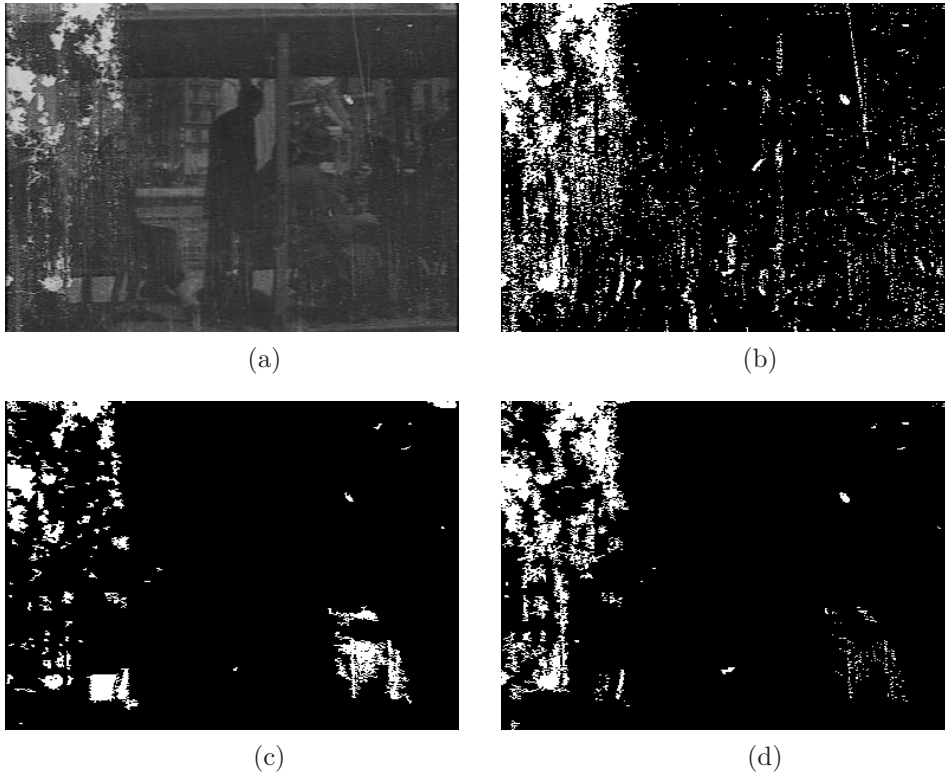


Figura 5.24: Detección de manchas claras: (a)Fotograma original de la secuencia *Taronja*. (b)Máscara obtenida mediante el método de la apertura temporal.(c)Máscara obtenida mediante la h -reconstrucción. (d)Máscara resultante del proceso global de detección.

Con este proceso se han seleccionado las manchas en función de su contraste y varianza espaciales, y de su variación temporal. También podría pensarse en seleccionarlas en función de su tamaño, eliminando aquellos objetos cuya área no superase un umbral; o puede que también nos interesasen las manchas con un área intermedia entre un máximo y un mínimo, para eliminar falsas alarmas debidas a objetos grandes en movimiento. Todo ello supondría una fase posterior a la detección. En los ejemplos mostrados no se ha realizado postprocesado, ya que eliminar las manchas de área menor de un determinado tamaño es irrelevante porque es imperceptible el resultado de restaurar o no manchitas muy pequeñas, y no hacerlo nos permite ahorrar tiempo de proceso. En cuanto al umbral de área máximo, tampoco se ha utilizado, ya que en los ejemplos de películas de nuestra base de datos existen manchas muy grandes que no serían restauradas.

5.4. Interpolación de manchas

Una vez finalizada la fase de detección y obtenida la máscara de manchas, el siguiente paso del proceso de restauración es la interpolación. Como se dijo al principio de este capítulo, el modelo que se asume para la secuencia original $y(x, y, t)$ es:

$$y(x, y, t) = (1 - d(x, y, t)) z(x, y, t) + d(x, y, t) c(x, y, t),$$

donde $z(x, y, t)$ es la secuencia con manchas, $d(x, y, t)$ la máscara obtenida en el proceso de detección y $c(x, y, t)$ la información que reemplazará a las manchas. En el proceso de interpolación lo que se pretende es obtener $c(x, y, t)$.

Para obtener esta información podemos utilizar diferentes tipos de métodos:

- Métodos temporales, que son los que se basan en la información que tenemos en fotogramas anteriores o siguientes al que se pretende restaurar.
- Métodos espaciales, que son los que interpolan utilizando la información del vecindario espacial de la zona a restaurar.
- También se puede utilizar una combinación de ambos.

5.4.1. Métodos temporales de interpolación

En la sección dedicada a la detección de manchas vimos tres métodos basados en la morfología matemática que detectaban y eliminaban zonas con gran variación temporal: la apertura (cierre) temporal, la apertura (cierre) por reconstrucción y la apertura (cierre) por λ -reconstrucción. Al aplicar estos métodos sobre la secuencia degradada se obtenía una secuencia en la que la intensidad de las manchas había cambiado utilizando información de fotogramas anteriores y posteriores al actual. Sin embargo, también se producían artefactos en zonas no deterioradas, por lo que en la fase de detección se determinó una máscara para cada fotograma que indicaba aquellos puntos que en la fase de interpolación debían ser rellenados con nuevas intensidades. Estas nuevas intensidades pueden ser extraídas, como hemos dicho, del resultado de aplicar una de las tres aperturas, si se está restaurando una mancha blanca o de uno de los tres cierres, si lo que se interpola es una mancha oscura.

A la hora de elegir cuál de estos métodos temporales se va a utilizar para rellenar las intensidades que nos faltan, hay que tener en cuenta las ventajas e inconvenientes de cada uno de ellos y en qué casos es mejor usar uno u otro. Vimos que el método que más defectos eliminaba, ya que tenía mayor probabilidad de detección, era la apertura (cierre) temporal, pero en el caso de objetos pequeños en movimiento los eliminaba o los partía. Por el contrario, las aperturas (cierres) por reconstrucción no tienen este problema, pero no eliminan bien las manchas. El punto intermedio son los operadores por λ -reconstrucción variando el valor de λ en función del movimiento. A la hora de rellenar, la apertura (cierre) por reconstrucción rellenan la mancha con el valor máximo que tiene el marcador en esa zona, por lo que si dicha zona está conectada a una zona clara de la imagen en algún punto, se transforma en una zona plana que puede resultar muy visible, como veremos en el ejemplo.

Así pues, como decíamos en la fase de detección, el funcionamiento de los métodos anteriores dependerá del movimiento de la escena, concretamente del movimiento de la zona donde se encuentre la mancha a restaurar:

- Si no existe movimiento, el método que mejor funciona es el de la apertura (cierre) temporal.
- Si existe movimiento, los operadores por λ -reconstrucción, aumentando el valor de λ si el movimiento es menor y disminuyendo dicho valor si el movimiento aumenta, funcionan mejor.

Comparación entre los métodos temporales

La figura 5.25 muestra el resultado de la interpolación de un fotograma dañado de la secuencia *Taronja* utilizando los tres métodos temporales. Como se ha dicho arriba, las manchas aparecen mucho mejor interpoladas utilizando la apertura temporal, ya que en la apertura con reconstrucción se convierten en una zona uniforme muy visible, no tanto como la mancha ya que la intensidad es menor, pero son muy visibles. Los problemas que tiene la apertura temporal en el caso de secuencias con movimiento, como es el caso de este ejemplo en el que la barra vertical del tranvía se partiría, se ha reducido gracias a la mejora introducida en la fase de detección que consigue reducir las falsas alarmas en la máscara. El resultado de la λ -reconstrucción es una solución intermedia entre la apertura temporal y la apertura por reconstrucción.

5.4.2. Métodos espaciales de interpolación

Además de estos métodos basados en la información temporal para la detección-interpolación, existen métodos de interpolación espaciales que pueden ser utilizados para rellenar las manchas en las películas. Estos métodos, dada la máscara de daños, restauran la imagen solamente usando información espacial del mismo fotograma. Uno de ellos, el utilizado por nosotros, es el que se explica a continuación.

Este método de interpolación espacial fue presentado por Prades en su tesis [113, 9]. Se trata de un algoritmo donde la interpolación se realiza siguiendo un orden que depende de la función distancia al borde de la zona dañada, D ; es decir, se sigue un procesamiento priorizado de los píxeles donde la función prioridad, que indica el orden en que se va a procesar un píxel, viene dada por la distancia de éste al borde de la zona dañada. A mayor distancia al borde, más tarde será restaurado el píxel. Este tipo de interpolación funciona bien cuando la zona a restaurar es uniforme, si se trata de un contorno no lo respeta y el resultado no es muy satisfactorio. Para solucionar esto, Prades propone un algoritmo de procesamiento priorizado respetando contornos.

El algoritmo de interpolación distingue dos tipos de regiones: zonas lisas, la interpolación de las cuales se realiza utilizando procesamiento priorizado; y zonas con contornos, para lo cual habrá que modelar el contorno e interpolar la zona siguiendo dicho modelo. Previamente a la interpolación se implementa una fase de detección de contornos.

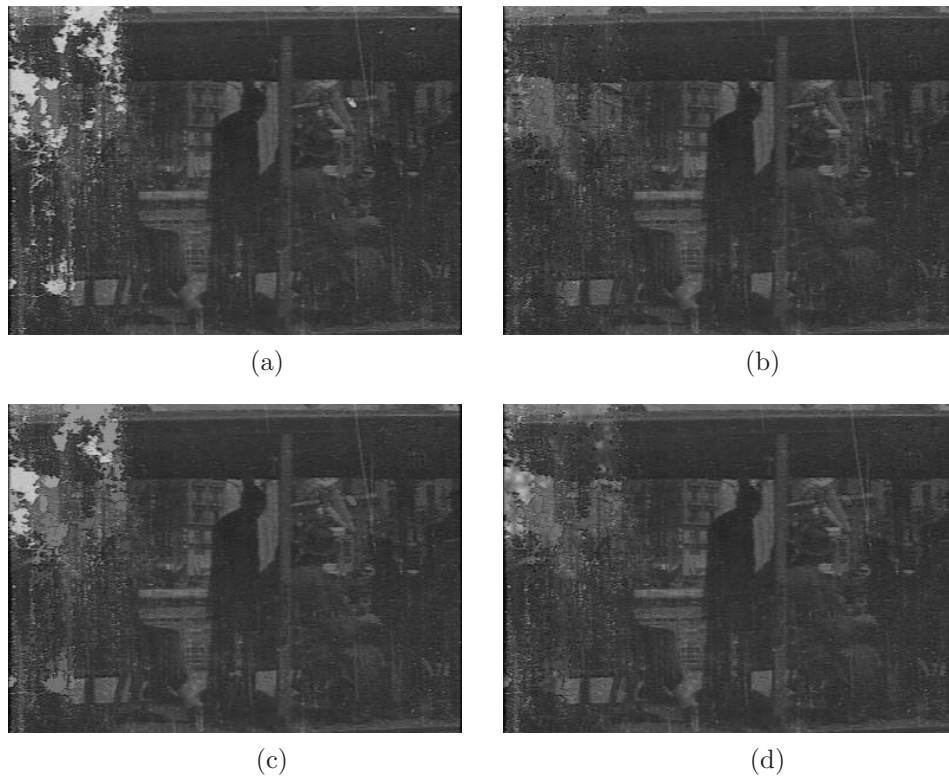


Figura 5.25: Interpolación de mancha mediante el método de interpolación espacial: (a)Fotograma original de la secuencia *Taronja*. (b)Restauración con la apertura temporal. (c)Fotograma una vez restaurado con la apertura por reconstrucción. Observar como las manchas no son eliminadas. (d)Fotograma una vez restaurada la mancha con la apertura por λ -reconstrucción con $\lambda = 15$. La restauración mejora con respecto a la apertura por reconstrucción.

El modelo de interpolación de contornos realiza una serie de hipótesis: que el contorno es estacionario, rectilíneo y de suavidad y contraste constantes. La capacidad del modelo para describir correctamente los contornos de una zona dañada es mayor cuanto menor es el tamaño de la zona, ya que cualquier contorno continuo puede aproximarse sin mucho error por uno rectilíneo de contraste y suavidad constantes.

Así pues la primera fase del algoritmo será detectar contornos para cada mancha. Si no existen la zona a interpolar es lisa. Si existen contornos, se estiman las direcciones de propagación y según éstas se interpola la zona dañada. Si la imagen a restaurar tiene de entrada buen contraste y buena calidad, para que se puedan detectar bien los contornos, los resultados son muy satisfactorios, como puede observarse en la figura 5.26. Sin embargo, en el caso que nos ocupa, la restauración en películas antiguas, se tiene un problema añadido, y es que las imágenes no tienen gran calidad ni buen contraste, y en la detección de contornos no se obtienen buenos resultados.

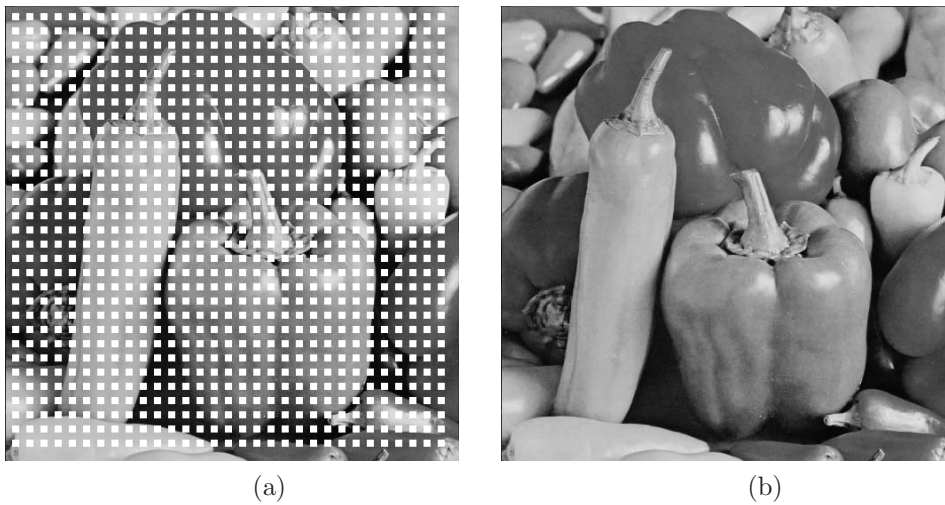


Figura 5.26: Restauración de una imagen fija mediante el método de restauración espacial (a)Imagen manchada. (b)Imagen restaurada.

5.4.3. Comparación entre métodos espaciales y temporales

Los métodos que se basan en la redundancia temporal tienen problemas cuando tenemos objetos en movimiento, sobre todo en las fronteras entre objetos en movimiento. Los métodos espaciales tienen problemas cuando la mancha a interpolar cruza uno o varios contornos. En la figura 5.27 se presenta un fotograma con una mancha simulada por la que pasa un contorno, y que pertenece a una secuencia que consiste en un trozo de sombrero de la imagen *Lena* que se mueve 2 píxeles en horizontal y 3 en vertical de un fotograma al siguiente, manteniéndose este desplazamiento constante para todos los fotogramas de la secuencia. Tenemos una zona a interpolar que es problemática para todos los métodos, y esta figura ilustra cómo se comportan. El resultado de la apertura temporal presenta un pequeño defecto en el borde restaurado, debido al movimiento del mismo. La apertura por reconstrucción no se comporta bien ya que la mancha está conectada con otra zona clara, por lo que se reconstruye como una zona plana con el valor del máximo del marcador en esa zona. La apertura por λ reconstrucción mejora el resultado anterior. En cuanto a los métodos espaciales, el que utiliza sólo procesamiento priorizado no reconstruye bien el contorno pero el que respeta los contornos lo reconstruye muy bien, ya que se trata de un contorno rectilíneo fácilmente modelable por el método.

La figura 5.28 muestra el resultado de la interpolación de unas manchas, las que se muestran en la máscara de daños, en un fotograma de la secuencia *Taronja* perteneciente a una película antigua. La mancha más grande se encuentra en un borde en movimiento, el borde del tranvía, y podemos ver un *zoom* de la zona de la mancha restaurada utilizando la apertura temporal y el procesamiento priorizado respetando contornos. Vemos que aunque las imágenes son de mala calidad, el resultado del método espacial no es tan poco satisfactorio como cabría esperar, y de hecho es mejor que la apertura temporal, porque ésta falla al intentar restaurar una zona en movimiento.

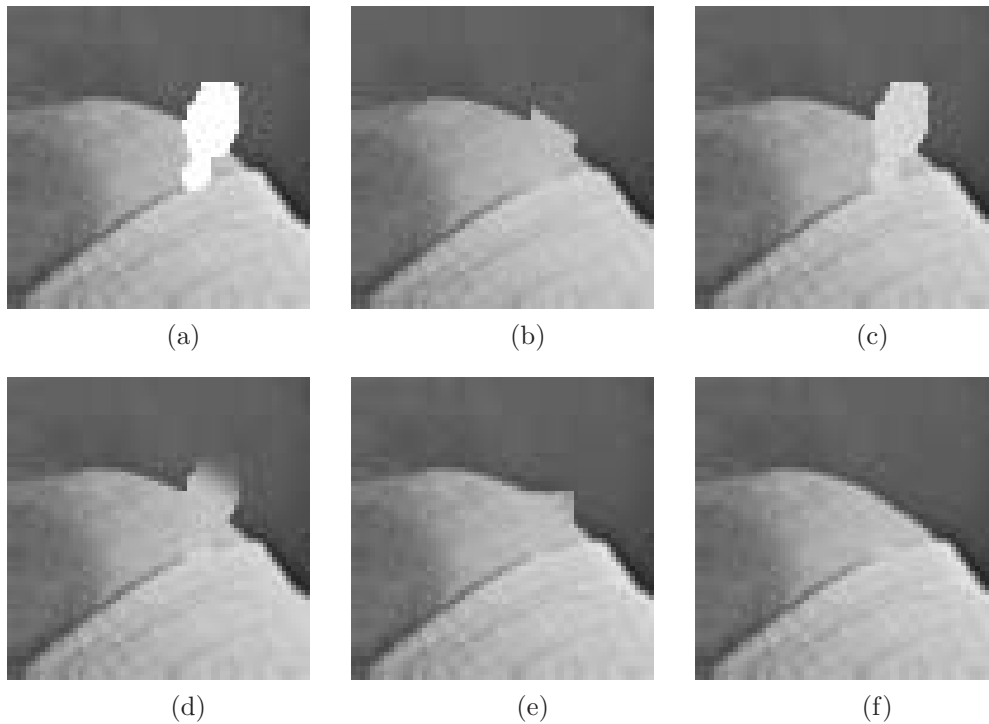


Figura 5.27: Restauración de manchas en un borde en movimiento: (a)Fotograma original de la secuencia en el que aparece la mancha. (b)Fotograma una vez restaurado usando la apertura temporal. Observar el pequeño defecto en el borde en movimiento. (c)Fotograma una vez restaurado usando la apertura por reconstrucción. La mancha se convierte en una zona plana de otro nivel de gris y muy visible. (d)Fotograma una vez restaurado usando la apertura por λ -reconstrucción con $\lambda = 20$. Se trata de una solución intermedia entre las dos anteriores. (e)Fotograma una vez restaurado usando el procesamiento priorizado. Observar el defecto en el borde. (f)Fotograma una vez restaurado usando el procesamiento priorizado respetando contorno. Observar que el borde se reconstruye satisfactoriamente, ya que se trata de un borde rectilíneo.

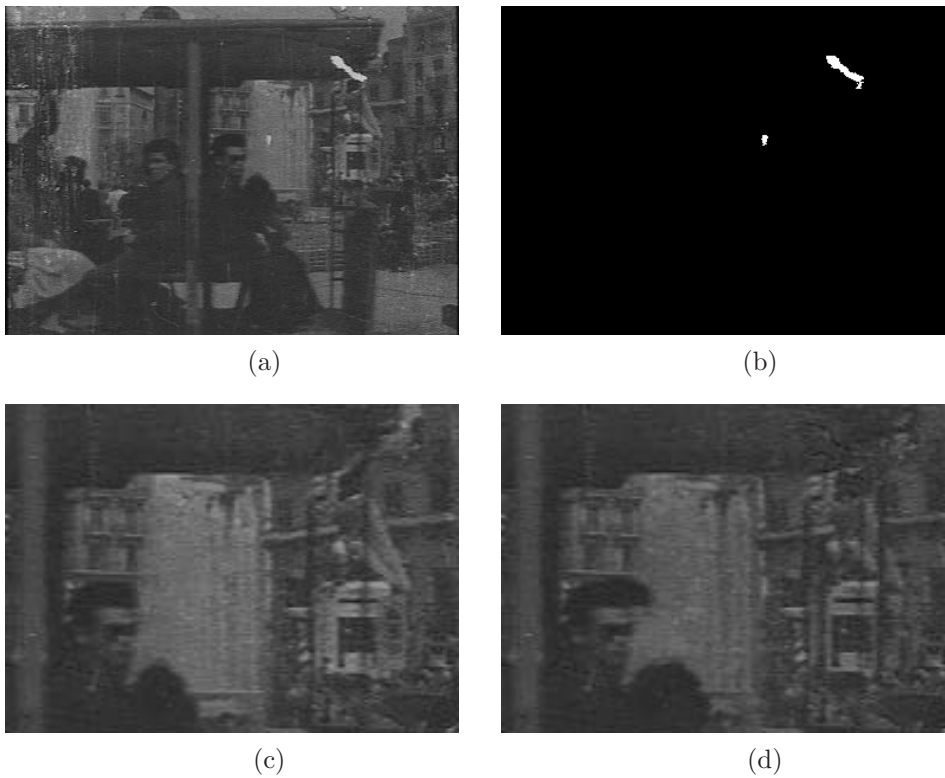


Figura 5.28: Interpolación de mancha mediante los métodos de interpolación espacial y temporal: (a)Fotograma original de la secuencia *Taronja*. (b)Máscara que indica las zonas a interpolar. (c)Fotograma una vez restauradas las manchas con el método de interpolación espacial, *zoom* de la zona de la mancha grande, que es la que está en una zona en movimiento, y en la que se puede ver la diferencia entre los métodos. (d)Fotograma una vez restaurada la mancha con la apertura temporal, *zoom* de la zona de la mancha.

5.5. Resultados obtenidos

La figura 5.29 muestra el resultado de la restauración de una secuencia mediante el método de la apertura temporal. La figura muestra a la izquierda fotogramas de la secuencia *Taronja*, y a la derecha los fotogramas restaurados utilizando la apertura temporal. En el proceso, se ha detectado primero la máscara de daños, utilizando la apertura temporal y mejorando la máscara eliminando zonas de bajo contraste y mucha varianza, mediante el esquema de la figura 5.23. Posteriormente, se ha interpolado la información con el resultado de la apertura temporal de tamaño 2, sólo en las zonas dañadas que indica la máscara.

5.6. Conclusiones

En este capítulo se presentan métodos de eliminación de defectos incorrelados en el tiempo que detectan e interpolan manchas oscuras y claras en un solo paso. Sin embargo, la restauración se ha dividido en dos: la detección y la interpolación. Con la detección se obtiene una máscara binaria que indica dónde se encuentran las manchas, y son precisamente esos píxeles los que van a cambiar su intensidad errónea por el resultado de uno de los métodos propuestos. Esto se hace así para respetar al máximo la información original y sana de la película y evitar que en las zonas correctas se produzcan artefactos.

Uno de los problemas principales que presentan los métodos de restauración de este tipo de defectos propuestos en la literatura, es la necesidad de la estimación de movimiento en la fase de detección y también en la de interpolación, aunque esta última no es tan problemática porque ya se sabe donde están las zonas dañadas y no se tienen en cuenta en la estimación. El problema de estimar el movimiento en la fase de detección es que las secuencias de las que disponemos, y las cuales tenemos que restaurar, tienen zonas, las manchas, donde falta la intensidad de los píxeles situados en esas posiciones, y por tanto, los métodos de estimación de movimiento fallan.

En este capítulo hemos realizado un estudio de tres métodos de detección de manchas que detectan zonas con gran variación temporal, basados en morfología matemática, que no necesitan estimar el movimiento para obtener la máscara de defectos. La conclusión extraída de este estudio es que el método de la apertura (cierre) temporal funciona bien si no existe movimiento ya que detecta y elimina, en la interpolación, más cantidad de manchas. Sin embargo, cuando existe movimiento, aparecen muchas falsas alarmas que se traducen en la rotura o eliminación de objetos, en resumen, en la aparición de artefactos tanto en zonas no dañadas (falsas alarmas) como en las dañadas si están en una zona en movimiento. El método de la apertura (cierre) por reconstrucción tiene el problema contrario, es decir, disminuye las falsas alarmas pero también las detecciones y, además, las manchas reconstruidas no se reconstruyen bien si están conectadas a zonas de la imagen de nivel de intensidad semejante. Una solución intermedia se propone utilizando la apertura (cierre) por λ -reconstrucción. Además, para reducir las falsas alarmas, más acusadas en el caso de la apertura (cierre) temporal, se ha propuesto un método de detección completo que no sólo tiene en cuenta la rápida variación temporal de las manchas, sino que además exige que la mancha sea una zona de alto contraste, utilizando un método basado en la h -reconstrucción, y de baja varianza

espacial.

A la hora de interpolar las manchas, el algoritmo utilizado en esta tesis sólo necesita una indicación de si el plano a restaurar tiene mucho o poco movimiento para aplicar el método que mejor funciona en cada caso: si existe poco movimiento lo mejor es utilizar un método temporal, pero si el movimiento es muy grande pero existe poco detalle (pocos contornos espaciales) lo mejor es el método espacial; en caso de mucho detalle y mucho movimiento podría ser conveniente una combinación (por ejemplo una media ponderada) del resultado de ambos tipos de métodos.

Con estos métodos, poco costosos computacionalmente, ya que están implementados de manera eficiente, se consiguen resultados satisfactorios, de los que han sido presentados unos ejemplos.



Figura 5.29: Secuencia *Taronja* (a)Original (b)Resultado de la restauración con apertura temporal.

Capítulo 6

Sistema Global

6.1. Introducción

En los capítulos anteriores de esta tesis se han presentado una serie de algoritmos para eliminar o minimizar diferentes defectos existentes en una película antigua, como son : variaciones indeseadas de intensidad, manchas y ruido de grano.

El problema con el que nos enfrentamos, una vez implementados los algoritmos que resuelven cada defecto, es el orden en que vamos a ejecutar cada uno de ellos, dentro del proceso global de restauración. Aunque a priori este problema parece trivial, no lo es tanto, ya que el resultado de corregir un determinado defecto puede afectar tanto de manera perjudicial como beneficiosa en la corrección de otro, o no afectar. Así pues, para determinar el orden en que vamos a procesar la secuencia a restaurar tendremos en cuenta qué algoritmos hacen que el resultado de su aplicación sea beneficioso para el éxito de la eliminación de otro defecto, antecediendo el primer método al segundo. A continuación se presenta un esquema global del proceso debidamente justificado, así como una serie de ejemplos que muestran secuencias originales y restauradas mediante el esquema y los algoritmos propuestos.

6.2. Diagrama de bloques del sistema

El proceso de restauración de una película se divide en tres fases, que son:

1. Digitalización de la película.

Esta primera fase consistirá en pasar las películas de formato analógico a formato digital. De esta forma podremos pasar a trabajar con ellas realizando, en primer lugar, una base de datos de las secuencias en la que se guarda información como duración, número de planos diferentes, efectos de transición, catálogo de daños que contienen, etc.

2. Segmentación de la secuencia en planos.

Una vez que la secuencia puede ser tratada por el ordenador, el primer paso es segmentarla en diferentes planos, entendiéndose por plano como el conjunto de fotogramas

que suceden entre que la cámara empieza y termina de grabar. La segmentación consistirá en detectar las diferentes transiciones que permiten el paso de un plano a otro, como: cortes (transiciones bruscas), fundidos entre imágenes, fundidos a un color (negro o blanco) y cortinillas. Es conveniente esta segmentación previa ya que todos los fotogramas de un plano tendrán características comunes, como la continuidad temporal de la intensidad, que nos van a permitir llevar a cabo el proceso de restauración. A la detección de transiciones y segmentación en planos se ha dedicado el capítulo 2 de esta tesis.

3. Eliminación de defectos en cada plano.

La figura 6.1 muestra el diagrama de bloques del proceso de restauración de cada plano de una secuencia perteneciente a una película antigua. Los bloques sombreados en la figura son los que han sido investigados e implementados en esta tesis. La ordenación de bloques viene determinada por la interacción entre los diferentes procesos, es decir, lo que afecta a un proceso realizar otro antes que él.

- Corrección del brillo: se realiza en primer lugar ya que mejora la coherencia temporal de la secuencia, es decir, elimina las diferencias entre píxeles de fotogramas consecutivos entre los que ocurre un cambio global e indeseado de intensidad. Mejorar esta coherencia será muy importante a la hora de restaurar defectos locales como rayas o manchas. Si intercambiásemos el orden de restauración de brillo y defectos locales, la corrección de brillo puede que mejorara aunque no significativamente, ya que la manchas o rayas ocupan un porcentaje pequeño de píxeles de la imagen, lo cual no es significativo en el cálculo y modificación del histograma. Por tanto, por un lado tenemos una mejora no muy significativa que supondría el orden de restauración “defectos locales-brillo” y por otro una aportación fundamental para la detección de manchas con el orden “brillo-defectos locales”, por lo que se realiza antes la corrección de brillo que la de defectos locales. En cuanto al resto de procesos, mejora el resultado de la estimación de movimiento para eliminar la vibración.
- Eliminación de manchas: se eliminan aquellas zonas conexas máximas o mínimas con incoherencia temporal. Este proceso beneficiará a la estimación de movimiento para estabilizar la secuencia. Se realiza, tanto manchas como rayas antes que el filtrado que elimina el ruido estacionario, porque los artefactos que se introduzcan en el proceso de eliminación de defectos locales afectarán sólo a un porcentaje de píxeles mientras que los que introduce el filtrado espacio-temporal influyen en todos los píxeles de la imagen por lo que todos los procesos que se realicen después se verán afectados por estos artefactos en gran medida.
- Eliminación de rayas: una forma de eliminar estos defectos podría ser detectar y posteriormente interpolar zonas conexas máximas o mínimas de la imagen con una determinada geometría, normalmente rayas verticales estrechas, con una coherencia temporal. La previa eliminación de manchas simplifica el problema de detección de rayas, ya que se reduce la cantidad de regiones máximas (mínimas) de la imagen. La eliminación de rayas también favorece a la corrección de la vibración.

- Eliminación de ruido de grano: se realiza después de la detección de manchas y rayas porque, como hemos dicho, los artefactos introducidos en este proceso afectan a todos los píxeles, y se arrastrarán en la interpolación de las manchas. Además como efecto del filtrado, para mantener la probabilidad de detección de manchas con respecto al orden inverso, habría que bajar el umbral en la obtención de la máscara, aumentándose las falsas alarmas.
- Eliminación de la vibración: aunque no está reflejado en esta memoria, ya se está trabajando en el tema de eliminación de vibración. Para ello, se estima el movimiento global entre fotogramas utilizando un método de bloques multiresolución, y se elige como movimiento global el que se obtiene para la mayoría de los bloques. Posteriormente se realiza una fase de procesado de los vectores de movimiento para diferenciar entre vibración brusca, es decir, movimiento global que nos interesa eliminar, y movimiento real de la escena, como por ejemplo un barrido de cámara. Por último se corrigen los fotogramas para eliminar la vibración detectada.

Para que el error en la estima del movimiento disminuya la calidad de las imágenes debe ser buena, sin ruido, sin manchas, sin parpadeo, es por esto por lo que la vibración se elimina en último lugar.

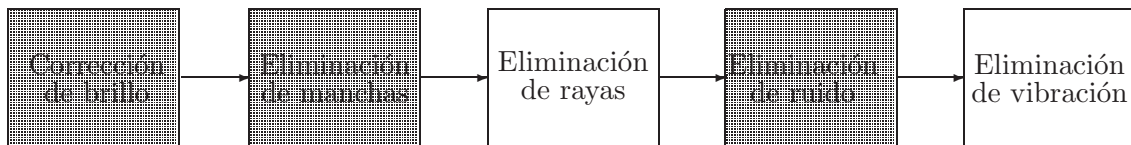


Figura 6.1: Esquema general del proceso de restauración para cada plano de la secuencia. Los bloques sombreados son los que han sido implementados en esta tesis.

En el esquema general de la figura 6.1, para el algoritmo global implementado en esta tesis, los diferentes procesos representados por cada bloque pueden ser activados o desactivados por el usuario en función de los tipos de defectos que observe en el plano a restaurar; incluso en el bloque de eliminación de manchas se puede elegir eliminar manchas blancas, negras o las dos. Este planteamiento se hace para que los artefactos que se introducen al eliminar un defecto no se introduzcan en la película intentando eliminar un defecto que de entrada no está presente. Además, el usuario debe introducir una serie de parámetros: tamaños de filtrados, valor de λ , cantidad de movimiento, umbrales, etc., que hacen que el funcionamiento del sistema, al igual que la mayoría de los existentes, no sea totalmente automático; aunque la fase de introducción de ciertos parámetros se podría eliminar y partir de unos por defecto, ya que muchos de los parámetros no dependen del contenido de las imágenes de la secuencia a restaurar; si se incluyen es para que el restaurador tenga el control en el proceso de restauración.

6.3. Resultados

La figura 6.2 muestra unos fotogramas de un plano de la secuencia *Taronja* originales y restaurados. En la restauración de este plano se han eliminado los siguientes defectos:

- Brillo: tamaño del filtrado temporal de los histogramas: 9
- Manchas blancas: apertura temporal, umbral máscara=15.
- Ruido: filtrado espacio-temporal de tamaño $3 \times 3 \times 5$.

Las figuras 6.3 y 6.4 muestran fotogramas originales y restaurados de un plano de la secuencia *Sanga*, de la película *Sangre y arena*. En la restauración de este plano se han eliminado los siguientes defectos:

- Brillo: tamaño del filtrado temporal de los histogramas: 9
- Manchas blancas: apertura con λ -reconstrucción ($\lambda = 20$), umbral máscara=15.
- Manchas negras: cierre con λ -reconstrucción ($\lambda = 20$), umbral máscara=15.
- Ruido: filtrado espacio-temporal de tamaño $3 \times 3 \times 5$.

En estas figuras puede observarse mejoras en la calidad de las imágenes una vez restauradas, sin embargo, la verdadera mejora se observa cuando se visualiza la secuencia a la velocidad que va a ser vista por el espectador (25 imágenes/segundo), ya que como hemos visto en los capítulos anteriores, hay algoritmos que basan su funcionamiento en las características del sistema visual humano, que va a ser el observador de la película. Secuencias originales y restauradas en formato “.avi” pueden encontrarse en la página WWW <http://gpiserver.dcom.upv.es/restoration.html>.

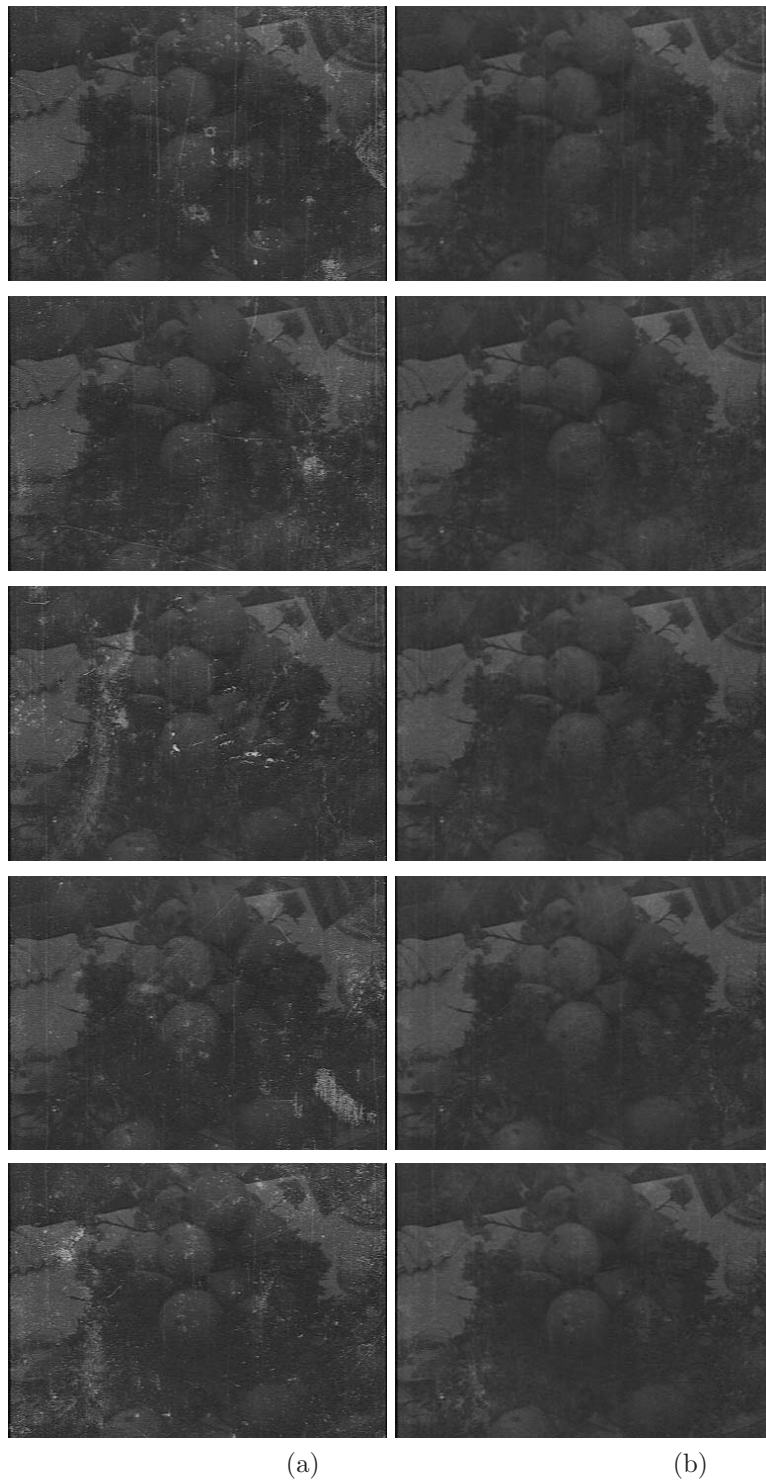


Figura 6.2: Secuencia *Taronja* (a)Original (b)Resultado de la restauración.



Figura 6.3: Secuencia *Sanga* (a)Original (b)Resultado del proceso de restauración.



Figura 6.4: Secuencia *Sanga* (a)Original (b)Resultado del proceso de restauración.

Capítulo 7

Conclusiones y líneas futuras

7.1. Introducción

En la presente tesis han sido desarrollados una serie de algoritmos destinados a corregir diferentes defectos que presentan las películas antiguas. De todos los posibles defectos a restaurar han sido implementados los métodos que corrigen las variaciones bruscas de intensidad, los que reducen el ruido estacionario y los que eliminan defectos incorrelados en el tiempo.

Un paso previo necesario a la eliminación de defectos es la segmentación de secuencias en planos, ya que una de las propiedades que se utilizará a la hora de la restauración será la correlación existente entre las imágenes que pertenecen al mismo plano. En esta tesis también se han desarrollado algoritmos para la segmentación en planos de las secuencias.

Para el desarrollo de los algoritmos que resuelven estos problemas se han tenido en cuenta las restricciones siguientes:

- Preservar la información correcta que hay en la película, en la medida de lo posible y, además, corregir sólo aquello que vaya a ser perceptible por el espectador, para lo que se ha tenido siempre presente la respuesta espacio-temporal del sistema visual humano.
- Intentar no realizar estimación de movimiento en ninguno de los algoritmos, ya que las imágenes de las que partimos están muy deterioradas, lo cual supone una causa de fallo muy importante en las técnicas de estimación de movimiento, además del notable incremento del coste computacional.
- Implementar algoritmos optimizados computacionalmente y sin necesidad de parámetros de diseño muy específicos, ya que ello posibilitará que el operador del sistema de restauración pueda ser una persona poco formada en el tema.

7.1.1. Conclusiones a la segmentación de secuencias

Para la segmentación de secuencias en planos se proponen métodos que se basan en técnicas adaptadas a la calidad de las imágenes de las que partimos. Un problema importante

que tienen las secuencias procedentes de películas antiguas son los cambios bruscos de luminosidad, hecho que provoca en los métodos de segmentación que aparecen en la literatura la aparición de falsas alarmas. Nuestros métodos, en el caso de la detección de cortes y fundidos, minimizan este problema [10] utilizando métricas que calculan solamente la varianza temporal de la luminosidad sin tener en cuenta la media de las imágenes.

En esta memoria, se han presentado los diferentes métodos que se han utilizado en esta tesis a la hora de segmentar las secuencias en planos, para detectar cortes, fundidos y cortinillas. Aunque se ha expuesto un método para cada tipo de transición, el trabajo más exhaustivo y minucioso se ha llevado a cabo en la detección de cortes. Los motivos que han llevado a esto son diversos:

- Las transiciones más frecuentes entre planos son los cortes, sobre todo en el caso que nos ocupa, las películas antiguas.
- Para restaurar las películas (objetivo de esta tesis) la condición necesaria para que los algoritmos sean efectivos es que haya continuidad de la información en el tiempo (que el nivel de gris de un píxel varíe lentamente), hecho que sólo no sucede cuando existe un corte (o en una zona dañada, la cual se restaurará). En el resto de transiciones la continuidad temporal se cumple, ya que se trata de cambios graduales.

Es por todo ello, como se deriva de los resultados obtenidos, para el cumplimiento de nuestro objetivo, mucho más exitosa la detección de cortes que la del resto de transiciones, y es por tanto muy importante que en esta detección se obtengan muy buenos resultados.

Nuestros métodos, sobre todo el de detección de cortes, que es el que más nos interesa, mejora los resultados con respecto a los que obtienen otros métodos propuestos en la literatura, incluso teniendo en cuenta que nuestro conjunto de secuencias de test es mucho peor, ya que se trata de secuencias procedentes de películas antiguas con imágenes de mala calidad.

7.1.2. Conclusiones de la corrección de defectos

Corrección del parpadeo

Para reducir el parpadeo se han propuesto dos métodos [104]: el primero está basado en una transformación afín de la intensidad para reducir las variaciones aleatorias y sinusoidales de la media y la varianza, calculando los parámetros de la transformación con un modelo de variación temporal de los dos estadísticos. El método consigue reducir estas variaciones, aunque los resultados visuales obtenidos no son los esperados. El segundo método presentado trata de mejorar los resultados visuales mediante una transformación no lineal basada en cambios en el histograma. El objetivo es transformar el histograma de cada imagen en un histograma de referencia calculado como la media del histograma del fotograma en restauración y el de sus vecinos.

Al comparar los resultados obtenidos con ambos métodos podemos observar una mejora de la apreciación visual de la secuencia cuando se utiliza el segundo método. En la restauración de la base de datos de secuencias utilizada en esta tesis, así como en el software final, se ha utilizado como método de eliminación del parpadeo el de cambio de histograma, debido a sus mejoras en la apreciación visual.

Reducción de ruido

El filtro diseñado en esta tesis para la reducción de ruido estacionario consigue el objetivo propuesto: eliminar ruido intentando conseguir que la degradación introducida no sea muy visible por el espectador. Se trata de un filtro lineal y FIR implementado de manera muy eficiente y basado en las propiedades del sistema visual humano, que reduce ruido por un factor considerable, emborronando sólo aquellos bordes espaciales donde el ojo no lo aprecie, en las zonas con movimiento. El factor de reducción conseguido es menor que el que obtiene un filtro espacial o temporal del mismo tamaño, pero también degrada menos la secuencia. El factor de reducción conseguido con nuestro filtro es comparable al obtenido por otros autores, pero la carga computacional es menor ya que no precisa estimación de movimiento y además está implementado de manera eficiente.

Sin embargo, este filtro no elimina las manchas, es decir ruido incorrelado temporalmente, con dimensiones mayores que las de la máscara espacial utilizada. Para solucionar este problema se implementa un método de restauración específico para este tipo de defectos.

Corrección de defectos incorrelados en el tiempo

Los métodos de eliminación de defectos incorrelados en el tiempo que han sido propuestos detectan e interpolan manchas oscuras y claras en un solo paso. Sin embargo, la restauración se ha dividido en dos fases: la detección y la interpolación. Con la detección se obtiene una máscara binaria que indica dónde se encuentran las manchas, y son precisamente esos píxeles los que van a cambiar su intensidad errónea por el resultado de uno de los métodos propuestos. Esto se hace así para respetar al máximo la información original y correcta de la película, y evitar que en las zonas correctas se produzcan artefactos.

Uno de los problemas principales que presentan los métodos de restauración de este tipo de defectos propuestos en la literatura es la necesidad de la estimación de movimiento en la fase de detección y también en la de interpolación, aunque esta última no es tan problemática porque ya se sabe donde están las zonas dañadas y no se tienen en cuenta en la estimación. El problema de estimar el movimiento en la fase de detección es que las secuencias de las que disponemos, y las cuales tenemos que restaurar, tienen zonas, las manchas, donde falta la intensidad correcta de los píxeles situados en esas posiciones, y por tanto, los métodos de estimación de movimiento fallan.

En esta tesis hemos realizado un estudio de tres métodos de detección de manchas que detectan zonas con gran variación temporal, basados en morfología matemática, que no necesitan estimar el movimiento para obtener la máscara de defectos. La conclusión extraída de este estudio es que el método de la apertura (cierre) temporal funciona bien si no existe movimiento ya que detecta y elimina, en la interpolación, más cantidad de manchas. Sin embargo, cuando existe movimiento, aparecen muchas falsas alarmas que se traducen en la rotura o eliminación de objetos, en resumen, en la aparición de artefactos tanto en zonas no dañadas (falsas alarmas) como en las dañadas si están en una zona en movimiento. El método de la apertura (cierre) con reconstrucción tiene el problema contrario, es decir, disminuye las falsas alarmas pero también las detecciones, y además las manchas reconstruidas no se reconstruyen bien si están conectadas a zonas de la imagen de semejante nivel de intensidad. Una

solución intermedia se propone utilizando la apertura (cierre) con λ -reconstrucción. Además, para reducir las falsas alarmas, más acusadas en el caso de la apertura (cierre) temporal, se ha propuesto un método de detección completo que no sólo tiene en cuenta la rápida variación temporal de las manchas, sino que además exige que la mancha sea una zona de alto contraste, utilizando un método basado en la h -reconstrucción, y de baja varianza espacial.

A la hora de interpolar las manchas, el algoritmo utilizado en esta tesis sólo necesita una indicación de si el plano a restaurar tiene mucho o poco movimiento, para aplicar el método que mejor funciona en cada caso: si existe poco movimiento, lo mejor es utilizar un método temporal. Si el movimiento es muy grande pero existe poco detalle (pocos contornos espaciales) lo mejor es el método espacial; en caso de mucho detalle y mucho movimiento, podría ser conveniente una combinación (por ejemplo una media ponderada) del resultado de ambos tipos de métodos.

Con estos métodos, poco costosos computacionalmente, ya que están implementados de manera eficiente, se consiguen resultados satisfactorios, de los que han sido presentados unos ejemplos.

Finalmente, se propone un algoritmo global que concatena los diferentes métodos presentados. Los diferentes bloques se concatenan de tal forma que los artefactos introducidos por la corrección de un determinado defecto no perjudiquen a la corrección de otros, sino por el contrario, un algoritmo antecede a otro si beneficia el resultado del segundo, o al menos no lo altera.

7.2. Líneas futuras

Posibles líneas futuras de investigación a seguir en el tema que compete a esta tesis serían:

- En cuanto a la segmentación de secuencias se podría seguir investigando para mejorar la detección de cortes, ya que es el tipo de transición que más problemas ocasiona en el proceso de segmentación.
- En cuanto a la eliminación de defectos, se podrían mejorar los métodos implementados en esta tesis en las siguientes líneas:
 - Estudio de los parámetros que nos permiten controlar el efecto del filtro de reducción de ruido. Para el filtro presentado vimos que los parámetros de diseño nos permitían establecer la velocidad máxima con la que se puede mover un objeto para que no sea emborronado, así como el emborronamiento máximo que sufre un objeto en la secuencia. Estos aspectos están relacionados con la respuesta del sistema visual humano, por lo que una línea futura de investigación podría ser hacer un estudio, con pruebas empíricas, de los valores óptimos de dichos parámetros para eliminar el máximo ruido posible sin que el espectador aprecie los artefactos introducidos por el filtro.
 - Estudio e implementación de otras técnicas de reducción de ruido que trabajen en función de la respuesta del sistema visual humano.

- Mejorar la interpolación de manchas, por ejemplo estimando el movimiento de la zona circundante a la de la mancha y encontrando una correspondencia de la misma en fotogramas adyacentes para, posteriormente, copiar la información de la zona de la mancha compensada en movimiento en la mancha en el fotograma actual.
- Otra línea futura de trabajo en la restauración de películas antiguas sería la corrección de defectos que no se han estudiado en esta tesis:
 - Eliminación de vibraciones indeseadas: aunque no está reflejado en esta memoria, ya se está trabajando en el tema de eliminación de vibración. Para ello, se estima el movimiento global entre fotogramas utilizando un método de bloques multiresolución, y se elige como movimiento global el que se obtiene para la mayoría de los bloques. Posteriormente se realiza una fase de procesado de los vectores de movimiento para diferenciar entre vibración brusca, es decir, el movimiento global que nos interesa eliminar, y movimiento real de la escena, como por ejemplo un barrido de cámara. Por último, se corrigen los fotogramas para eliminar la vibración detectada.
 - Eliminación de rayas: una forma de eliminar estos defectos podría ser detectar y posteriormente interpolar zonas conexas máximas o mínimas de la imagen con una determinada geometría y orientación, normalmente rayas verticales estrechas, con gran correlación temporal.
 - Detección e interpolación de fotogramas perdidos: para resolver este problema, el cual ya está siendo tratado por nosotros, se estudia la trayectoria del movimiento de los objetos de la escena para detectar un salto brusco en la misma. Para interpolar un fotograma se utilizan fotogramas anteriores y posteriores a los que hay que interporlar realizando una media ponderada de los mismos compensados en movimiento.

Apéndice A

Descripción de las secuencias utilizadas

A.1. Introducción

En este anexo se describirán las secuencias utilizadas para las pruebas de los algoritmos. Cómo ya indicamos en la introducción al proyecto, provienen de dos fuentes diferentes:

- Películas antiguas facilitadas por la Canal 9 (TV valenciana) procedentes de la filmoteca de Valencia (películas en formato SVHS que posteriormente se han digitalizado).
- Películas antiguas comerciales originales en formato VHS.
- Secuencias en formato *MPEG* procedentes de las bases de datos documentales de Radiotelevisión Española (películas actuales).

Estas secuencias están almacenadas y disponibles como listas de ficheros comprimidos en *JPG*. Para procesarlas sido filtradas y diezmadas x 8. Estas secuencias *dc* también se encuentran almacenadas y disponibles.

A.2. Secuencias y su descripción

- **DRAMA**: Fragmento de la serie de T.V. “Pepa y Pepe” (T.V.E).
- **MOVIE**: Fragmento de la película “La sombra del ciprés es alargada”(1995).
- **NEWSB**: Fragmento de “Informe Semanal” (T.V.E).
- **NEWSA**: Fragmento de un Telediario de T.V (T.V.E).
- **BASKET**: Fragmento de un partido de baloncesto: Real Madrid-Estudiantes (T.V.E).
- **CYCLING**: Fragmento de una etapa de ciclismo: Vuelta a España (T.V.E).

- **ALCOI:** Fragmento de una grabación documental sobre la empresa textil en Alcoi, de los años 40 (Canal 9 TV).
- **MALVALOCA:** Fragmento de la película “Malvaloca” (1942).
- **CHAPLIN:** Fragmento de la película “Charlot cambia de oficio” (1915).
- **TARONJA:** Fragmento de una grabación documental sobre ”Miss Naranja”, de los años 40 (Canal 9 TV).
- **SANGA:** Fragmento de la película “Sangre y arena” (1941).
- **SANGB:** Fragmento de la película “Sangre y arena”.
- **SANGC:** Fragmento de la película “Sangre y arena”.
- **TORBELLINO:** Fragmento de la película “Torbellino”.

Secuencia	Fotogramas	Efecto
DRAMA 3012 fotogramas	364-365	Corte
	665-666	Corte
	708-709	Corte
	1093-1094	Corte
	1147-1148	Corte
	1510-1511	Corte
	1665-1666	Corte
	1836-1837	Corte
	2081-2082	Corte
	2504-2505	Corte
2970-2971	Corte	
MOVIE 3010 fotogramas	282-283	Corte
	519-520	Corte
	605-606	Corte
	738-739	Corte
	998-999	Corte
	1119-1120	Corte
	1284-1285	Corte
	1322-1323	Corte
	1409-1410	Corte
	1514-1515	Corte
	1535-1536	Corte
	1910-1911	Corte
	2116-2117	Corte
2693-2694	Corte	

NEWSB 1497 fotogramas	522-523-524-525-526 624-625 743-744 1247-1248 1285-1286	Fundido Corte Corte Corte Corte
NEWSA 1907 fotogramas	37-38 75-76 97-98 116-117 130-131 152-153 168-169 282-303 334-341 557-558 760-761 875-876 1054-1055 1265-1266 1296-1297 1328-1329 1373-1374 1515-1516 1639-1653	Corte Corte Corte Corte Corte Corte Corte Cortinilla Horizontal Fundido Corte Corte Corte Corte Corte Corte Corte Corte Corte Cortinilla Vertical
BASKET 1502 fotogramas	21-22 166-167 223-224 309-310 408-409 513-514 1197-1198 1228-1229 1275-1276	Corte Corte Corte Corte Corte Corte Corte Corte Corte
CYCLING 1998 fotogramas	731-740 789-799 882-893 1312-1317 1931-1938	Fundido Fundido (aparece cartel") Fundido (desaparece cartel") Fundido Fundido
ALCOI 783 fotogramas	7-8-9 17-18 101-102 159-160	<i>Corte (con frame intermedio)</i> Corte Corte Corte

TARONJA 5450 fotogramas	2317-2318	Corte
	2444-2445	Corte
	2551-2552	Corte
	2652-2653	Corte
	2750-2751	Corte
	2834-2835	Corte
	2909-2910	Corte
	2981-2982	Corte
	3196-3197	Corte
	3320-3321	Corte
	3414-3415	Corte
	3515-3516	Corte
	3579-3580	Corte
	3687-3688	Corte
	3754-3755	Corte
	3872-3873	Corte
	3946-3947	Corte
	4025-4026	Corte
	4198-4199	Corte
	4337-4338	Corte
	4448-4449	Corte
	4523-4524	Corte
	4640-4641	Corte
	4734-4735	Corte
	4804-4805	Corte
	4814-4815	Corte
	4880-4881	Corte
	4928-4929	Corte
4987-4988	Corte	
5070-5071	Corte	
5197-5198	Corte	
5314-5315	Corte	
SANGA 78 fotogramas	29-30	Corte
SANGB 1877 fotogramas	12-13	Corte
	78-79	Corte
	178-179	Corte
	452-453	Corte
	773-774	Corte
	901-902	Corte
	996-1011	Fundido a negro
	1161-1162	Corte

	1347-1348 1479-1480 1526-1538 1638-1656 1831-1832 1891-1892	Corte Corte Fundido a negro Fundido a negro Corte Corte
SANGC 1231 fotogramas	9-10 170-171 829-830 928-929 1142-1143 1172-1173	Corte Corte Corte Corte Corte Corte
TORBELLINO 2214 fotogramas	0-56 135-145 423-430 434-435 928-929 673-697 835-836 1297-1306	fundido a negro + fundido de negro Fundido Fundido Corte Corte Cortinilla horizontal Corte Cortinilla vertical

Índice de figuras

1.1. Variación de brillo en dos fotogramas consecutivos de una secuencia.	3
1.2. Defecto correlado en el tiempo: raya.	4
1.3. Defecto incorrelado en el tiempo.	4
2.1. Ejemplo de transición abrupta: (a)Último fotograma de un plano. (b)Primer fotograma del plano siguiente.	11
2.2. Ejemplo de transición gradual (fundido con dos fotogramas intermedios): (a)Último fotograma de un plano. (b)Fotograma intermedio del fundido. (c)Fotograma intermedio del fundido. (d)Primer fotograma del plano siguiente.	12
2.3. Imagen “entera” dividida en bloques frente a la imagen dc	24
2.4. Imagen de <i>Lena</i> : (a)Imagen “entera” de 512×512 píxeles. (b)Imagen dc de tamaño 64×64 píxeles.	25
2.5. Diferencia de dos fotogramas consecutivos en la secuencia <i>Drama</i> : (a)Imagen diferencia de los fotogramas 363 y 364, pertenecientes al mismo plano. (b)Imagen diferencia de los fotogramas 364 y 365, que constituyen una transición abrupta.	27
2.6. d_c calculado para la secuencia <i>Drama</i>	28
2.7. Fotogramas de la secuencia <i>Chaplin</i>	30
2.8. Módulo de la diferencia para una subsecuencia de <i>Chaplin</i> . La falsa transición aparece marcada con un “*”.	31
2.9. $d_c(n)$ y $d(n)$ de 2500 fotogramas de la secuencia <i>Chaplin</i> . El * muestra dónde existe un verdadero corte.	32
2.10. Fotogramas 2320 y 2321 de la secuencia <i>Chaplin</i> . Véase el pico correspondiente en $d_c(n)$ en la figura 2.9, y cómo no aparece dicho pico en la misma posición en $d(n)$	33
2.11. Fotogramas 1555 y 1556 de la secuencia <i>Chaplin</i> . Véase los picos correspondientes en $d_c(n)$ y $d(n)$ en la figura 2.9.	33
2.12. Fotogramas 1313 y 1314 de la secuencia <i>Chaplin</i>	34
2.13. Fotogramas 2072 y 2073 de la secuencia <i>Chaplin</i>	34
2.14. $\rho_n(i, j)$ para los fotogramas de la figura 2.10. Negro representa -1, blanco 1 y gris medio 0.	35
2.15. ρ para los fotogramas de la figura 2.13.	35
2.16. $d_\rho(n)$ para la secuencia <i>Chaplin</i> . El * aparece en los fotogramas donde ocurren cortes reales.	36

2.17. Residuo del cierre de $d_\rho(n)$ para la secuencia <i>Chaplin</i> . El * aparece en los fotogramas donde ocurren cortes reales.	37
2.18. Fotogramas 2461 y 2462 de la secuencia <i>Chaplin</i>	37
2.19. Resultados de la apertura y el cierre sobre funciones unidimensionales: (a) Apertura con EE de tamaño 5. (b) Cierre con EE de tamaño 5. (c) Comparación del efecto de la apertura con dos EE diferentes. (d) Comparación del efecto del cierre con dos EE diferentes.	39
2.20. Aplicación de la apertura para la localización de los cortes en <i>Newsb</i> : (a) En rojo, aparece $d_c(n)$ y en azul el resultado de su apertura con $EE = 2$. (b) Diferencia entre $d_c(n)$ y la apertura.	41
2.21. Ejemplo de transición abrupta utilizando un fotograma intermedio.	41
2.22. Ejemplo de transición utilizando dos fotogramas, (b) y (c), intermedios: (a) Fotograma 355. (b) Fotograma 356. (c) Fotograma 357. (d) Fotograma 358 de la secuencia <i>Alcoi</i>	42
2.23. Aplicación de la apertura para la localización de los cortes en <i>Alcoi</i> : (a) Señal $d_c(X, Y)$. (b) Residuo entre $d_c(n)$ y la apertura con $EE = 2$. (c) Residuo entre la apertura con $EE = 2$ y la apertura con $EE = 3$. (d) Residuo entre la apertura con $EE = 3$ y la apertura con $EE = 4$	43
2.24. Resultado del cierre sobre la secuencia <i>Alcoi</i> : (a) $d_c(n)$. (b) Cierre de $d_c(n)$ con un $EE = 9$	44
2.25. Resultado de la apertura sobre la secuencia <i>Alcoi</i> : (a) $d_\rho(n)$. (b) Apertura de $d_\rho(n)$ con un $EE = 9$	45
2.26. Aplicación del cierre para la localización de los cortes en <i>Alcoi</i> (los cierres se aplican sobre la apertura de $d_\rho(n)$ con $EE = 9$): (a) Señal $d_\rho(n)$. (b) Residuo entre el cierre con $EE = 2$ y la apertura con $EE = 9$. (c) Residuo entre el cierre con $EE = 3$ y el cierre con $EE = 2$. (d) Residuo entre el cierre con $EE = 4$ y el cierre con $EE = 3$	46
2.27. Modelo de elección del umbral según Neyman-Pearson.	48
2.28. (a) Transiciones: fdp <i>normal</i> . (b) No transiciones: fdp <i>exponencial</i>	49
2.29. Estudio estadístico para la elección del u_{dc} óptimo: (a) fdp para las transiciones y las no transiciones. (b) P_T frente a P_{NT} . (c) P_T frente a u_{dc} . (d) P_{NT} frente a u_{dc}	50
2.30. Proceso de obtención del fundido de dos fuentes vídeo 1 y 2.	51
2.31. Ejemplo de fundido con cuatro fotogramas intermedios: (a) Fotograma inicial, A . (b) Fotograma intermedio, $C_1 = \frac{4}{5} \times A + \frac{1}{5} \times B$. (c) Fotograma intermedio, $C_2 = \frac{3}{5} \times A + \frac{2}{5} \times B$. (d) Fotograma intermedio, $C_3 = \frac{2}{5} \times A + \frac{3}{5} \times B$. (e) Fotograma intermedio, $C_4 = \frac{1}{5} \times A + \frac{4}{5} \times B$. (f) Fotograma final.	52
2.32. Curva de la varianza en un fundido.	53
2.33. Varianza de la secuencia <i>Newsb</i> : (a) Varianza para todos los fotogramas. El fundido está localizado entre las líneas verticales. (b) Varianza en la región del fundido.	54
2.34. Cortes ortogonales de una secuencia, para el píxel (i, j) : (a) Corte Horizontal: fila x_0 - tiempo. (b) Corte Vertical: tiempo - columna y_0	55

2.35. Corte vertical (para la componente R) de la imagen corte horizontal de una secuencia.	56
2.36. Resultado de calcular $d_{\rho}^{fundidos}(n)$ para la secuencia <i>Newsb</i>	57
2.37. Proceso de realización de una cortinilla con dos fuentes vídeo 1 y 2.	58
2.38. Ejemplo de cortinilla vertical (desplazándose hacia la derecha), con cuatro fotogramas intermedios.	59
2.39. Ejemplo de cortinilla horizontal (desplazándose hacia abajo), con cuatro fotogramas intermedios.	60
2.40. Dos fotogramas de la cortinilla vertical de <i>Newsa</i> : (a)Fotograma 1645. (b)Fotograma 1646. (c)Imagen diferencia de los fotogramas 1645 y 1646. (d)Imagen anterior convertida a grises.	61
2.41. Proyección por columnas de la imagen diferencia entre dos fotogramas de una cortinilla vertical.	62
2.42. Reconstrucción geodésica: (a)Proyección por columnas de la imagen diferencia entre dos fotogramas de una cortinilla vertical. Es la señal de referencia. (b)Señal marcador. (c)Resultado de la reconstrucción.	63
2.43. Proyecciones por columnas de tres imágenes diferencia entre fotogramas de una cortinilla vertical.	64
2.44. Resultado de búsqueda de cortinillas verticales para una subsecuencia de <i>Newsa</i> : (a)Imagen “tira” para proyecciones por columnas. (b)Residuo de la apertura horizontal de tamaño 20.	65
2.45. Localización de cortinillas verticales en la subsecuencia de <i>Newsa</i> : (a)Imagen “tira” después de procesado. (b)Segmentación mediante umbral de la imagen de (a).	68
2.46. Ejemplo de secuencia que produce un falso positivo al detectar cortinillas verticales.	70
2.47. Localización de cortinillas verticales de la figura 2.46: (a)Imagen “tira” inicial. (b)Imagen “tira ” después de todo el procesado.	71
3.1. Dos fotogramas consecutivos un plano de la secuencia <i>Chaplin</i>	80
3.2. Doce primeros fotogramas del plano de la secuencia <i>Chaplin</i> del fot. 1314 al 1429.	82
3.3. Doce primeros fotogramas del plano de la secuencia <i>Chaplin</i> del fot. 2073 al 2292.	83
3.4. Media y varianza del plano de <i>Chaplin</i> de la figura 3.2.	83
3.5. Media y varianza del plano de <i>Chaplin</i> de la figura 3.3.	84
3.6. Espectro de la media de <i>Chaplin</i> de la figura 3.2. Se ha marcado con un * el máximo.	85
3.7. Componentes lineal y sinusoidal de la media y la varianza del plano de <i>Chaplin</i> de la figura 3.2.	86
3.8. Componentes lineal y sinusoidal de la media y la varianza del plano de <i>Chaplin</i> de la figura 3.3.	86
3.9. Media y varianza originales y corregidas del plano de la figura 3.2.	88
3.10. Media y varianza originales y corregidas del plano de la figura 3.3.	88

3.11. Fotogramas de la figura 3.2 corregidos con el método de variación temporal.	89
3.12. Fotogramas de la figura 3.3 corregidos con el método de variación temporal.	89
3.13. Histogramas de la imágenes de la figura 3.1.	91
3.14. Histograma de referencia.	91
3.15. Relación E/S del método de cambio de histograma.	92
3.16. Resultado del método de cambio de histograma aplicado sobre los fotogramas de la figura 3.1.	92
3.17. Fotogramas de la figura 3.2 corregidos con el método de cambio de histograma.	93
3.18. Fotogramas de la figura 3.3 corregidos con el método de cambio de histograma.	94
4.1. Ejemplo de un plano de la secuencia <i>Taronja</i>	96
4.2. Modelo simplificado de la respuesta espacio-temporal del sistema visual humano.	105
4.3. Respuesta espacio-temporal del sistema visual humano (- - -) y respuesta del filtro a implementar (—).	105
4.4. Respuesta en frecuencia del filtro espacio-temporal.	106
4.5. Corte por el plano $f_y - f_t$ de la respuesta en frecuencia del filtro espacio-temporal ideal. La zona sombreada corresponde a la banda eliminada del filtro.	106
4.6. Espectro espacio-temporal de la secuencia $s(y, t) = \cos(2\pi f_{y0} y)$	107
4.7. Espectro espacio-temporal de la secuencia $s(y, t) = \cos(2\pi f_{y0}(y - vt))$	108
4.8. Ancho de banda espacial en función de la velocidad para el filtro espacio-temporal.	109
4.9. Corte por el plano $f_y - f_t$ de la respuesta en frecuencia de un filtro temporal.	110
4.10. Ancho de banda espacial en función de la velocidad para el filtro temporal.	110
4.11. Respuesta en frecuencia del filtro espacial, $H_1(e^{j\omega_x}, e^{j\omega_y}, e^{j\omega_t})$	111
4.12. Respuesta en frecuencia del filtro temporal, $H_2(e^{j\omega_x}, e^{j\omega_y}, e^{j\omega_t})$	112
4.13. Respuesta en frecuencia del filtro espacio-temporal, $H_3(e^{j\omega_x}, e^{j\omega_y}, e^{j\omega_t})$	112
4.14. Plano $f_y - f_t$ de la respuesta ideal $H(e^{j\omega_x}, e^{j\omega_y}, e^{j\omega_t})$	114
4.15. Plano $f_y - f_t$ de la respuesta real $TF\{h[x, y, t]\}$	114
4.16. Proceso para filtrar la secuencia con $h[x, y, t]$	115
4.17. Bordes espaciales del filtrado espacio-temporal.	116
4.18. Bordes temporales del filtrado espacio-temporal.	116
4.19. Factor de reducción de ruido en función del tamaño del filtrado.	119
4.20. Comparación entre filtrados: (a)Secuencia original ruidosa: carphone+ruido. (b)Filtrado espacio-temporal ($M \times N = 3 \times 3$ y $L = 9$). (c)Filtrado temporal ($L = 9$). (d)Filtrado espacial ($M \times N = 3 \times 3$). Para comparar el resultado del filtrado espacio-temporal (b) y el temporal (c) podemos fijarnos en la cortinilla del fondo, que es estática, y en ambos resultados está nítida, mientras que la cara del personaje, que es una zona en movimiento es la que está emborronada, mucho más en el caso del filtrado temporal (c). Al comparar ahora con el filtrado espacial, vemos que éste emborrona por igual zonas estáticas (cortinilla) y dinámica (cara), y este emborronamiento es el máximo que introduce el filtrado espacio-temporal (caso cuando existe movimiento); esto puede comprobarse en la cara del personaje, que para ambos filtros (espacial y espacio-temporal) está igual de borrosa.	121

4.21. Secuencia <i>Taronja</i> (a)Original (b)Resultado filtro espacio-temporal ($M \times N = 5 \times 5$ y $L = 5$).	123
4.22. Secuencia <i>Taronja</i> (a)Original (b)Resultado filtro espacio-temporal ($M \times N = 5 \times 5$ y $L = 5$). En el primer fotograma se observa que la mancha original no ha desaparecido después del filtrado.	124
5.1. Fotogramas consecutivos de un plano de la secuencia <i>Taronja</i> . Mancha en el fotograma central.	126
5.2. Esquema general del proceso de detección de manchas.	133
5.3. Obtención de máscara de manchas claras mediante el método de la apertura temporal.	136
5.4. Obtención de máscara de manchas oscuras mediante el método del cierre temporal.	137
5.5. Objeto que se despaza una cantidad igual a D en horizontal. Posición en los fotogramas I_{n-1} , I_n e I_{n+1} . El área sombreada es la intersección del objeto entre dos fotogramas adyacentes.	138
5.6. Resultado del cierre temporal de tamaño 2 sobre la secuencia de la figura 5.5 para el fotograma I_n . Vemos como el objeto se ha partido en dos ya que su desplazamiento por fotograma es mayor que la mitad de su tamaño.	138
5.7. Detección de manchas claras: (a)Fotograma original de la secuencia <i>Taronja</i> . (b)Mismo fotograma una vez realizada la apertura temporal. Observar como desaparecen las manchas blancas, pero la barra vertical del tranvía se rompe por la mitad. (c)Residuo de la apertura con el fotograma original . (d)Máscara: residuo tras el umbral de 30.	139
5.8. Detección de manchas oscuras: (a)Fotograma original de la secuencia <i>Sanga</i> (película Sangre y arena). (b)Mismo fotograma una vez realizado el cierre temporal. (c)Residuo del cierre con el fotograma original . (d)Máscara: residuo tras un umbral de 30.	140
5.9. Apertura por reconstrucción.	142
5.10. Obtención de máscara de manchas claras mediante el método de la apertura por reconstrucción.	143
5.11. Obtención de máscara de manchas oscuras mediante el método del cierre por reconstrucción.	144
5.12. Obtención de máscara de manchas claras mediante el método de la apertura por reconstrucción con dilatación del marcador.	145
5.13. Obtención de máscara de manchas oscuras mediante el método del cierre por reconstrucción con erosión del marcador.	145

- 5.14. Detección de manchas claras: (a)Fotograma original de la secuencia *Taronja*, I_{n-1} . (b)Fotograma original de la secuencia *Taronja*, I_n . (c)Fotograma original de la secuencia *Taronja*, I_{n+1} . (d)Marcador obtenido según el proceso de la figura 5.12. Se puede observar que las manchas no aparecen en el marcador, por lo que no aparecerán en la reconstrucción con la misma intensidad que en la imagen de referencia, y por tanto, aparecerán en el residuo. Por otro lado, la barra del tranvía en la imagen I_n tiene algún punto en común con dicha barra en la imagen marcador, por lo que se reconstruirá. (e)Resultado de la apertura por reconstrucción de la imagen (b) con el marcador (d). (c)Residuo de la apertura con el fotograma original. (d)Máscara: residuo tras un umbral de 30. Observar cómo se detectan todas las manchas blancas, y no aparece el artefacto en la barra del tranvía. 147
- 5.15. Detección de manchas oscuras: (a)Fotograma original de la secuencia *Sanga*, I_{n-1} . (b)Fotograma original de la secuencia *Sanga*, I_n . (c)Fotograma original de la secuencia *Sanga*, I_{n+1} . (d)Marcador obtenido según el proceso de la figura 5.13. Se puede observar que las manchas no aparecen en el marcador. (e)Resultado del cierre por reconstrucción de la imagen (b) con el marcador (d). Se puede observar que muchas manchas son reconstruidas, ya que están conectadas a zonas oscuras de la imagen. (c)Residuo de la apertura con el fotograma original. (d)Máscara: residuo tras un umbral de 15. Hay manchas que no son detectadas ya que han sido reconstruidas al nivel de intensidad de la zona oscura a la que están conectadas, y dicha intensidad es similar a la de la mancha. 148
- 5.16. (a) Apertura por reconstrucción. (b) Apertura por λ -reconstrucción $\lambda = 1$. (c)Apertura por λ -reconstrucción $\lambda = 5$ 150
- 5.17. Detección de manchas oscuras: (a)Fotograma original de la secuencia *Sanga*. (b)Mismo fotograma una vez realizado el cierre por λ -reconstrucción con $\lambda = 5$. (c)Residuo del cierre con el fotograma original. (d)Máscara: residuo tras un umbral de 15. Observar como se detectan más manchas negras que en el caso del cierre por reconstrucción, aunque no todas. Nótese también como aparece la falsa alarma en el brazo del personaje. 151
- 5.18. Detección de manchas oscuras: (a)Fotograma original de la secuencia *Sanga*. (b)Mismo fotograma una vez realizado el cierre por λ -reconstrucción con $\lambda = 20$. (c)Residuo del cierre con el fotograma original. (d)Máscara: residuo tras un umbral de 15. Observar como se detectan más manchas negras que en el caso de $\lambda = 5$, aunque también aparecen más puntos aislados, falsa alarma debida al ruido; además la falsa detección debida al brazo se incrementa. . . . 152
- 5.19. Comparación entre métodos de detección de manchas oscuras: (a)Fotograma original de la secuencia *Sanga*. (b)Máscara obtenida con el método del cierre temporal. (c)Máscara obtenida con el método del cierre por reconstrucción. (d)Máscara obtenida con el método del cierre por λ -reconstrucción con $\lambda = 5$. 153
- 5.20. h -reconstrucción 155
- 5.21. Obtención de máscara de zonas claras mediante la apertura con h -reconstrucción. 155

5.22. Detección de zonas claras mediante h -reconstrucción: (a)Fotograma original de la secuencia *Taronja*. (b)Resultado de la h -reconstrucción con $h=70$. En esta imagen se han eliminado todas las zonas claras con menor contraste que 70, y las que tienen mayor contraste se han transformado en una zona plana de ese valor. (c)Residuo de la h -reconstrucción. (d)Umbral del residuo de valor 1. Aparecen todas las zonas que en la imagen h -reconstruida tienen alguna diferencia con la original. (e) Umbral del residuo de valor 70. Máximos de la imagen original con un contraste mayor o igual que 70. (f) Resultado de la reconstrucción de la imagen obtenida en (d) utilizando como marcador la obtenida en (f). Zonas de la imagen original con un contraste superior a 70. 156

5.23. Procedimiento global de obtención de máscara de manchas claras. 157

5.24. Detección de manchas claras: (a)Fotograma original de la secuencia *Taronja*. (b)Máscara obtenida mediante el método de la apertura temporal.(c)Máscara obtenida mediante la h -reconstrucción. (d)Máscara resultante del proceso global de detección. 158

5.25. Interpolación de mancha mediante el método de interpolación espacial: (a)Fotograma original de la secuencia *Taronja*. (b)Restauración con la apertura temporal. (c)Fotograma una vez restaurado con la apertura por reconstrucción. Observar como las manchas no son eliminadas. (d)Fotograma una vez restaurada la mancha con la apertura por λ -reconstrucción con $\lambda = 15$. La restauración mejora con respecto a la apertura por reconstrucción. 161

5.26. Restauración de una imagen fija mediante el método de restauración espacial (a)Imagen manchada. (b)Imagen restaurada. 162

5.27. Restauración de manchas en un borde en movimiento: (a)Fotograma original de la secuencia en el que aparece la mancha. (b)Fotograma una vez restaurado usando la apertura temporal. Observar el pequeño defecto en el borde en movimiento. (c)Fotograma una vez restaurado usando la apertura por reconstrucción. La mancha se convierte en una zona plana de otro nivel de gris y muy visible. (d)Fotograma una vez restaurado usando la apertura por λ -reconstrucción con $\lambda = 20$. Se trata de una solución intermedia entre las dos anteriores. (e)Fotograma una vez restaurado usando el procesamiento priorizado. Observar el defecto en el borde. (f)Fotograma una vez restaurado usando el procesamiento priorizado respetando contorno. Observar que el borde se reconstruye satisfactoriamente, ya que se trata de un borde rectilíneo. 163

5.28. Interpolación de mancha mediante los métodos de interpolación espacial y temporal: (a)Fotograma original de la secuencia *Taronja*. (b)Máscara que indica las zonas a interpolar. (c)Fotograma una vez restauradas las manchas con el método de interpolación espacial, *zoom* de la zona de la mancha grande, que es la que está en una zona en movimiento, y en la que se puede ver la diferencia entre los métodos. (d)Fotograma una vez restaurada la mancha con la apertura temporal, *zoom* de la zona de la mancha. 164

5.29. Secuencia *Taronja* (a)Original (b)Resultado de la restauración con apertura temporal. 167

6.1.	Esquema general del proceso de restauración para cada plano de la secuencia. Los bloques sombreados son los que han sido implementados en esta tesis.	171
6.2.	Secuencia <i>Taronja</i> (a)Original (b)Resultado de la restauración.	173
6.3.	Secuencia <i>Sanga</i> (a)Original (b)Resultado del proceso de restauración.	174
6.4.	Secuencia <i>Sanga</i> (a)Original (b)Resultado del proceso de restauración.	175

Índice de cuadros

2.1.	Resultados de algunos algoritmos comunes en segmentación de vídeo	21
2.2.	Resultados de la detección de cortes para vídeo comprimido	21
2.3.	Resultados de la detección de cortes con el método 1.	74
2.4.	Resultados de la detección de cortes con el método 2.	75
2.5.	Valores de acierto, precisión, τ , % de correctos y % de falsos para los métodos 1 y 2 de detección de cortes.	75
2.6.	Valores de acierto, precisión, τ , % de correctos y % de falsos para los métodos 1 y 2 de detección de cortes para el conjunto de películas actuales.	76
2.7.	Valores de acierto, precisión, τ , % de correctos y % de falsos para los métodos 1 y 2 de detección de cortes para el conjunto de películas antiguas.	76
2.8.	Resultados de la detección de fundidos.	76
2.9.	Resultados de la detección de cortinillas.	77
4.1.	Comparación del factor de reducción de ruido (en dB): filtro espacio-temporal $M \times N \times 3$ y filtro espacial $M \times N$	120
4.2.	Comparación del factor de reducción de ruido (en dB): filtro espacio-temporal $5 \times 5 \times L$ y filtro temporal L	120

Bibliografía

- [1] Archangel by snell & wilcox. the ultimate in real-time restoration of film-originated archives. <http://www.snellwilcox.com/internet/reference/brochures/archangel.pdf>.
- [2] Aurora. outils de restauration pour le futur. <http://www.ina.fr/Recherche/Aurora/aurora.fr.html>.
- [3] Brava. <http://www.ina.fr/recherche/projets/encours/brava/>.
- [4] The digital film restoration process. <http://www.vcpc.univie.ac.at/activities/projects/FRAME/>.
- [5] Edward H. Adelson and James R. Bergen. The extraction of spatio-temporal energy in human and machine vision. In *Proceedings of the Workshop on Motion: Representation and Analysis*, pages 151–155, Charleston, SC, May 7-9 1986.
- [6] G. Ahanger and T.D.C. Little. A survey of technologies for parsing and indexing digital video. *Journal of Visual Communications and Image Representations*, 7(1):28–43, 1996.
- [7] P. Aigrain and P. Joly. The automatic real-time analysis of film editing and transition effects and its applications. *Comput. and Graphics.*, 18(1):93–103, 1994.
- [8] Philippe Aigrain, HongJiang Zhang, and Dragutin Petkovic. Content-based representation and retrieval of visual media: A state-of-the-art review. *Multimedia Tools and Applications*, 3(3):179–202, 1996.
- [9] A. Albiol and J. Prades. Patch filling using prioritized image processing. In *Proc. of the IASTED International Conference on Signal and Image Processing (SIP98)*, Las Vegas (NV), october 27-31 1998.
- [10] Antonio Albiol, Valery Naranjo, and Jesús Angulo. Low complexity cut detection in the presence of flicker. In IEEE, editor, *Proc. of International Conference of Image Processing 2000*, 2000.
- [11] M. B. Alp and Y.Ñeuvo. 3-dimensional median filters for image sequence processing. In IEEE, editor, *Proc. of IEEE Int. Conf. Acoust., Speech, Signal Process.*, volume 4, pages 2917–2920, Toronto, Canada, May 1991.

- [12] A. Amer and H. Schröder. A new video noise reduction algorithm using spatial subbands. In *Internationally Conference Electronics, Circuits and Systems (ICECS'96)*, Rodes, 1996.
- [13] J. Angulo, A. Albiol, and V.Ñaranjo. Automatic video partitioning through cut and gradual effects detection using morphological operators. *Signal Processing, EURASIP*, To be published.
- [14] Jesús Angulo. Segmentación temporal de secuencias de vídeo. Proyecto fin de carrera, Escuela Técnica Superior de Ingenieros de Telecomunicación. Universidad Politécnica de Valencia., Octubre 1999. Director Antonio Albiol.
- [15] G. R. Arce. Multistage order statistic filters for image sequence processing. *IEEE Trans. Acoust., Speech., Signal Process.*, 39:1146–1163, 1991.
- [16] F. Arman, A. Hsu, and M.Y. Chiu. Feature management for large video databases. In *SPIE Storage and Retrieval for Image and Video Databases*, pages 2–12, 1993.
- [17] F. Arman, A. Hsu, and M.Y. Chiu. Image processing on compressed data large video databases. In *First ACM International Conference Multimedia*, Anaheim, CA, August 1993.
- [18] S. Armstrong, A. C. Kokaram, and P. J. Rayner. Non-linear interpolation of missing image data using min-max functions. In *Proc. of the 1997 IEEE Workshop on nonlinear signal and image processing (NSIP'97)*, Mackinac Island, Michigan, 7-11 September 1997.
- [19] Steven Armstrong. *Film and video restoration using rank-order models*. PhD thesis, King's College, University of Cambridge, June 1999.
- [20] John S. Boreczky and Lawrence A. Rowe. Comparasion of video shot boundary detection techniques. In *SPIE Conf. on Visual Communications and Image Proc.*, 1996.
- [21] D. Borghys and M. Acheroy. Noise reduction in image sequences whith sparse temporal sampling. In *Proc. of the IASTED International Conference on Signal and Image Processing (SIP95)*, Las Vegas (NV), novembre 20-23 1995.
- [22] P. Bouthemy, C. Garcia, R. Ronfard, G. Tziritas, E. Veneau, and D. Zugaj. Scene segmentation and image feature extraction for video indexing and retrieval. In *Third International Conference on Visual Information Systems (VISUAL'99)*, Amsterdam, june 1999.
- [23] P. Bouthemy, M. Gelgon, and F. Ganansia. A unified approach to shot change detection and camera motion characterization. Report interne of IRISA 1148, IRISA, 1997.
- [24] J. M. Boyce. Noise reduction of image sequence using adaptive motion compensated frame averaging. In IEEE, editor, *Proc. of IEEE Int. Conf. Acoust., Speech, Signal Process.*, volume 3, pages 461–464, San Francisco, CA, March 1992.

- [25] J. Brailean, R.P. Kleihorst and S. Efstratiadis, A.K. Katsaggelos, and R.L. Lagendijk. Noise reduction filters for dynamic image sequences: a review. *Proceedings of the IEEE*, 83(9):1270–1292, september 1995.
- [26] James C. Brailean and Aggelos K. Katsaggelos. Recursive displacement estimation and restoration of noisy-blurred image sequences. In *Proc. of the IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, volume 5, pages 273–276, Minneapolis, Minnesota, USA, Apr. 27-30 1993.
- [27] J.C. Brailean and A. K. Katsaggelos. Simultaneous recursive displacement estimation and restoration of noisy-blurred image sequences. *IEEE Transactions on Image Processing*, 4(9):1236–1251, September 1995.
- [28] R. Brunelli, O. Mich, and C.M. Modena. A survey on the automatic indexing of video data. *Journal of Visual Communication and Image Representation*, 10:78–112, 1999.
- [29] D. Cano and M. Benard. *Image sequence processing and dynamic scene analysis*, chapter 3D Kalman filtering of image sequences. Springer-Verlag, 1983.
- [30] Jesús Bescós Cano. *Segmentación temporal de secuencias de vídeo*. PhD thesis, ETSI Telecomunicación. Universidad Politécnica de Madrid, Valencia, 2001. Directores: Guillermo Cisneros y José Manuel Menéndez.
- [31] M. Ceccarelli, A. Hanjalic, and R. L. Lagendijk. A sequence analysis system for video databases. In *5th Workshop on Time-Varying Image Processing and Moving Object Recognition*, Florence (IT), September 1996.
- [32] Cheuk L. Chan, A. K. Katsaggelos, and A.V. Sahakian. Linear-quadratic noise-smoothing filters for quantum-limited images. *IEEE Transactions on Image Processing*, 4(9):1328–1333, September 1995.
- [33] J. Chen, C. M. Taskiran, Alberto Albiol, C. A. Bouman, and E. J. Delp. Vibe: A video indexing and browsing environment. In *Proceedings of the SPIE Conference on Multimedia Storage and Archiving Systems IV*, volume 3846, pages 148–164, Boston, USA, September 20-22 1999.
- [34] J-H Chenot, J. O. Drewery, and D. Lyon. Restoration of archived television programmes for digital broadcasting. In *Proceedings of the International Broadcasting Convention 1998 (IBC 98)*, volume 3846, pages 26–31, Amsterdam, The Netherlands, September 1998.
- [35] M.Ñ. Chong, P. Liu, W. B. Goh, and D. Krishnan. A new spatio-temporal mrf model for the detection of missing data in image sequences. In IEEE, editor, *Proceedings of International Conference on Acoustics Speech and Signal Processing*, volume 4, pages 2977–2980, April 1997.
- [36] Man-Nang Chong, S.Kalra, D. Krishnan, and A. Laud. Computerised motion picture restoration software: Revival.
<http://www.geocities.com/Hollywood/Academy/9772/whatsnew/revival.html>.

- [37] A. Dailianas, R. Allen, and P. England. Comparisons of automatic video segmentation algorithms. In *Proceedings, SPIE Photonics East'95: Integration Issues in Large Commercial Media Delivery Systems.*, Philadelphia, Octubre 1995.
- [38] Commission Supérieure Technique de l'Image et du Son (CST). *Guide de la conservation des films.* CST, Paris, 1995.
- [39] Commission Supérieure Technique de l'Image et du Son (CST). *La restauration numérique des films cinématographiques.* CST, Paris, 1997.
- [40] F. Dekeyser, P. Pérez, and P. Bouthemy. Restoration of noisy, blurred, undersampled image sequences using a parametric motion model. In *Inter. Symposium on Image and Video Communications, ISIVC 2000*, Rabat, Maroc, April 2000.
- [41] Fabien Dekeyser, Patrick Bouthemy, and Patrick Pérez. Spatio-temporal wiener filtering of image sequences using a parametric model. In IEEE, editor, *ICIP 2000*, 2000.
- [42] Claire-Hélène Demarty. *Segmentation et structuration d'un document vidéo pour la caractérisation et l'indexation de son contenu sémantique.* PhD thesis, Ecole Nationale Supérieure des Mines de Paris, Paris, janvier 2000.
- [43] Eric Dubois. The sampling and reconstruction of time-varying imagery with application in video systems. *Proceedings of the IEEE*, 73(4):502–522, April 1985.
- [44] Eric Dubois and Janusz Konrad. Motion estimation and motion-compensated filtering of video signals. In *Proc. of the IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, volume 1, pages 95–98, Minneapolis, Minnesota, USA, Apr. 27-30 1993.
- [45] Rakesh Dugad and Narendra Ahuja. Noise reduction in video by joint spatial and temporal processing. *Submitted for review to IEEE Trans. on Circuits and Systems for Video Technology, September, 2001.*
- [46] Rakesh Dugad and Narendra Ahuja. Performance enhancement and analysis of a video denoising scheme. *Submitted for review to IEEE Trans. on Circuits and Systems for Video Technology, September, 2001.*
- [47] Rakesh Dugad and Narendra Ahuja. Video denoising by combining kalman and wiener estimates. In IEEE, editor, *ICIP-99*, Japan, October 1999.
- [48] Peter Eaves. Traditional film restoration techniques in a digital world. In *First Annual Spring Film Conference*, number 7, Los Angeles, California, March 1997.
- [49] W. A. C. Fernando, C.N. Canagarajah, and D.R.Bull. Fade and dissolve detection in uncompressed and compressed video sequences. In IEEE, editor, *ICIP-99*, 1999.
- [50] W. A. C. Fernando, C.N. Canagarajah, and D.R.Bull. Wipe scene change detection in video sequences. In IEEE, editor, *ICIP-99*, 1999.

- [51] Etienne Decenciere Ferrandiere. Motion picture restoration using morphological tools. In *International Symposium on mathematical morphology (ISMM)*, pages 361–368, 1996.
- [52] Etienne Decenciere Ferrandiere. *Restauration automatique de films anciens*. PhD thesis, Ecole Nationale Supérieure des Mines de Paris, Paris, December 1997.
- [53] Sergei V. Fogel. The estimation of velocity vector fields from time-varying image sequences. *Computer vision ,graphics and image processing*, 53(3):253–287, 1991.
- [54] U. Gargi, R. Kasturi, and S. Antani. Performance characterization and comparison of video indexing algorithms. In *Conference of Computer Vision and Pattern Recognition (CVPR)*, 1998.
- [55] I. A. Glover and P. M. Grant. *Digital Communications*. Prentice-Hall, 1998.
- [56] W. B. Goh, M.Ñ. Chong, S. Kalra, and Dilip Krishnan. Bi-directional 3d auto-regressive model approach to motion picture restoration. In *ICASSP*, pages 2277–2280, Atlanta, Georgia, 7-9 May 1996.
- [57] Cristina Gomila. *Mise en correspondance de partitions en vue du suivi d'objets*. PhD thesis, Ecole Nationale Supérieure des Mines de Paris, Paris, Septembre 2001.
- [58] R. C. Gonzalez and P. Wintz. *Digital image processing*. Addison-Welsey, Massachusetts, 1987.
- [59] M. Grimaud. A new measure of contrast: dynamics. In *SPIE, Image Algebra and Morphological Image Processing III*, volume 1769, pages 292–305, San Diego, 1992.
- [60] A. Hampapur, R. Jain, and T. Weymouth. Digital video segmentation. In *Proc. 2nd ACM Intl. Conf. on Multimedia*, pages 357–364, 1994.
- [61] Neal R. Harvey and Stephen Marshall. Film restoration using soft morphological filters. In *Proc. of the 6th International Conference on Image Processeing and its Applications (IPA '97)*, pages 279–282, Dublin, Ireland, july 1997.
- [62] Y.J. Zhang H.B.Lu and Y.R.Yao. Robust gradual scene change detection. In IEEE, editor, *ICIP-99*, 1999.
- [63] Kien A. Hua and JungHwan Oh. Detecting video shot boundaries up to 16 times faster. In *The 8th ACM International Multimedia Conference*, pages 385–387, LA, CA., Oct.30-Nov.3 2000.
- [64] Kien A. Hua, JungHwan Oh, and Khanh Vu. A content-based scene change detection and clasiffication thecnique using background tracking. In *IS and T/SPIE Conference on Multimedia Computing and Networking 2000*, pages 254–265, San Jose CA., 24-28 Jan. 2000.

- [65] Kien A. Hua, JungHwan Oh, and Khanh Vu. Non-linear approach to shot boundary detection. In *S and T/SPIE Conference on Multimedia Computing and Networking 2001*, pages 1–12, San Jose CA., 22-25 Jan. 2001.
- [66] P. Joly. Segmentation en plans: Les effets ... des effets de transition. http://www.irit.fr/ACTIVITES/EQ_AMI/PUBLI/GT3/, 1999.
- [67] K. Jostschulte and A. Amer. A new cascade spatio-temporal noise reduction scheme for interlaced video. In *International Conference on Image Processing*, pages 438–439, Chicago, Illinois, 4-7 october 1998.
- [68] K. Jostschulte, A. Amer, M.Schu, and H. Schröder. Perception adaptative temporal tv-noise reduction using contour preserving techniques. 4(3):1091–1096, 1998.
- [69] K. Jostschulte, A. Amer, M.Schu, and H. Schröder. A subband based spatio-temporal noise reduction technique for interlaced video signals. In *Internationally Conference on Consumer Electronics*, pages 438–439, Los Angeles, 2-4 june 1998.
- [70] Showbhik Kalra, M.N.Chong, and Dilip Krishnan. A new auto-regresive (ar) model-based algorithm for motion picture restoration. In *ICASSP*, volume 4, pages 2557–2560, Munich, Germany, 21-24 April 1997.
- [71] John R. Kender and Boon-Lock Yeo. Video scene segmentation via continuous video coherence. Research report, IBM, 1997.
- [72] J. Kim and J. W. Woods. Spatio-temporal adaptive 3-d kalman filter for video. *IEEE Trans. on Image Process.*, 6(3):414–423, March 1997.
- [73] Sang Hyun Kim and Rae-Hong Park. A novel approach to scene change detection using a cross entropy. In IEEE, editor, *ICIP-00*, 2000.
- [74] Richard P. Kleihorst, Reginald L. Lagendijk, and J. Biemond. Noise reduction of severely corrupted image sequences. In *Proc. of the IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, volume 5, pages 293–296, Minneapolis, Minnesota, USA, Apr. 27-30 1993.
- [75] Richard P. Kleihorst, Reginald L. Lagendijk, and J. Biemond. An order-statistics supported non-linear filter with application to image sequence filtering. In *Proc. of the IEEE Winter Workshop on Nonlinear Digital Signal Processing*, pages 1.2–2.1 to 1.2–2.5, Tampere, Finland, Jan. 17-20 1993.
- [76] Richard P. Kleihorst, Reginald L. Lagendijk, and J. Biemond. An efficient spatio-temporal os-filter for gamma-corrected video signals. In *Proc. of the IEEE International Conference on Image Processeing (ICIP'94)*, Austin, Texas, august 1994.
- [77] Richard P. Kleihorst, Reginald L. Lagendijk, and J. Biemond. Noise reduction of image sequences using motion compensation and signal descomposition. *IEEE Transactions on Image Processing*, 4(3):274–284, 1995.

- [78] Richard P. Kleihorst, Reginald L. Lagendijk, and J. Biemond. An adaptative order-statistic noise filter for gamma-corrected image sequences. *IEEE Transactions on Image Processing*, 6(10), 1997.
- [79] Richard Petrus Kleihorst. *Noise filtering of image sequences*. PhD thesis, University of Delft, 1994.
- [80] J. C. Klein. *Conception et réalisation d'une unité logique pour l'analyse quantitative d'images*. PhD thesis, Université de Nancy, 1976.
- [81] V. Kobla, D. Doermann, and A. Rosenfeld. Compressed domain video segmentation. V. Kobla, D. Doermann, and A. Rosenfeld, Compressed domain video segmentation, CfAR Technical Report CAR-TR-839 (CS-TR-3688) , 1996., 1996. cite-seer.nj.nec.com/vikrant96compressed.html.
- [82] A. C. Kokaram, R.D. Morris, W.J. Fitzgerald, and P.J. Rayner. Interpolation of missing data in image sequences. *IEEE Trans. Image Processing*, 4(11):1509–1519, November 1995.
- [83] A. C. Kokaram, R.D. Morris, W.J. Fitzgerald, and P.J. Rayner. Detection of missing data in image sequences. *IEEE Trans. Image Processing*, 4(11):1496–1508, November 1995.
- [84] Anil Kokaram. *Motion Picture Restoration*. Springer, London, 1998.
- [85] Anil C. Kokaram. Reconstruction of severely degraded image sequence. In *9th International conference on image analysis and processing (ICIAP)*, pages 773–780, Florence, Italy, September 1997.
- [86] Anil C. Kokaram and Simon J. Godsill. A system for reconstruction of missing data in image sequences using sampled 3d ar models and mrf motion priors. In *Proceedings of the European Conference on Computer Vision*, pages 613–624, april 1996.
- [87] Anil C. Kokaram and Simon J. Godsill. Joint detection, interpolation, motion an parameter estimation for image sequences with missing data. In *Proceedings of the International conference image processing*, volume 2, pages 191–194, Santa Barbara, CA, USA, 26-29 October 1997.
- [88] Anil C. Kokaram and Simon J. Godsill. Joint noise reduction, motion estimation, missing data reconstruction and model parameter estimation for degraded motion pictures. In *Proceedings of the SPIE Conference on Bayesian Inference for Inverse Problems*, San Diego, USA, july 1998.
- [89] Anil Cristopher Kokaram. *Motion picture restoration*. PhD thesis, Churchill College, University of Cambridge, May 1993.
- [90] Irena Koprinska and Sergio Carrato. Temporal video segmentation: A survey. cite-seer.nj.nec.com/378900.html, 2000.

- [91] Reginal L. Lagendijk, Peter M. B. van Roosmalen, and J. Biemond. *Video Enhancement and Restoration*, chapter 3. Image and Video Processing. Academic Press, 2000.
- [92] H. C. Liu and G. L. Zick. Automatic determination of scene changes in mpeg compressed video. In *ISCAS-IEEE International Symposium on Circuits and Systems*, 1995.
- [93] J. Llach and P. Salembier. Analysis of video sequences: Table of contents and index creation. In *Int. Workshop on Very Low Bitrate Video, VLBV'99*, October 1999.
- [94] Ze-Nian Li Mark S. Drew and Xiang Zhong. Video dissolve and wipe detection via spatio-temporal images of chromatic histogram differences. In IEEE, editor, *ICIP-00*, 2000.
- [95] Ferrán Marqués. *Multiresolution image segmentation based on compound random fields: Application to image coding*. PhD thesis, DTSC Universitat Politècnica de Catalunya, Barcelona, Decembre 1992. Director: Antoni Gasull.
- [96] Simon Masnou. Disocclusion: a variational approach using level lines. *IEEE Transactions on Image Processing*, 11(2):68–76, February 2002.
- [97] G. Matheron. *Eléments pour une théorie des milieux poreux*. Masson, Paris, 1967.
- [98] G. Matheron. *Random sets and integral geometry*. John Wiley and Sons, New York, 1975.
- [99] J. Meng, Y. Juan, and S.-F. Chang. Scene change detection in a mpeg compressed video sequence. In *IST/SPIE Symposium Proceedings*, volume 2419, pages 14–25, San Jose, California, February 1995.
- [100] José Manuel Mossi. *Contribución a la detección y análisis de microcalcificaciones en mamografías mediante tratamiento digital de imagen*. PhD thesis, ETSI Telecomunicación. Universidat Politècnica de Valencia, Valencia, 1998. Director: Antonio Albiol.
- [101] H. Muller-Seelich, W. Plaschzug, P. Schallauer, S. Postzman, and W. Haas. Digital restoration of 35 mm films. volume 1, pages 255–265, Louvain-la-Neuve, Belgium, 1996.
- [102] M. J. Nadenau and S. K. Mitra. Blotch and scratch detection in image sequences based on rank ordered differences. In *Proceedings of the 5th International Conference on Time-Varying Image Processing and Moving Object Recognition*, Florence, Italy, 5-6 September 1996.
- [103] A. Nagasaka and Y. Tanaka. *Visual Database Systems II*, chapter Automatic Video Indexing and Full-Video Search for Object Appearances, pages 113–127. Elsevier Science Publishers, 1982.
- [104] Valery Naranjo and Antonio Albiol. Flicker reduction in old films. In IEEE, editor, *Proc. of International Conference of Image Processing 2000*, 2000.

- [105] C. W. Ngo, T. C. Pong, and R.T.Chin. Camera breaks detection by partitioning of 2d spatio-temporal images in mpeg. In *IEEE Multimedia Systems (ICMCS'99)*, 1999.
- [106] H.Ñorell, B. Oelmann, and Youzhi Xu. Spatio-temporal noise reduction asic for real-time video processing. In *Proceedings of 2000 IEEE Noedic Signal Processing Symposium*, Kolmården,Sweden, 13-15 June 2000.
- [107] JungHwan Oh and Kien A. Hua. A practical technique for video decomposing. In *Proceedings of The 11th Conference of The Chinese- American Scholars Association of Florida*, pages 38–45, Orlando,FL., 21-22 July. 2000.
- [108] Toshiaki Ohuchi, Takamasa Seto, Takashi Komatsu, and Takahiro Saito. A robust method of image flicker correction for heavily corrupted old film sequences. In IEEE, editor, *ICIP 2000*, 2000.
- [109] M. Pardàs, Jean Serra, and Luís Torres. Connectivity filters for image sequences. In *SPIE Symposium on Image Algebra and Morphological Image Processing III*, volume 1769, pages 318–329, 1992.
- [110] N. Patel and I. Sethi. Compressed video processing for cut detection. In *IEE Proceedings - Vision, Image and Signal Processing*, volume 143, pages 315–323, October 1996.
- [111] N. Patel and I. Sethi. Video shot detection and characterization for video databases. *Pattern Recognition*, 30:583–592, 1997.
- [112] A. Patti, A. Tekalp, and M. Sezan. A new motion compensated reduced order model kalman filter for space-varying restoration of progressive and interlaced video. 7(4):543–554, April 1998.
- [113] José Prades. *Interpolación de imágenes mediante algoritmos no lineales*. PhD thesis, ETSI Telecomunicación. Universidae Politécnica de Valencia, Valencia, 2000. Director: Antonio Albiol.
- [114] William K. Pratt. *Digital image processing*. John Wiley, 1991.
- [115] P. S. Richardson and D. Suter. Restoration of historical film for digital compression: A case of study. In *ICIP-95*, pages II 49–52, Washintong D.C., October 1995. IEE,1995.
- [116] P. M. B. Roosmalen, A. C. Kokaram, and J. Biemond. Improved blotch detection by postprocessing. In *Proc. of IEEE Signal Processing Symposium SPS'98*, pages 223–226, Leuven, Belgium, 1998.
- [117] P. M. B. Roosmalen, A. C. Kokaram, and J. Biemond. Noise reduction of image sequences as preprocessing for mpeg2 encoding. In *Signal Processing IX*, pages 2253–2256, Rodos, Greece, September 1998.
- [118] P. M. B. Roosmalen, A. C. Kokaram, and J. Biemond. Fast high quality interpolation of missing data in image sequences using a controlled pasting scheme. In *Proc. of ICASSP'99*, 1999.

- [119] Peter M. B. Roosmalen, J. Biemod, and R. J. Lagendijk. Restoration and storage of film and video archive material. *Signal Processing for Multimedia*, pages 167–191, 1999.
- [120] Peter M. B. Roosmalen, R. J. Lagendijk, and J. Biemod. *Time-varying image processing and moving object recognition*, volume 4, chapter Flicker reduction in old film sequences, pages 11–18. Elsevier Science, 1997.
- [121] Peter M. B. Roosmalen, R. J. Lagendijk, and J. Biemod. Correction of intensity flicker in old film sequences. *IEEE Transactions on Circuits and Systems for Video Technology*, December 1999.
- [122] Peter Van Roosmalen. *Restoration of archived film and video*. PhD thesis, University of Delft, 1999.
- [123] L. Rosenthaler, A. Wittmann, A. Günzl, and R. Gschwind. Restoration of old movie films by digital image processing. In *IMAGE'COM 96*, Bordeaux, France, May 1996. Available from <http://foto.unibas.ch/research/paper3/filmrest.html>.
- [124] Takahiro Saito, Takashi Komatsu, Toshiaki Ohuchi, and Takamasa Seto. Image processing for restoration of heavily-corrupted old film sequences. In IEEE, editor, *Proceedings of the International Conference on Pattern Recognition: ICPR'00*, 2000.
- [125] Takahiro Saito, Takashi Komatsu, Toshiaki Ohuchi, and Tomohisa Hoshi. Practical nonlinear filtering for removal of blotches from old film. In *ICIP 99*. IEEE, 1999.
- [126] Philippe Salembier and Jean Serra. Flat zones filtering, connected operators and filters by reconstruction. *IEEE Trans. on Image Processing*, 3(8):1153–1160, august 1995.
- [127] Peter Schallauer, Axel Pinz, and Werner Haas. Automatic restoration algorithms for 35 mm films. *Videre: Journal of Computer Vision Research*, 1(3):60–85, Summer 1999.
- [128] Louis L. Scharf. *Statistical signal processing: detection, estimation, and time series analysis*. Addison-Wesley, 1991.
- [129] Jean Serra. *Image analysis and mathematical morphology*. Academic Press, London, 1982.
- [130] Jean Serra. *Image analysis and mathematical morphology, vol. II: Theoretical advances*. Academic Press, London, 1988.
- [131] Jean Serra and Philippe Salembier. Connected operators and pyramids. In *Proceedings of SPIE. Image Algebra and Mathematical Morphology 93*, San Diego, July 1993.
- [132] I. K. Sethi and Nilesh Patel. A statistical approach to scene change detection. In *IS and T/SPIE Proc.: Storage and Retrieval for Image and Video Databases III*, volume 2420, 1995.
- [133] Takamasa Seto, Toshiaki Ohuchi, Takashi Komatsu, and Takahiro Saito. Selective sharpness enhancement of heavily corrupted old film sequences. In IEEE, editor, *ICIP 2000*, 2000.

- [134] M.I. Sezan, M.K.Özkan, and S.V. Fogel. Temporally adaptive filtering of noisy image sequences using a robust motion estimation algorithm. In *Proc. of the IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, volume 4, pages 2429–2432, Toronto, Canada, May 14–17 1991.
- [135] B. Shahraray. Scene change detection and content-based sampling of video sequences. In *Digital Video Compression: Algorithms and Technologies*, volume SPIE-2419, pages 2–13, Feb. 1995.
- [136] B. Shen, D. Li, and I.K. Sethi. Cut detection via compressed domain edge extraction. In *IEEE Workshop on Nonlinear Signal and Image Processing*, September 1997.
- [137] K. Shen and E. J. Delp. A fast algorithm for video parsing using mpeg compressed sequences. In IEEE, editor, *Proceeding IEEE International Conference of Image Processing*, volume III, pages 252–255, Washington, D.C., Octubre 1995.
- [138] J. Signes. Réduction du bruit dans les séquences d’images par filtrage compensé du mouvement. Technical report 93-03, INRS-Télécommunications, January 1993.
- [139] Thomas Sikora. Mpeg digital video-coding standars. *IEEE Signal Processing Magazine*, pages 82–100, September 1997.
- [140] D. Suter. Film restoration and processing. Emerge report, June 1997. Available from <http://www.batman.eng.monash.edu.au/admin/emerge.html>.
- [141] D. Suter and P. Richardson. Historical film restoration and video coding. In *PCS’96*, pages 389–394, Melbourne, Aust, March 1996. http://www.batman.eng.monash.edu.au/suter_publications/newpcs.ps.gz.
- [142] A. Murat Tekalp. *Digital Video Processing*, chapter 5. Optical Flow Methods, pages 72–116. Prentice Hall, 1995.
- [143] A. Murat Tekalp. *Digital Video Processing*. 0-130190075-7. Prentice-Hall, 1995.
- [144] Livio Tenze, Giovanni Ramponi, and Sergio Carrato. Blotches correction and contrast enhancement for old film pictures. In IEEE, editor, *ICIP-00*, 2000.
- [145] Ba Tu Truong, Chitra Dorai, and Svetha Venkatesh. Improved fade and dissolve detection for reliable video segmentation. In IEEE, editor, *ICIP-00*, 2000.
- [146] Christian J. van den Branden Lambrecht. *Perceptual models and architectures for video coding applications*. PhD thesis, École Polytechnique Fédérale de Lausanne, Lausanne, 1996.
- [147] Christian J. van den Branden Lambrecht. A working spatio-temporal model of the human visual system for image restoration and quality assessment applications. In *Proceedings of ICASSP 96*, 1996.

- [148] P. M. B. van Roosmalen, S. J. P. Westen, R.L. Lagendijk, and J. Biemond. Noise reduction for image sequences using an oriented pyramid thresholding technique. In *Proc. of ICIP-96*, volume I, pages 375–378, Lausanne, Switzerland, September 1996.
- [149] N. Vasconcelos and A. Lippman. A bayesian video modeling framework for shot segmentation and content characterization. In *Workshop on CAIVL CVPR'97 San Juan Puerto Rico 1997 IEEE*, 1997.
- [150] Raymond Veldhuis. *Restoration of lost samples in digital signals*. 0-775198-2. Prentice-Hall, London, 1990.
- [151] Luc Vincent. Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms. *IEEE Trans. on Image Processing*, 2(2), April 1993.
- [152] Yao Wang, Jörn Ostermann, and Ya-Quin Zhang. *Video processing and Communications*. 0-13017547-1. Prentice-Hall, Upper Saddle River, New Jersey, 2002.
- [153] S. J. P. Westen, R. L. Lagendijk, and J. Biemond. Spatio-temporal model of human vision for digital video compression. In *Proceedings of the SPIE Conference on Human Vision and Electronic Imaging*, San Jose, USA, 1997.
- [154] B. Yeo and B. Liu. Rapid scene analysis on compressed video. *IEEE Transactions on Circuits and Systems for Video Technology*, 5(6):533–544, December 1995.
- [155] Boon-Lock Yeo and Bede Liu. On the extraction of dc sequence from mpeg compressed video. In *Proceedings of the International conference image processing*, Washington DC, USA, 23-26 October 1995.
- [156] R. Zabih, J. Miller, and K. Mai. A feature-based algorithm for detecting and classifying scene breaks. In *Proc. of ACM Multimedia '95*, pages 189–200, San Francisco, November 1995.
- [157] Ramin Zabih, Justin Miller, and Kevin Mai. A feature-based algorithm for detecting and classifying production effects. *Multimedia Systems*, 7(2):119–128, 1999.
- [158] H. J. Zhang, A. Kankanhalli, and S. W. Smoliar. Automatic partitioning of full motion video. *Multimedia Systems*, 1:10–28, 1993.
- [159] H.J. Zhang, C. Y. Low, and S.W. Smoliar. Video parsing and browsing using compressed data. *Multimedia Tools Appl.*, 1:91–113, 1995.
- [160] M. K. Özkan, A. Tanju Erdem, M. I. Sezan, and A. Murat Tekalp. Efficient multiframe wiener restoration of blurred and noisy image sequences. *IEEE Transactions on Image Processing*, 1(4):453–476, October 1992.